UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Head Office: Università degli Studi di Padova

Department of Chemical Sciences

Ph.D. COURSE IN: Molecular Sciences
CURRICULUM: Pharmaceutical Sciences
CYCLE XXXV

APPLICATION, EVALUATION, AND IMPROVEMENT OF COMPUTATIONAL METHODOLOGIES IN DRUG DISCOVERY

Coordinator: Ch.mo Prof. Leonard Jan Prins
Supervisor: Ch.mo Prof. Stefano Moro

Ph.D Student: Davide Bassani

ABSTRACT

The application of computational approaches in drug discovery has been consolidated in the last decades. These families of techniques are usually grouped under the common name of "Computer-Aided Drug Design" (CADD), and they now constitute one of the pillars in the drug discovery pipelines in many academic and industrial environments. Their implementation has been demonstrated to tremendously improve the speed of the early discovery steps, allowing to proficiently and rationally choose proper compounds for a desired therapeutic need among the extreme vastity of the drug-like chemical space.

Moreover, the application of CADD approaches allows the rationalization of biochemical and interactive processes of pharmaceutical interest at the molecular level. As a result of this, computational tools are now extensively used also in the field of rational 3D design and optimization of chemical entities starting from the structural information of the targets, which can be experimentally resolved or can be also obtained with other computer-based techniques.

In this Ph.D. work, we extensively applied state-of-the-art Computer-Aided Drug Design methods in different scenarios of pharmaceutical and biological interest, highlighting their great potential and their benefits, but also discussing their actual limitations and eventual weaknesses. This work combines the practical implementation of computational approaches with the development of their methods, about which we focused mainly on the field of novel filtration and scoring methods for candidate selection.

Summary

INTRODUCTION

1. The benefits of computational methods for drug discovery.

1.1. The drug discovery pipeline and the problem of candidate selection.

The drug discovery process is an extremely money- and time-consuming procedure which is necessary to guarantee the safety and the quality of novel therapeutical entities entering the market. It has been reported that a single novel small molecule can require up to 14 years and more than one billion dollars in the several steps which bring from the target assessment up to the regulatory approval[1][2]. Moreover, the failure risk in the pharmaceutical scenario is known to be one of the highest in the industry. Indeed, it has been estimated that just one or two out of 10000 screened molecules are able to effectively become drugs[3].

Another major challenge in the medicinal field is the tremendously extended chemical space forming the "drug-like" environment. It has been calculated that the number of small molecules included in such a concept would be roughly 10^{60}[4], which is a number higher than the seconds of life of the entire Universe. Such a vast chemical space is totally unfeasible to explore, and this is even more true from an experimental perspective. Indeed, even if the most technological High-Throughput Screening (HTS) methods nowadays are able to evaluate the on-target activity of hundreds of thousands of compounds/week[5], their capacity would never reach the order of magnitude of the potential candidates for that specific biological entity. This led medicinal chemists, at the end of the 20^{th} century, to elaborate smarter methodologies for candidate selection. Specifically, one of the first ideas that came up was to exploit computers to perform "virtual" screenings prior to the experimental ones. This approach, which was called "High-Throughput Virtual Screening" (HTVS) represents still one of the main applications of computational methodologies in drug discovery[6]. Indeed, the capacity of the virtual screening depends essentially on the computer power of the infrastructure exploited for the purpose, and it is obviously much faster and cheaper than the preparation and execution of experimental assays. As a demonstration of this, with the actual combinations of software and hardware, the evaluation of billions of compounds/day is achievable. In the latest decades, many academic and industrial groups extensively put efforts into the improvement of these methods, making them now one of the pillars in the drug discovery pipeline, especially in the early discovery phases.

1.2. The application of computational methods in drug discovery.

To go a little bit more into detail, the drug discovery process can be divided into five main steps[7]. The first is the "discovery and development" phase, which includes hit identification, hit-to-lead

(H2L), and lead optimization. The first of these consists in highlighting some molecular candidates with a good activity profile against the desired target, but presenting pharmacokinetic (PK) or pharmacodynamic (PD) limitations. Talking in terms of on-target potency, the hit compounds are usually in the micromolar (μM) range of activity, and they are tendentially not so selective. Even with all these pharmaceutical problems, the hit compounds are very important to give some hints to the drug design teams, being very useful starting points for further modification[8][9].

The second passage consists of the hit-to-lead optimization phase. In this step, the hit compounds are modified with different methodologies in order to improve their on-target activity and selectivity, always keeping their PK/PD profile under strict evaluation[10]. After this process, the optimized molecular candidates take the name of "lead" compounds, and are usually very active (in the nanomolar range for what concerns potency) and reasonably selective. These compounds then enter the second main step, which is the pre-clinical experimental phase, and are so tested in animal and organoid models to assess their safety and efficacy. The third phase, which is also the longest, consists of human clinical trials. These are divided into three main sections (I, II, and III), each with a different endpoint and with increasing patients enrolled in the tests. Just after a positive outcome of the therapy with the new candidate in the clinical phase III, the commercialization of the drug can be asked the regulatory agencies (e.g., EMEA for Europe, FDA for the USA), determining the opening of the fourth phase of drug development. After this, the fifth and last step consists of post-market drug safety monitoring.

Even if the pre-clinical and clinical trials constitute the longest and most expensive part of the drug discovery pipeline, not so much can be done to reduce them, mainly because of the extremely delicate outcomes in terms of safety and efficacy that they provide. On the other side, the steps which lead from the hits up to the lead compounds can be extensively optimized, and that is the space in which computational design approaches really found their main application. Indeed, other than exponentially improving the number of virtual compounds that can be evaluated on a daily basis, these methodologies offer also the possibility to deeply analyse the patterns in the chemical data under evaluation, and on top of that, they make the rational design of such entities much more than achievable[11]. Indeed, the advances in spectroscopic techniques, together with the tremendous improvement in computer graphics, allowed to bring the visual inspection of proteins, ligands, and biologically-relevant complexes in the routines of drug design groups[12]. With such computational approaches, is so possible to effectively design new molecular candidates, and for

this reason, the techniques of this family have been gathered under the common name of "Computer-Aided Drug Design" (CADD).

1.3. The main methodology branches in CADD.

The ways in which a computational chemist approaches a pharmacological problem can be multiple, but the main underlying factor discriminating the procedure is the quantity of data available on the topic examined. Indeed, one of the most important determinants is the presence of experimental structural information on the target of interest[11], which can be obtained nowadays with various techniques, among which the most relevant are for sure Nuclear Magnetic Resonance (NMR), X-Ray Crystallography (XR), and Cryogenic Electron Microscopy (cryo-EM)[13]. If such data are available, then the approach chosen by the scientist addresses usually the application of an ensemble of computational methods which take advantages of this, such as molecular docking and molecular dynamics. Because of this, those techniques fall under the family of the "Structure-Based Drug Design" (SBDD) approaches. On the other hand, when no experimental information about the target three-dimensional structure is available, then two main possibilities are open to the CADD scientist. The first consists of searching for close homologs of the target of interest with the purpose of creating a computational model of it (also known as the *homology model*), which would then be tested for structural reliability and used with SBDD techniques[14]. The other main option is to exploit just the information coming from the ligands which have been tested on the target, extrapolating from them enough information to build reliable quantitative structure-activity relation (QSAR) models[15]. These approaches were among the first used in rational drug design and now fall into the category of the "Ligand-Based Drug Design" (LBDD) techniques. This family comprises methods such as pharmacophore search[16] and matched molecular pair analysis[17], and even if their application in the years has lowered, giving more and more space to the SBDD techniques, they are still widely used in several scenarios. Finally, the latest advances in computer science, together with the exponential increase of the application of machine learning and artificial intelligence approaches, gave another powerful instrument in the hands of the CADD scientists[18]. Indeed, when huge amounts of data are available for a defined context (which can both regard the target and the ligands), these approaches can be proficiently used for the proper prediction of molecular properties of pharmaceutical relevance. Moreover, the creation and maintenance of an "intelligent" algorithm has brought in the most recent years to let such "computational brains" create novel chemical structures, with an approach known as *de novo drug design*[19].

2. Ligand-Based Drug Design (LBDD).

By far the most used approaches in the dawn of rational drug design, these techniques rely just on structural information of the molecular structures tested on the desired target. The main goal of such methods is to identify patterns in the data that can be extrapolated in order to guide the further steps to take in terms of drug discovery. Those patterns are usually identified as "Quantitative Structure-Activity Relationship" (QSAR) models, and should allow us to obtain a discrete and quantitative correlation between chemical moieties and pharmacological outcomes. Some of the techniques that are mostly used for the Ligand-based determination of these interconnections are cheminformatics[20] (e.g., matched molecular pair analysis), Ligand-based pharmacophore search, and Free-Wilson Analysis[21]. Some other very famous equations for QSAR modeling were advanced by Hansch, Hammet, and Taft[22]. Even if these instruments are still widely used, their main disadvantage remains related to the low level of generalization that they can provide. Indeed, they tend to work just on high congeneric series of ligands, and in some cases, they require great amounts of experimental data to provide reliable results. Moreover, these approaches don't take into consideration the conformational freedom of the ligands, focusing just on the 2D representation of the molecules considered. Together with the rising importance of structure-based approaches, and of "three-dimensionality" in general, also the conformational properties of the ligands have been taken into strong consideration by the LBDD methods. One of the main examples is the generation of "3D pharmacophores", which take into account both the atomic and conformational features of the molecules in order to build proper "3D-QSAR" models[23].

3. Structure-Based Drug Design (SBDD).

With the exponential increase in the availability of three-dimensional structures of proteins and nucleic acids, which started roughly in the 2000s years, the trend in the methodologies in computational drug design moved towards other techniques which could take into account also the three-dimensional interactive features of the molecules in respect to the target. Indeed, the prior knowledge of the biological target conferred a huge advantage to the scientists, that could develop novel chemical species based on its binding site characteristics. All the methods based on this kind of data fall into the family of "Structure-Based Drug Design" (SBDD)[24], and they are by far the most used approaches in computational drug design nowadays. Moreover, while twenty years ago

some complex membrane protein structures were not considered feasible to determine experimentally, the modern technology of cryo-EM allowed the reliable resolution of some of those complex systems[13], extending further the applicability domain of SBDD.

3.1. Molecular Docking

Maybe the most exploited technique in modern computer-aided drug design, molecular docking, consists of the determination of the best conformation with which a molecule binds to another to form a stable complex. The name of the technique comes from the very first program which exploited it, which was called “DOCK”, and was proposed by Kuntz in 1982[25]. In a pharmaceutical context, these methods are extensively used in order to screen millions or billions of small molecules against a biological target of interest (e.g., a protein or a nucleic acid). It is important to mention that molecular docking requires prior knowledge of the binding site location on the target. A molecular docking algorithm consists of two main parts, the conformational search algorithm, and the scoring function[26]. The first has the purpose to search the conformational space of the ligand considered to find a state which fits within the binding site, while the second ranks the different conformations in order to prioritize the most reliable. Scoring functions operate based on equations that take into account the conformational strain, the electrostatics, and the steric hindrance of the ligand in its bound state. Three different main types of scoring functions are available today, force field-based scoring functions, empirical scoring functions, and knowledge-based scoring functions. In the first of these kinds, the energy of the system is evaluated using a force field. On the other hand, the empirical scoring functions consist of different terms representing different intermolecular interactions, and each term is modeled by using experimental values for the interaction related to it. The top-ranked poses by these functions are the ones that are most close to the experimental values of reference. Lastly, knowledge-based scoring functions rely on statistical analyses of the most observed interactions between a specific ligand’s atom type and a particular protein’s atom type. The top-ranked poses are the ones more similar to what is statistically retrievable in the high-quality X-Ray databases[27].

Hundreds of different molecular docking algorithms exist nowadays, each coupling different search algorithms and scoring functions. Even if they are different, they can be grouped into different families based on the way in which they operate to find the “bound” state of the ligand. Some famous families are constituted by genetic algorithms (among which the program GOLD is one of the most known and used[28]), systematic algorithms (such as the program Glide, developed and

distributed by the company Schrödinger[29]), and Ant-Colony Optimization algorithms (like the program PLANTS[30], extensively used in this Ph.D. work together with the two already mentioned software)[31]. Another classification for molecular docking is related to the degrees of freedom considered in the calculation. Indeed, in "rigid docking", both the ligand and the protein are kept rigid. The "flexible-ligand docking" approach, on the other hand, allows the ligand to explore different conformational states, keeping the target rigid[32]. Then, the "semi-flexible" or "Induced-Fit" approach consists in taking into account both the ligand's and the binding site residues' conformational spaces[33], avoiding that small clashes with a rigid side chain could impair the selection of reasonable docking poses. In the last method, which is the "ensemble" docking, the molecular docking is executed against an ensemble of protein conformations, usually coming from molecular dynamics simulations. In this way, the full flexibility of the protein can be indirectly taken into account[34]. Of course, the computational power required for such approaches increases with the increase of the complexity which has to be considered.

As already mentioned, molecular docking is extensively used in the early phases of drug discovery, from hit identification up to lead optimization. It is applied both to identify novel chemical candidates for pharmacological testing and to help rationalize experimental data to a molecular level. Nevertheless, molecular docking is essentially a "static" approach, which considers just the final state of the ligand-target system, and is performed mainly in vacuum. Indeed, even if water molecules are included in the docking calculations, this passage has to be done explicitly[35], and this requires structural information coupled with molecular dynamics data. As a result of all of this, the main problem of molecular docking is the high false-positive ratio of the compounds selected by the scoring functions[36], which induced CADD scientists to investigate ways to filter the poses produced by molecular docking algorithms using other approaches. Nowadays, many of these methods are available. One of the simplest is called "consensus docking", and relies on the principle that, by exploiting different docking programs based on different algorithms, the molecules prioritized by all of them will have a lower probability of being false positives[37][38]. The success of such an approach has been extensively demonstrated in the literature. Other methods, called "post-docking" techniques, further filter the poses produced by molecular docking based on certain molecular features[39]. One of the examples for the first case is the implementation of a three-dimensional pharmacophore, in which the most relevant 3D features for the interaction with the target are embedded. In this case, only the docking poses which are able to respect these boundaries are kept, discarding the others[40]. Regarding the second kind of technique, one

example is the implementation of molecular dynamics as a “post-docking” approach[41]. Basically, the poses in which the molecules are able to keep the interaction pattern with the protein for a longer simulation time are referred to as more “kinetically stable”, while the others are de-prioritized. A novel technique exploiting this approach, but also implementing a temperature increase with simulation time, has recently been developed by our lab and is known as “Thermal Titration Molecular Dynamics” (TTMD)[42], and we will discuss it in the following chapter.

3.2. Molecular Dynamics

Molecular Dynamics (MD) is a computational technique used to investigate the dynamic behavior of a chemical and/or biological entity with time. The method consists of the iterative resolution of Newton’s Equation of motion in order to continuously predict the atomic positions of the molecules considered with respect to the other during the simulation time[43]. Molecular Dynamics is used for various purposes in drug discovery, from the simple dynamic evaluation of a system to mechanistic understanding of a molecular process, but also as a well-known post-docking filtration system[44] (Figure 1). The main advantage of this method is that, oppositely to molecular docking, the system considered is free to move, and is so more “realistic” if we consider the environment in which the real biochemical processes will happen. Moreover, molecular dynamics can be executed using explicit solvent models (e.g., TIP3P[45]), in which the contribution of each single water molecule is taken into account. This way is possible to get a better understanding of the role of water molecules in target stabilization as well as in ligand-target recognition. The first and main drawback of molecular dynamics is for sure related to the computational power required for its implementation in the pipeline. Indeed, depending on the simulation time and on the dimensions of the system to evaluate, MD can require from tenth to thousands of time/molecule more than molecular docking, and for this reason, its application has to be limited to a lower number of compound usually (e.g., in the post-docking approaches). Other than this, MD simulations rely on molecular mechanics and force fields in order to extrapolate atomic velocities with time, and if this increases exponentially the speed of the simulations with respect to the quantum-based methods, it also carries several approximations that have to be known and taken into account by the CADD scientists[46]. First, the polarizability is not conceived in force-field-based MD simulations, every molecule of the system has fixed bond lengths and partial charges, and no bond can be created or broken (made an exception for QM/MM methods, which include a focused region calculated at the QM level). In the latest years, the continuous increase in computational power of the modern

hardware architecture is allowing quantum-mechanical calculations to be more and more affordable, and for this reason probably leading to a new era in computational drug discovery[47].

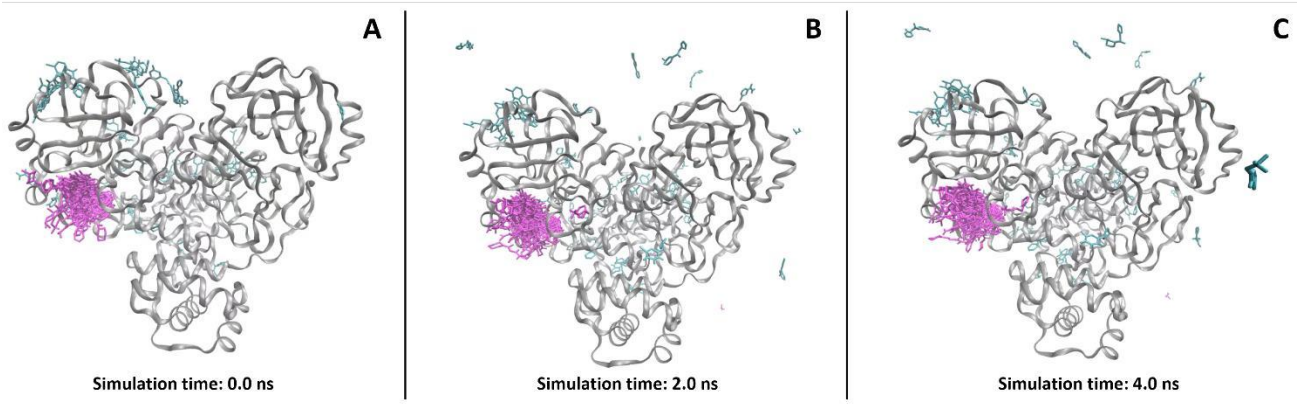


Figure 1. *Example of the evolution of a system (from A to C) using Molecular Dynamics simulations, taken from a recent study published by our lab*[48]. *As can be seen, each of the SARS-CoV-2 M^{pro} crystallographic ligands starts from a defined position (the crystallographic one) at the beginning of the simulation. Then, after the MD is started, the molecules outside the catalytic pocket (coloured in cyan), which are tendentially more exposed to the solvent, are more prone to lose the initial conformation and, eventually, the binding site itself. Oppositely, the compounds which are crystallized in the catalytic pocket (depicted in magenta) are more prone to keep their initial position during the simulation, assessing to bind stronger to the protein.*

3.2.1. Molecular Dynamics as a post-docking approach

As already mentioned, one of the main applications of Molecular Dynamics in the drug discovery pipeline consists of its exploitation in the discrimination of molecules after a molecular docking run. Indeed, while molecular docking gives just a static representation of the binding event, focusing on the final state only, MD is on the other side able to evaluate the dynamic stability of such conformation in the binding site. In MD-based post-docking, the poses resulting from the docking calculations are then used as the starting point for Molecular Dynamics, which samples diverse short simulations (usually very few nanoseconds long) for the complex considered[49]. The parameters usually tackled with this perspective are of a geometric kind, such as the root-mean-square deviation (RMSD) and the root-mean-square fluctuation (RMSF) of atomic positions. While the first describes how much a molecular entity (e.g., the ligand) displaces from its initial position during the simulation, the second quantifies the magnitude of the displacement from the most frequent position, indicating, indeed, the “fluctuation” of the entity itself. Even if such parameters are easy to calculate and compare, they often offer a poor description of the binding quality. Indeed, small changes in RMSD can bring the ligand to completely lose its interaction pattern during the simulation, while higher deviations in RMSD could be just due to some flexible moieties which are exposed to the solvent, and are so able to freely fluctuate in it. In order to overcome these

limitations, other metrics should be considered to evaluate MD-based post-docking replicas. One of the examples of this is the tackling of protein-ligand interaction fingerprints (PLIFs), which can be compared for all the MD frames, allowing to evaluate the effective quality of the interactions described by the molecular docking poses[50]. With this perspective, the molecule with the most proficient interaction patterns will keep the PLIFs during the MD simulations, while weaker-interacting compounds will tendentially lose the contacts which stabilize their bound conformation.

3.2.2. Thermal Titration Molecular Dynamics (TTMD)

One of the limitations of MD-based post-docking methods is related to the fact that, in many situations, some nanoseconds of simulation are not enough to discriminate potentially good from potentially weak binders[51]. Indeed, in many scenarios, both kinds of ligands keep low RMSD and RMSF values through the MD replicas, and their PLIFs are also equally maintained. To overcome this limitation, we recently proposed a new method, named “Thermal Titration Molecular Dynamics” (TTMD), which proficiently takes advantage of the “concept of temperature” in molecular mechanics and Molecular Dynamics with the goal of classifying ligands based on their on-target affinity[42]. Indeed, in classic MD, the temperature is simply a value used to scale the atomic velocities with time, and so is not related to the “real life” concept of temperature, which heavily influences all the biochemical processes. This allows MD simulations to be performed at temperatures above 350K or even 400K without seeing any unfolding event taking place, while in the experimental setup those values would totally compromise the experiment. Said this, the TTMD method consists, starting from a protein-ligand complex, in executing MD replicas in which, for every TTMD-step (T_i), the temperature of the system is increased to a certain defined amount of degrees. While this process takes place, the PLIFs are monitored, as well as the RMSD of the protein backbone (which can be used as a metric to effectively assesses the protein integrity during the simulations). The TTMD experiment may end in two different ways: when the PLIFs are totally lost, or after a user-defined simulation time. In its first implementation, this method was applied to 4 different case studies (CK1δ, CK2, PDK2, and SARS-CoV-2 M^{pro}), taking 5 different crystal structures each. The potencies of the ligands in such complexes were known experimentally. The TTMD experiment was set up with a temperature ramp from 300K to 450K, with an increase of 10K every 10ns simulation. In all the cases considered, TTMD was effectively able to clearly distinguish the potent nanomolar ligands from the weaker micro- to millimolar ones (Figure 2). The success of this simple approach opens the possibility of its application in MD-based post-docking approaches, given

its intrinsic advantage to overcome the eventual inability of classical MD to discriminate ligands in a reduced timescale.

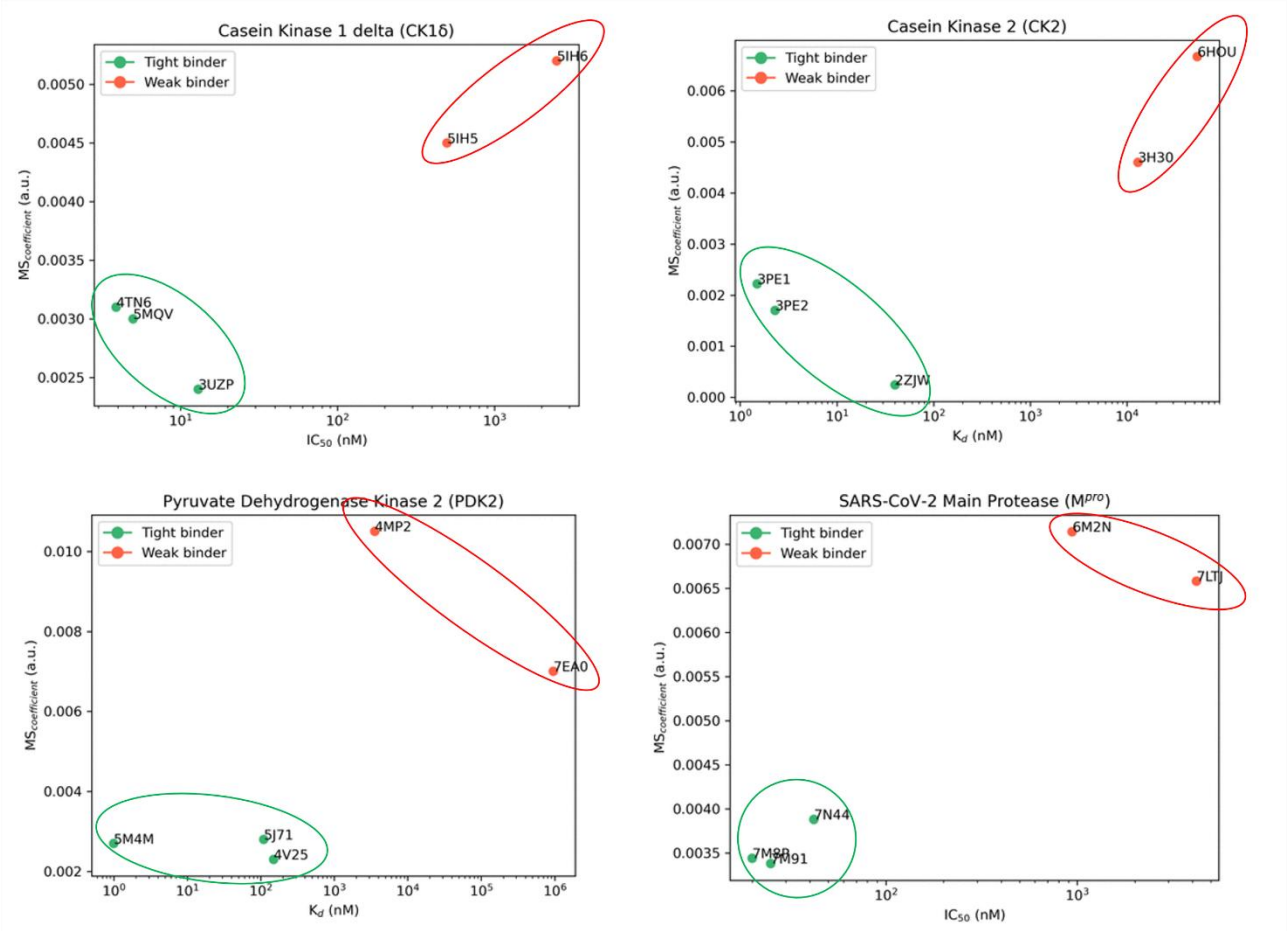


Figure 2. *The results of the first published application of Thermal Titration Molecular Dynamics (TTMD). As it can be seen, in all the case studies considered, the method was able to efficiently discriminate the nanomolar ligands (indicated with the green dots and highlighted with the green circles) from the micro- and millimolar ones (depicted with red dots and circled in red). The MS coefficient, which was used for the classification, depends on the ability of each molecule to preserve its Protein-Ligand Interaction Fingerprints (PLIFs) during the TTMD experiment (a more detailed and mathematical explanation of its derivation is reported in the publication).*

4. Supervised Molecular Dynamics

Supervised Molecular Dynamics (SuMD) consists of an MD approach that is designed to describe a "rare" molecular binding event in a reduced timescale[52]. In the specific case of SuMD, the event considered is the target-ligand recognition process, which would require simulations in the timescales of microseconds in order to be described by a classical MD simulation. This is due to the huge amount of time that the ligand would pass simply fluctuating in the free solvent, not taking any contact with the target and, even less, with the binding site. Indeed, classical MD is referred to as a "low-sampling" approach, because the force fields are able to sample just very partially the potential energy surface of the system. Usually, in order to overcome these limitations, the

computational election approaches are the Markov State Models and the Enhanced Sampling techniques. In the first case, the MD simulation is considered an ensemble of microstates, which are independent of one another. The algorithm then calculates the transition probability matrix, which allows for deriving the probability for the system to pass from one microstate to another[53]. On the other hand, enhanced sampling methods consist of the perturbation of the potential energy surface of the system, allowing the escape from local minima[54].
SuMD, on the other hand, is a technique in which no perturbation of the potential energy surface takes place. It is based on a supervision algorithm that evaluates the distance between the ligand and its binding site on the target in an iterative fashion. Specifically, the ligand is placed at a distance from the binding site in which there is no possibility of interaction at the beginning of the simulation (at least 40 Å away), and then a series of small MD simulations are sampled. At the end of each one of those, the distance between the ligand and the binding site is calculated, and just if the value is lower with respect to the distance at the beginning of the small MD simulation, the final coordinates are kept and another MD is started from those. Otherwise, the initial coordinates are restored, and the MD simulation is repeated. When the ligand reaches a defined distance threshold (5 Å, by default), the supervision is shut down and the simulation continues as a classical MD, allowing the ligand to relax into the binding site. In the end, the final "SuMD trajectory" is obtained by merging all the small MD simulations[55]. The great advantage of the method is that is able to describe an event such as the target-ligand recognition process in the nanoseconds timescale other than microseconds (typical of classical MD), and moreover, this is accomplished without the introduction of energetic biases. This technique has already been extensively used on different targets, such as G Protein-Coupled Receptors (GPCRs)[56][57] (see Figure 3), proteases[58], and kinases[59], considering as ligands small molecules, peptides[60], and aptamers[61]. The SuMD analysis allows to get a visual representation of the most relevant residues for the interaction with the target at each step of the simulation, providing very relevant information with respect to how the approaching process influences the binding, and also regarding the eventual "meta-binding sites" present on the target.

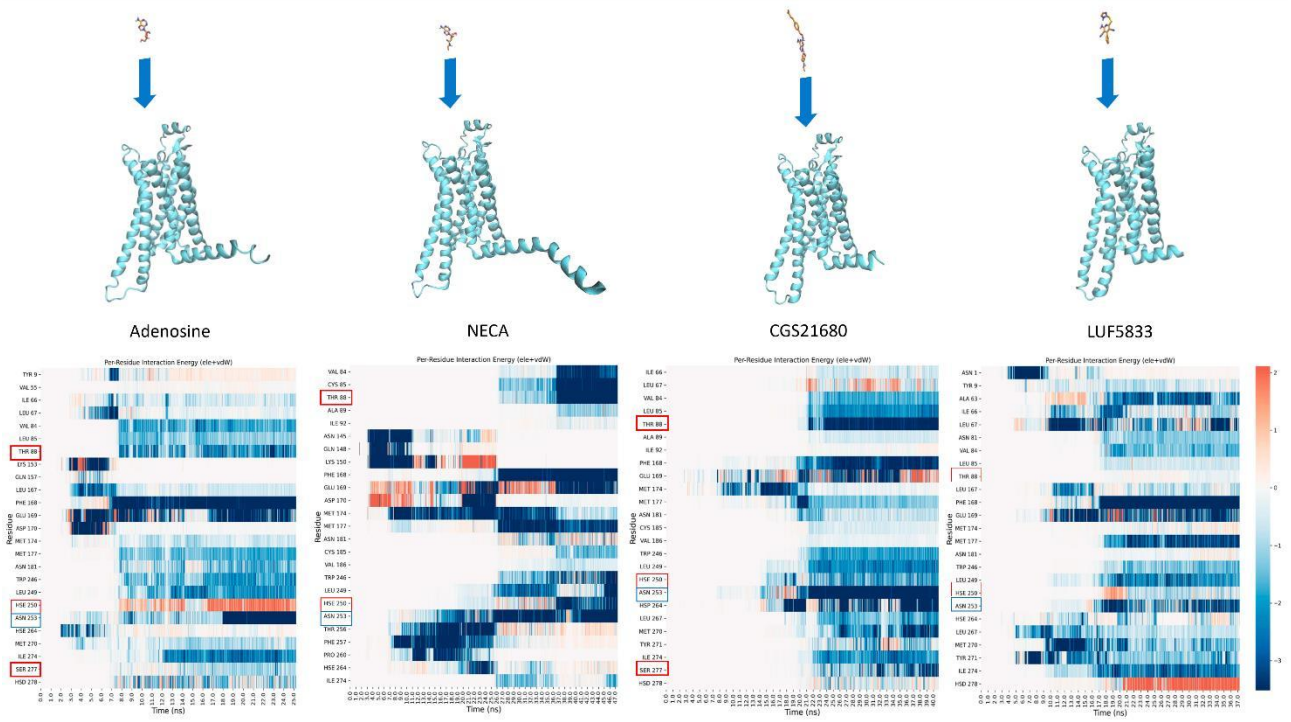


Figure 3. *Results of the application of Supervised Molecular Dynamics (SuMD) to the elucidation of the differences in the ligand binding paths between full agonists (Adenosine, NECA, and CGS21680) and a non-ribosidic partial agonist (LUF5833) of adenosine A_{2A} receptor. The panels on the upper part of the figure represent the initial configuration of each SuMD simulation, with the ligand placed away from the orthosteric binding site. The plots in the lower part depict the outcomes of the time-based per-residue interaction analysis, in which the summation of electrostatic and van der Waals contributions for each of the 25 most contacted protein residues is reported for each frame of the trajectories produced. In this case study, SuMD highlighted the main differences between the ligand-protein recognition paths of full agonists and LUF5833, which is able to exert a partial agonism in respect to $A_{2A}R$ even if important residues such as Thr88 and Ser277 are not recruited directly by this specific compound. An explanation of the molecular reasons behind this behaviour is reported in the publication*[62].

References

[1] L. Martin, M. Hutchens, and C. Hawkins, “Clinical trial cycle times continue to increase despite industry efforts,” *Nat. Rev. Drug Discov.*, vol. 16, no. 3, pp. 157–157, Mar. 2017, doi: 10.1038/nrd.2017.21.

[2] S. Simoens and I. Huys, “R&D Costs of New Medicines: A Landscape Analysis,” *Front. Med.*, vol. 8, Oct. 2021, doi: 10.3389/fmed.2021.760762.

[3] “EFPIA The Pharmaceutical Industry in Figures Key Data 2021.” https://www.efpia.eu/media/602709/the-pharmaceutical-industry-in-figures-2021.pdf.

[4] R. S. Bohacek, C. McMartin, and W. C. Guida, “The art and practice of structure-based drug design: A molecular modeling perspective,” *Med. Res. Rev.*, vol. 16, no. 1, pp. 3–50, Jan. 1996, doi: 10.1002/(SICI)1098-1128(199601)16:1<3::AID-MED1>3.0.CO;2-6.

[5] P. Szymański, M. Markowicz, and E. Mikiciuk-Olasik, “Adaptation of High-Throughput Screening in Drug Discovery—Toxicological Screening Tests,” *Int. J. Mol. Sci.*, vol. 13, no. 1, pp. 427–452, Dec. 2011, doi: 10.3390/ijms13010427.

[6] S. Moro, M. Bacilieri, and F. Deflorian, “Combining ligand-based and structure-based drug design in the virtual screening arena,” *Expert Opin. Drug Discov.*, vol. 2, no. 1, pp. 37–49, Jan. 2007, doi: 10.1517/17460441.2.1.37.

[7] “The Drug Development Process.” https://www.fda.gov/patients/learn-about-drug-and-device-approvals/drug-development-process.

[8] T. Zhu *et al.*, “Hit Identification and Optimization in Virtual Screening: Practical Recommendations Based on a Critical Literature Analysis,” *J. Med. Chem.*, vol. 56, no. 17, pp. 6560–6572, Sep. 2013, doi: 10.1021/jm301916b.

[9] J. Hughes, S. Rees, S. Kalindjian, and K. Philpott, “Principles of early drug discovery,” *Br. J. Pharmacol.*, vol. 162, no. 6, pp. 1239–1249, Mar. 2011, doi: 10.1111/j.1476-5381.2010.01127.x.

[10] G. M. Keserű and G. M. Makara, “Hit discovery and hit-to-lead approaches,” *Drug Discov. Today*, vol. 11, no. 15–16, pp. 741–748, Aug. 2006, doi: 10.1016/j.drudis.2006.06.016.

[11] S. P. Leelananda and S. Lindert, “Computational methods in drug discovery,” *Beilstein J. Org. Chem.*, vol. 12, pp. 2694–2718, Dec. 2016, doi: 10.3762/bjoc.12.267.

[12] Z. Dauter and A. Wlodawer, “Progress in protein crystallography,” *Protein Pept. Lett.*, vol. 23, no. 3, pp. 201–210, Feb. 2016, doi: 10.2174/0929866523666160106153524.

[13] X. Benjin and L. Ling, “Developments, applications, and prospects of cryo-electron microscopy,” *Protein Sci.*, vol. 29, no. 4, pp. 872–882, Apr. 2020, doi: 10.1002/pro.3805.

[14] C. N. Cavasotto and S. S. Phatak, “Homology modeling in drug discovery: current trends and applications,” *Drug Discov. Today*, vol. 14, no. 13–14, Jul. 2009, doi: 10.1016/j.drudis.2009.04.006.

[15] A. Cherkasov *et al.*, “QSAR Modeling: Where Have You Been? Where Are You Going To?,” *J. Med. Chem.*, vol. 57, no. 12, pp. 4977–5010, Jun. 2014, doi: 10.1021/jm4004285.

[16] S.-Y. Yang, “Pharmacophore modeling and applications in drug discovery: challenges and recent advances,” *Drug Discov. Today*, vol. 15, no. 11–12, pp. 444–450, Jun. 2010, doi: 10.1016/j.drudis.2010.03.013.

[17] C. Kramer, J. E. Fuchs, S. Whitebread, P. Gedeck, and K. R. Liedl, “Matched Molecular Pair Analysis: Significance and the Impact of Experimental Uncertainty,” *J. Med. Chem.*, vol. 57, no. 9, pp. 3786–3802, May 2014, doi: 10.1021/jm500317a.

[18] F. Zhong *et al.*, “Artificial intelligence in drug design,” *Sci. China Life Sci.*, vol. 61, no. 10, pp. 1191–1204, Oct. 2018, doi: 10.1007/s11427-018-9342-2.

[19] V. D. Mouchlis *et al.*, “Advances in De Novo Drug Design: From Conventional to Machine Learning Methods,” *Int. J. Mol. Sci.*, vol. 22, no. 4, p. 1676, Feb. 2021, doi: 10.3390/ijms22041676.

[20] B. F. Begam and J. S. Kumar, “A Study on Cheminformatics and its Applications on Modern Drug Discovery,” *Procedia Eng.*, vol. 38, pp. 1264–1275, 2012, doi: 10.1016/j.proeng.2012.06.156.

[21] H. Kubinyi, “Free Wilson Analysis. Theory, Applications and its Relationship to Hansch Analysis,” *Quant. Struct. Relationships*, vol. 7, no. 3, pp. 121–133, 1988, doi: 10.1002/qsar.19880070303.

[22] O. Silakari and P. K. Singh, “QSAR: Descriptor calculations, model generation, validation and their application,” in *Concepts and Experimental Protocols of Modelling and Informatics in Drug Design*, Elsevier, 2021, pp. 29–63.

[23] J. Verma, V. Khedkar, and E. Coutinho, “3D-QSAR in Drug Design - A Review,” *Curr. Top. Med. Chem.*, vol. 10, no. 1, pp. 95–115, Jan. 2010, doi: 10.2174/156802610790232260.

[24] A. C. Anderson, “The Process of Structure-Based Drug Design,” *Chem. Biol.*, vol. 10, no. 9, pp. 787–797, Sep. 2003, doi: 10.1016/j.chembiol.2003.09.002.

[25] I. D. Kuntz, J. M. Blaney, S. J. Oatley, R. Langridge, and T. E. Ferrin, “A geometric approach to macromolecule-ligand interactions,” *J. Mol. Biol.*, vol. 161, no. 2, pp. 269–288, Oct. 1982, doi: 10.1016/0022-2836(82)90153-X.

[26] X.-Y. Meng, H.-X. Zhang, M. Mezei, and M. Cui, “Molecular Docking: A Powerful Approach for Structure-Based Drug Discovery,” *Curr. Comput. Aided-Drug Des.*, vol. 7, no. 2, pp. 146–157, Jun. 2011, doi: 10.2174/157340911795677602.

[27] J. Li, A. Fu, and L. Zhang, “An Overview of Scoring Functions Used for Protein–Ligand Interactions in Molecular Docking,” *Interdiscip. Sci. Comput. Life Sci.*, vol. 11, no. 2, pp. 320–328, Jun. 2019, doi: 10.1007/s12539-019-00327-w.

[28] G. Jones, P. Willett, R. C. Glen, A. R. Leach, and R. Taylor, “Development and validation of a genetic algorithm for flexible docking 1 1Edited by F. E. Cohen,” *J. Mol. Biol.*, vol. 267, no. 3, Apr. 1997, doi: 10.1006/jmbi.1996.0897.

[29] R. A. Friesner *et al.*, “Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy,” *J. Med. Chem.*, vol. 47, no. 7, Mar. 2004, doi: 10.1021/jm0306430.

[30] O. Korb, T. Stützle, and T. E. Exner, “PLANTS: Application of Ant Colony Optimization to Structure-Based Drug Design,” 2006.

[31] P. H. M. Torres, A. C. R. Sodero, P. Jofily, and F. P. Silva-Jr, “Key Topics in Molecular Docking for Drug Design,” *Int. J. Mol. Sci.*, vol. 20, no. 18, p. 4574, Sep. 2019, doi: 10.3390/ijms20184574.

[32] S.-Y. Huang, “Comprehensive assessment of flexible-ligand docking algorithms: current effectiveness and challenges,” *Brief. Bioinform.*, vol. 19, no. 5, pp. 982–994, Sep. 2018, doi: 10.1093/bib/bbx030.

[33] C. A. Sotriffer, “Accounting for Induced-Fit Effects in Docking: What is Possible and What is Not?,” *Curr. Top. Med. Chem.*, vol. 11, no. 2, pp. 179–191, Jan. 2011, doi: 10.2174/156802611794863544.

[34] R. E. Amaro *et al.*, “Ensemble Docking in Drug Discovery,” *Biophys. J.*, vol. 114, no. 10, pp. 2271–2278, May 2018, doi: 10.1016/j.bpj.2018.02.038.

[35] B. C. Roberts and R. L. Mancera, “Ligand−Protein Docking with Water Molecules,” *J. Chem. Inf. Model.*, vol. 48, no. 2, pp. 397–408, Feb. 2008, doi: 10.1021/ci700285e.

[36] N. Deng *et al.*, “Distinguishing Binders from False Positives by Free Energy Calculations: Fragment Screening Against the Flap Site of HIV Protease,” *J. Phys. Chem. B*, vol. 119, no. 3, pp. 976–988, Jan. 2015, doi: 10.1021/jp506376z.

[37] G. Poli and T. Tuccinardi, “Consensus Docking in Drug Discovery,” *Curr. Bioact. Compd.*, vol. 16, no. 3, pp. 182–190, Jun. 2020, doi: 10.2174/1573407214666181023114820.

[38] D. R. Houston and M. D. Walkinshaw, “Consensus Docking: Improving the Reliability of Docking in a Virtual Screening Context,” *J. Chem. Inf. Model.*, vol. 53, no. 2, pp. 384–390, Feb. 2013, doi: 10.1021/ci300399w.

[39] G. Rastelli and L. Pinzi, “Refinement and Rescoring of Virtual Screening Results,” *Front. Chem.*, vol. 7, Jul. 2019, doi: 10.3389/fchem.2019.00498.

[40] M. L. Peach and M. C. Nicklaus, “Combining docking with pharmacophore filtering for improved virtual screening,” *J. Cheminform.*, vol. 1, no. 1, p. 6, Dec. 2009, doi: 10.1186/1758-2946-1-6.

[41] H. Alonso, A. A. Bliznyuk, and J. E. Gready, “Combining docking and molecular dynamic simulations in drug design,” *Medicinal Research Reviews*. 2006, doi: 10.1002/med.20067.

[42] M. Pavan, S. Menin, D. Bassani, M. Sturlese, and S. Moro, “Qualitative Estimation of Protein–Ligand Complex Stability through Thermal Titration Molecular Dynamics Simulations,” *J. Chem. Inf. Model.*, Oct. 2022, doi: 10.1021/acs.jcim.2c00995.

[43] S. A. Hollingsworth and R. O. Dror, “Molecular Dynamics Simulation for All,” *Neuron*, vol. 99, no. 6, Sep. 2018, doi: 10.1016/j.neuron.2018.08.011.

[44] M. De Vivo, M. Masetti, G. Bottegoni, and A. Cavalli, “Role of Molecular Dynamics and Related Methods in Drug Discovery,” *J. Med. Chem.*, vol. 59, no. 9, May 2016, doi: 10.1021/acs.jmedchem.5b01684.

[45] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, “Comparison of simple potential functions for simulating liquid water,” *J. Chem. Phys.*, vol. 79, no. 2, pp. 926–935, Jul. 1983, doi: 10.1063/1.445869.

[46] J. D. Durrant and J. A. McCammon, “Molecular dynamics simulations and drug discovery,” *BMC Biol.*, vol. 9, no. 1, Dec. 2011, doi: 10.1186/1741-7007-9-71.

[47] C. Gorgulla, A. Jayaraj, K. Fackeldey, and H. Arthanari, “Emerging frontiers in virtual drug discovery: From quantum mechanical methods to deep learning approaches,” *Curr. Opin. Chem. Biol.*, vol. 69, p. 102156, Aug. 2022, doi: 10.1016/j.cbpa.2022.102156.

[48] D. Bassani, M. Pavan, G. Bolcato, M. Sturlese, and S. Moro, “Re-Exploring the Ability of Common Docking Programs to Correctly Reproduce the Binding Modes of Non-Covalent Inhibitors of SARS-CoV-2 Protease Mpro,” *Pharmaceuticals*, vol. 15, no. 2, p. 180, Jan. 2022, doi: 10.3390/ph15020180.

[49] H. Alonso, A. A. Bliznyuk, and J. E. Gready, “Combining docking and molecular dynamic simulations in drug design,” *Med. Res. Rev.*, vol. 26, no. 5, pp. 531–568, Sep. 2006, doi: 10.1002/med.20067.

[50] M. Pavan, S. Menin, D. Bassani, M. Sturlese, and S. Moro, “Implementing a Scoring Function Based on Interaction Fingerprint for Autogrow4: Protein Kinase CK1δ as a Case Study,” *Front. Mol. Biosci.*, vol. 9, Jul. 2022, doi: 10.3389/fmolb.2022.909499.

[51] A. C. Pan, H. Xu, T. Palpant, and D. E. Shaw, “Quantitative Characterization of the Binding and Unbinding of Millimolar Drug Fragments with Molecular Dynamics Simulations,” *J. Chem. Theory Comput.*, vol. 13, no. 7, pp. 3372–3377, Jul. 2017, doi: 10.1021/acs.jctc.7b00172.

[52] D. Sabbadin and S. Moro, “Supervised Molecular Dynamics (SuMD) as a Helpful Tool To Depict GPCR–Ligand Recognition Pathway in a Nanosecond Time Scale,” *J. Chem. Inf. Model.*, vol. 54, no. 2, pp. 372–376, Feb. 2014, doi: 10.1021/ci400766b.

[53] B. E. Husic and V. S. Pande, “Markov State Models: From an Art to a Science,” *J. Am. Chem. Soc.*, vol. 140, no. 7, pp. 2386–2396, Feb. 2018, doi: 10.1021/jacs.7b12191.

[54] R. C. Bernardi, M. C. R. Melo, and K. Schulten, “Enhanced sampling techniques in molecular dynamics simulations of biological systems,” *Biochim. Biophys. Acta - Gen. Subj.*, vol. 1850, no. 5, pp. 872–877, May 2015, doi: 10.1016/j.bbagen.2014.10.019.

[55] A. Cuzzolin *et al.*, “Deciphering the Complexity of Ligand-Protein Recognition Pathways Using Supervised Molecular Dynamics (SuMD) Simulations,” *J. Chem. Inf. Model.*, vol. 56, no. 4, pp. 687–705, 2016, doi: 10.1021/acs.jcim.5b00702.

[56] D. Sabbadin, A. Ciancetta, G. Deganutti, A. Cuzzolin, and S. Moro, “Exploring the recognition pathway at the

human A 2A adenosine receptor of the endogenous agonist adenosine using supervised molecular dynamics simulations," *Medchemcomm*, vol. 6, no. 6, pp. 1081–1085, 2015, doi: 10.1039/C5MD00016E.

[57] G. Bolcato, M. Pavan, D. Bassani, M. Sturlese, and S. Moro, "Ribose and Non-Ribose A2A Adenosine Receptor Agonists: Do They Share the Same Receptor Recognition Mechanism?," *Biomedicines*, vol. 10, no. 2, p. 515, Feb. 2022, doi: 10.3390/biomedicines10020515.

[58] M. Pavan, G. Bolcato, D. Bassani, M. Sturlese, and S. Moro, "Supervised Molecular Dynamics (SuMD) Insights into the mechanism of action of SARS-CoV-2 main protease inhibitor PF-07321332," *J. Enzyme Inhib. Med. Chem.*, vol. 36, no. 1, pp. 1645–1649, Jan. 2021, doi: 10.1080/14756366.2021.1954919.

[59] S. K. Panday, M. Sturlese, V. Salmaso, I. Ghosh, and S. Moro, "Coupling Supervised Molecular Dynamics (SuMD) with Entropy Estimations To Shine Light on the Stability of Multiple Binding Sites," *ACS Med. Chem. Lett.*, vol. 10, no. 4, pp. 444–449, Apr. 2019, doi: 10.1021/acsmedchemlett.8b00490.

[60] V. Salmaso, M. Sturlese, A. Cuzzolin, and S. Moro, "Exploring Protein-Peptide Recognition Pathways Using a Supervised Molecular Dynamics Approach," *Structure*, vol. 25, no. 4, pp. 655-662.e2, Apr. 2017, doi: 10.1016/j.str.2017.02.009.

[61] M. Pavan, D. Bassani, M. Sturlese, and S. Moro, "Investigating RNA–protein recognition mechanisms through supervised molecular dynamics (SuMD) simulations," *NAR Genomics Bioinforma.*, vol. 4, no. 4, Oct. 2022, doi: 10.1093/nargab/lqac088.

[62] G. Bolcato, M. Pavan, D. Bassani, M. Sturlese, and S. Moro, "Ribose and Non-Ribose A2A Adenosine Receptor Agonists: Do They Share the Same Receptor Recognition Mechanism?," *Biomedicines*, vol. 10, no. 2, p. 515, Feb. 2022, doi: 10.3390/biomedicines10020515.

SCIENTIFIC PUBLICATIONS

Overview of the Scientific publications

1. Molecular Docking

1.1. G. Bolcato et al., "A Computational Workflow for the Identification of Novel Fragments Acting as Inhibitors of the Activity of Protein Kinase CK1δ," Int. J. Mol. Sci., vol. 22, no. 18, p. 9741, Sep. 2021, doi: 10.3390/ijms22189741

In this work, we applied an in-house designed computational workflow for the identification of novel fragments binding to Casein Kinase 1δ. One of the main filtration steps was based on the application of consensus molecular docking, a technique whose purpose is to diminish the false positive rate given by docking approaches. It is worth noting that MD-based refinement was also applied afterwards.

1.2. D. Bassani, M. Pavan, G. Bolcato, M. Sturlese, and S. Moro, "Re-Exploring the Ability of Common Docking Programs to Correctly Reproduce the Binding Modes of Non-Covalent Inhibitors of SARS-CoV-2 Protease Mpro," Pharmaceuticals, vol. 15, no. 2, p. 180, Jan. 2022, doi: 10.3390/ph15020180.

In this methodological work, we analyzed the performance of molecular docking programs in reproducing the experimental conformations of the available non-covalent small molecule crystallographic ligands of SARS-CoV-2 main protease (M^{pro}). Our results show how the algorithms perform well when dealing with compounds characterized by a low solvent exposed binding mode, with a special focus on the molecules crystallized inside the catalytic pocket. Molecular Dynamics simulations further validated our analysis.

1.3. D. Bassani, M. Pavan, M. Sturlese, and S. Moro, "Sodium or Not Sodium: Should Its Presence Affect the Accuracy of Pose Prediction in Docking GPCR Antagonists?," Pharmaceuticals, vol. 15, no. 3, p. 346, Mar. 2022, doi: 10.3390/ph15030346.

This article has the purpose of analyzing the importance of the consideration of the presence of the negative allosteric sodium ion, always present in the transmembrane (7TM) region of inactive class-A GPCRs, when performing molecular docking-based virtual screenings. Our outcomes suggest that such a presence should be taken into account only when the position of the ion is

obtainable from experimental data, and even in that case the increase in the performance of the docking algorithms seems to be not so relevant.

1.4. Pavan, M., Menin, S., Bassani, D., Sturlese, M., & Moro, S. (2022). Implementing a Scoring Function Based on Interaction Fingerprint for Autogrow4: Protein Kinase CK1δ as a Case Study. Frontiers in Molecular Biosciences, 0, 629. https://doi.org/10.3389/FMOLB.2022.909499

This article describes the implementation of a novel scoring function into the genetic *de novo design* program Autogrow4, taking Casein Kinase 1δ as the case study. The novel method replaces the VINA scoring function with a ranking system based on pharmacophore filtration and protein-ligand interaction fingerprint (PLIFs) similarities with the crystallographic ligands. We have shown how the proposed methodology leads to compounds which have a lower molecular weight (a useful feature for the hit-to-lead phase) and higher shape and electrostatic similarities with the CK1δ ligands.

2. Molecular Dynamics

2.1. Pavan, M., Bolcato, G., Bassani, D., Sturlese, M. & Moro, S. Supervised Molecular Dynamics (SuMD) Insights into the mechanism of action of SARS-CoV-2 main protease inhibitor PF-07321332. J. Enzyme Inhib. Med. Chem. 36, 1646–1650 (2021).

In this article, we performed a prospective study for the analysis of the ligand-protein recognition path of PF-07321332 (now known as Nirmatrelvir) and SARS-CoV-2 main protease (M^{pro}) by exploiting Supervised Molecular Dynamics (SuMD). Our experiments led to a binding conformation fairly superimposable to the later published crystallographic pose of the ligand, validating the potential of SuMD for the description of molecular binding events.

2.2. Bissaro, M. et al. Inspecting the Mechanism of Fragment Hits Binding on SARS-CoV-2 M pro by Using Supervised Molecular Dynamics (SuMD) Simulations. ChemMedChem cmdc.202100156 (2021) doi:10.1002/cmdc.202100156.

In the present study, we analyzed the performance of the High-Throughput Supervised Molecular Dynamics (HT-SuMD) methodology in reproducing the crystallographic poses of the 23 early-published fragments bound to SARS-CoV-2 main protease (M^{pro}). The article outlines the strength

and limitations of the technique, and also purposing novel methods for the refinement of the results given by HT-SuMD.

2.3. Bolcato, G., Pavan, M., Bassani, D., Sturlese, M., & Moro, S. (2022). Ribose and Non-Ribose A2A Adenosine Receptor Agonists: Do They Share the Same Receptor Recognition Mechanism? Biomedicines 2022, Vol. 10, Page 515, 10(2), 515. https://doi.org/10.3390/BIOMEDICINES10020515

In this work, we exploited Supervised Molecular Dynamics (SuMD) simulations in order to describe the differences between the ligand-protein recognition paths of classical ribosidic full agonists of A_{2A} receptor (adenosine, CGS21680, and NECA) and one of its non-ribosidic partial agonists (LUF5833). Our analysis highlights the importance of the role of water molecules in defining the partial activation of the receptor by the non-ribosidic compound examined.

2.4. Pavan, M., Menin, S., Bassani, D., Sturlese, M., & Moro, S. (2022). Qualitative Estimation of Protein–Ligand Complex Stability through Thermal Titration Molecular Dynamics Simulations. Journal of Chemical Information and Modeling. https://doi.org/10.1021/ACS.JCIM.2C00995

In this article, we propose the very first application of Thermal Titration Molecular Dynamics (TTMD), a novel MD-based method for ligand activity classification based on MD residence time. In all the four case studies reported (CK1δ, CK2, PDK2, and SARS-CoV-2 M^{pro}), the method (which is based on progressive temperature increase with the simulation time) was able to distinguish the nanomolar to the micro- and millimolar ligands, showing its potential for candidate selection if implemented in the drug discovery pipeline.

2.5. M. Pavan, D. Bassani, M. Sturlese, and S. Moro, “Investigating RNA–protein recognition mechanisms through supervised molecular dynamics (SuMD) simulations,” NAR Genomics Bioinforma., vol. 4, no. 4, Oct. 2022, doi: 10.1093/nargab/lqac088.

In this article, we present the first application of Supervised Molecular Dynamics (SuMD) for the description of the aptamer-protein recognition process. In all the case studies considered, SuMD was able to lead to a reasonable solution for the binding, even considering the increased difficulty

due to the conformational freedom of this class of binders, which is higher with respect to the ligands on which SuMD was usually applied on before.

3. Application of CADD for the rationalization of phenomena of pharmaceutical relevance

3.1. G. Sartore, D. Bassani, E. Ragazzi, P. Traldi, A. Lapolla, and S. Moro, "In silico evaluation of the interaction between ACE2 and SARS-CoV-2 Spike protein in a hyperglycemic environment," Sci. Rep., vol. 11, no. 1, p. 22860, Dec. 2021, doi: 10.1038/s41598-021-02297-w.

In this work, we used molecular modeling to get an analysis of the interaction patterns of SARS-CoV-2 Spike protein and human ACE2 receptor in a simulated hyperglycaemic environment, typical of diabetic patients.

3.2. D. Bassani, E. Ragazzi, A. Lapolla, G. Sartore, and S. Moro, "Omicron Variant of SARS-CoV-2 Virus: In Silico Evaluation of the Possible Impact on People Affected by Diabetes Mellitus," Front. Endocrinol. (Lausanne)., vol. 13, Mar. 2022, doi: 10.3389/fendo.2022.847993.

Similarily to the previously cited article, in this work, we exploited molecular modeling in order to evince the molecular basis for the different outcomes of SARS-CoV-2 Omicron variant infection in diabetic patients. For this, the interaction pattern of Spike protein and human ACE2 receptor was analyzed, highlighting the differences between the wild-type and the Omicron SARS-CoV-2 variants.

3.3. Matteo Pavan, Davide Bassani, Mattia Sturlese & Stefano Moro (2022) Bat coronaviruses related to SARS-CoV-2: what about their 3CL proteases (MPro)?, Journal of Enzyme Inhibition and Medicinal Chemistry, 37:1, 1077-1082, DOI: 10.1080/14756366.2022.2062336

In this article, we focused our attention on the analysis of the M^{pro} proteins typical of novel coronaviruses recently discovered by Temmam et al. in Laos. Our molecular modeling operations highlighted how, even if some of these viral entities have shown an increased affinity for the human ACE2 receptor, their main protease is always highly conserved, with very few mutations in non-catalytically relevant locations. This allowed us to conclude that focusing on the development of novel protease inhibitors could be one of the best solutions for the treatment of novel viral variants

in the future.

3.4. Matteo Pavan, Davide Bassani, Mattia Sturlese & Stefano Moro (2022) From the Wuhan-Hu-1 strain to the XD and XE variants: is targeting the SARS-CoV-2 spike protein still a pharmaceutically relevant option against COVID-19?, Journal of Enzyme Inhibition and Medicinal Chemistry, 37:1, 1704-1714, DOI: 10.1080/14756366.2022.2081847

In the present work, we presented an analysis of the differences between the most relevant SARS-CoV-2 variants at the molecular level, focusing our attention on Spike protein and SARS-CoV-2 main protease (M^{pro}). The outcomes of the study show how, while Spike has mutated several times in the latest 18 months, heavily changing its receptor-binding domain (RBD), M^{pro} has just barely mutated, and always in locations not relevant for its catalytic activity. This work has the purpose to inform the scientific community, and encouraging the development of novel SARS-CoV-2 M^{pro} inhibitors as the main pharmaceutically relevant option for SARS-CoV-2 infection treatment.

3.5. E. Heilmann et al., "SARS-CoV-2 3CL pro mutations selected in a VSV-based system confer resistance to nirmatrelvir, ensitrelvir, and GC376," Sci. Transl. Med., Oct. 2022, doi: 10.1126/scitranslmed.abq7360.

In this article, we applied molecular modeling for the rationalization of the reasons behind the development of resistance to Nirmatrelvir by some ad-hoc created SARS-CoV-2 M^{pro} variants. The mutations were experimentally generated on an engineered chimeric vesicular stomatitis virus (VSV), and the resistance to Nirmatrelvir was assessed in the laboratory with different methodologies.

4. Reviews

4.1. D. Bassani, M. Pavan, S. Federico, G. Spalluto, M. Sturlese, and S. Moro, "The Multifaceted Role of GPCRs in Amyotrophic Lateral Sclerosis: A New Therapeutic Perspective?," Int. J. Mol. Sci., vol. 23, no. 9, p. 4504, Apr. 2022, doi: 10.3390/ijms23094504.

In this work, we present an overview of the involvement of some very relevant GPCRs in the onset and development of Amyotrophic Lateral Sclerosis (ALS). For each of the targets analysed, a brief

description is provided, as well as the most relevant literature supporting its implications in the mentioned degenerative pathology.

4.2. M. Pavan, D. Bassani, G. Bolcato, M. Bissaro, M. Sturles, and S. Moro, “Computational strategies to identify new drug candidates against neuroinflammation.,” Curr. Med. Chem., vol. 29, Feb. 2022, doi: 10.2174/0929867329666220208095122.

In this review, we present a comprehensive analysis of the current computational approaches useful for the discovery and development of novel molecular candidates for neuroinflammation treatment. The methodologies, as well as their strength and limitations, are extensively discussed.

Supervised Molecular Dynamics (SuMD) Insights into the mechanism of action of SARS-CoV-2 main protease inhibitor PF-07321332

Matteo Pavan, Giovanni Bolcato, **Davide Bassani**, Mattia Sturlese, Stefano Moro.

Pavan, M., Bolcato, G., Bassani, D., Sturlese, M. & Moro, S. Supervised Molecular Dynamics (SuMD) Insights into the mechanism of action of SARS-CoV-2 main protease inhibitor PF-07321332. J. Enzyme Inhib. Med. Chem. 36, 1646–1650 (2021).

Abstract

The chemical structure of PF-07321332, the first orally available Covid-19 clinical candidate, has recently been revealed by Pfizer. No information has been provided about the interaction pattern between PF-07321332 and its biomolecular counterpart, the SARS-CoV-2 main protease (M^{pro}). In the present work, we exploited Supervised Molecular Dynamics (SuMD) simulations to elucidate the key features that characterize the interaction between this drug candidate and the protease, emphasizing similarities and differences with other structurally related inhibitors such as Boceprevir and PF-07304814. The structural insights provided by SuMD will hopefully be able to inspire the rational discovery of other potent and selective protease inhibitors.

1. Introduction

The Covid-19 pandemic, caused by a single-stranded RNA betacoronavirus known as SARS-CoV-2, has caused the death of more than 3 million people around the world since its outbreak in December 2019[1], [2]. Despite the impressive cooperative effort promoted by the international community and by medicinal chemists around the world[3], [4], to date, there is only one drug approved by the Food and Drug Administration (FDA) for the treatment of Covid-19 patients.

Remdesivir, a polymerase inhibitor initially conceived to target Ebola Virus, proved to be efficient in shortening the recovery time in adult patients hospitalized with Covid-19[5], [6] and has therefore been granted Emergency Use Authorization (EUA). Unfortunately, due to its pharmacokinetic profile, this drug has to be administered intravenously in a hospital setting, thereby limiting its use for Covid-19 treatment on a massive scale. The first attempts to face this lack of pharmacological tools to contrast the Covid-19 pandemic involved the repurposing of antiviral drugs designed for the

treatment of other virus-related illnesses against Covid-19: this approach, despite being very appealing from a timescale perspective[7], did not bring any significant results, with several clinical trials showing little to no efficacy of those active principles against SARS-CoV-2[8].

Meanwhile, the early release to the scientific community of the crystallographic structure of the SARS-CoV-2 main protease (M^{pro}) (PDB ID: 6LU7), caused a shift in the attention of researchers around the world towards the Structure-Based approach to the rational design of new potential protease inhibitors[9]. Among all the different chemical entities developed to target the main protease, PF-07321332 is, to date, the first and only orally available COVID-19 antiviral clinical candidate.

Designed amid the pandemic, the structure of PF-07321332 was unveiled by Pfizer on April 6th at the American Chemical Society Spring 2021 meeting[10]. This compound, which has recently entered clinical phase I, was developed to target SARS-CoV-2 main protease, thereby impairing the virus's ability to reproduce itself, and it is intended as a pharmacological tool to prevent the development of COVID-19 in people who have been exposed to the pathogen. Even though the compound structure has been revealed, no further information has been provided yet about the way PF-07321332 interacts with the main protease active site, except for the fact that it reacts reversibly with a cysteine residue located in the binding site.

In this perspective computational investigation, we exploited Supervised Molecular Dynamics (SuMD)[11], an emerging protocol allowing to decipher at an atomic level of detail the recognition process between two molecular entities, to sample and characterize a putative binding pathway for PF-07321332. As described in the original publication, SuMD simulations fully consider both the protein flexibility and the contribution of the solvent molecules, which are explicitly simulated, throughout the binding process. As shown by previous scientific works[12], [13], this makes it possible to overcome the limitations of traditional techniques such as molecular docking when working on challenging targets such as M^{pro}, whose active site is relatively shallow, plastic and solvent exposed[14].

2. Methods

2.1 Software overview

For every general molecular modeling operation, such as protein and ligand structure preparation, MOE suite (Molecular Operating Environment, version 2019.01[15]) was used, exploiting an 8 CPU (Intel Xeon E5-1620 3.50 GHz) Linux Workstation. Molecular Dynamics simulations were carried out with ACEMD[16] (version 3.3.0), which is based upon OpenMM[17] (version 7.4.0), on a cluster composed of 20 NVIDIA GPUs.

2.2 Structure preparation

The crystallographic structure of the unliganded M^{pro} was retrieved from the Protein Data Bank (PDB ID: 7K3T). At first, the active functional dimer of the protease was restored applying the symmetric crystallographic transformation to each asymmetric unit. Residues with alternative conformation were assigned to the one with the highest occupancy. The Protonate3D tool was then used to add missing hydrogen atoms, evaluating the most probable protonation state for each titratable residue at pH 7.4. Finally, each non-protein residues (e.g.: water, co-solvents, etc.) were removed before successive steps. The ligand structure was prepared exploiting tautomers, fixpka, and molcharge tools from the QUACPAC OpenEye[18] software suite to assign the most probable tautomeric and protomeric state at pH 7 and ligand partial charges according to the MMFF94 force field. Three-dimensional coordinates were generated with Corina Classic[19].

2.3 Molecular Dynamics system setup

The simulated system contained both the protein and the ligand structure prepared as described in the previous section, with the ligand positioned at least 30 Å away from the nearest receptor atoms. For system parametrization, the combination of Amber ff14SB and General Amber Force Field (GAFF) was used to describe each component of the simulation box.

The system was explicitly solvated in a cubic TIP3P[20] water box with 15 Å padding and neutralized with the addition of Na+/Cl- ions until a 0.154 M concentration was reached. Prior to the simulation, 1000 steps of energy minimization with the conjugated-gradient algorithm were performed. A two-step equilibration stage was carried out in the following way: the first step consisted of 0.1 ns of simulation in the canonical ensemble (NVT) with harmonic positional restraints applied both on the protease and ligand atoms using a 5 Kcal mol^{-1} $Å^{-2}$ force constant, the second step consisted of 0.5

ns of simulation in the isothermal-isobaric ensemble (NPT) with the same harmonic positional restraints applied only on protein alpha carbons and ligand atoms. For each simulation, an integration timestep of 2 fs was used. To constrain bonds involving hydrogen atoms the M-SHAKE algorithm was used. A 9.0 Å cutoff was applied for the calculation of Lennard-Jones interactions, while electrostatic interactions were computed exploiting the particle-mesh Ewald method (PME). The temperature was maintained at the constant value of 310K by the Langevin thermostat, with a friction coefficient of 0.1 ps^{-1}. During the second equilibration stage, the pressure was maintained constant at 1.0 atm utilizing a Monte Carlo barostat.

2.4 Supervised Molecular Dynamics (SuMD) simulation

SuMD code is written in Python 2.7 and exploits the ProDy[21] package to perform geometrical supervision upon the ligand-binding process. This supervision allows to reduce the timescale, hence shrinking the computational effort, that is required to sample the ligand-biomolecular target recognition process to the range of nanoseconds, instead of the usual hundreds of nanoseconds or microseconds that are required by unbiased molecular dynamics (MD) simulations. The entire SuMD derived trajectory is composed by short unbiased 600 ps MD simulation runs (NVT ensemble, T= 310 K) with the ACEMD3 software: at the end of each simulation (the so-called “SuMD-step”), the distance between the center of mass of the ligand and the binding site is computed at five different points, picked at regular time intervals, and fitted into a linear function evaluated by a tabu-like algorithm. Only those SuMD-steps whose computed slope is negative (indicating that the ligand is approaching the binding site) are retained. Every time a SuMD-step is rejected (positive slope), the simulation is restarted from the previous productive step by randomly assigning the atomic velocities. The supervision algorithm is switched off after the distance between the center of mass of the ligand and the binding site drops below 5 Å: from that point on the simulation continues as a classical MD simulation.

3. Results

In our computational study, we exploited Supervised Molecular Dynamics simulations to obtain a putative binding pathway between PF-07321332 and the SARS-CoV-2 Main Protease (M^{pro}) catalytic site. A total amount of 36 ns of SuMD simulation time proved sufficient to sample the entire

recognition trajectory, from the starting unbound state to the final predicted protein-ligand complex.

As can be seen in Video1, PF-07321332 reaches M^{pro} active site after about 7 ns of simulation time, making its first contacts with Leu141, Asp 142, Gln189, and Glu166. Leu141 and Asp142 are part of the oxyanion loop (residues 138-145), which lines the binding pocket of Glutamine P1 and is assumed to stabilize the tetrahedral acyl transition state[14]. Glu166 is a key residue located in the middle of the binding site: mutagenesis studies carried out on SARS-CoV M^{pro} (which has 96% sequence identity with SARS-CoV-2 M^{pro} and is identical at the binding site level[12]) showed that this residue plays a key role in linking the dimer interface with the substrate-binding site[22]. Gln189 is located at the boundary of the S3 site and is assumed to be one of the key interactors with SARS-CoV-2 M^{pro} inhibitors, as well as Glu166[23]. Asn142 and Gln189, located on opposite sides at the

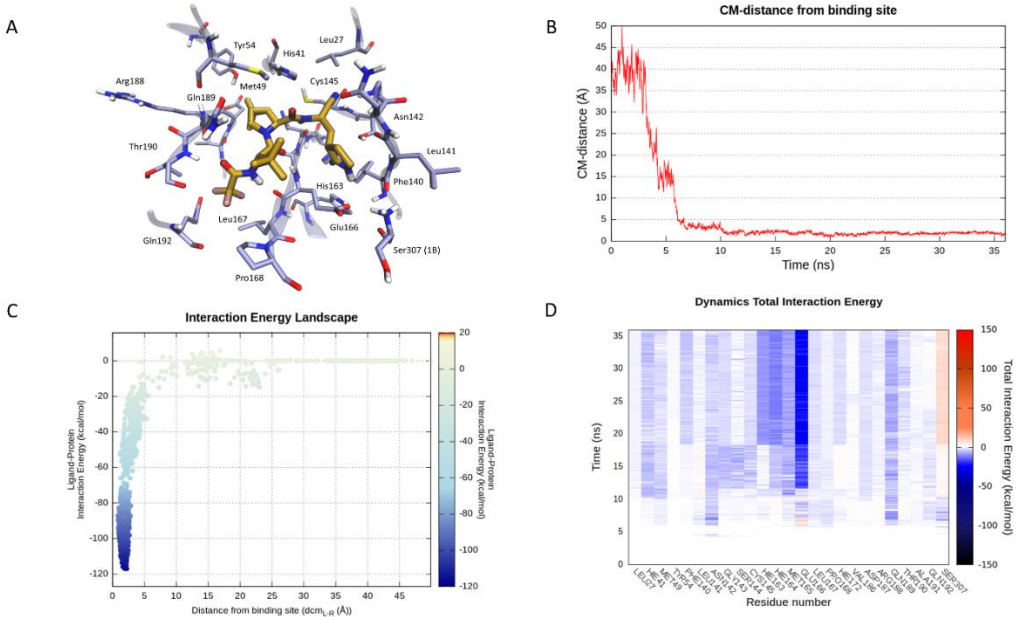


Figure 1. *This panel encompasses the recognition pathway between PF-07321332 and the SARS-CoV-2 main protease predicted by SuMD.* ***(A)*** *PF-07321332 conformation within the binding site, sampled in the last SuMD trajectory frame (orange). Binding site residues within 4 Å of the ligand are depicted in ice-blue.* ***(B)*** *Profile of the distance between the center of mass of the ligand and the M^{pro} catalytic site during SuMD simulation.* ***(C)*** *Interaction Energy Landscape describing the protein-ligand binding pathway; values are arranged according to distances between the center of mass of the ligand the one of the M^{pro} catalytic site.* ***(D)*** *Dynamic total interaction energy (sum of electrostatic and van der Waals contribution) computed for the 25 most contacted residues throughout the SuMD trajectory.*

boundary of the binding sites, seem to serve as electrostatic recruiters for the ligand, exploiting their polar and flexible sidechains to maneuver the entrance of the ligand into the core region of the binding site. Glu166 appears to instead serve as an electrostatic anchor that tightly hooks the

middle portion of the ligand with the central region of the binding site, facilitating the formation of further interactions with residues such as His 164.

After the tri-fluoro-acetamide moiety of the compound establishes contact with the side chain of Gln189, the cyclopropyl-proline moiety occupies the central portion of the binding site, establishing a series of coordinated hydrogen bonds with the backbone of His164 and Glu166 and orientating the cyclopropyl group towards the hydrophobic S2 pocket, delimited by the side chains of His41, Met49, Tyr54, and Met165. Meanwhile, the pyrrolidone moiety is inserted in the S1 pocket, interacting with key residues of the oxyanion loop such as Asn142, Gly143, and Ser144, before undergoing a conformational rearrangement around the 18 ns simulation time mark which allows the carbonyl of the pyrrolidone to establish a hydrogen bond with His163. This interaction has been flagged as a conserved interaction across several deposited structures of non-covalent inhibitors[24]. Moreover, this interaction is conserved across all possible substrate peptide crystal structures, where the interacting group is the sidechain of the Glutamine P1 residue[25].

Subsequently, the pyrrolidone moiety rearrangement also allows the reactive nitrile group to face the catalytic Cys145, making it possible to reach the final covalent-bound state which cannot be described through molecular mechanics. Finally, in the final conformation, the tri-fluoro acetamide moiety is fully inserted in the S4 subpocket, establishing two additional hydrogen bonds with the backbone of Thr190 and Glu166.

4. Discussion

Intriguingly, the binding mode proposed by the SuMD simulation for PF-07321332 is fairly superimposable to the ones of other two covalent protease inhibitor, Boceprevir (PDB ID: 6WNP) and PF-00835231 (PDB ID: 6XHM) which share common structural features with the oral candidate, validating the hypothesis that they could also share an overall similar interacting pattern (Figure 2).

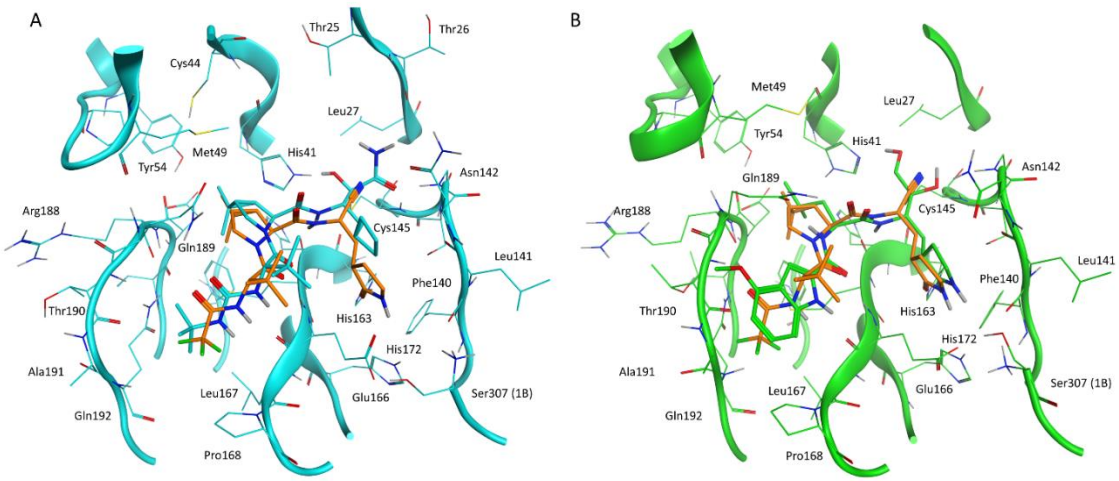


Figure 2. *This panel illustrates the similarities between PF-07321332 conformation in the final SuMD trajectory frame and the crystallographic complexes of two structurally related covalent inhibitors of SARS-CoV-2 M^{pro}: Boceprevir and PF-00835231 (active metabolite of PF-07304814).* ***(A)*** *superposition between the binding mode predicted by SuMD for PF-07321332 (orange) and the crystallographic complex of Boceprevir within the catalytic site of SARS-CoV-2 M^{pro} (cyan, PDB ID: 6WNP).* ***(B)*** *superposition between the binding mode predicted by SuMD for PF-07321332 (orange) and the crystallographic complex of PF-00835231 within the catalytic site of SARS-CoV-2 M^{pro} (green, PDB ID: 6XHM)*

Boceprevir is a protease inhibitor originally developed for the Hepatitis C Virus (HCV) NS3 protease[26]. It shares many common structural features with PF-07321332, such as the cyclopropyl proline residue at P2 and the alanine at the P3 position but has a different reactive group (α-ketoamide), a cyclobutyl alanine at P1, and a tert-butyl carbamate capping moiety at P4. From a binding mode point of view, the most prominent difference between the newly developed inhibitor and Boceprevir regards the hydrogen bond with His163 (absent in Boceprevir complex with the protease) which, as previously mentioned, is a crucial interaction also for natural peptidic substrates.

PF-07304814 is a Phase I clinical candidate originally developed by Pfizer in 2002-2003 against SARS-CoV and repurposed for SARS-CoV-2 due to the aforementioned similarities between the two viruses proteases[27]. The compound contains a hydrolyzable phosphate group which enhances its solubility and is cleaved by alkaline phosphatases in tissue releasing the active compound PF-00835231. The main limiting factor for this candidate is that, unlike its successor PF-07321332, it has to be administered intravenously, making it less appealing for massive distribution and relegating its usage to hospital settings. From a structural point of view, this latter compound is less similar to PF-07321332 compared to Boceprevir, but still retains the key features concerning its binding mode with the MPro active site. The only conserved structural feature between the two inhibitors developed by Pfizer is the pyrrolidone group at the P1 position, which establishes a

hydrogen bond with His163. The reactive group, in this case, is an aldehyde, the same as for Boceprevir. The hydrophobic residue at P2, in this case, is a leucine, which is the most recurrent amino acid that can be found at the P2 position in natural substrate peptides (included the N-term of M^{pro} itself)[25], while the P3 terminal residue is a 4-methoxyl indole group, which interacts through a hydrogen bond with the backbone of Glu166. Additional interaction occurs at the P1' subsite, where the two hydroxyl groups (one of which is formed upon reaction between the aldehyde group and Cys145 sidechain) form hydrogen bonds with Cys145 backbone and His41 sidechain.

Overall, PF-07321332 appears to have combined the strong points of both Boceprevir and PF-07304814 in a single molecular entity, showing that it is possible to repurpose the knowledge acquired in previous drug development campaigns on different virus proteases to rationally design SARS-CoV-2 M^{pro} inhibitors suitable for advancement to clinical phases, hence addressing the need for a quick response against a widespread disease like Covid-19. Moreover, the combination of innovative computational strategies such as SuMD with experimental data coming from X-Ray Crystallography could provide useful structural insights to stir the rational development of antiviral drugs in a more rational and less time-consuming way.

5. Conclusions

In this computational study, we employed Supervised Molecular Dynamics (SuMD) to investigate the recognition process between PF-07321332, the first orally available Covid-19 antiviral candidate to reach clinical phase I, and its biological target, SARS-CoV-2 main protease (M^{pro}).

About 36 ns of SuMD simulations proved sufficient to sample a putative binding process, allowing to simulate the whole approaching path from the unbound state to the final protein-ligand complex. SuMD simulations suggest a possible role in the first stages of the recruitment of the ligand for residues such as Leu141, Asp 142, Gln189, and Glu166, which have already been acknowledged as crucial residues for the binding of both natural and synthetic substrates.

Finally, the binding mode predicted by SuMD for PF-07321332 is quite similar for other structurally related protease inhibitors, namely Boceprevir and PF-07304814, which could also share a similar binding pathway.

References

[1] J. Guarner, “Three Emerging Coronaviruses in Two Decades: The Story of SARS, MERS, and Now COVID-19,” *American Journal of Clinical Pathology*, vol. 153, no. 4. Oxford University Press, pp. 420–421, Mar. 2020, doi: 10.1093/ajcp/aqaa029.

[2] “COVID Live Update: 163,750,604 Cases and 3,394,311 Deaths from the Coronavirus - Worldometer.” .

[3] L. Zhang and Y. Liu, “Potential interventions for novel coronavirus in China: A systematic review,” *Journal of Medical Virology*, vol. 92, no. 5. John Wiley and Sons Inc., pp. 479–490, May 2020, doi: 10.1002/jmv.25707.

[4] D. L. Heymann and N. Shindo, “COVID-19: what is next for public health?,” *The Lancet*, vol. 395, no. 10224. Lancet Publishing Group, pp. 542–545, Feb. 2020, doi: 10.1016/S0140-6736(20)30374-3.

[5] G. Kokic *et al.*, “Mechanism of SARS-CoV-2 polymerase stalling by remdesivir,” *Nat. Commun.*, vol. 12, no. 1, pp. 1–7, Dec. 2021, doi: 10.1038/s41467-020-20542-0.

[6] J. H. Beigel *et al.*, “Remdesivir for the Treatment of Covid-19 — Final Report,” *N. Engl. J. Med.*, vol. 383, no. 19, pp. 1813–1826, Nov. 2020, doi: 10.1056/nejmoa2007764.

[7] D. Mani, A. Wadhwani, and P. T. Krishnamurthy, “Drug Repurposing in Antiviral Research: A Current Scenario,” *J. Young Pharm.*, vol. 11, no. 2, pp. 117–121, May 2019, doi: 10.5530/jyp.2019.11.26.

[8] S. G. Viveiros Rosa and W. C. Santos, “Clinical trials on drug repositioning for COVID-19 treatment,” *Rev. Panam. Salud Publica/Pan Am. J. Public Heal.*, vol. 44, p. e40, May 2020, doi: 10.26633/RPSP.2020.40.

[9] Z. Jin *et al.*, “Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors,” *Nature*, vol. 582, no. 7811, pp. 289–293, Jun. 2020, doi: 10.1038/s41586-020-2223-y.

[10] “Pfizer unveils its oral SARS-CoV-2 inhibitor.” .

[11] D. Sabbadin and S. Moro, “Supervised Molecular Dynamics (SuMD) as a Helpful Tool To Depict GPCR–Ligand Recognition Pathway in a Nanosecond Time Scale,” *J. Chem. Inf. Model.*, vol. 54, no. 2, pp. 372–376, Feb. 2014, doi: 10.1021/ci400766b.

[12] G. Bolcato, M. Bissaro, M. Pavan, M. Sturlese, and S. Moro, “Targeting the coronavirus SARS-CoV-2: computational insights into the mechanism of action of the protease inhibitors lopinavir, ritonavir and nelfinavir,” *Sci. Rep.*, vol. 10, no. 1, p. 20927, Dec. 2020, doi: 10.1038/s41598-020-77700-z.

[13] M. Bissaro, G. Bolcato, M. Pavan, D. Bassani, M. Sturlese, and S. Moro, “Inspecting the mechanism of fragment hit binding on SARS-CoV-2 Mpro by using supervised molecular dynamics (SuMD) simulations.,” *ChemMedChem*, Apr. 2021, doi: 10.1002/cmdc.202100156.

[14] E. Fornasier *et al.*, “A novel conformational state for SARS-CoV-2 main protease,” *bioRxiv*, p. 2021.03.04.433882, Mar. 2021, doi: 10.1101/2021.03.04.433882.

[15] “Molecular Operating Environment (MOE), 2019.01; Chemical Computing Group ULC, 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2021.” .

[16] M. J. Harvey, G. Giupponi, and G. De Fabritiis, “ACEMD: Accelerating Biomolecular Dynamics in the Microsecond Time Scale,” *J. Chem. Theory Comput.*, vol. 5, no. 6, Jun. 2009, doi: 10.1021/ct9000685.

[17] P. Eastman *et al.*, “OpenMM 7: Rapid development of high performance algorithms for molecular dynamics,” *PLOS Comput. Biol.*, vol. 13, no. 7, p. e1005659, Jul. 2017, doi: 10.1371/journal.pcbi.1005659.

[18] “QUACPAC 2.1.1.0: OpenEye Scientific Software, Santa Fe, NM.” .

[19] J. Sadowski, J. Gasteiger, and G. Klebe, “Comparison of Automatic Three-Dimensional Model Builders Using 639 X-ray Structures,” *J. Chem. Inf. Comput. Sci.*, 1994, doi: 10.1021/ci00020a039.

[20] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, “Comparison of simple potential functions for simulating liquid water,” *J. Chem. Phys.*, vol. 79, no. 2, pp. 926–935, Jul. 1983, doi: 10.1063/1.445869.

[21] A. Bakan, L. M. Meireles, and I. Bahar, “ProDy: Protein dynamics inferred from theory and experiments,” *Bioinformatics*, vol. 27, no. 11, pp. 1575–1577, 2011, doi: 10.1093/bioinformatics/btr168.

[22] S. C. Cheng, G. G. Chang, and C. Y. Chou, “Mutation of glu-166 blocks the substrate-induced dimerization of SARS coronavirus main protease,” *Biophys. J.*, vol. 98, no. 7, pp. 1327–1336, Apr. 2010, doi: 10.1016/j.bpj.2009.12.4272.

[23] B. Goyal and D. Goyal, “Targeting the Dimerization of the Main Protease of Coronaviruses: A Potential Broad-Spectrum Therapeutic Strategy,” *ACS Combinatorial Science*, vol. 22, no. 6. American Chemical Society, pp. 297–305, Jun. 2020, doi: 10.1021/acscombsci.0c00058.

[24] Y. L. Weng, S. R. Naik, N. Dingelstad, M. R. Lugo, S. Kalyaanamoorthy, and A. Ganesan, “Molecular dynamics and in silico mutagenesis on the reversible inhibitor-bound SARS-CoV-2 main protease complexes reveal the role of lateral pocket in enhancing the ligand affinity,” *Sci. Rep.*, vol. 11, no. 1, p. 7429, Dec. 2021, doi: 10.1038/s41598-021-86471-0.

[25] W. Rut *et al.*, “Substrate specificity profiling of SARS-CoV-2 main protease enables design of activity-based probes for patient-sample imaging,” *bioRxiv*, p. 2020.03.07.981928, Mar. 2020, doi: 10.1101/2020.03.07.981928.

[26] F. G. Njoroge, K. X. Chen, N.-Y. Shih, and J. J. Piwinski, “Challenges in Modern Drug Discovery: A Case Study of Boceprevir, an HCV Protease Inhibitor for the Treatment of Hepatitis C Virus Infection,” *Acc. Chem. Res.*, vol. 41, no. 1, pp. 50–59, Jan. 2008, doi: 10.1021/ar700109k.

[27] B. Boras *et al.*, “Discovery of a Novel Inhibitor of Coronavirus 3CL Protease as a Clinical Candidate for the Potential Treatment of COVID-19.,” *bioRxiv Prepr. Serv. Biol.*, p. 2020.09.12.293498, Sep. 2020, doi: 10.1101/2020.09.12.293498.

Inspecting the Mechanism of Fragment Hits Binding on SARS-CoV-2 M^{pro} by Using Supervised Molecular Dynamics (SuMD) Simulations

Maicol Bissaro, Giovanni Bolcato, Matteo Pavan, **Davide Bassani**, Mattia Sturlese and Stefano Moro

Bissaro, M. *et al.* Inspecting the Mechanism of Fragment Hits Binding on SARS-CoV-2 M pro by Using Supervised Molecular Dynamics (SuMD) Simulations. *ChemMedChem* cmdc.202100156 (2021) doi:10.1002/cmdc.202100156.

Abstract

Computational approaches supporting the early characterization of fragment molecular recognition mechanism represent a valuable complement to more expansive and low-throughput experimental techniques. In this retrospective study, we have investigated the geometric accuracy with which high-throughput supervised molecular dynamics simulations (HT-SuMD) can anticipate the experimental bound state for a set of 23 fragments targeting the SARS-CoV-2 main protease. Despite the encouraging results herein reported, in line with those previously described for other MD-based posing approaches, a high number of incorrect binding modes still complicate HT-SuMD routine application. To overcome this limitation, fragment pose stability has been investigated and integrated as part of our in-silico pipeline, allowing us to prioritize only the more reliable predictions.

1. Introduction

Fragment-based drug discovery (FBDD) has progressively established as a game-changing approach to navigate the chemical space in the drug discovery pipelines, both on academic and industrial early discovery stages[1][2][3]. By definition, fragments are low molecular weight organic molecules able to recognize a target of therapeutic interest in a mild affinity range and with a poor selectivity profile[4]. Intriguingly, the screening of small sized fragment libraries in place of conventional larger ones has proven to provide better coverage of the chemical diversity and higher hit rates[5][6]. The identification of such weak binders, however, strictly depends on the implementation of biophysical screening techniques, such as X-Ray Crystallography (XRC), Nuclear Magnetic Resonance (NMR), surface plasmon resonance (SPR), or Thermal Shift Assay (TSA)[1][7][8]. Anyway, broad differences exist among such methods and each of them suffers unique limitations in the challenging identification of reliable fragment; indeed the agreement in the hits identified is surprisingly

limited[9][10][11]. Besides, only XRC and NMR offer the possibility to investigate the binding mode of weak binders. In light of this, structure-based computational strategies have increasingly gained appeal[12][13][14]. As highlighted in a recent review, during the last decade Molecular Dynamics (MD) simulations have been extensively applied also in the FBDD field, providing an atomistic insight on the fragment-receptor binding mechanisms, with a femtosecond temporal resolution[15]. From this perspective, we recently developed HT-SuMD, a computational protocol exploiting supervised MD simulations to perform the screening of a small fragments library in a competitive timescale[16]. The performance of the protocol in prioritizing the most promising fragment binders was compared with NMR-based screening, against the oncological protein target Bcl-xL. Despite the notable agreement with NMR in identifying the most promising hits, the lack of structural data prevent the assessment of HT-SuMD accuracy in fragments binding mode prediction, which would represent a valuable set of information to guide the subsequent hit to lead (H2L) optimization steps. In this methodological study, we have therefore retrospectively investigated the accuracy of HT-SuMD simulations in reproducing the experimental binding mode of several fragment-protein complexes, exploiting the 3- C-like main protease (Mpro) of the novel SARS-CoV-2 coronavirus as a relevant case study. Following indeed the dramatic spread of the COVID-19 pandemic, a collaborative XRC fragment screening against the protein Mpro has timely offered to the scientific community valuable structural information to accelerate the rational design of new protease inhibitors[17][18][19]. For this validation study in detail, among the 71 fragments targeting the catalytic site of Mpro originally identified by the XRC screening, only the 23 presenting a reversible mechanism of recognition were taken into consideration, due to the impossibility of modeling covalent reactivity through classical molecular mechanics (MM) force fields[20][21].

2. Results and Discussion

2.1 Characterization of fragment-receptor complexes

The high-quality Mpro crystallographic structures were collected from the Protein Data Bank database (PDB ID are reported in Table 1 of SI) and prepared by applying symmetric transformation to each asymmetric unit, thus recreating the original functional dimer[22]. A visual inspection of the catalytic clefts has revealed how the 23 non-covalent fragments comprehensively explore most protease binding subsites (S1, S2, S3, and S1') while providing decent coverage of chemical diversity.

Besides, Mpro catalytic cleft is easily accessible from the bulk solvent and hence suitable to SuMD studies, as recently demonstrated for a couple of Mpro inhibitors[23]. The complexity, as well as the plasticity of the Mpro binding pocket, made this test case particularly challenging, the reason why an MD-based stability characterization of all the experimental-solved crystallographic complexes was performed, before investigating HT-SuMD accuracy in the fragment posing process. For this purpose, the AMBER14SB force field was combined with the general amber force field (GAFF) to parameterize respectively the protein biopolymers and the small organic fragments[24][25]. To ensure results robustness, 5 trajectories each 20 ns long were collected for all Mprocomplexes, resulting in a total of 2.3 µs of conventional MD simulations. The content of information extrapolated from a single trajectory has been hence doubled by simply repeating the analysis against the two distal and independent catalytic sites of the homodimeric SARS-CoV-2 Mpro. To characterize the geometric

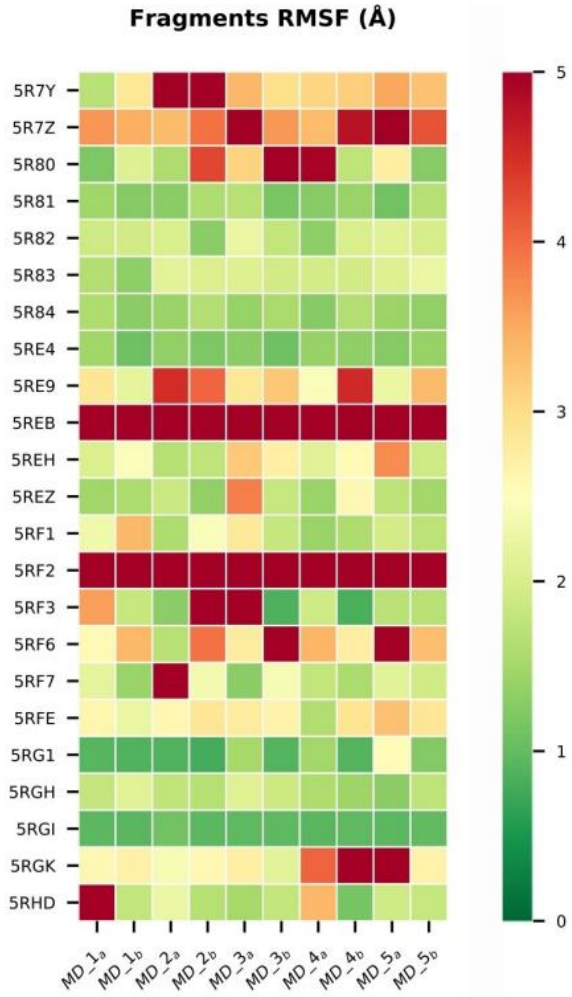


***Figure 1**. Fragment stability assessed by classical MD of the 23 crystallographic complexes. For each MD simulation collected (x-axis) starting from the crystallographic ligand-receptor complexes (y-axis), the pose stability value of the fragment is herein reported through a heatmap representation. The colorimetric scale, from green to red, quantitatively represents the RMSF computed for each ligand heavy atoms (0 to 5 Å scale). The MD simulation were carried out on each subunit of the Mpro functional dimer resulting in two set (labelled a and b) for each of the 5 runs.*

stability of the experimental-solved fragment complexes the root-mean-square fluctuation (RMSF) of ligands heavy atoms has been chosen as a metric, then summarizing the results through a heatmap representation, as reported in Figure 1. The colorimetric scale helps in differentiating those fragments which maintained the original binding mode during all the collected replicates (green color), from others undergoing a neat perturbation of the recognition modality (yellow color) or that even experience a spontaneous unbinding event, repetitively leaving the catalytic cleft (red color). Interestingly, a strong correlation was identified between the topological localization of the fragments and their RMSFavg, with those ligands occupying the highly flexible S2 subsite also showing the more pronounced propensity in losing the experimental solved binding mode. This information not only offers valuable insights for the H2L optimization phase but also opens up questions about the suitability of MD-based approaches for the posing of ligands characterized by such limited structural stability.

2.2 Fragments posing through HT-SuMD

HT-SuMD protocol has been applied to investigate the binding mechanism of the 23 non-covalent fragments against the unliganded crystal structure of the SARS-CoV-2 Mpro (PDB ID 6YB7). As accurately described in the original paper, HT-SuMD manages the preparation, collection, and analysis of multiple SuMD simulations in an automatic modality, only requiring the binding pocket localization as initial information. SuMD, briefly, exploiting a tabu-like supervision algorithm that monitors in times variations in the ligand-protein binding site distances, could be considered an enhanced sampling approach improving the efficiency with which rare events, such as binding, are described[26][27]. For each fragment investigated, a solvated MD simulation box has been set up (a detailed description is reported on supplementary materials) and equilibrated after distancing the ligand at least 30 Å away from the protein catalytic cleft, to avoid premature intramolecular interactions. Also in this case, as an attempt to increase the robustness of the results, 10 SuMD replicates have been collected, resulting in a total of 6.3 μs of simulation time. The ensemble of 230 trajectories describing different fragment binding pathways has been later geometrically discretized through DBSCAN, a density-based clustering algorithm, which allows all the most populated ligand-protein states to emerge from the background noise[28][29]. In detail, a cluster is initialized if it contains at least 25 similar fragments conformations, which therefore differ from each other by no

more than 1.5 Å. Finally, each binding mode was qualitatively evaluated using the MM/GBSA approach to approximate the ligand-protein free energy of binding, thus allowing to perform a ranking of the predicted poses[30]. The accuracy of the predictions was assessed by comparing each cluster of fragment conformations identified with the respective crystallographic reference, computing the root-mean-square deviations (RMSD) of non-hydrogen atomic coordinates. The results obtained for the 23 Mpro crystallographic inhibitors have been extensively reported in the supplementary information (SI_HT-SuMD.xlsx) and graphically summarized in Figure 2, exploiting a colorimetric map to differentiate the correctness of the posing protocol. More specifically, for each fragment, the minimum RMSD (RMSDmin) and the average RMSD (RMSDavg) values for the best cluster, i. e. the cluster closer to the crystallographic reference, were reported then comparing the

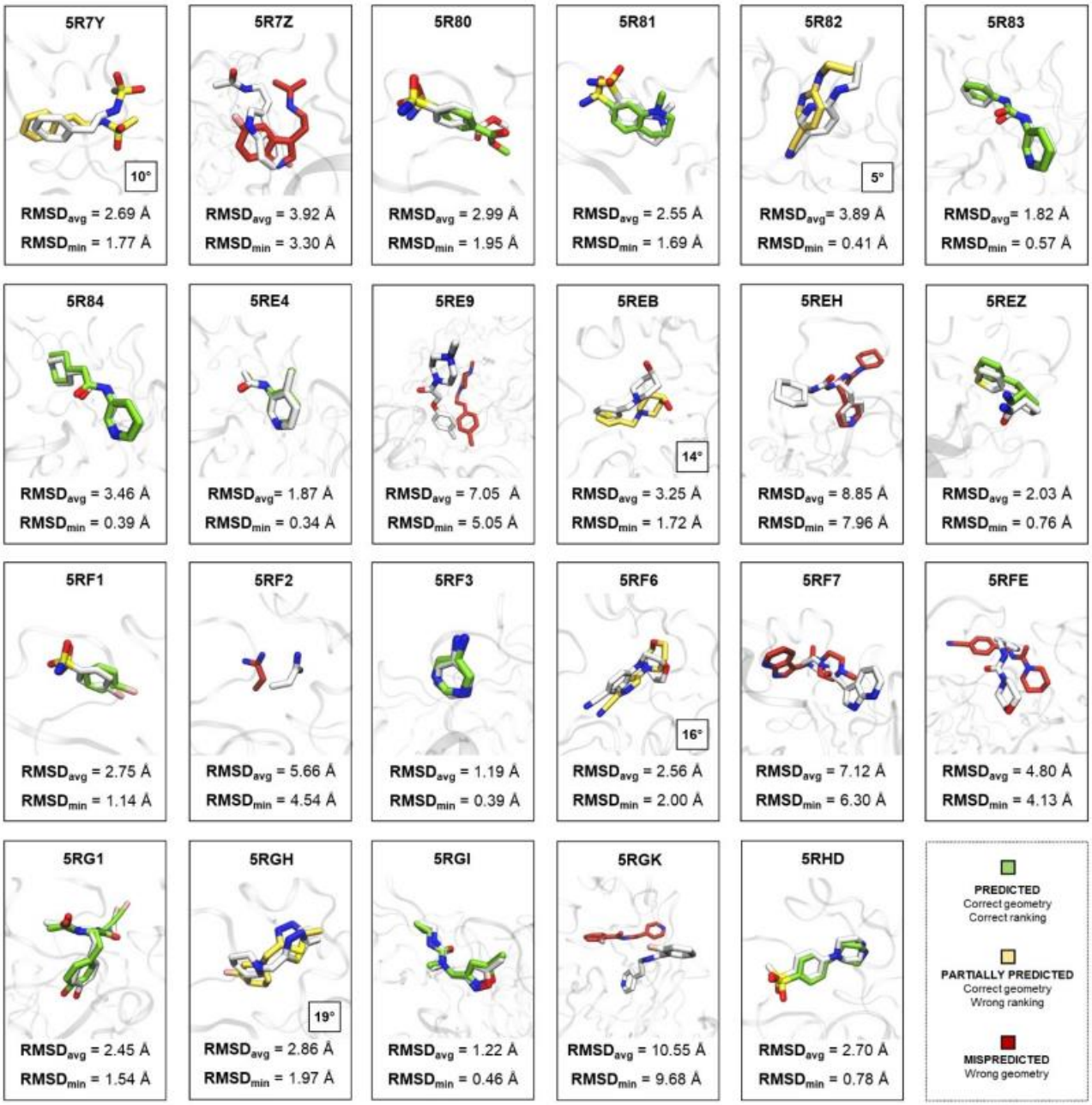


Figure 2. *The results of the HT-SuMD posing protocol have been herein summarized. For each of the 23 fragments investigated the cluster of ligand conformations closes to the experimentally solved binding mode was reported, measuring the accuracy of the prediction through the RMSDavg and RMSDmin values of the selected cluster. The crystallographic reference has been rendered in white color, while the HT-SuMD predicted binding modes have been differentiated in green, yellow, and red color, following the criteria described in the legend. In the case of partially predicted fragments, in which a good binding geometry was retrieved but erroneously ranked, the magnitude of the error has been underlined reporting the incorrect ranking position.*

predicted binding mode with the experimental one. The fragment posing exercise was considered correctly achieved if the RMSDmin of the cluster selected falls below the cut-off value of 2 Å.

For 11 fragments out of 23, representing almost half of the considered cases, the protocol was able to identify and correctly rank the experimental binding mode (green-coloured molecules). Among these, the most noteworthy case is represented by the fragment with the PDB ID 5RGI, the only one targeting the S1' subsite. HT-SuMD posing approach, fully exploring the conformational flexibility of the receptor, was able to reproduce the fragment crystallographic binding mode in an extremely accurate way, with an RMSDmin value of 0.46 Å. This result is impressive since, in the unliganded Mpro structure chosen in this study, the S1' pocket, due to a different orientation of the residues composing the catalytic dyad (H41 and C145), is initially inaccessible.

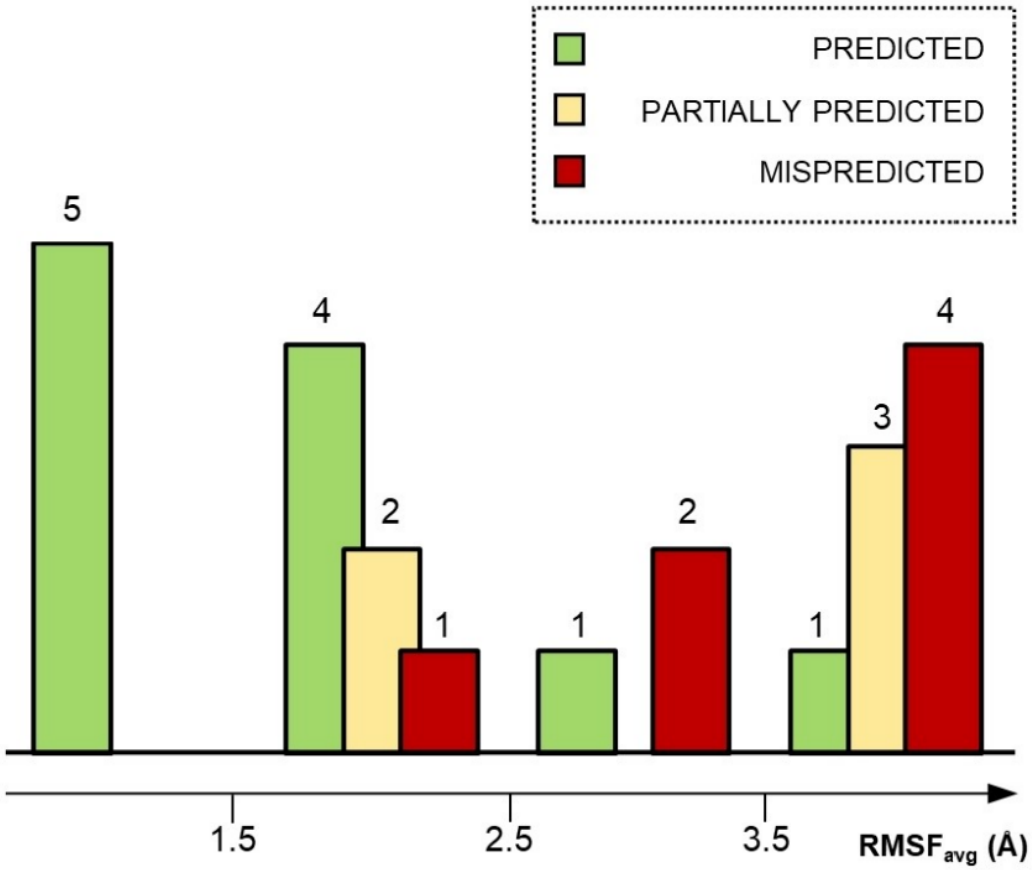


***Figure 3**. HT-SuMD predictions have been analyzed and related to the average fragment pose stability values (RMSFavg) showed by each original crystallographic complex during the MD-based characterization study.*

For the remaining 12 fragments, an in-depth analysis highlighted two orthogonal reasons underneath the HT-SuMD based posing failures. In 5 cases the MM/GBSA-based scoring method was unable to prioritize the experimental binding mode, even if it was exhaustively sampled by SuMD simulations (yellow-colored molecules). The incorrect ranking position was then reported in Figure 2 within a squared box, to underline the magnitude of the scoring error. This disagreement

may be caused by limitations affecting the MM models, as errors in the fragments force field parameters or, more intriguingly, the crystallographic structures could capture only one of the possible accommodation states that the ligand can explore within the binding site[31]. In the other 7 cases instead, the experimental conformation was never sampled (red-colored molecules), suggesting possible MD-sampling issues that may be addressed by widening the number of SuMD replicates performed for each compound, however increasing the computational cost of our approach. The accuracy of HT-SuMD protocol, therefore, with 48% of correct binding mode predictions is greater than non-native docking-based protocols reported in the literature and in line with that of other MD based fragments posing approaches[32][33]. It appears however evident how the posing of fragments still represents a tough pharmaceutical challenge, in particular, as suggested by Verdonk, for those characterized by a low-ligand efficiency (LE). Even our computational approach, in about half of the examined cases, fails to return a reliable result making its routine application very complex in a pharmaceutical drug discovery context.

To elucidate the applicability domain of HT-SuMD and better understand the limitations related to the implementation of MD-based protocols for the fragment binding modes prediction, we have therefore investigated if the fragment pose stability, a geometric-dynamic property, could impact the predictivity of our method. The fragment pose stability retraces the concept behind the structural stability criterion that has recently been discussed also by Barril's research group, as a complement to more traditional thermodynamic-based approaches in the identification of correct fragment-receptor binding mode[34]. HT-SuMD outcomes have therefore been compared, as reported in Figure 3, with the average values of atomic coordinates fluctuation (i. e. RMSFavg) respectively showed by each crystallographic fragment in the classical MD study previously discussed. Intriguingly, a clear pattern is noticeable since almost the totality of the correctly predicted binding modes (9/11) has been recovered for those fragments characterized by marked structural stability, with an RMSFavg value lower than 2.5 Å. Above this empirical cut-off, consistently most of the incorrect predictions concentrate, thus corroborating the existence of an inverse relationship linking together the stability of a crystallographic final state and the ability to correctly anticipate it through MD-based approaches, as our protocol configure.

2.3 Fragment poses stability as a confidence metric.

The relationship described above could therefore be exploited to drive the analysis and the interpretation of HT-SuMD results, providing an observable with which distinguish reliable binding

modes predictions from decoys. To test this hypothesis, the results collected through HT-SuMD posing protocol were retrospectively evaluated simulating a real screening scenario, in which crystallographic references are not available. Hence, for each of the 23 Mpro fragments previously investigated through HT-SuMD, the binding mode with the lowest MM/ GBSA score was blindly selected, regardless of whether or not it corresponds to the original experimental pose. Then, multiple classical MD simulations 20 ns long were started from the predicted final states, to characterize their relative fragment pose stability. Results of this study have been summarized in Figure 4, sorting the data concerning the RMSFSuMD values, or the average fluctuations of SuMD-predicted binding poses, computed on the fragment's heavy atoms. A first interesting aspect to underline is how almost the totality of the correct binding modes anticipated by HT-SuMD (green-colored molecules) only undergoes a mild conformational perturbation during classical MD simulations, in agreement with the results described in the first part of the manuscript for the crystallographic complexes. On the contrary, incorrect binding mode (yellow and red-colored molecules) in most of the cases experience great lability when refined through MD simulations, sometimes even culminating in a spontaneous unbinding event of the fragment.

These observations corroborate the initial hypothesis, suggesting how a combination of HT-SuMD protocol for the posing of fragments with classical MD simulation for the refinement of results could represent an optimal operative pipeline, which allows overcoming some of the previously discussed methodological limitations. In this specific case indeed, the implementation of a geometric-dynamic property, namely the RMSFSuMD, results extremely useful to qualitatively estimate the reliability of the in-silico predicted poses.

Observing the ranking reported in Figure 3, as the structural stability of the HT-SuMD predicted binding mode decreases, a worsening in posing accuracy occurs contextually. Intriguingly, also, in this case, 2.5 Å configure as a valuable empirical threshold which allows us to prioritize all the 11 correct fragment binding mode predictions. However, it is worth noting how the same cut-off is also responsible for the incorporation of three false positives, predictions characterized by remarkable structural stability, but which are nevertheless geometrically far from the crystallographic reference. For what concerns the fragment belonging to the PDB ID 5R7Y complex, HT-SuMD protocol has probably prioritized a metastable binding mode anticipating the experimental one, that has been nevertheless sampled through MD simulations but incorrectly scored by MM/GBSA. In the other two cases (PDB ID 5REH and 5RGK) the misprediction affects two fragments sharing a similar structure and interactivity. In the specific case of the 5REH complex, the HT-SuMD posing protocol

has prioritized an alternative binding mode in which the pyridine portion of the fragment is correctly predicted, reproducing the key hydrogen bond interaction with H163 residue, while the remaining flexible portion is erroneously accommodated in the subsite S2 causing, as indicated in Figure 2, the high RMSD value of the cluster. This aspect is particularly interesting in the FBDD context, considering how the mild affinity profile characterizing these compounds could determine multiple recognition modes.

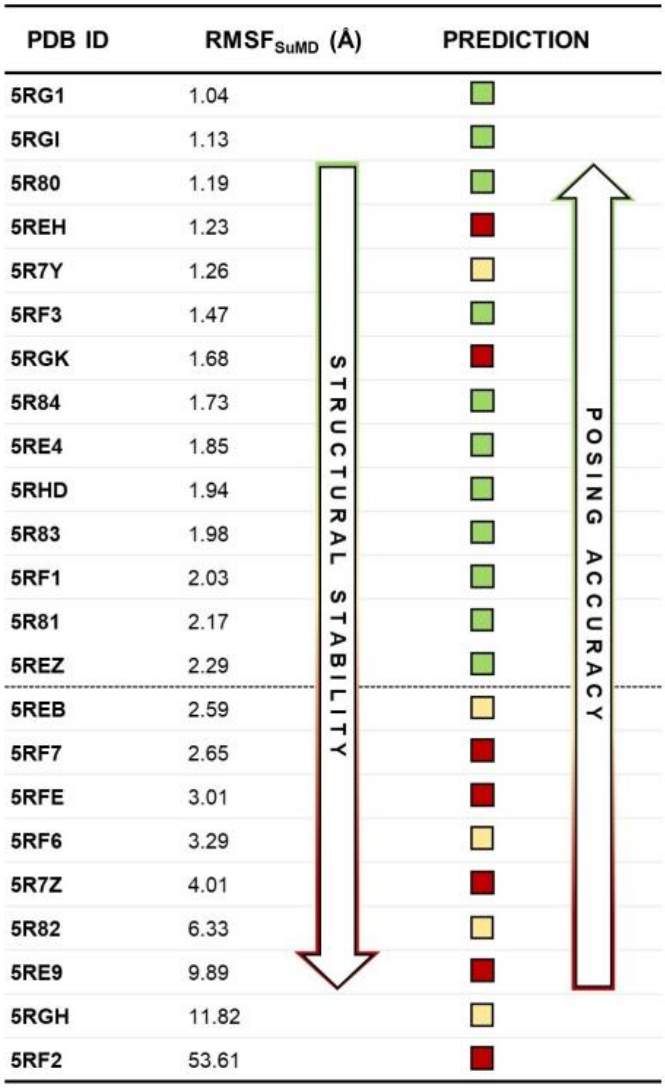


Figure 4. *HT-SuMD predicted binding modes (i. e. the cluster of fragments conformations characterized by the lowest MM/GBSA value) have undergone an MD-based refinement step. The fragment poses stability of each prediction, measured as the RMSFSuMD, has been exploited to rank HT-SuMD results, allowing in this way to efficiently prioritizing the correct binding modes at the expense of the incorrect ones. The dashed line delimits the empirical cut-off of 2.5 Å used to discriminate the reliability ofHT-SuMD posing prediction.*

3. Conclusion

The elucidation of fragment binding modes in the early stages of FBDD campaigns still represents a tough medicinal chemistry task, which can be mitigated by the concomitant application of in-silico approaches. In this work, we have therefore investigated the geometric accuracy with which our recently developed computational protocol can reproduce experimentally solved fragment-receptor complexes. For this purpose, the XRC structures of 23 non-covalent fragments targeting SARSCoV-2 Mpro, a pharmaceutical hot target in this actual COVID-19 pandemic, were exploited. HT-SuMD, as summarized in Figure 5, samples for each fragment multiple binding trajectories (Box 1), which are subsequently geometrically discretized through DBSCAN clustering and energetically evaluated using the MM/ GBSA approach (Box 2). Our methodology was able to recover and prioritize in almost half of the cases taken into consideration (48%) the original fragment bound geometry, with an accuracy comparable to that described for other MD-based posing approaches.

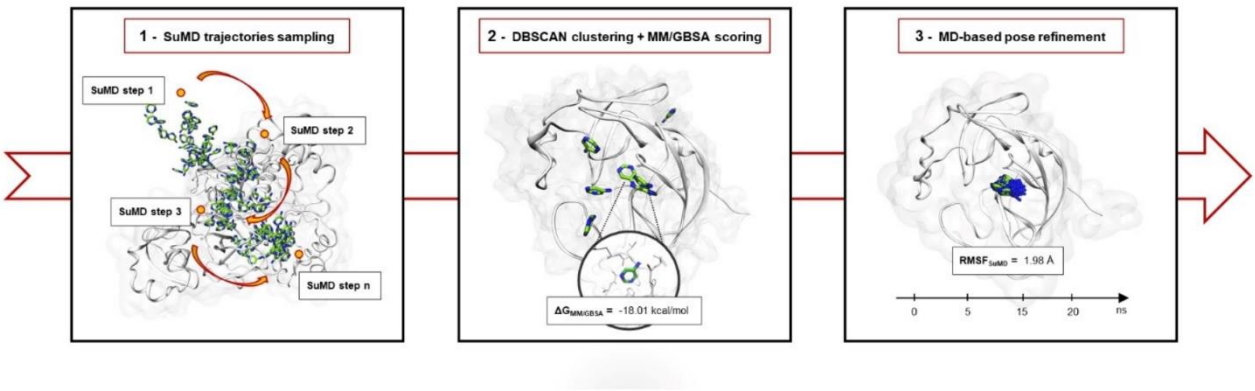


***Figure 5**. HT-SuMD protocol for the posing of fragments mainly consists in three operative steps, that are respectively summarized in this graphical workflow. In detail, supervised MD simulations are exploited to sample multiple binding trajectories for all the fragments analyzed (1), then DBSCAN clustering algorithm allows to identify of the most populated ligand conformation, which is energetically evaluated using MM/GBSA scoring method (2). The in-silico predicted binding modes finally undergo an MD-based refinement step, using the RMSFSuMD as a metric to qualitatively characterize the posing reliability.*

Intriguingly, a clear correlation has been identified between HT-SuMD posing accuracy and the stability of the respective crystallographic complexes, with most of the correct binding modes predictions retrieved for those fragments characterized by a low RMSFavg. In light of this aspect, a refinement step of HT-SuMD results through classical MD simulations has become an integrative part of our posing protocol (Figure 5– Box 3). More specifically, the structural stability of the predicted binding mode, i. e. the RMSFSuMD, has been exploited and validated as a metric to qualitatively estimate the reliability of each single in-silico prediction. In this way, it was possible to effectively rank and prioritize the 11 correct HT-SuMD binding poses while discharging the ones

characterized by a marked instability that was mainly revealed as incorrect predictions. This concept is exemplified in Video1 (supplementary information), reporting how MM/GBSA, a thermodynamic-based approach, fails in distinguishing a correct form and incorrect fragment binding pose, while the subsequent MD refinement steps allow highlighting a marked difference between the two different predictions, in terms of RMSFSuMD.

Despite these preliminary encouraging results, which must be certainly consolidated with further case studies, an improvement in the fragment posing accuracy is however still desirable. From this perspective, the ever-increasing computing power that will be available in the next years coupled with the continuous optimization of the conformational sampling algorithm, as well as the force fields model used, could pave the way for the development of more accurate fragment posing protocols, that could massively impact many in-silico FBDD pipelines.

References

[1] C. W. Murray and D. C. Rees, “The rise of fragment-based drug discovery.,” *Nat Chem*, vol. 1, no. 3, pp. 187–192, Jun. 2009, doi: 10.1038/nchem.217.

[2] D. A. Erlanson, S. W. Fesik, R. E. Hubbard, W. Jahnke, and H. Jhoti, “Twenty years on: the impact of fragments on drug discovery,” *Nat. Rev. Drug Discov.*, vol. 15, no. 9, pp. 605–619, Sep. 2016, doi: 10.1038/nrd.2016.109.

[3] C. Jacquemard and E. Kellenberger, “A bright future for fragment-based drug discovery: what does it hold?,” *Expert Opin. Drug Discov.*, vol. 14, no. 5, pp. 413–416, May 2019, doi: 10.1080/17460441.2019.1583643.

[4] F. Giordanetto, C. Jin, L. Willmore, M. Feher, and D. E. Shaw, “Fragment Hits: What do They Look Like and How do They Bind?,” *J. Med. Chem.*, vol. 62, no. 7, pp. 3381–3394, Apr. 2019, doi: 10.1021/acs.jmedchem.8b01855.

[5] R. J. Hall, P. N. Mortenson, and C. W. Murray, “Efficient exploration of chemical space by fragment-based screening.,” *Prog Biophys Mol Biol*, vol. 116, no. 2–3, pp. 82–91, Dec. 2014, doi: 10.1016/j.pbiomolbio.2014.09.007.

[6] P. J. Hajduk and J. Greer, “A decade of fragment-based drug design: strategic advances and lessons learned,” *Nat. Rev. Drug Discov.*, vol. 6, no. 3, pp. 211–219, Mar. 2007, doi: 10.1038/nrd2220.

[7] B. J. Davis and S. D. Roughley, “Fragment-Based Lead Discovery,” 2017, pp. 371–439.

[8] R. Ma, P. Wang, J. Wu, and K. Ruan, “Process of Fragment-Based Lead Discovery—A Perspective from NMR,” *Molecules*, vol. 21, no. 7, p. 854, Jul. 2016, doi: 10.3390/molecules21070854.

[9] B. J. Davis and D. A. Erlanson, “Learning from our mistakes: The ‘unknown knowns’ in fragment screening,” *Bioorg. Med. Chem. Lett.*, vol. 23, no. 10, pp. 2844–2852, May 2013, doi: 10.1016/j.bmcl.2013.03.028.

[10] J. Schiebel *et al.*, “Six Biophysical Screening Methods Miss a Large Proportion of Crystallographically Discovered Fragment Hits: A Case Study,” *ACS Chem. Biol.*, vol. 11, no. 6, pp. 1693–1701, Jun. 2016, doi: 10.1021/acschembio.5b01034.

[11] J. Wielens *et al.*, “Parallel Screening of Low Molecular Weight Fragment Libraries,” *J. Biomol. Screen.*, vol. 18, no. 2, pp. 147–159, Feb. 2013, doi: 10.1177/1087057112465979.

[12] C. Sheng and W. Zhang, “Fragment Informatics and Computational Fragment-Based Drug Design: An Overview and Update,” *Med. Res. Rev.*, vol. 33, no. 3, pp. 554–598, May 2013, doi: 10.1002/med.21255.

[13] J. Mortier, C. Rakers, R. Frederick, and G. Wolber, “Computational Tools for In Silico Fragment-Based Drug Design,” *Curr. Top. Med. Chem.*, vol. 12, no. 17, pp. 1935–1943, Sep. 2012, doi: 10.2174/156802612804547371.

[14] L. R. de Souza Neto *et al.*, “In silico Strategies to Support Fragment-to-Lead Optimization in Drug Discovery,” *Frontiers in Chemistry*, vol. 8. Frontiers Media S.A., p. 93, Feb. 2020, doi: 10.3389/fchem.2020.00093.

[15] M. Bissaro, M. Sturlese, and S. Moro, “The rise of molecular simulations in fragment-based drug design (FBDD): an overview,” *Drug Discov. Today*, vol. 25, no. 9, pp. 1693–1701, Sep. 2020, doi: 10.1016/j.drudis.2020.06.023.

[16] F. Ferrari *et al.*, “HT-SuMD: making molecular dynamics simulations suitable for fragment-based screening. A comparative study with NMR,” *J. Enzyme Inhib. Med. Chem.*, vol. 36, no. 1, pp. 1–14,

Jan. 2021, doi: 10.1080/14756366.2020.1838499.

[17] “No Title.” .

[18] A. Douangamath *et al.*, “Crystallographic and electrophilic fragment screening of the SARS-CoV-2 main protease,” *bioRxiv*, p. 2020.05.27.118117, May 2020, doi: 10.1101/2020.05.27.118117.

[19] N. Z. T. C. M. Consortium, H. Achdout, A. Aimon, E. Bar-David, H. Barr, A. Ben-Shmuel, J. Bennett, M. L. Bobby, J. Brun, S. BVNBS, M. Calmiano, A. Carbery, E. Cattermole, J. D. Chodera, A. Clyde, J. E. Coffland, G. Cohen, J. Cole, A. Contini, L. Cox, M. Cvitkov, “COVID Moonshot: Open Science Discovery of SARS-CoV-2 Main Protease Inhibitors by Combining Crowdsourcing, High-Throughput Experiments, Computational Simulations, and Machine Learning.”

[20] M. De Vivo, M. Masetti, G. Bottegoni, and A. Cavalli, “Role of Molecular Dynamics and Related Methods in Drug Discovery,” *J. Med. Chem.*, vol. 59, no. 9, May 2016, doi: 10.1021/acs.jmedchem.5b01684.

[21] V. Salmaso and S. Moro, “Bridging molecular docking to molecular dynamics in exploring ligand-protein recognition process: An overview,” *Frontiers in Pharmacology*, vol. 9, no. AUG. Frontiers Media S.A., Aug. 2018, doi: 10.3389/fphar.2018.00923.

[22] H. M. Berman, “The Protein Data Bank http://www.rcsb.org/pdb/,” *Nucleic Acids Res.*, vol. 28, no. 1, pp. 235–242, 2000, doi: 10.1093/nar/28.1.235.

[23] G. Bolcato, M. Bissaro, M. Pavan, M. Sturlese, and S. Moro, “Targeting the coronavirus SARS-CoV-2: computational insights into the mechanism of action of the protease inhibitors lopinavir, ritonavir and nelfinavir,” *Sci. Rep.*, vol. 10, no. 1, p. 20927, Dec. 2020, doi: 10.1038/s41598-020-77700-z.

[24] J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, and C. Simmerling, “ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB.,” *J Chem Theory Comput*, vol. 11, no. 8, pp. 3696–3713, Aug. 2015, doi: 10.1021/acs.jctc.5b00255.

[25] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, “Development and testing of a general amber force field,” *J. Comput. Chem.*, vol. 25, no. 9, pp. 1157–1174, Jul. 2004, doi: 10.1002/jcc.20035.

[26] D. Sabbadin and S. Moro, “Supervised Molecular Dynamics (SuMD) as a Helpful Tool To Depict GPCR–Ligand Recognition Pathway in a Nanosecond Time Scale,” *J. Chem. Inf. Model.*, vol. 54, no. 2, Feb. 2014, doi: 10.1021/ci400766b.

[27] A. Cuzzolin *et al.*, “Deciphering the Complexity of Ligand-Protein Recognition Pathways Using Supervised Molecular Dynamics (SuMD) Simulations,” *J. Chem. Inf. Model.*, vol. 56, no. 4, pp. 687–705, 2016, doi: 10.1021/acs.jcim.5b00702.

[28] F. Pedregosa *et al.*, “Scikit-learn: Machine learning in Python,” *J. Mach. Learn. Res.*, 2011.

[29] X. X. Ester, M., H. P. Kriegel, J. Sander, *A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise in Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining*. Portland: AAAI press, 1996.

[30] and P. A. K. D.A. Case, H.M. Aktulga, K. Belfon, I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, G.A. Cisneros, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, G. Giambasu, M.K. Gilson, H. Gohlke, A.W. Goetz, R. Harris, S. Izadi, S.A. Izmailov, C. Jin, K. Ka, “Amber 2021,” 2020.

[31] D. L. Mobley and K. A. Dill, “Binding of Small-Molecule Ligands to Proteins: ‘What You See’ Is Not Always ‘What You Get,’” *Structure*, vol. 17, no. 4, pp. 489–498, Apr. 2009, doi: 10.1016/j.str.2009.02.010.

[32] M. L. Verdonk, I. Giangreco, R. J. Hall, O. Korb, P. N. Mortenson, and C. W. Murray, “Docking

Performance of Fragments and Druglike Compounds," *J. Med. Chem.*, vol. 54, no. 15, pp. 5422–5431, Aug. 2011, doi: 10.1021/jm200558u.

[33] N. M. Lim, M. Osato, G. L. Warren, and D. L. Mobley, "Fragment Pose Prediction Using Non-equilibrium Candidate Monte Carlo and Molecular Dynamics Simulations," *J. Chem. Theory Comput.*, 2020, doi: 10.1021/acs.jctc.9b01096.

[34] M. Majewski and X. Barril, "Structural Stability Predicts the Binding Mode of Protein–Ligand Complexes," *J. Chem. Inf. Model.*, vol. 60, no. 3, pp. 1644–1651, Mar. 2020, doi: 10.1021/acs.jcim.9b01062.

A Computational Workflow for the Identification of Novel Fragments Acting as Inhibitors of the Activity of Protein Kinase CK1δ

Giovanni Bolcato, Eleonora Cescon, Matteo Pavan, Maicol Bissaro, **Davide Bassani**, Stephanie Federico , Giampiero Spalluto, Mattia Sturlese and Stefano Moro

G. Bolcato *et al.*, "A Computational Workflow for the Identification of Novel Fragments Acting as Inhibitors of the Activity of Protein Kinase CK1δ," *Int. J. Mol. Sci.*, vol. 22, no. 18, p. 9741, Sep. 2021, doi: 10.3390/ijms22189741

Abstract

Fragment-Based Drug Discovery (FBDD) has become, in recent years, a consolidated approach in the drug discovery process, leading to several drug candidates under investigation in clinical trials and some approved drugs. Among these successful applications of the FBDD approach, kinases represent a class of targets where this strategy has demonstrated its real potential with the approved kinase inhibitor Vemurafenib. In the Kinase family, protein kinase CK1 isoform δ (CK1δ) has become a promising target in the treatment of different neurodegenerative diseases such as Alzheimer's disease, Parkinson's disease, and amyotrophic lateral sclerosis. In the present work, we set up and applied a computational workflow for the identification of putative fragment binders in large virtual databases. To validate the method, the selected compounds were tested in vitro to assess the CK1δ inhibition.

1. Introduction

1.1. Protein Kinase CK1δ

Protein kinase CK1δ belongs to the family of CK1 Kinases (Casein Kinase 1), which in turn belongs to the class of Ser-Thr Kinases; seven isoforms of this family were identified in mammals: α, β, γ1, γ2, γ3, δ, ε. All the isoforms of CK1 are constitutionally active and they exhibit activity in monomeric form, They present a highly conserved catalytic domain (unlike in N and terminal C domains), they utilize ATP as a phosphate group donator and they are generally independent of the presence of a cofactor [1].

CK1δ and the other isoforms of the family of CK1 can phosphorylate Ser or Thr residues in sequences such as *(P)Ser/Thr-X_{1-2}-Ser/Thr,* where *(P)Ser/Thr* indicates a Ser or Thr pre-phosphorylated residue

[2]; CK1, therefore, needs the substrate to be already phosphorylated. Nevertheless, it has been demonstrated that a set of amino acids with acidic character in the direction of the N-terminal with respect to Ser/Thr target residue or an acidic residue in position 3 (preferably Asp) can provide for the lack of the pre-phosphorylated amino acid [3][4]. This allows CK1 to act also as a Priming Kinase activating the substrate towards a second enzyme by phosphorylation. Currently, about 140 substrates (in vitro or in vivo) recognized by CK1 have been described [1].

The activity of CK1 isoforms is regulated in different ways. Phosphorylation is the principal strategy adopted for the activity regulation of this family of kinases. CK1δ is phosphorylated by kinases such as Akt, PKA, PKCα, CLK2, and Chk1. Moreover, CK1δ can also be subjected to auto-phosphorylation [1],[5],[6]. Another fundamental aspect in the CK1δ activity regulation is the subcellular compartmentalization, operated through the binding to intracellular structures and other proteins [7],[8]. One last mechanism reported in the literature for the CK1δ regulation is the formation of homodimers [9],[10].

CK1δ, together with other CK1 isoforms, has been correlated to several neurodegenerative processes [11]; in particular, CK1 seems implied in tauopathies, among which Alzheimer's disease (AD) is the most representative one.

AD is associated with several cellular processes. The first mechanism described is correlated to Tau protein, which after phosphorylation tends to come off from the microtubules forming aggregates at a cytoplasmatic level, leading to cellular damage. A second mechanism implies instead production and accumulation, with consequent cellular death, of the β-amyloid peptide. This is produced by the cut of its precursor APP (Amyloid Precursor Protein) by β-secretase 1 and γ-secretase enzymes. The implications of CK1 isoforms in pathogenetic processes at the root of Alzheimer's disease are many. In general, CK1δ proves to be overexpressed in brain tissue, up to 30 times in patients affected by Alzheimer's disease [12],[13].

Concerning Tau protein, initially, it was observed how CK1 turns out to be associated with fibrillar masses of hyperphosphorylated Tau protein (Paired Helical filaments) [14]; in particular, CK1δ seems to be accumulated within these fibrillar masses [15]. Later it was demonstrated how CK1δ can phosphorylate Tau protein causing its separation from microtubules; the residues of Tau phosphorylated by CK1δ are Ser202, Thr205, Ser396, and Thr404 [11],[16]. As regards β-amyloid peptide, it was described how this can stimulate the activity of CK1 and CK2 (employing casein as a substrate) [17]. Likewise, there is evidence that CK1 activity would be proportionally correlated to β-amyloid peptide production, since in presence of constitutionally active CK1 forms the amount of

this peptide increases, whereas it decreases in presence of CK1 inhibitors. CK1 interference seems to take place along with the γ-secretase enzyme [18], but it is more likely correlated to CK1ε isoform, than to CK1δ[19].

As regards Parkinson's Disease, it has been observed how CK1 isoforms phosphorylated Ser129 of α-synuclein [11],[20].

Amyotrophic lateral sclerosis (ALS) is another neurodegenerative disease where CK1δ plays a role. Indeed, CK1δ phosphorylated TDP-43 (TransActivate Response DNA Binding Protein 43) at many different residues. TDP-43 is the principal component of the protein aggregates observed in the pathogenesis of ALS [21],[22].

1.2. Fragment-Based Drug Discovery (FBDD) Principles.

FBDD is a strategy used in drug discovery that has gained popularity both in the industrial and academic contexts. In a typical FBDD process a library of polar low molecular weight compounds is screened against a specific target. Usually, the screening is performed by biophysical methods including X-ray crystallography, nuclear magnetic resonance (NMR), thermal shift assay, and surface plasmon resonance (SPR). One of the key factors in the FBDD success is the smaller size of the fragment-like chemical space compared to the size of the drug-like one. The size of the drug-like chemical space has been estimated at around 10^{60} compounds, many orders of magnitude greater than that of the fragment-like compounds' chemical space [23]. This means that, through the screening of fragments, the portion of chemical space sampled is larger than the one sampled with the screening of drug-like molecules. This will promisingly also allow the attainment of innovative scaffolds for drug candidates.

Despite the hit fragments having typically a low affinity, they could be turned into a lead compound that efficiently binds the target. Fragments, having a low molecular weight, establish few interactions with the target; however, the combination of multiple fragments by linking and merging or by decorating them with adequate functional groups (fragment growing) allows the development of specific and more affine compounds.

1.3. Fragment-Based Drug Discovery and Kinase Inhibitors

Concerning the identification of kinase inhibitors through an FBDD approach, X-ray crystallography has also been largely employed because kinases represent a class of protein that provides good results with this technique.

The most outstanding example of kinase inhibitors derived from an FBDD approach is Vemurafenib (inhibitor of BRAF) which is an approved drug for the treatment of metastatic melanoma [24]. The discovery of vemurafenib started with an enzymatic assay screening of a fragment library. The hit compounds identified were analyzed through X-ray crystallography, using the structural information so obtained one fragment was chosen for optimization leading at the end to Vemurafenib [25]. Another notable example is Asciminib an allosteric inhibitor of BCR-ABL1 tyrosine kinase, now in phase III clinical trial for resistant chronic myeloid leukemia. This compound was identified from an NMR-based fragment screening; the fragment hits identified were then optimized using In Silico methodologies, X-ray crystallography, and NMR [26],[27].

Many other Kinase inhibitors derived from FBDD approaches are in clinical trials; for a comprehensive review of FBDD derived drugs that have been approved or which are in clinical trials see [28].

An interesting observation is that the fragments identified often bind at the hinge region of the kinase and maintain this binding mode also in the mature compound. For this reason, the library of compounds tested in the present work has been focalized, using in silico methodologies described in the next sections, to be composed of putative hinge-binding fragments.

1.4. Computational Methods in FBDD

Since the dawn of FBDD, computational chemistry has played a major role in both fragments' hit identification and in the process of fragment optimization. The MCSS (multiple copy simultaneous search) algorithm [29] was a pioneering method for the study of fragment binding modes in a protein site. Another method for fragment posing based on grand canonical Monte Carlo (GCMC) has been reported [30].

Over the years many in silico methods have been proposed non only for fragment placement prediction but also to aid the fragment optimization process. Software like LUDI [31], HOOK [32], CAVEAT [33], RECORE [34], and many others have been developed for this purpose. Additionally, Schrodinger [35] and CCG [36] implement in their software suites many tools to aid the fragment optimization process.

Molecular dynamics (MD)-based tools represent the most advanced in silico techniques used in FBDD. The first application of MD to FBDD was the refinement of docking poses, a method note as post-docking [37]. More advanced protocols have also been developed. Nonequilibrium candidate

Monte Carlo (NCMC) is an algorithm that has been applied to enhance the sampling of fragment binding modes [38]; this method has been successfully applied to FBDD [39]. Another promising approach is the application of Markov state models to MD simulations, which has proved its potential to FBDD [40]. Recently, Supervised Molecular Dynamics (SuMD) [41] has been applied as a fragment screening tool [42].

Molecular docking has also become a routinely used tool in FBDD. While the conformational sampling performs by docking protocols is generally effective in reproducing the correct pose for a ligand, the scoring functions frequently fail in valuating this pose [43], this is especially true for Fragment-like compounds for which many doubts have been raised about the docking applicability [44]. This said, to make the docking results more reliable a consensus docking approach was used [45], and instead of the scoring function, the poses were evaluated using a pharmacophore model. A post-docking refinement of the poses was then performed. A detailed explanation of the computational workflow adopted in the present work is reported in Section 4.1, Section 4.2, and Section 4.3.

2. Results

2.1. Computational Results

A library of around 272,000 commercially available fragment compounds was screened in silico using an integrated structure-based approach based on different techniques such as molecular docking, molecular dynamics (MD), and pharmacophore filter. The workflow adopted is reported in Figure 1.

At first, three independent docking-based virtual screenings were performed in parallel exploiting three different protocols: PLANTS-ChemPLP, GOLD-ChemScore, and Glide-SP. PLANTS exploits an Ant-Colony Optimization (ACO) algorithm, GOLD a genetic one while Glide performs an exhaustive search. The choice of these three protocols was made to evaluate the virtual library with three orthogonal search algorithms, to minimize the false-positive rate to which traditional docking-based virtual screenings are prone. At the end of each virtual screening, a total of about 13.6 M poses (50 per ligand) was obtained for each protocol. The choice to generate such a great number of poses for each ligand was taken in order not to rely on the scoring function ability to prioritize the best binding mode for each compound, since fragments can have multiple binding modes that are similar

from an energetic and qualitative point of view and are therefore difficult to distinguish for scoring functions that are trained upon mature, lead-like, compounds.

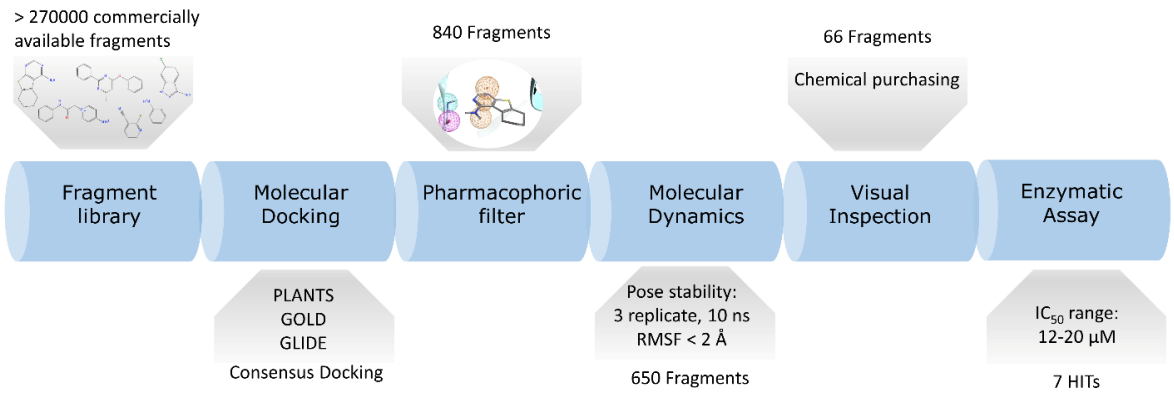


Figure 1. *Schematic representation of the workflow adopted in the present work. First the fragments are retrieved from several vendors libraries. After proper preparation, the database is docked using three different docking protocols. the resulting poses have been filtered using a pharmacophore model and only the molecule that fit the model for each protocol have been retained. The poses of these molecules were further refined using MD to assess the stability of the binding mode. the molecules that appear to be stable were finally selected trough visual inspection.*

To filter this huge amount of ligand conformations and retain only the most interesting compounds, we decided to exploit the structural knowledge provided by the 23 Ck1d protein–ligand complexes deposited in the Protein Data Bank and create a pharmacophore filter. This pharmacophore model was built to retain those features which are vital for the interaction with the hinge region of the kinase since these features are the most commonly found across the structures. The pharmacophore included three features, two of them to guarantee the interaction with Leu85 (a hydrogen bond donor and a hydrogen bond acceptor) and the presence of and a feature for an aromatic ring also in the proximity of the hinge region.

The pharmacophore filter was then applied independently on each pose database generated by the three different docking protocols. Exploiting an approach known as consensus docking, the three libraries containing those ligand conformations that fit the pharmacophore model were merged, retaining only those found within each dataset. After this consensus filtering, only 840 docking poses were left.

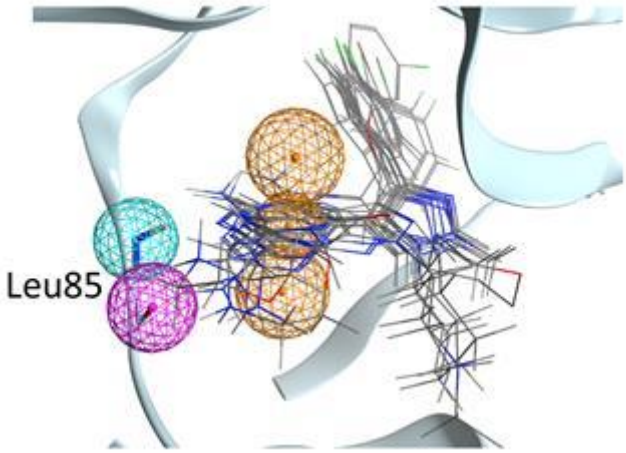


Figure 2. *Representation of the pharmacophore model used in the present work. Some representative crystallographic ligands are displayed (not all for clarity). The Pharmacophore model is formed by an aromatic ring (the three orange spheres define the position and its orientation) and two hydrogen bonds with the backbone of Leu85 (an acceptor and one donor).*

To further filter out those poses characterized by unstable binding modes, a post-docking molecular dynamics refinement was performed (three replicates, 10 ns each). The average Root Mean Squared Fluctuation of atomic positions (RMSF) across the three replicates was used as a cutoff to eliminate those poses characterized by conformational instability over time. After filtering out those ligand conformations with RMSF > 2Å, 650 stable poses were maintained.

With the intent of prioritizing the most interesting compound for in vitro assays, each pose was carefully manually examined. After this visual inspection [46] step, 66 fragments were finally selected to be purchased and tested. The structure of all the 66 fragment compounds tested are reported in supplementary Table S1, while the pose of each of them resulted from the VS pipeline is reported in Video S1.

2.2. Enzymatic Assay Results

Fragments were tested against CK1δ using a luminescent-based assay. Compounds were evaluated at a fixed concentration of 100 µM(see Figure 3) and those that showed a kinase residual activity lower than 40% were tested also at a fixed concentration of 40 µM (see Figure 4).

IC_{50} values were calculated for compounds with a residual kinase activity lower than 40%. Compounds **37**, **38**, **52**, **59**,**62** and **63** showed IC_{50} values in the micromolar range of 12.71 µM (9.57–16.80), 20.49 µM (17.46–24.08), 13.50 µM (12.47–14.62), 13.92 µM (11.89–16.29), 18.15 µM (16.78–19.64) and 24.86 µM (21.46–28.92), respectively. Remarkably, compound **28** shows a half-

maximal inhibitory concentration of 3.31 µM (2.67–4.12). The IC_{50} curves for the seven hits are reported on SI. The value of IC50 is based on the average of three independent measurements.

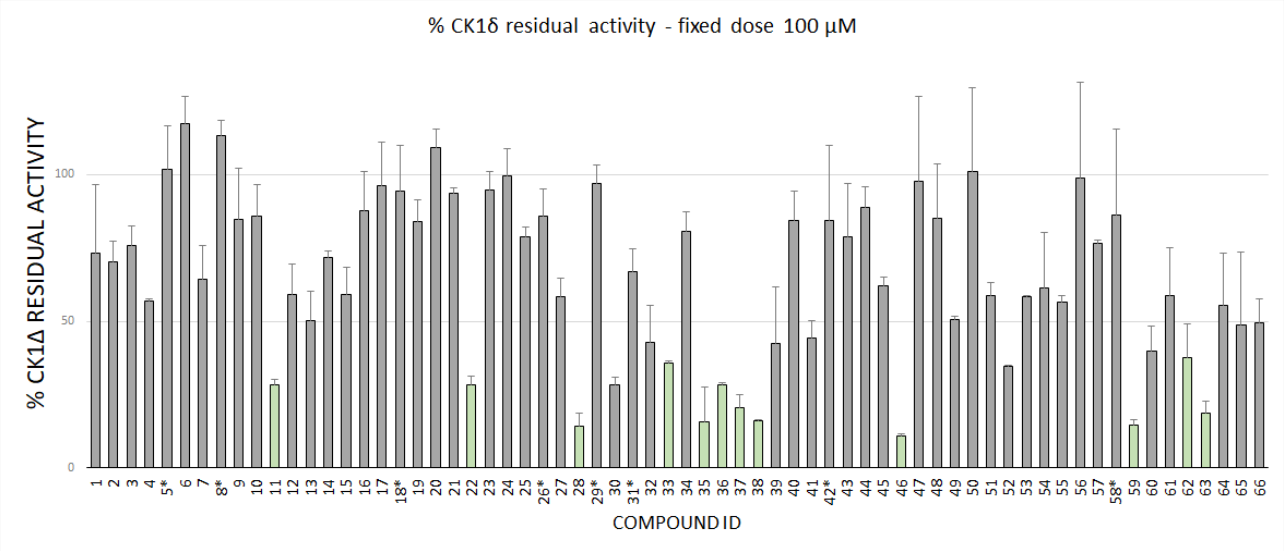


Figure 3. *CK1δ residual activity at a concentration of 100 µM of the ligand under examination.The molecules marked with a star has been tested at 50 µM due to solubility issues.*

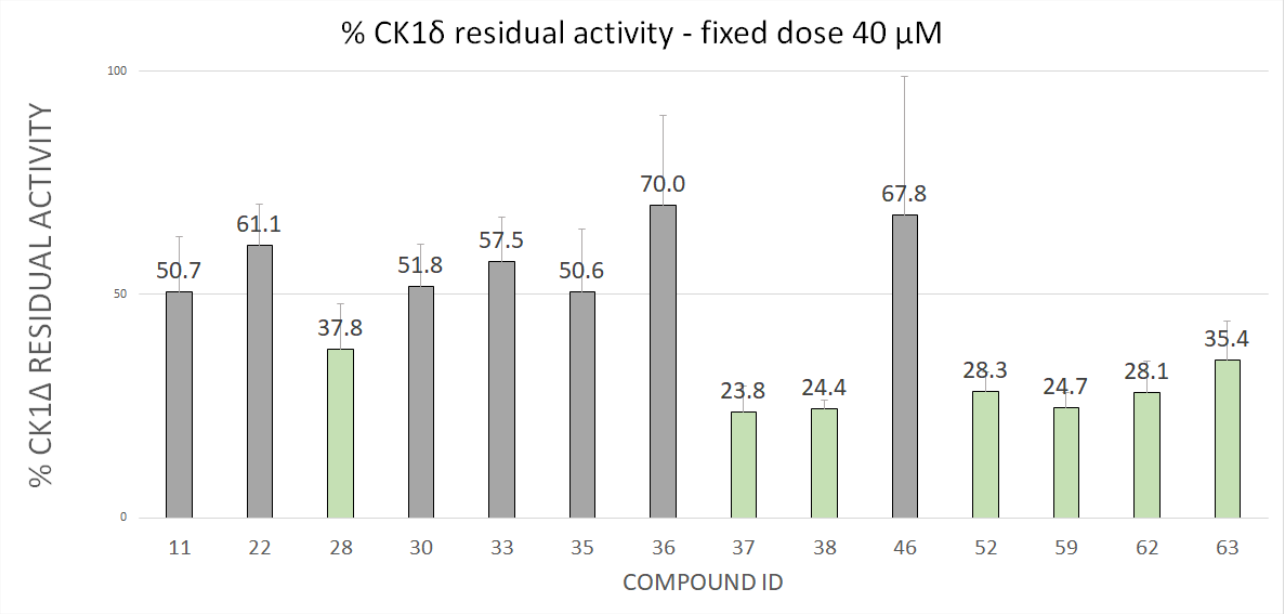


Figure 4. *CK1δ Residual activity at a concentration of 40 µM of the ligands that showed a residual activity of less than 40% at 100 µM.*

2.3. Molecular Recognition Studies of the Most Promising Fragment

To shed light on the possible recognition mechanism of the most effective inhibitor, compound **28** (IC_{50} = 3.31 µM) was investigated by mean of Supervised Molecular Dynamics simulations (SuMD). The primary scope was to assess if the hypothesized bound state obtained by our computational protocol was also accessible by simulating the fragment association from the unbound state without

any information about the ligand conformation. Since in our VS-pipeline the pharmacophoric filter plays a primary goal in defining the bound geometries, its validation by using a more articulated technique based on MD and in which the water molecules need to be displaced by the fragment to reach the hinge region would provide the reliability of the binding mode.
A complete recognition pathway of the length of 15 ns is reported in Video S2 (SI). Compound **28** showed three steps during the recognition, with two stable states (Figure 5A).

A pivotal role in the first phases (around 1 ns time mark) of the ligand recruitment within the binding site is played by Asp149, which acts as an electrostatic recruiter for the amino-thiophene moiety of the ligand. By contrast, the vicinal residue Lys38 hampers the ligand entrance into the core portion of the binding site due to the electrostatic repulsion between the charged amino group of the amino acid side chain and the non-charged amino group of the ligand. The balance in attraction and repulsion between the flexible side chains of these two amino acids located at the boundary of the binding site is depicted also by the large energetic funnel shown in Figure 5A at around 10 Å with regard to the distance between the centers of mass of the binding site and the ligand (dcm_{L-R}).

Afterwards, the binding pathway is characterized by two stable ligand conformations within the binding site. The first state (S1) occurred at a dcm_{L-R} distance of 4.5 Å, with the ligand interacting with the backbone of Leu85 through its amino-thiophene moiety and the morpholine moiety oriented towards the external part of the binding site (solvent-exposed), while the second one (S2) at a dcm_{L-R} distance of 1.5 Å is characterized by a bivalent hydrogen bond with Leu85 and the morpholine moiety of the ligand buried within the hydrophobic selectivity pocket defined by Met80, Met82, Ile23 and the alkyl portion of the Lys38 side chain. Although these two states are characterized by similar interaction energy values (according to the AMBER forcefield), their energetic funnels have different shapes: the final state (S2) shows a narrower profile than the S1 state, suggesting that the pharmacophore binding mode (S2) has a higher stability than S1. Furthermore, the final bound state nicely retraced the pose obtained with the VS pipeline, validating both the pharmacophore model used in this work and the binding mode proposed by molecular docking for this compound (Figure 5B)

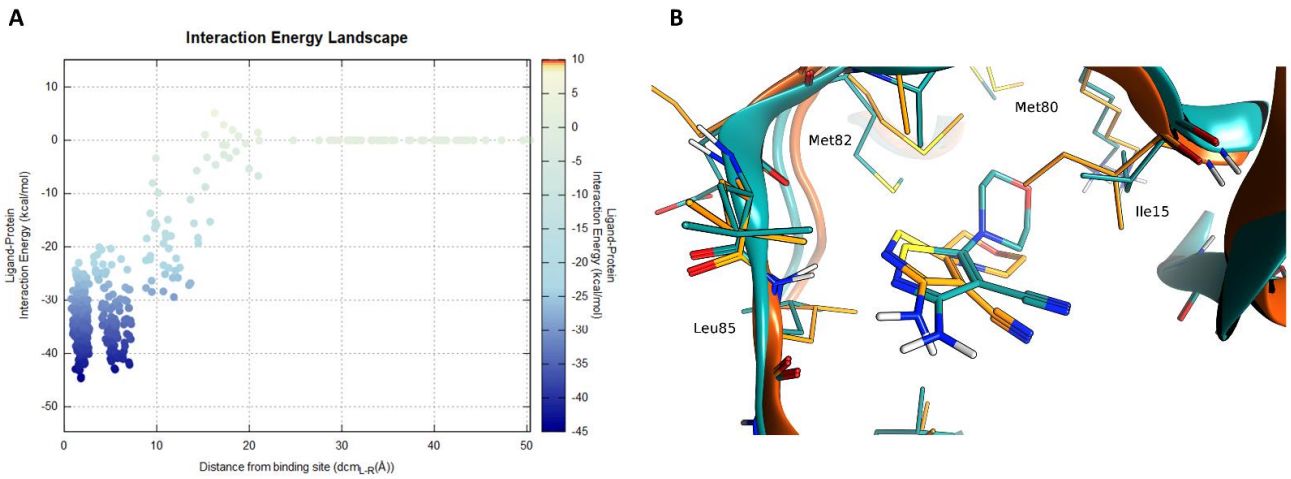
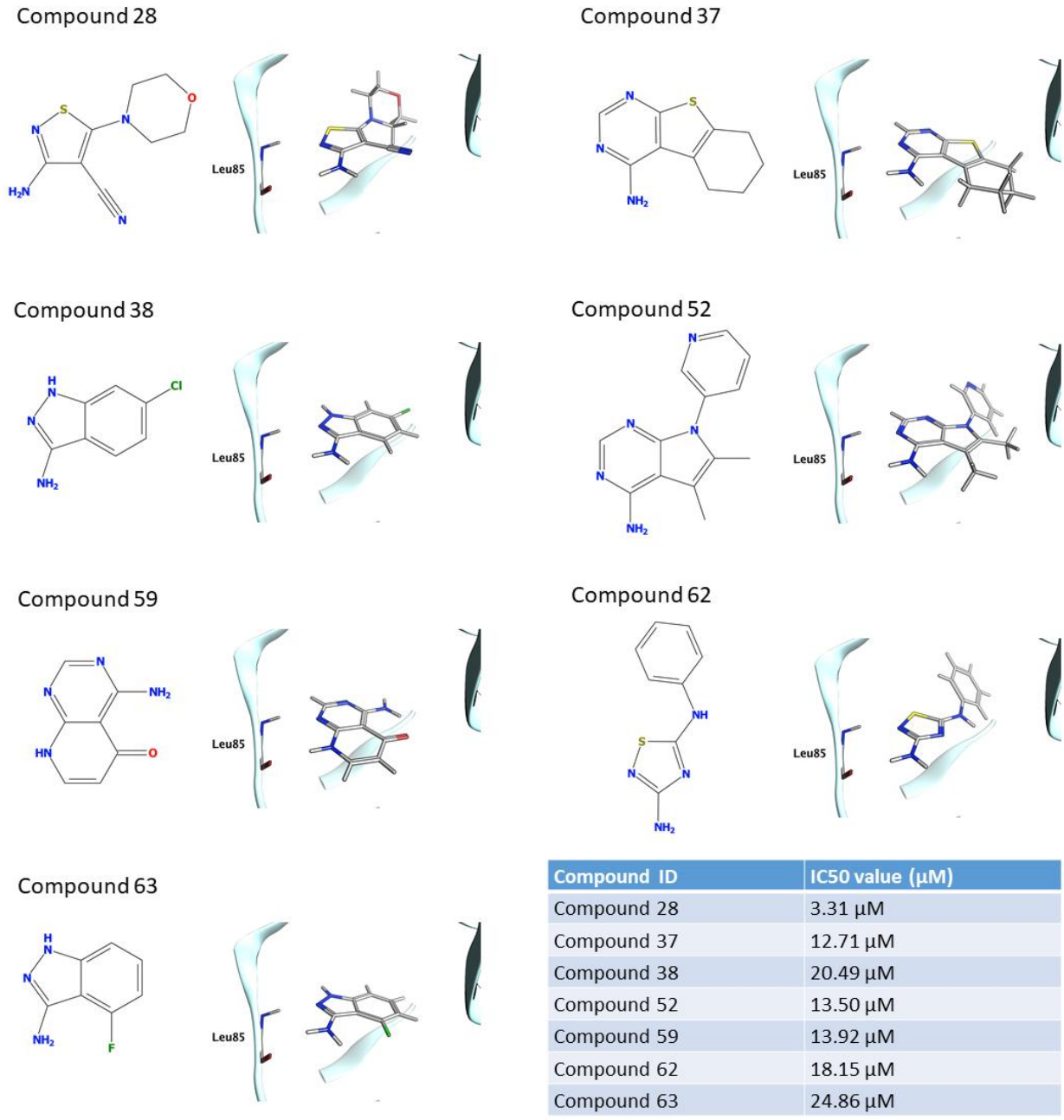


Figure 5. *SuMD simulation of compound **28.** In panel (**A**) the interaction energy landscape is reported for the recognition trajectory displaying the ligand–protein interaction energy plotted against the distances between the protein–ligand center of mass. In panel (**B**), the superposition of the VS-pose (cyan) for compound **28** against the lowest energy frame from the SuMD trajectory (orange).*

3. Discussion

The seven fragments that were characterized by calculating the IC_{50} showed a noticeable chemical diversity including scaffolds spanning from one to three nitrogen-containing fused rings. The poses of the seven hits as obtained in the VS are reported in Figure 6. All the fragments logically share the common interaction pattern required by the pharmacophore filter. Interestingly, compounds **28**, **37**, **38**, **52**, **62**, and **63** showed a similar interaction scheme in which an aromatic amine moiety was able to establish a hydrogen bond with the carbonyl oxygen of the Leu85 backbone while a further hydrogen bond between the Leu85 backbone amide is guaranteed by aromatic nitrogen in ortho to the amine group. Compounds **37**, **52**, and **59** share a conserved pyrimidine ring that is part of different fused systems. Compound **59** also has the pyrimidine ring in a different orientation: it restores the hydrogen bond donor by its fused pyridone ring. Compounds **38** and **63** present the same scaffold. To assess the novelty of the identified fragments, a substructure search was performed against ChEMBL using the main ring recognized by the pharmacophore as a query; except for compounds **38** e **52**, which resulted in **34** and **20** already known CK1δ inhibitors, for all the remaining hits none known inhibitors were found sharing the principal ring. The 3-amino-indazole scaffold of compound **38** was found in a multikinase inhibitor (CHEMBL1999931) with a Ki of 316.23 nM [47]. For compound **52** a couple of ligands with low uM activity were found; in particular CHEMBL2000114 with a Ki of 1 uM arose from the same kinome scan from Abbott Labs [47]. Additionally, compound GSK1838705A showed the same scaffold of **52**, in this case the Ki reported

is 3.5 uM but it is a residual activity since the compound is a potent inhibitor of ALK kinase (IC_{50} = 0.5 nM) [48].



Compound ID	IC50 value (µM)
Compound 28	3.31 µM
Compound 37	12.71 µM
Compound 38	20.49 µM
Compound 52	13.50 µM
Compound 59	13.92 µM
Compound 62	18.15 µM
Compound 63	24.86 µM

Figure 6. *The structure and binding mode for the seven compounds for which the IC50 value is reported. The value of IC50 is based on the average of three independent measurements.*

4. Materials and Methods

4.1. Molecular Modelling and Docking

The virtual library used in this work was obtained through the merging of different libraries of commercially available compounds designed for FBDD. The vendors are Asinex, Chembridge, Enamine, Life Chemicals, Maybridge, Otava, Timtec, Vitas. The total number of fragments in the merged library is about 272,000 virtual compounds.

The merged library was prepared to be suitable for the Docking-Based Virtual Screening. This preparation consists of the following steps: the tautomeric state enumeration for each compound and determination of the most probable tautomer (for each molecule at the three most tautomeric states was retained), the most probable ionization state at pH 7.4 calculation, the atomic partial charge calculation (using MMFF94 force field), the 3D coordinates generation. All these steps were performed using QUACPAC of the Openeye suite [49] except for the 3D coordinated generation for which Corina Classic was used [50].

The protein used both for Docking and for MD simulation was prepared using MOE. The preparation consists of the removal of the crystallographic water molecules and other solvent molecules together with ions and the ligand. The correct protonation state for each residue at pH 7.4 was calculated with the Protonate3D tool of MOE.

For the Consensus Docking strategy, three different Molecular Docking protocols were used. To make the results more robust, the three docking protocols chosen rely on search algorithms of different types. The Molecular Docking Protocols are PLANTS [51] which is based on an Ant Colony Optimization algorithm, GOLD [52],[53] which employs a genetic algorithm, and Glide [54],[55] which use a systematic searching approach. The Scoring Functions adopted are CHEMPLP for PLANTS, ChemScore for GOLD, and Glide SP for Glide. For each fragment 50 poses were generated using each Docking Protocol even if the termination criteria and the nature of the algorithms did not always provide 50 poses, in particular for Glide SP.

Similarity and substructure searches were performed with MOE using the ChEMBL29 database.

4.2. Pharmacophore Modeling

Each ensemble of poses (one for each docking protocol) was then filtered using a pharmacophore model. This pharmacophore model was calculated using MOE: all the holo crystal structures available on the PDB for human CK1δ were superposed and the common features of each ligand

were analyzed. The list of complexes included 23 complexes with PDB ID: 3UYT, 3UZP, 4HGT, 4HNF, 4KB8, 4KBA, 4KBC, 4KBK, 4TN6, 4TW9, 4TWC, 5IH5, 5IH6, 5MQV, 5OKT, 5W4W, 6F1W, 6F26, 6GZM, 6HMP, 6HMR, 6RCG, 6RCH.

Since the ligands present in the crystal structures are drug-like molecules, it is difficult that a fragment can comply with all the common features observed in the crystal structures. For this reason (and because as stated above the first fragment identified in an FBDD process of a kinase inhibitor is a hinge binding fragment) the pharmacophore model was built using only the features involved in the interaction with the hinge region of the kinase. The model included three features: one hydrogen bond donor and one hydrogen bond acceptor to guarantee the interaction with the backbone of Leu85 (Figure 2). The last feature represents an aromatic ring also in the proximity of the hinge region. Only the molecule that has passed the Pharmacophore filtering for each protocol hwas retained (*consensus*).

4.3. Molecular Dynamics

The molecules retained after the consensus filtering were subjected to a post-docking refinement. The docking pose used in this step is the one obtained from Glide. All the simulations were carried out using ACEMD3 [56] with ff14SB as force field [57], the system preparation was conducted with MOE concerning protein preparation and with the use of AmberTools14 for the simulation box preparation.

For each complex, a simulation box was prepared: the protein was immersed in an explicit TIP3P [58] solvent box, with an ionic strength of 0.154 M obtained using Na^+/Cl^-. The protein is 15Å away from the border of the box.

Using the conjugate gradient method, the system energy was minimized for 500 steps; after this minimization the system was equilibrated in two stages. The first equilibration consists of 1ns of NVT simulation with harmonic positional constraints of 1 kcal mol^{-1} $Å^{-2}$ on the protein. In the second equilibration step, which consists in this case of 1ns of NPT simulation, the constraints of 1 kcal mol^{-1} $Å^{-2}$ were applied only on the α carbons of the protein. After the equilibration for each protein–pose complex, three NVT trajectories of 10 ns were produced. The average RMSF of the ligand during these three replicas was calculated and if this value was greater than 2Å the molecule was discarded. A Supervised Molecular Dynamics[59],[41] simulation was performed to gain some insights into the binding process of the most potent fragment (Compound **28**). SuMD is an MD-based method

developed to investigate molecular binding events without energetic biases. The algorithm is based on the supervision of the ligand–protein binding site center of mass distance during a classical short MD simulation. At the end of each small simulation (SuMD step), this distance is measured: if it has shortened during the SuMD step, the simulation continues with another SuMD step, otherwise, it is stopped, and the simulation restarts from the previous set of coordinates. The fragment was placed 30 Å away from the protein. Each SuMD step was set to 300 ps.

4.4. Enzymatic Assay

Compounds were evaluated towards CK1δ (aa 1-294, Merck Millipore) with the KinaseGlo® luminescence assay (Promega) following procedures reported in the literature [22]. In detail, luminescent assays were performed in white 96-well plates, using the following buffer: 50 mM HEPES (pH 7.5),1 mM EDTA, 1 mM EGTA, and 15 mM $MgCl_2$. Compound PF-670462 (IC50 = 14 nM) was used as a positive control for CK1δ [60] and DMSO/buffer solution was used as a negative control. In a typical assay, 10 µL of inhibitor solution (dissolved in DMSO at 10 mM concentration and diluted in assay buffer to the desired concentration) and 10 µL (16 ng) of enzyme solution were added to each well, followed by 20 µL of assay buffer containing 0.1% casein substrate and 4 µM ATP. The final DMSO concentration in the reaction mixture did not exceed 1%.

After 60 min of incubation at 30 °C, the enzymatic reactions were stopped with 40 µL of KinaseGlo® reagent (Promega). The luminescence signal (relative light unit, RLU) was recorded after 10 min at 25 °C using Tecan Infinite M100. Fixed-dose experiments were performed at 100 µM and for more potent compounds also at 40 µM. Two independent experiments were performed in duplicate and the corresponding residual activity of CK1δ was obtained. Data were analyzed using Excel and reported as the mean of the two experiments with standard deviation. For IC_{50} determination ten different inhibitor concentrations ranging from 100 to 0.026 µM were used and each point was assessed in duplicate. IC_{50} values are the mean of three independent experiments and 95% confidence limits were also reported. Data were analyzed using GraphPad Prism software (version 8.0).

5. Conclusions

In the present work to find new potential CK1δ inhibitors, we elaborated a computational workflow for the identification of candidate hinge binding fragments. This workflow consists of the generation of a large number of poses for each compound of a virtual library of commercially available

fragments using three different Docking protocols. These poses were filtered using a pharmacophore model and only the fragment for which each docking protocol was able to produce a pose that fits the model was retained (consensus docking). In the next, step each protein-fragment complex that passed the previous filter was subjected to an MD-driven post-docking refinement to inspect the geometric stability of the pose. Finally, some fragments were manually selected among the group that demonstrated a good performance in the post-docking refinement; to validate the method these fragments were tested using an enzymatic assay test to assess the CK1δ residual activity, and for the most promising candidates, the IC_{50} value was determined, with a value in the low micromolar range. Five of seven fragments showed novel scaffolds for CK1δ, confirming that the proposed pipeline could be particularly useful to identify novel structures.

References

[1] U. Knippschild *et al.*, “The CK1 Family: Contribution to Cellular Stress Response and Its Role in Carcinogenesis,” *Front. Oncol.*, vol. 4, May 2014, doi: 10.3389/fonc.2014.00096.

[2] F. Meggio, J. W. Perich, E. C. Reynolds, and L. A. Pinna, “A synthetic β-casein phosphopeptide and analogues as model substrates for casein kinase-1, a ubiquitous, phosphate directed protein kinase,” *FEBS Lett.*, 1991, doi: 10.1016/0014-5793(91)80614-9.

[3] V. Pulgar, O. Marin, F. Meggio, C. C. Allende, J. E. Allende, and L. A. Pinna, “Optimal sequences for non-phosphate-directed phoshorylation by protein kinase CK1 (casein kinase-1) - A re-evaluation,” *Eur. J. Biochem.*, 1999, doi: 10.1046/j.1432-1327.1999.00195.x.

[4] O. MARIN, F. MEGGIO, S. SARNO, M. ANDRETTA, and L. A. PINNA, “Phosphorylation of synthetic fragments of inhibitor-2 of protein phosphatase-1 by casein kinase-1 and -2: Evidence that phosphorylated residues are not strictly required for efficient targeting by casein kinase-1,” *Eur. J. Biochem.*, 1994, doi: 10.1111/j.1432-1033.1994.tb19037.x.

[5] J. Bischof, S. J. Randoll, N. Süßner, D. Henne-Bruns, L. A. Pinna, and U. Knippschild, “CK1δ Kinase Activity Is Modulated by Chk1-Mediated Phosphorylation,” *PLoS One*, 2013, doi: 10.1371/journal.pone.0068803.

[6] P. R. Graves and P. J. Roach, “Role of COOH-terminal phosphorylation in the regulation of casein kinase Iδ,” *J. Biol. Chem.*, 1995, doi: 10.1074/jbc.270.37.21689.

[7] D. M. Milne, P. Looby, and D. W. Meek, “Catalytic activity of protein kinase CK1δ (casein kinase 1 δ) is essential for its normal subcellular localization,” *Exp. Cell Res.*, 2001, doi: 10.1006/excr.2000.5100.

[8] P. Xu *et al.*, “Structure, regulation, and (patho-)physiological functions of the stress-induced protein kinase CK1 delta (CSNK1D),” *Gene*, vol. 715, Oct. 2019, doi: 10.1016/j.gene.2019.144005.

[9] K. L. Longenecker, P. J. Roach, and T. D. Hurley, “Crystallographic studies of casein kinasc I δ: Toward a structural understanding of auto-inhibition,” *Acta Crystallogr. Sect. D Biol. Crystallogr.*, 1998, doi: 10.1107/s0907444997011724.

[10] H. Hirner *et al.*, “Impaired CK1 delta activity attenuates SV40-induced cellular transformation in vitro and mouse mammary carcinogenesis in Vivo,” *PLoS One*, 2012, doi: 10.1371/journal.pone.0029709.

[11] D. I. Perez, C. Gil, and A. Martinez, “Protein kinases CK1 and CK2 as new targets for neurodegenerative diseases,” *Medicinal Research Reviews*. 2011, doi: 10.1002/med.20207.

[12] N. Ghoshal *et al.*, “A new molecular link between the fibrillar and granulovacuolar lesions of Alzheimer’s disease,” *Am. J. Pathol.*, 1999, doi: 10.1016/S0002-9440(10)65219-4.

[13] K. Yasojima, J. Kuret, A. J. Demaggio, E. McGeer, and P. L. McGeer, “Casein kinase 1 delta mRNA is upregulated in Alzheimer disease brain,” *Brain Res.*, 2000, doi: 10.1016/S0006-8993(00)02200-9.

[14] J. Kuret, G. S. Johnson, D. Cha, E. R. Christenson, A. J. DeMaggio, and M. F. Hoekstra, “Casein kinase 1 is tightly associated with paired-helical filaments isolated from Alzheimer’s disease brain,” *J. Neurochem.*, 1997, doi: 10.1046/j.1471-4159.1997.69062506.x.

[15] C. Schwab, A. J. Demaggio, N. Ghoshal, L. I. Binder, J. Kuret, and P. L. McGeer, “Casein kinase 1 delta is associated with pathological accumulation of tau in several neurodegenerative diseases,” *Neurobiol. Aging*, 2000, doi: 10.1016/S0197-4580(00)00110-X.

[16] G. Li, H. Yin, and J. Kuret, “Casein Kinase 1 Delta Phosphorylates Tau and Disrupts Its Binding to Microtubules,” *J. Biol. Chem.*, 2004, doi: 10.1074/jbc.M314116200.

[17] A. Chauhan, V. P. S. Chauhan, N. Murakami, H. Brockerhoff, and H. M. Wisniewski, “Amyloid β-protein stimulates casein kinase I and casein kinase II activities,” *Brain Res.*, 1993, doi: 10.1016/0006-8993(93)90479-7.

[18] M. Flajolet, G. He, M. Heiman, A. Lin, A. C. Nairn, and P. Greengard, “Regulation of Alzheimer’s disease amyloid-β formation by casein kinase I,” *Proc. Natl. Acad. Sci. U. S. A.*, 2007, doi: 10.1073/pnas.0611236104.

[19] N. Höttecke *et al.*, “Inhibition of γ-secretase by the CK1 inhibitor IC261 does not depend on CK1δ,” *Bioorganic Med. Chem. Lett.*, 2010, doi: 10.1016/j.bmcl.2010.02.110.

[20] J. Kosten *et al.*, “Efficient modification of alpha-synuclein serine 129 by protein kinase CK1 requires phosphorylation of tyrosine 125 as a priming event,” *ACS Chem. Neurosci.*, 2014, doi: 10.1021/cn5002254.

[21] T. Nonaka *et al.*, “Phosphorylation of TAR DNA-binding protein of 43 kDa (TDP-43) by truncated casein kinase 1δ triggers mislocalization and accumulation of TDP-43,” *J. Biol. Chem.*, 2016, doi: 10.1074/jbc.M115.695379.

[22] I. G. Salado *et al.*, “Protein kinase CK-1 inhibitors as new potential drugs for amyotrophic lateral sclerosis,” *J. Med. Chem.*, vol. 57, no. 6, pp. 2755–2772, Mar. 2014, doi: 10.1021/jm500065f.

[23] R. J. Hall, P. N. Mortenson, and C. W. Murray, “Efficient exploration of chemical space by fragment-based screening.,” *Prog Biophys Mol Biol*, vol. 116, no. 2–3, pp. 82–91, Dec. 2014, doi: 10.1016/j.pbiomolbio.2014.09.007.

[24] K. T. Flaherty, U. Yasothan, and P. Kirkpatrick, “Vemurafenib,” *Nat. Rev. Drug Discov.*, 2011, doi: 10.1038/nrd3579.

[25] G. Bollag *et al.*, “Vemurafenib: The first drug approved for BRAF-mutant cancer,” *Nature Reviews Drug Discovery*. 2012, doi: 10.1038/nrd3847.

[26] D. Romero, “Initial results with asciminib in CML,” *Nature Reviews Clinical Oncology*. 2020, doi: 10.1038/s41571-019-0324-z.

[27] J. Schoepfer *et al.*, “Discovery of Asciminib (ABL001), an Allosteric Inhibitor of the Tyrosine Kinase Activity of BCR-ABL1,” *J. Med. Chem.*, 2018, doi: 10.1021/acs.jmedchem.8b01040.

[28] D. A. Erlanson, S. W. Fesik, R. E. Hubbard, W. Jahnke, and H. Jhoti, “Twenty years on: The impact of fragments on drug discovery,” *Nature Reviews Drug Discovery*, vol. 15, no. 9. Nature Publishing Group, pp. 605–619, Sep. 2016, doi: 10.1038/nrd.2016.109.

[29] A. Miranker and M. Karplus, “Functionality maps of binding sites: A multiple copy simultaneous search method,” *Proteins Struct. Funct. Bioinforma.*, 1991, doi: 10.1002/prot.340110104.

[30] M. Clark, S. Meshkat, G. T. Talbot, P. Carnevali, and J. S. Wiseman, “Fragment-based computation of binding free energies by systematic sampling,” *J. Chem. Inf. Model.*, 2009, doi: 10.1021/ci900132r.

[31] H. J. Böhm, “The computer program LUDI: A new method for the de novo design of enzyme inhibitors,” *J. Comput. Aided. Mol. Des.*, 1992, doi: 10.1007/BF00124387.

[32] M. B. Eisen, D. C. Wiley, M. Karplus, and R. E. Hubbard, “HOOK: A program for finding novel molecular architectures that satisfy the chemical and steric requirements of a macromolecule binding site,” *Proteins Struct. Funct. Bioinforma.*, 1994, doi: 10.1002/prot.340190305.

[33] G. Lauri and P. A. Bartlett, “CAVEAT: A program to facilitate the design of organic molecules,” *J. Comput. Aided. Mol. Des.*, 1994, doi: 10.1007/BF00124349.

[34] P. Maass, T. Schulz-Gasch, M. Stahl, and M. Rarey, “Recore: A fast and versatile method for scaffold

hopping based on small molecule crystal structure conformations," *J. Chem. Inf. Model.*, 2007, doi: 10.1021/ci060094h.

[35] "Schrödinger Release 2020-4: Maestro, Schrödinger, LLC, New York, NY, 2020." .

[36] "Chemical Computing Group ULC, Molecular Operating Environment (MOE), 2019.01. 1010 Sherbrooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2019." .

[37] H. Alonso, A. A. Bliznyuk, and J. E. Gready, "Combining docking and molecular dynamic simulations in drug design," *Medicinal Research Reviews*. 2006, doi: 10.1002/med.20067.

[38] S. C. Gill *et al.*, "Binding Modes of Ligands Using Enhanced Sampling (BLUES): Rapid Decorrelation of Ligand Binding Modes via Nonequilibrium Candidate Monte Carlo," *J. Phys. Chem. B*, 2018, doi: 10.1021/acs.jpcb.7b11820.

[39] N. M. Lim, M. Osato, G. L. Warren, and D. L. Mobley, "Fragment Pose Prediction Using Non-equilibrium Candidate Monte Carlo and Molecular Dynamics Simulations," *J. Chem. Theory Comput.*, 2020, doi: 10.1021/acs.jctc.9b01096.

[40] S. M. Linker, A. Magarkar, J. Köfinger, G. Hummer, and D. Seeliger, "Fragment Binding Pose Predictions Using Unbiased Simulations and Markov-State Models," *J. Chem. Theory Comput.*, 2019, doi: 10.1021/acs.jctc.9b00069.

[41] D. Sabbadin and S. Moro, "Supervised Molecular Dynamics (SuMD) as a Helpful Tool To Depict GPCR–Ligand Recognition Pathway in a Nanosecond Time Scale," *J. Chem. Inf. Model.*, vol. 54, no. 2, Feb. 2014, doi: 10.1021/ci400766b.

[42] F. Ferrari *et al.*, "HT-SuMD: making molecular dynamics simulations suitable for fragment-based screening. A comparative study with NMR," *J. Enzyme Inhib. Med. Chem.*, vol. 36, no. 1, pp. 1–14, Jan. 2021, doi: 10.1080/14756366.2020.1838499.

[43] L. Chaput and L. Mouawad, "Efficient conformational sampling and weak scoring in docking programs? Strategy of the wisdom of crowds," *J. Cheminform.*, 2017, doi: 10.1186/s13321-017-0227-x.

[44] L. R. de Souza Neto *et al.*, "In silico Strategies to Support Fragment-to-Lead Optimization in Drug Discovery," *Frontiers in Chemistry*, vol. 8. Frontiers Media S.A., p. 93, Feb. 2020, doi: 10.3389/fchem.2020.00093.

[45] D. R. Houston and M. D. Walkinshaw, "Consensus Docking: Improving the Reliability of Docking in a Virtual Screening Context," *J. Chem. Inf. Model.*, vol. 53, no. 2, pp. 384–390, Feb. 2013, doi: 10.1021/ci300399w.

[46] A. Fischer, M. Smieško, M. Sellner, and M. A. Lill, "Decision Making in Structure-Based Drug Discovery: Visual Inspection of Docking Results," *J. Med. Chem.*, vol. 64, no. 5, pp. 2489–2500, Mar. 2021, doi: 10.1021/acs.jmedchem.0c02227.

[47] J. T. Metz, E. F. Johnson, N. B. Soni, P. J. Merta, L. Kifle, and P. J. Hajduk, "Navigating the kinome," *Nat. Chem. Biol.*, vol. 7, no. 4, pp. 200–202, 2011, doi: 10.1038/nchembio.530.

[48] S. P *et al.*, "GSK1838705A inhibits the insulin-like growth factor-1 receptor and anaplastic lymphoma kinase and shows antitumor activity in experimental models of human cancers," *Mol. Cancer Ther.*, vol. 8, no. 10, pp. 2811–2820, Oct. 2009, doi: 10.1158/1535-7163.MCT-09-0423.

[49] "QUACPAC 2.1.1.0: OpenEye Scientific Software, Santa Fe, NM." .

[50] J. Sadowski, J. Gasteiger, and G. Klebe, "Comparison of Automatic Three-Dimensional Model Builders Using 639 X-ray Structures," *J. Chem. Inf. Comput. Sci.*, 1994, doi: 10.1021/ci00020a039.

[51] O. Korb, T. Stützle, and T. E. Exner, “PLANTS: Application of Ant Colony Optimization to Structure-Based Drug Design,” 2006.

[52] G. Jones, P. Willett, R. C. Glen, A. R. Leach, and R. Taylor, “Development and validation of a genetic algorithm for flexible docking 1 1Edited by F. E. Cohen,” *J. Mol. Biol.*, vol. 267, no. 3, Apr. 1997, doi: 10.1006/jmbi.1996.0897.

[53] M. L. Verdonk, J. C. Cole, M. J. Hartshorn, C. W. Murray, and R. D. Taylor, “Improved protein-ligand docking using GOLD.,” *Proteins*, vol. 52, no. 4, pp. 609–623, Sep. 2003, doi: 10.1002/prot.10465.

[54] T. A. Halgren *et al.*, “Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening,” *J. Med. Chem.*, 2004, doi: 10.1021/jm030644s.

[55] M. Sándor, R. Kiss, and G. M. Keseru, “Virtual fragment docking by glide: A validation study on 190 protein-fragment complexes,” *J. Chem. Inf. Model.*, 2010, doi: 10.1021/ci1000407.

[56] M. J. Harvey, G. Giupponi, and G. De Fabritiis, “ACEMD: Accelerating Biomolecular Dynamics in the Microsecond Time Scale,” *J. Chem. Theory Comput.*, vol. 5, no. 6, Jun. 2009, doi: 10.1021/ct9000685.

[57] J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, and C. Simmerling, “ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB.,” *J Chem Theory Comput*, vol. 11, no. 8, pp. 3696–3713, Aug. 2015, doi: 10.1021/acs.jctc.5b00255.

[58] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, “Comparison of simple potential functions for simulating liquid water,” *J. Chem. Phys.*, vol. 79, no. 2, pp. 926–935, Jul. 1983, doi: 10.1063/1.445869.

[59] A. Cuzzolin *et al.*, “Deciphering the Complexity of Ligand-Protein Recognition Pathways Using Supervised Molecular Dynamics (SuMD) Simulations,” *J. Chem. Inf. Model.*, vol. 56, no. 4, pp. 687–705, 2016, doi: 10.1021/acs.jcim.5b00702.

[60] L. Badura *et al.*, “An inhibitor of casein kinase Iε induces phase delays in circadian rhythms under free-running and entrained conditions,” *J. Pharmacol. Exp. Ther.*, 2007, doi: 10.1124/jpet.107.122846.

In silico evaluation of the interaction between ACE2 and SARS-CoV-2 Spike protein in a hyperglycemic environment

Giovanni Sartore, **Davide Bassani**, Eugenio Ragazzi, Pietro Traldi, Annunziata Lapolla and Stefano Moro.

G. Sartore, D. Bassani, E. Ragazzi, P. Traldi, A. Lapolla, and S. Moro, "In silico evaluation of the interaction between ACE2 and SARS-CoV-2 Spike protein in a hyperglycemic environment," *Sci. Rep.*, vol. 11, no. 1, p. 22860, Dec. 2021, doi: 10.1038/s41598-021-02297-w.

Abstract

The worse outcome of COVID-19 in people with diabetes could be related to the non-enzymatic glycation of human ACE2, leading to a more susceptible interaction with virus Spike protein. We aimed to evaluate, through a computational approach, the interaction between human ACE2 receptor and SARS-CoV-2 Spike protein under different conditions of hyperglycemic environment. A computational analysis was performed, based on the X-ray crystallographic structure of the Spike Receptor-Binding Domain (RBD)-ACE2 system. The possible scenarios of lysine aminoacid residues on surface transformed by glycation were considered: 1) on ACE2 receptor; 2) on Spike protein; 3) on both ACE2 receptor and Spike protein. In comparison to the native condition, the number of polar bonds (comprising both hydrogen bonds and salt bridges) in the poses considered are 10, 6, 6, and 4 for the states ACE2/Spike both native, ACE2 native/Spike glycated, ACE2 glycated/Spike native, ACE2/Spike both glycated, respectively. The analysis highlighted also how the number of non-polar contacts (in this case, van der Waals and aromatic interactions) significantly decreases when the lysine aminoacid residues undergo glycation. Following non-enzymatic glycation, the number of interactions between human ACE2 receptor and SARS-CoV-2 Spike protein is decreased in comparison to the unmodified model. The reduced affinity of the Spike protein for ACE2 receptor in case of non-enzymatic glycation may shift the virus to multiple alternative entry routes.

Abbreviations: AGE: Advanced Glycation End products; ACE1: Angiotensin-Converting Enzyme-1; ACE2: Angiotensin-Converting Enzyme-2; CD26: cluster of differentiation 26; DPP4: Dipeptidyl peptidase 4; GRP78: glucose-regulated protein 78; MOE:Molecular Operating Environment; NRP1: Neuropilin-1; PDB: Protein Data Bank; RAGE:Receptors for Advanced Glycation End products; RBD: Receptor Binding Domain; TMPRSS2: Transmembrane Protease Serine-2.

1. Introduction

Type 2 diabetes mellitus has been considered a risk factor for acquiring the SARS-CoV-2 infection [1]. Increased morbidity in Type 2 diabetes mellitus has been documented since the initial spread of the pandemy and people with preexisting Type2 diabetes have an increased need for medical intervention [2]. Meta-analyses and literature reviews have confirmed that patients with diabetes have higher risk of COVID-19 disease severity and mortality [3-6]. The reason for this phenomenon is still under debate [6-9]. Among the hypotheses available, one considers that people affected by diabetes have an increased risk of severe COVID-19 disease due to an imbalance between ACE1 and ACE2 activity [1] which leads to pro-inflammatory responses, predisposing to the cytokine storm syndrome [1, 10]. Moreover, the excess of adipose tissue, typical of type-2 diabetes mellitus, is associated with increased macrophage and T-cell activation, together with increased proinflammatory cytokine production [11]. Regarding COVID-19 in patients affected by diabetes, little attention seems to be paid to the mechanisms of SARS-CoV-2 interaction with access sites. In a previous paper, we hypothesized that a worse outcome of COVID-19 in people with diabetes could be related to the non-enzymatic glycation of human ACE2, which could trigger the activity of ACE2 to a more susceptible interaction with virus Spike protein [12]. Recent findings that about half of hospitalized patients with Type 2 diabetes mellitus present a myocardial damage have suggested a specific role played by glycated ACE2 receptor [13]. On the other side, “glycosilated” ACE2 receptor has been documented [14] and glycosylation status has been considered as a possible determinant in SARS-CoV-2 infection susceptibility [15]. With the term “glycation”, we refer here to the modification of lysine caused by the Maillard reaction followed by the rearrangement of the Amadori products obtained [16, 17]. Advanced Glycation End products (AGEs), which are produced by glycation of cellular molecules, including proteins, have been linked to increased COVID-19 risk factors [9]. Based on our hypothesis [12], an upregulation of ACE2, due to its non-enzymatic glycation, together with a variation of the protein tertiary structure due to the aforementioned aminoacidic modifications, was suggested as a pathogenetic mechanism of SARS-CoV-2 negative outcome in diabetes. The present work aimed to evaluate, through a computational approach, the interaction between human ACE2 receptor and SARS-CoV-2 Spike protein under different conditions of the hyperglycemic environment, which has been shown to influence the non-enzymatic glycation of the lysine residues of the aforementioned proteins. Looking deeply into the complexes deposited in the PDB (Protein Data Bank) such as 6LZG or

6M0J, considering the 34 lysine residues present in the extracellular portion of the ACE2 enzyme, a conformational change caused by non-enzymatic glycation could be possible. In addition, it is important to remember that also Spike protein has several lysine residues present on its surface, and a change in its tertiary structure due to glycation should also be considered.

2. Methods

To assess if the glycation of the lysine aminoacid residues of ACE2 receptor or Spike protein Receptor Binding Domain (RBD) could affect their interaction, computational analysis was performed. First of all, the X-Ray crystallographic structure of the Spike RBD-ACE2 system was downloaded from Protein Data Bank (PDB: 6M0J, X-ray resolution: 2.45Å, DOI: 10.2210/pdb6M0J/pdb) [18]. This system was properly prepared with Molecular Operating Environment (MOE) Structure Preparation Tool. The missing loops were rebuilt exploiting the MOE loop builder application. This program is able to create small missing parts of protein structures based on their sequence. Each loop created is then subjected to a multi-stage energy optimization, firstly aiming to remove the clashes in its structure and then to minimize it in the overall system, which is kept fixed with the exeption of the 3 nearest residues in both sides of the loop examined. The orientation of residues with alternative conformational states was chosen based on the occupancy. The hydrogen atoms were added with the MOE Protonate 3D tool, and the same program was used to assess the most probable protonation state at pH 7.4. The protonation and flip states are taken from a database of states which is built in MOE. Among the states present in this database, the Protonate 3D program selects the most probable for each aminoacid evaluating specifically its own environment. The protonation predictions exploit a Generalized Born implicit solvation model, which is able to take into account the long-range interactions and the solvation effects (for these calculations, in this experiment, the salt concentration parameter was set at 0.15 mol/L). The added hydrogen atoms were minimized using AMBER10:EHT [19] force field implemented in MOE [20]. It is important to underline that, in the present study, we considered the system without its glycosylated chains since glycan chains are very difficult to reproduce and are out of our main aim. A discussion of the role of glycans is reported in Supplementary Material 1. To effectively compare the interactivity between the two protein interfaces in different glycation states, we decided to divide the computational study into three different parts. Firstly, we considered ACE2 receptor with lysine aminoacid residues transformed by glycation, then Spike protein with

lysine aminoacid residues transformed by glycation, and finally we analyzed the situation with all lysines of the system subjected to glycation. The analysis of each system was carried out both by visual inspection and by the intermolecular interactions count, exploiting the "GetContacts" tool (https://getcontacts.github.io/). As "glycation" we refer to the modification of lysine due to the Maillard reaction followed by the rearrangement of the Amadori products obtained [16, 17] , and the derivative that we consider is the cyclic amino sugar, as represented in Figure 1.

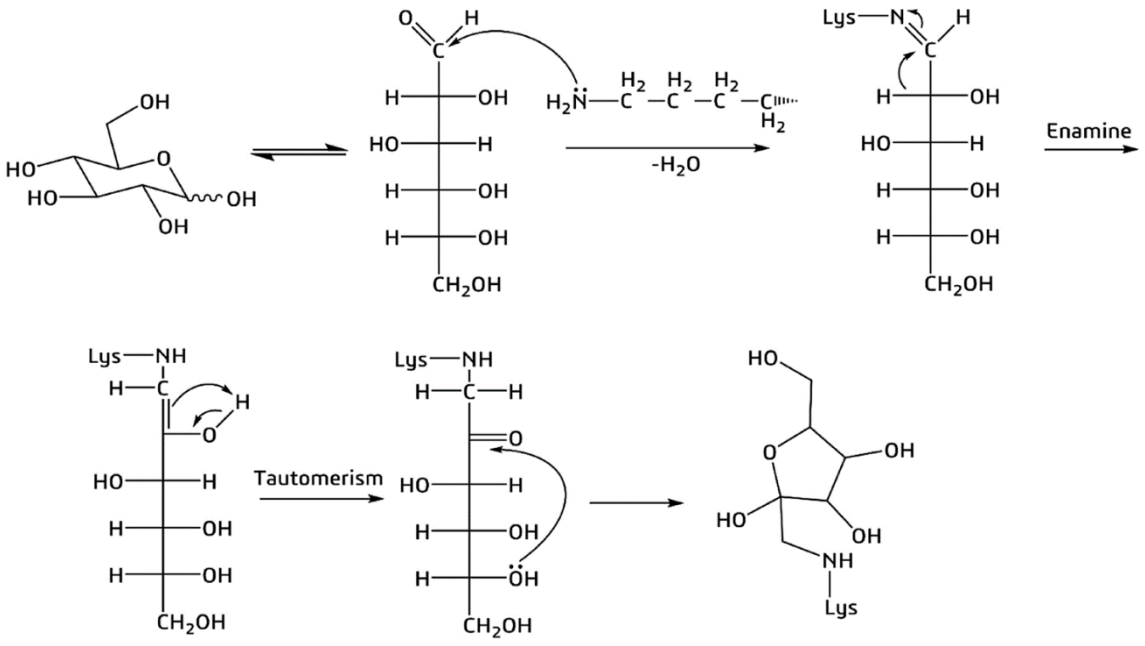


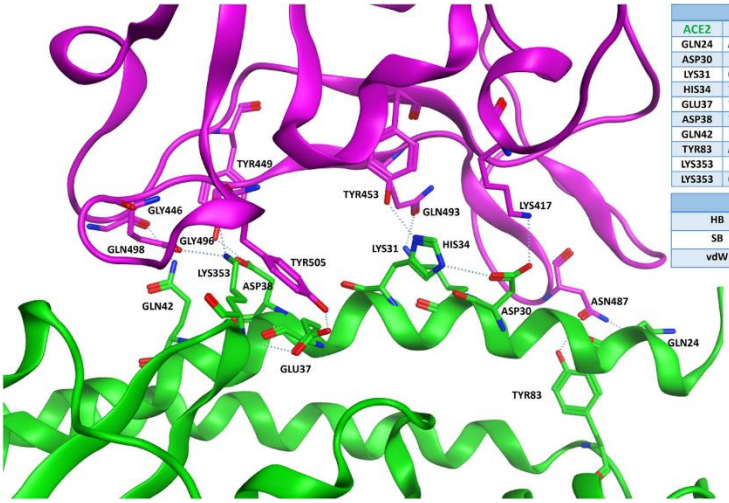
Figure 1. *Mechanism of non-enzymatic glycation of a lysine amino acid residue by D-glucose. The final product is the cyclic amino sugar that we considered for the study.*

Both proteins of interest, SARS-CoV-2 Spike protein, and human ACE2 receptor, have lysine aminoacid residues on their surface, so all these sites were considered for glycation. In order to obtain images of easier interpretation, even if we considered all lysine aminoacid residues of the system, here we just report the results for the ones involved in the interaction. To effectively analyze how interactivity changes with respect to glycation, we used the MOE protein-protein docking tool. The ACE2 protein was treated as the receptor and the RBD of Spike as the ligand. The ACE2 residues considered as the "binding site" were chosen according to Deganutti et al. [21], to reproduce a proper binding situation between the two entities. For each system, 100 docking poses were generated, ranked, and visually inspected. The pose which could better reproduce the situation observable in 6M0J was selected and used for further analysis. In order to obtain the glycated lysine residues for the study, the modified side

chain was manually build using the Molecular Operating Environment “Builder” tool. The glycated residues obtained were then selectively minimized with the AMBER10:EHT [19] implemented in MOE.

3. Results

Figure 2 shows the well-known interaction of Spike protein RBD (violet) with the native ACE2 receptor (green). The contacts established between the two proteins are reported in two different tables, one for polar interactions (namely, hydrogen bonds and salt bridges) and one for non-polar relationships (mainly van der Waals interactions). To get a better visual effect, just the residues involved in the polar interactions are labelled in the 3D image reported, together with the dashed lines indicating the polar contacts themselves. For the same reason, in the table regarding the non-polar relationships, just the aminoacid couples which interact between the proteins are indicated.



Polar interactions detected			
ACE2	Spike	Interaction type	Distance (Å)
GLN24	ASN487	HB	2.688
ASP30	LYS417	SB	2.905
LYS31	GLN493	HB	2.669
HIS34	TYR453	HB	2.737
GLU37	TYR505	HB	2.648
ASP38	TYR449	HB	2.592
GLN42	GLY446	HB	2.840
TYR83	ASN487	HB	2.788
LYS353	GLY496	HB	2.776
LYS353	GLN498	HB	2.773

Legend	
HB	hydrogen bond
SB	salt bridge
vdW	van der Waals interaction

Non-polar interactions detected		
ACE2	Spike	Interaction type
GLN24	ALA475	vdW
GLN24	ASN487	vdW
THR27	PHE456	vdW
THR27	TYR489	vdW
PHE28	TYR489	vdW
ASP30	LYS417	vdW
LYS31	GLN493	vdW
HIS34	TYR453	vdW
HIS34	LEU455	vdW
GLU35	GLN493	vdW
GLU37	TYR505	vdW
ASP38	TYR449	vdW
TYR41	GLN498	vdW
TYR41	THR500	vdW
TYR41	ASN501	vdW
GLN42	GLY446	vdW
GLN42	TYR449	vdW
GLN42	GLN498	vdW
LEU79	PHE486	vdW
MET82	PHE486	vdW
TYR83	PHE486	vdW
TYR83	ASN487	vdW
ASN330	THR500	vdW
LYS353	GLY496	vdW
LYS353	ASN501	vdW
LYS353	GLY502	vdW
LYS353	TYR505	vdW
GLY354	GLY502	vdW
ASP355	THR500	vdW
ARG357	THR500	vdW

Figure 2. *The typical situation of viral Spike protein RBD (PDB: 6M0J, violet-colored) binding to ACE2 receptor (green). The tables on the right report the polar and the non-polar interactions in which the residues on the interface are involved. For the polar interactions, also the distance between the interacting atoms (measured in angstroms) is reported in the table. To get a better visual representation of the contacts, just the residues engaged in the polar interactions are labelled in the 3D image on the left, while all the non-polar interactions are omitted.*

The possible occurrence of glycation on ACE2 receptor (Figure 3A) as suggested for diabetes condition, still allows a relevant binding of Spike protein RBD (orange) to the modified ACE2 (blue) site. Spatial differences in the binding, represented by the distance between the carbonyl oxygen atoms of Spike RBD SER494 and ACE2 HIS34, are shown in Figure 3B.

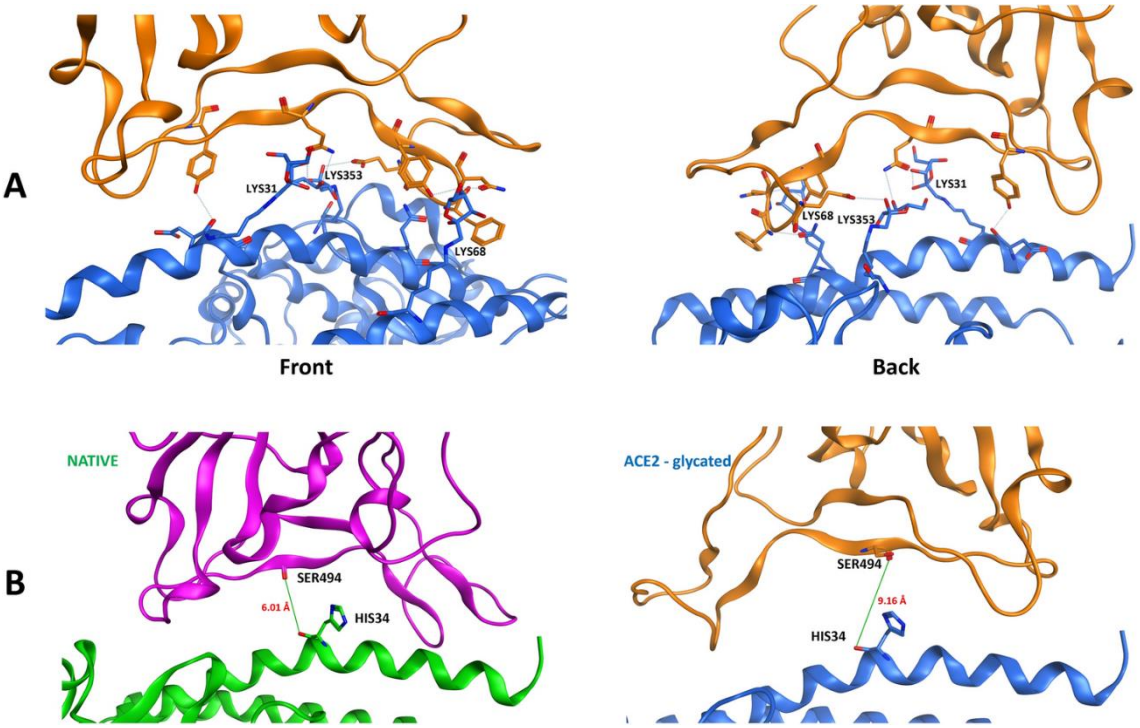
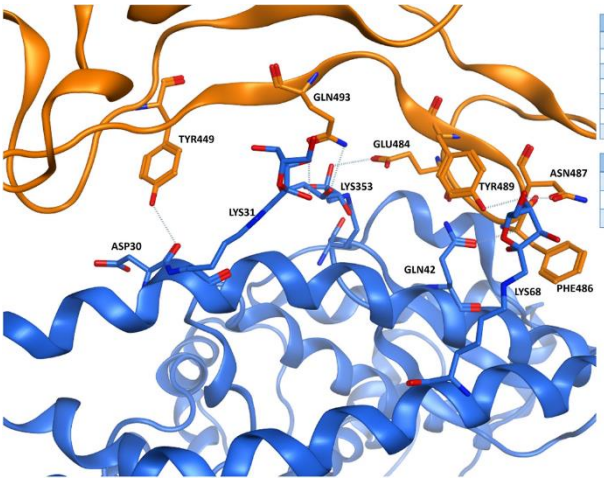


Figure 3. ***A.*** *Front and back highlight of the viral Spike protein (orange) binding to glycated ACE2 receptor (blue). The lysine amino acid residues subjected to glycation on ACE2 receptor are labelled in the image.* ***B.*** *Difference in the distances between the first two cases considered. On the left: the classical interaction in PDB 6M0J, in which the distance between the carbonyl oxygen atoms of ACE2 receptor (green) HIS34 and Spike protein RBD (violet) SER494 is 6.01Å. On the right: the interaction between glycated ACE2 receptor (blue) and native Spike protein RBD (orange). In this case, the distance between the carbonyl oxygen of ACE2 receptor (blue) HIS34 and Spike protein RBD (orange) SER494 is 9.16Å.*

Figure 4 presents the intermolecular interaction analysis of glycated ACE2 (blue) interaction with the native viral Spike RBD (not glycated, orange). As shown in the table, both the polar and the non-polar contacts between the proteins are decreased in comparison to the unmodified model of Figure 2. Moreover, the distances between the atoms involved in the polar bonds tend to increase, and the only salt bridge establishing the interaction is lost.



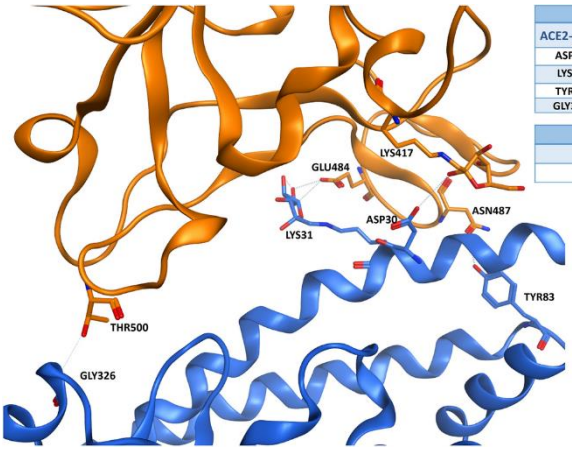
Polar interactions detected			
ACE2-glyco	Spike	Interaction type	Distance (Å)
ASP30	TYR449	HB	3.108
GLN42	PHE486	HB	3.450
LYS68	ASN487	HB	2.607
LYS68	TYR489	HB	2.668
LYS353	GLU484	HB	3.165
LYS353	GLN493	HB	3.152

Legend	
HB	hydrogen bond
vdW	van der Waals interaction
PS	pi-stacking

Non-polar interactions detected		
ACE2-glyco	Spike	Interaction type
SER19	THR500	vdW
LYS27	GLN498	vdW
ASP30	GLY446	vdW
ASP30	TYR449	vdW
LYS31	GLN493	vdW
HIS34	TYR449	vdW
HIS34	TYR449	PS
TYR41	GLU484	vdW
GLN42	GLY485	vdW
GLN42	PHE486	vdW
LEU45	GLY485	vdW
ASN61	PHE486	vdW
LYS353	GLU484	vdW
LYS353	GLN493	vdW

Figure 4. *Viral Spike protein RBD (PDB: 6M0J, orange-colored) is binding to glycated ACE2 receptor (blue). The tables on the right report the polar and the non-polar interactions in which the residues on the interface are involved. For the polar interactions, also the distance between the interacting atoms (measured in angstroms) is reported in the table. To get a better visual representation of the contacts, just the residues engaged in the polar interactions are labelled in the 3D image on the left, while all the non-polar interactions are omitted.*

In the hypothesis that glycation may involve also Spike protein, a further evaluation was done. As shown in Figure 5, the interaction analysis of glycated ACE2 receptor (blue) interaction with the glycated viral Spike RBD (orange) indicates a further decrease in the number of polar interactions involved, while the number of non-polar contacts is does not change significantly.



Polar interactions detected			
ACE2-glyco	Spike-glyco	Interaction type	Distance (Å)
ASP30	LYS417	HB	3.439
LYS31	GLU484	HB	2.622
TYR83	ASN487	HB	2.873
GLY326	THR500	HB	2.914

Legend	
HB	hydrogen bond
vdW	van der Waals interaction

Non-polar interactions detected		
ACE2-glyco	Spike-glyco	Interaction type
GLU23	LYS417	vdW
LYS26	LYS417	vdW
THR27	LYS417	vdW
THR27	TYR489	vdW
PHE28	TYR489	vdW
ASP30	LYS417	vdW
ASP30	PHE456	vdW
LYS31	GLU484	vdW
LYS31	TYR489	vdW
HIS34	LEU455	vdW
LEU79	PHE486	vdW
TYR83	PHE486	vdW
TYR83	ASN487	vdW
GLY326	THR500	vdW
LYS353	TYR449	vdW
LYS353	GLN493	vdW
LYS353	GLY496	vdW

Figure 5. *Glycated viral Spike protein RBD (orange) binding to glycated ACE2 receptor (blue). The tables on the right report the polar and the non-polar interactions in which the residues on the interface are involved. For the polar interactions, also the distance between the interacting atoms (measured in angstroms) is reported in the table. To get a better visual representation of the contacts, just the residues engaged in the polar interactions are labelled in the 3D image on the left, while all the non-polar interactions are omitted.*

Further details of the molecular interactions in the different hypotheses of glycation, including the interaction of glycated Spike protein with native ACE2 (Figure S2.1), are reported in Supplementary material 2.

4. Discussion

As a result of our *in silico* study, it appears that the glycation of the lysine aminoacid residues present both in Spike and ACE2 proteins lead to a loss of interaction between them, as is depictable with the analysis of the number of both polar and non-polar contacts. Considering all the possible scenarios of glycation of ACE2 receptor and/or Spike protein, it derives that the number of polar bonds in the poses considered are 10, 6, 6, and 4 for the states ACE2/Spike both native, ACE2 native/Spike glycated, ACE2 glycated/Spike native, ACE2/Spike both glycated, respectively. The number of non-polar contacts shows a very important decrease if the native model (in which 30 different aminoacid pairs establish non-polar interactions between the proteins) is compared with anyone of the others, while it does not show huge changes among the variously glycated models. Indeed, 14 aminoacid pairs which stabilize non-polar interactions can be depicted for the ACE2 glycated/Spike native system, 19 for the ACE2 native/Spike glycated model and 17 in the case in which both ACE2 and Spike proteins have lysines glycated. Looking deeply to the polar interactions, it is possible to evaluate an increase in the medium distance between the atoms involved in the contacts passing from the native model to the various glycated ones, suggesting a progressive interaction weakening. From data of the present analysis it could be concluded that non-enzymatic glycation of both ACE2 and Spike proteins exerts a detrimental effect on their interaction, independently on the fact that one or both proteins are involved in the glycation process. The analysis reported suggests that ACE2 glycation has a higher influence on the overall interactivity, mainly because the highest losses in non-polar interactions are depictable in the systems in which ACE2 has its lysine residues glycated.

Even if no experimental validation has been brought in the present work, further researches can benefit of the analysis here provided. Indeed, the results of the present study concur to enhance the knowledge about both the link between SARS-CoV-2 infection and diabetes and on SARS-CoV-2 Spike protein interaction with human ACE2 [22] at the molecular level. Angiotensin-Converting Enzyme-2 (ACE2) is the most investigated site of SARS-CoV-2 access to the host cells [23, 24]. SARS-CoV-2, compared to SARS-CoV, is characterized by a tighter bound

to ACE2 [18, 23, 25], which can explain the high transmission rate of the SARS-CoV-2 virus [25]. The initial hypothesis that an increased risk of worse outcomes in patients with Type 2 diabetes mellitus is due to increased virus interaction with ACE2 sites following glycation [12], seems however to be excluded, based on the present analysis. Therefore, other possible explanations to disease severity of SARS-CoV-2 infection in diabetic patients may be provided. New additional evidence has suggested that alternative viral entry molecules are involved, both at viral and cellular level [26]. The interaction of SARS-CoV-2 with human cells is also dependent on Transmembrane Protease Serine-2 (TMPRSS2), a protein whose expression is regulated by androgen expression [27, 28], and that helps the virus enter into the host [24, 26, 29, 30]. Moreover, it has been demonstrated that COVID-19 infection causes damage to the pancreatic tissue mediated by TMPRSS2 [31], further suggesting a pivotal role of the virus in the disease outcome in diabetic conditions. The reduced affinity of the Spike protein for the ACE2 receptor in the case of non-enzymatic glycation, as found in the present study, may shift the virus to alternative entry routes, and TMPRSS2 could be one of these. A common feature of viruses is their ability of developing new effective ways to penetrate into host cells when their main entry mechanism results to be slowed or impaired. In this context, another described route of SARS-CoV-2 entry into cells, Neuropilin-1 (NRP1), could be considered [26, 30, 32-34]. NRP1 has been demonstrated as an additional SARS CoV-2 infection mediator, in particular linked to the neurological aspects found in COVID 19 [33], but also because it is strictly involved in SARS-CoV-2 binding in lung tissue [35] and in mediating diabetic nephropathy [36]. NRP1 has been found significantly up-regulated following SARS CoV-2 infection [35]. The widespread presence of NRP1 in epithelial cells makes it a feasible entry pathway for the virus, and therefore in diabetes condition, when the glycated ACE2 receptor has lost at least part of its binding affinity for virus, NRP1 receptor, as well as TMPRSS2, could play a preferential, alternative, role of SARS-CoV-2 entry receptor. Other proposed entry routes for SARS CoV-2 include Dipeptidyl peptidase 4 (DPP4), also known as cluster of differentiation 26 (CD26), an ectopeptidase found in many tissues, such as lung and kidney, and involved in several physiological processes and diseases of the immune system [26, 37]. A soluble form (sCD26) has been also described in blood, acting as protective factor against virus entry; sCD26 is reduced in diabetes, therefore conditioning an increased risk of infection in this population [37]. Different affinity for glycated ACE2 receptor may shift the virus entry through CD26. This mechanism could be involved in the reduced COVID-19 mortality observed with sitagliptin, a DDP-4 inhibitor drug used in Type 2

diabetes mellitus, by increasing the soluble form of DPP-4/sCD26 [38]. Another investigated receptor for SARS-CoV-2 infection is represented by the transmembrane glycoprotein CD147 (Basigin 2), expressed in pathological tissues and in inflammation [26]. ACE2 and CD147 activities as entry pathways for SARS-CoV-2 are co-regulated, leading to downregulation following virus exposure [39]. Therefore, since CD147 expression is upregulated by high glucose concentrations and AGEs [40, 41], this alternative pathway may assume a pivotal role in conditioning virus binding to cells in diabetes. Also glucose-regulated protein 78 (GRP78) and other receptors have been suggested as potential alterative receptors for SARS-CoV-2 entry into cells [42]. The existence of multiple pathways for virus binding can explain the altered susceptibility to COVID-19 in diabetes, after the present *in silico* analysis suggesting decreased affinity for virus of glycated ACE2 receptor. The present finding that glycation of ACE2 receptor reduces the affinity for Spike protein supports also the hypothesis that a downregulation of ACE2, observed after SARS-CoV-2 infection, leads to accumulation of angiotensin II and related metabolites [43], conditioning the acute respiratory distress typical of COVID-19. A role of Spike protein glycation, also possible in diabetes, and explored in the present study, might be hypothesized in conditioning the ability of SARS-CoV-2 to interact with the several possible access routes of the virus. The scenario of possible interactions between viral Spike and host structures should also consider the fact that in diabetes the decomposition of glycated amino sugars, generated through the Maillard reaction, leads to intermediates of Advanced Glycation End (AGE) products, such as glyoxal and methylglyoxal [16, 44]. These highly reactive carbonyl compounds, and others, formed also by degradation of glucose itself, can interact, even with greater reactivity, with proteins implicated in the virus access, and might further enhance the severity of SARS-CoV-2 infection in people affected by diabetes [9]. These considerations, supported by the present results, could help to explain why hyperglycemia worsens the prognosis of COVID-19, which has been also linked to the system of receptors for advanced glycation end products (RAGE) [45]. Moreover, a putative role of glycated hemoglobin, which is elevated in diabetic patients, has been suggested as an important factor for COVID-19 infection and mortality [46]. Evaluation of the possible reactive components of glucose metabolism may deserve further investigation. An experimental evaluation of the hypothesis regarding glycated ACE2 and Spike interactions in diabetes goes beyond the purpose of the present work, being a strict computational approach, undertaken following observations of clinical and epidemiological data in the COVID-19 pandemy. The further hypotheses of a modified affinity

of Spike protein in diabetes for the several above discussed alternative viral access mechanisms which might become relevant also after Spike protein mutations, already documented [47, 48], have not been considered in the present study, and may deserve a future specific in silico evaluation followed by a possible experimental validation. In conclusion, the present analysis supports the hypothesis that glycation, consequent to hyperglycemia in patients affected by diabetes, could have a role in the SARS-CoV-2 infection, possibly modulating other binding sites for SARS-CoV-2 access into the body.

References

[1] Bornstein, S.R., Dalan, R., Hopkins, D., Mingrone, G. & Boehm, B.O. Endocrine and metabolic link to coronavirus infection. *Nat. Rev. Endocrinol.* **16**, 297–298 (2020).

[2] Zhu, L. *et al.* Association of blood glucose control and outcomes in patients with COVID-19 and pre-existing Type 2 diabetes. *Cell Metab*. **31**, 1068–1077.e3; 10.1016/j.cmet.2020.04.021 (2020).

[3] Wu, Z.-H., Tang, Y. & Cheng, Q. Diabetes increases the mortality of patients with COVID-19: A meta-analysis. *Acta Diabetol*. **58**, 139–144 (2020).

[4] Varikasuvu, S.R., Dutt, N., Thangappazham, B. & Varshney, S. Diabetes and COVID-19: A pooled analysis related to disease severity and mortality. *Prim. Care Diabetes* **15**, 24–27 (2021).

[5] Norouzi, M., *et al.* Type-2 Diabetes as a Risk Factor for Severe COVID-19 Infection. *Microorganisms* **9**, 1211; 10.3390/microorganisms9061211 (2021).

[6] Abramczyk, U. & Kuzan, A. What Every Diabetologist Should Know about SARS-CoV-2: State of Knowledge at the Beginning of 2021. *J. Clin. Med*. **10**, 1022; 10.3390/jcm10051022 (2021).

[7] Shah, H., Khan, M.S.H., Dhurandhar, N.V. & Hegde, V. The triumvirate: why hypertension, obesity, and diabetes are risk factors for adverse effects in patients with COVID-19. *Acta Diabetol*. **15**, 1–13 (2021).

[8] Liao, Y.H., Zheng, J.Q., Zheng, C.M., Lu, K.C. & Chao, Y.C. Novel Molecular Evidence Related to COVID-19 in Patients with Diabetes Mellitus. *J Clin. Med.* **9**, 3962; 10.3390/jcm9123962 (2020).

[9] Sellegounder, D., Zafari, P., Rajabinejad, M., Taghadosi, M. & Kapahi, P. Advanced glycation end products (AGEs) and its receptor, RAGE, modulate age-dependent COVID-19 morbidity and mortality. A review and hypothesis. *Int. Immunopharmacol*. **98**, 107806; 10.1016/j.intimp.2021.107806 (2021).

[10] Zamorano Cuervo, N. & Grandvaux, N. ACE2: Evidence of role as entry receptor for SARS-CoV-2 and implications in comorbidities. *Elife* **9**, e61390; 10.7554/eLife.61390 (2020).

[11] Tsalamandris, S. *et al.* The Role of Inflammation in Diabetes: Current Concepts and Future Perspectives. *Eur. Cardiol*. **14**, 50-59 (2019).

[12] Sartore, G., Ragazzi, E., Faccin, L. & Lapolla, A. A role of glycation and methylation for SARS-CoV-2 infection in diabetes? *Med. Hypotheses* **144**, 110247; 10.1016/j.mehy.2020.110247 (2020).

[13] D'Onofrio, N. *et al.* Glycated ACE2 receptor in diabetes: open door for SARS-COV-2 entry in cardiomyocyte. *Cardiovasc. Diabetol.* **20**, 99; 10.1186/s12933-021-01286-7 (2021).

[14] Towler, P. *et al*. ACE2 X-ray structures reveal a large hingebending motion important for inhibitor binding and catalysis. *J. Biol. Chem.* **279**, 17996–8007 (2004).

[15] Chen, L., *et al.* The effects of chloroquine and hydroxychloroquine on ACE2 related coronavirus pathology and the cardiovascular system: An evidence based review. *Function* (Oxford, England) **1**, zqaa012; 10.1093/function/zqaa012 (2020).

[16] Lapolla, A., Traldi, P. & Fedele, D. Importance of measuring products of non-enzymatic glycation of proteins. *Clin. Biochem.* **38**, 103-115 (2005).

[17] Brito-Arias, M. The Chemistry of Oxidative Stress and Glycoxidation As Risk Factors for Developing Degenerative Disease. *Ann. Rev. Resear.* **4**, 555648; https://juniperpublishers.com/arr/pdf/ARR.MS.ID.555648.pdf (2019).

[18] Lan, J., *et al.* Structure of the SARS-CoV 2 Spike receptor-binding domain bound to the ACE2 receptor. *Nature* **581**, 215–220 (2020).

[19] Case, D.A. *et al. Amber 10* (University of California, San Francisco, 2008). DOI:10.13140/RG.2.2.32447.92320

[20] Molecular Operating Environment (MOE), 2019.01; Chemical Computing Group ULC, 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2021.

[21] Deganutti, G., Prischi, F. & Reynolds, C.A. Supervised molecular dynamics for exploring the druggability of the SARS-CoV-2 spike protein. *J. Comput. Aided Mol. Des.* **35**, 195-207 (2021).

[22] de Andrade, J., Gonçalves, P.F.B. & Netz, P.A. Why Does the Novel Coronavirus Spike Protein Interact so Strongly with the Human ACE2? A Thermodynamic Answer. *Chembiochem.* **22**, 865-875 (2021).

[23] Walls, A.C., Park, Y.J., Tortorici, M.A., Wall, A., McGuire, A.T. & Veesler, D. Structure, function, and antigenicity of the SARS-CoV-2 Spike glycoprotein. *Cell* **181**, 281–292.e6; 10.1016/j.cell.2020.02.058 (2020).

[24] Kothandaraman, N., *et al*. COVID-19 endocrinopathy with hindsight from SARS. *Am. J. Physiol. Endocrinol. Metab*. **320**, E139-E150 (2021).

[25] Liu, Y., Gayle, A.A., Wilder-Smith, A. & Rocklov, J. The reproductive number of COVID-19 is higher compared to SARS coronavirus. *J. Travel Med.* **27**, taaa021; 10.1093/jtm/taaa021 (2020).

[26] Masre, S.F., Jufri, N.F., Ibrahim, F.W. & Abdul Raub, S.H. Classical and alternative receptors for SARS-CoV-2 therapeutic strategy. *Rev. Med. Virol*. e2207 (2020).

[27] Montopoli, M. *et al.* Androgen-deprivation therapies for prostate cancer and risk ofinfection by SARS-CoV-2: a population-based study (N=4532). *Ann. Oncol.* **31**, 1040–1045 (2020).

[28] Mollica, V., Rizzo, A. & Massari, F. The pivotal role of TMPRSS2 in coronavirus disease 2019 and prostate cancer. *Future Oncol.* **16**, 2029–2033 (2020).

[29] Iwata-Yoshikawa, N., Okamura, T., Shimizu, Y., Hasegawa, H., Takeda, M. & Nagata, N. TMPRSS2 Contributes to Virus Spread and Immunopathology in the Airways of Murine Models after Coronavirus Infection. *J Virol*. **93**, e01815-18; 10.1128/JVI.01815-18 (2019).

[30] Katopodis, P. *et al.* COVID 19 and SARS CoV 2 host cell entry mediators: Expression profiling of TMRSS4 in health and disease. *Int. J. Mol. Med*. **47**, 64; 10.3892/ijmm.2021.4897 (2021).

[31] Patel, K.P. *et al.* Gastrointestinal, hepatobiliary, and pancreatic manifestations of COVID-19. *J Clin. Virol*. **128**, 104386; 10.1016/j.jcv.2020.104386 (2020).

[32] Daly, J.L. *et al.* Neuropilin-1 is a host factor for SARS-CoV-2 infection. *Science* **370**, 861-865 (2020).

[33] Davies, J. *et al*. Neuropilin 1 as a new potential SARS CoV 2 infection mediator implicated in the neurologic features and central nervous system involvement of COVID 19. *Mol. Med. Rep.* **22**, 4221-4226 (2020).

[34] Kyrou, I., Randeva, H.S., Spandidos, D.A. & Karteris, E. Not only ACE2-the quest for additional host cell mediators of SARS-CoV-2 infection: Neuropilin-1 (NRP1) as a novel SARS-CoV-2 host cell entry mediator implicated in COVID-19. *Signal Transduct. Target Ther*. **6**, 21; 10.1038/s41392-020-00460-9 (2021).

[35] Cantuti-Castelvetri, L. *et al.* Neuropilin-1 facilitates SARS-CoV-2 cell entry and infectivity. *Science* **370**, 856-860 (2020).

[36] Mourad, D., Azar, N.S. & Azar, S.T. Diabetic Nephropathy and COVID-19: The Potential Role of Immune Actors*. Int. J. Mol. Sci*., **22**, 7762; 10.3390/ijms22157762 (2021).

[37] Raha, A.A. *et al*. Investigation of CD26, a potential SARS-CoV-2 receptor, as a biomarker of age and pathology. *Biosci. Rep.* **40**, BSR20203092; 10.1042/BSR20203092 (2020).

[38] Nauck, M.A. & Meier, J.J. Reduced COVID-19 Mortality With Sitagliptin Treatment? Weighing the Dissemination of Potentially Lifesaving Findings Against the Assurance of High Scientific Standards. *Diabetes Care* **43**, 2906-2909 (2020).

[39] Fenizia, C. *et al.* SARS-CoV-2 entry: At the Crossroads of CD147 and ACE2. *Cells* **10**, 1434; 10.3390/cells10061434 (2021).

[40] Bao, W. *et al.* Monocyte CD147 is induced by advanced glycation end products and high glucose concentration: possible role in diabetic complications. *Am. J. Physiol. Cell Physiol.* **299**, C1212-1219 (2010).

[41] Radzikowska, U. *et al.* Distribution of ACE2, CD147, CD26, and other SARS-CoV-2 associated molecules in tissues and immune cells in health and in asthma, COPD, obesity, hypertension, and COVID-19 risk factors. *Allergy* **75**, 2829–2845 (2020).

[42] Shahriari Felordi, M., Memarnejadian, A., Najimi, M. & Vosough, M. Is there any alternative receptor for SARS-CoV-2? *Cell J.* **23**, 247-250 (2021).

[43] Han, T.C., Harhay, M.O., Brown, T.S., Cohen, J.B. & Mohareb, A.M. Is there an association between COVID-19 mortality and the renin-angiotensin system? A call for epidemiologic investigations. *Clin. Infect. Dis*. **71**, 870–874 (2020).

[44] Thornalley, P.J., Langborg, A. & Minhas, H.S. Formation of glyoxal, methylglyoxal and 3-deoxyglucosone in the glycation of proteins by glucose. *Biochem. J.* **344 Pt 1**, 109-116 (1999).

[45] Dozio, E. *et al.* Soluble Receptor for Advanced Glycation End Products and Its Forms in COVID-19 Patients with and without Diabetes Mellitus: A Pilot Study on Their Role as Disease Biomarkers. *J. Clin. Med.* **9**, 3785; 10.3390/jcm9113785 (2020).

[46] Wang, J. & Meng, W. COVID-19 and diabetes: the contributions of hyperglycemia. *J. Mol. Cell Biol*. **12**, 958-962 (2021).

[47] Focosi, D. & Maggi, F. Neutralising antibody escape of SARS-CoV-2 Spike protein: Risk assessment for antibody-based Covid-19 therapeutics and vaccines. *Rev. Med. Virol.* 1-21; 10.1002/rmv.2231 (2021).

[48] Liu, H. *et al*. The basis of a more contagious 501Y.V1 variant of SARS-COV-2. *Cell Res.* **31**, 720-722 (2021).

Re-Exploring the Ability of Common Docking Programs to Correctly Reproduce the Binding Modes of Non-Covalent Inhibitors of SARS-CoV-2 Protease M^{pro}

Davide Bassani, Matteo Pavan, Giovanni Bolcato, Mattia Sturlese and Stefano Moro

D. Bassani, M. Pavan, G. Bolcato, M. Sturlese, and S. Moro, "Re-Exploring the Ability of Common Docking Programs to Correctly Reproduce the Binding Modes of Non-Covalent Inhibitors of SARS-CoV-2 Protease Mpro," *Pharmaceuticals*, vol. 15, no. 2, p. 180, Jan. 2022, doi: 10.3390/ph15020180.

Abstract

In the latest decades, molecular docking has imposed itself as one of the most used approaches to computational drug discovery. Several docking benchmarks have been published, comparing the performance of different algorithms in respect to a molecular target of interest, usually evaluating their ability in reproducing the experimental data, which in most cases comes from X-ray structures. In this study, we elucidated the variation of the performance of three docking algorithms, GOLD, Glide, and PLANTS, in replicating the coordinates of the crystallographic ligands of Sars-COV-2 main protease (M^{pro}). Through the comparison of the data coming from docking experiments and the values derived from the calculation of the solvent exposure of the crystallographic ligands, we highlighted the importance of this last variable for docking performance. Indeed, we underlined how an increase in the percentage of the ligand surface exposed to the solvent in a crystallographic complex makes it harder for the docking algorithms to reproduce its conformation. We further validated our hypothesis through molecular dynamics simulations, showing that the less stable protein-ligand complexes (in terms of root-mean-square deviation and root-mean-square fluctuation) tend to derive from the cases in which the solvent exposure of the ligand in the starting system is higher.

1. Introduction

In the 1980s, with the first study provided by Kuntz et. al[1], the computational technique of molecular docking had its birth. The efficiency, speed, and robustness of this method make its presence a constant in every structure-based drug discovery pipeline[2]. To give a brief explanation, molecular docking consists of a multi-step computational process that aims to find the best conformation of a molecule to bind to another to form create a stable complex[3]. In the field of medicinal chemistry, as is deductible, its main application is finding the best molecules to bind in a firm way to the desired target (a protein, a nucleic acid, etc.). The algorithm starts with the exploration of the conformations space of the ligands (exploiting the so-called "search algorithm"). The conformations (called "poses") are then classified by a "scoring function", which attributes a numeric value to the goodness of the interaction according to energetical and/or geometrical function. After a series of iterations, the final conformations are presented from the program to the user, ranked by the internal scoring function[4]. In the last 30 years, many docking programs have been developed. Among them, GOLD[5] (a genetic docking algorithm developed by the Cambridge Crystallographic Data Center - CCDC), Glide[6] (a systematic docking program released under license by Schrödinger), AutoDock[7] (a non-commercial genetic algorithm from The Scripps Research Institute), AutoDock VINA[8] (created by the same organization and released for non-commercial use), and PLANTS [9] (an algorithm based on an "Ant Colony Optimization" method) have gained popularity.

The performance of molecular docking programs can be measured by evaluating their ability to reproduce the experimental structural data, such as the crystallographic coordinates of a ligand into its binding site[10]. This ability has been evaluated in several benchmarks[11,12] to rank the performance of different algorithms regarding a specific target, usually using as the key parameter the root-mean-square deviation (RMSD) between the coordinates of the different poses given by the program and the crystallographic ones.

The ability to reproduce a crystallographic conformation strongly relies on different factors. First, the geometrical characteristics of the binding site, like extension and shape play a very important role; it is known that the performance of the algorithms has been improved to dock molecules in

“cavities” or “pockets” rather than surfaces of proteins[13]. Second, the nature and the dimensions of the ligand are also crucial. Indeed, very small ligands may explore different places in a binding site, and the interactions that they can establish are usually few, reducing the “synergism” which could induce a molecule to keep a peculiar shape into a pocket[14]. On the other hand, even if drug-like molecules generally have higher conformational freedom, their dimensions force them to be oriented into a site in a more conserved way, so they have less roto-translational freedom.

In this study, we examine the ability of three docking programs characterized by diverse conformational sampling algorithms to efficiently reproduce the crystallographic pose of different ligands bound in different sites of a protein. To accomplish this task, a target of which several crystal structures are solved with the ligands located in different sites of the macromolecule itself is needed. To this scope, we considered a very recent and relevant target in the current pharmaceutical scenario, such as the SARS-CoV-2 main protease (M^{pro}).

In the last couple of years, with the pandemic spreading of the SARS-CoV-2 virus, the world of medical sciences had found itself fighting a new and dangerous adversary[15,16]. This biological entity, which is part of the coronavirus family, has been demonstrated to cause a pulmonary infection which eventually leads to serious complications, as witnessed by the high number of deaths that have already been linked to it (more than 5 million, at the present day[17]). The replication cycle of this virus strongly relies on the activity of its main protease (known as M^{pro} or $3CL^{pro}$, a crystallographic structure example is reported in Figure 1)[18]. Indeed, this protein is responsible for the cleavage of the propeptide transcribed by the viral genome. This way, the formation of all the functional proteins for the building of new virions takes place, and so the viral infection can proceed. Even if many molecules have been shown to bind to M^{pro}[19] and inhibit its activity, and even if a molecule is currently in phase III clinical trial for this purpose (PF-07321332, from Pfizer[20,21]) no drug has already been approved by the European Medicinal Agency for the treatment of COVID-19. Computational methods have already proven to be beneficial in the research for new potential inhibitors for M^{pro}, as literature witnesses[22,23]. In this work, we decided to implement a molecular docking-based approach

relying on the programs GOLD, Glide, and PLANTS. These algorithms are considered “orthogonal” because they are characterized by diverse placing and scoring algorithms to obtain the best solution to the “protein-ligand posing problem”. Each of these programs has been used to dock each of the different non-covalent ligands of the various crystal structures of M^{pro}, and this allowed us to evaluate the factors which influence the variability in reproducing the crystallographic poses. A self-docking protocol similar to the one herein reported had already been developed by our laboratory, with the name “DockBench”. This program has been implemented with success in several workflows, as literature assesses[22,23]. In this study, a slightly modified version of that tool was used, which exploits only three docking programs at the present moment but can expand the analysis of the results obtained.

Looking at the docking benchmarking protocols on M^{pro}, a remarkable study has already been conducted and published by Zev et al.[26]. In that specific work, 6 different docking programs were evaluated in their performance in reproducing the M^{pro} non-covalent ligands’ crystallographic poses, and 3 of those algorithms have also been compared in their ability to correctly place M^{pro} covalent ligands into their proper binding site. In our work, we decided to expand the considerations brought by that study, evaluating specifically how docking performance changes in respect of the crystallographic data that has to be reproduced.

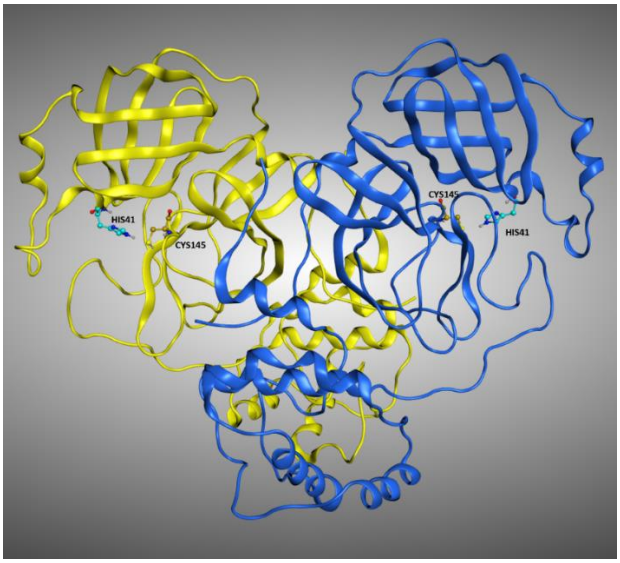


Figure 1. *Representation of the crystal structure of M^{pro} (PDB:7L10). The two monomers composing the protein are colored differently, while the residues of the catalytic dyad, Cys145, and His41, are labeled in each of the monomers.*

Indeed, we considered in our calculations parameters like the solvent exposure of the ligand and the influence of the crystallographic water molecules in docking calculations, focusing our evaluations just on non-covalent M^{pro} ligands. We executed the experiment in two different scenarios, one which excluded the crystallographic waters from the calculation, and one which induced the docking programs to consider them. After that, we compared the docking results with the percentage of solvent exposure of the crystallographic pose of the ligand, successfully confirming that a higher solvent exposure tendentially reflects a worsening in the ability to reproduce the crystallographic pose by the algorithms (that, as already mentioned, are better trained for "cavities" rather than "surfaces"). To further investigate this aspect, we expanded our computational analysis by performing a molecular dynamics (MD) experiment, in which each crystallographic ligand was left free to move for 5ns (3 replicas per system). This approach (known as "MD post-docking") has already become part of our computational protocol[25,26] and is based on the fact that the conformations of the ligands which are less prone to be displaced from their initial position during the simulation are related to higher stability and binding strength with the target. In the case presented, this principle was applied directly to the crystallographic conformations of the ligands rather than to docking poses. This was done because the goal wasn't to select the most promising molecules in binding to a specific region of the protein, but instead, the objective was to elucidate which are the features of the crystallographic ligands that tend to guarantee a tighter binding with the receptor. Our evaluation demonstrated that the molecules bound to the orthosteric pocket of M^{pro} keep their position much stronger than the molecules crystallized on other sites, further validating our solvent exposure-based theory. A representation of the M^{pro} ligands crystallized in the various sites of the protein is given in Figure 2.

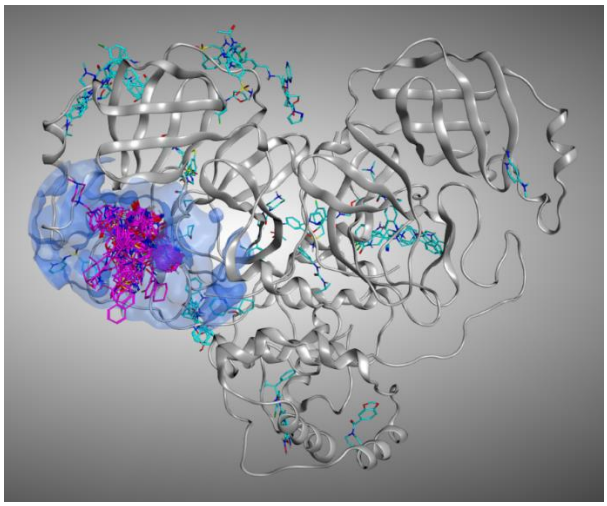


Figure 2. *Representation of all the crystallographic ligands of M^{pro} superposed. To give a better view, just one protein structure is represented (the one coming from PDB:7L10). The ligands which are crystallized inside the catalytic pocket are colored in magenta, while all the small molecules crystallized outside the orthosteric binding site are colored in cyan.*

2. Results and Discussion

2.1. "Scenario 1" – docking calculations without considering the crystallographic water molecules

The results of our docking protocol for this section (which are numerically reported in the Supplementary materials file "Selfdocking_scenario1.csv") are graphically represented with colormaps. All the colormaps present in this study are based on a colorimetric scale delineating the RMSD values, starting from 0 Å, which corresponds to a molecular docking pose exactly superposable to the crystallographic one (maximum docking performance, represented by the dark blue color), and reaching values of 5 Å or higher (minimum docking performance, all represented by the dark red color), corresponding to a very low level of overlap between the coordinates of the pose produced and the ones of the crystallographic conformation. The colormaps in Figure 3 show the self-docking results obtained on the different M^{pro} crystal structures in the case in which water molecules are not considered in the calculation. As is depicted, the RMSD values were far lower for all the complexes in which the crystallographic ligand is located in the orthosteric pocket.

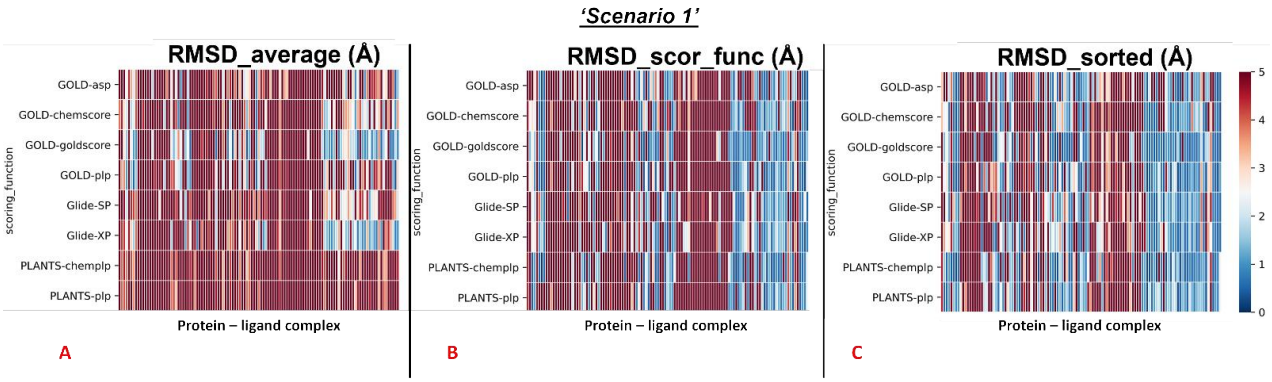


Figure 3. *Colormaps represent the results of the self-docking experiments in the case in which the crystallographic water molecules are not considered during the docking runs. The three plots illustrate respectively: A) the results coming from the average of the RMSDs of all the poses for each docking run; B) the results derived just from the RMSD between the crystallographic ligand coordinates and the pose classified as the best from the scoring function; C) the results of the self-docking experiments if just the pose showing the best RMSD value between its coordinates and the crystallographic ones are retained. The x-axis lists all the different protein-ligand complexes, which are plotted against the different pairs docking program-scoring function used for this study, reported in the y-axis.*

To give a better resolution of this, we have separated each map into two different colormaps, one grouping all the 78 proteins in which the ligand is located into the catalytic pocket, and one including all other cases (41 complexes).

We analyzed the data coming from the calculations, and we computed that, looking at all the complexes with all the different couples docking program-scoring functions, the average values of all the RMSDs obtained was 5.76Å ("RMSD_average"). Looking at the average of the RMSDs coming from the poses which were scored as the best ones from the algorithms' scoring functions ("RMSD_scor_func"), the value was 5.10Å. If the lowest RMSD values only are taken into account for each docking run ("RMSD_sorted"), the mean of the values was 3.70Å.

The average values have also been calculated separately for all the complexes in which the crystallographic ligand is located into the catalytic pocket, and for all other cases. The colormaps for these different conditions are reported in Figure 4 and Figure 5.

First, the analysis focused on the complexes having the crystallographic ligand located within the orthosteric pocket. For this set of systems, we calculated the average RMSD value of all the poses ("RMSD_average"), which was revealed to be 4.54Å. Then, also the average of the RMSD values coming from the poses which were ranked with the best score from the scoring functions was computed ("RMSD_scor_func"), and its value was 3.43Å. Finally, the average RMSD value of the

poses with the lowest RMSD in each run was calculated (“RMSD_sorted”), and its measure was 2.45Å.

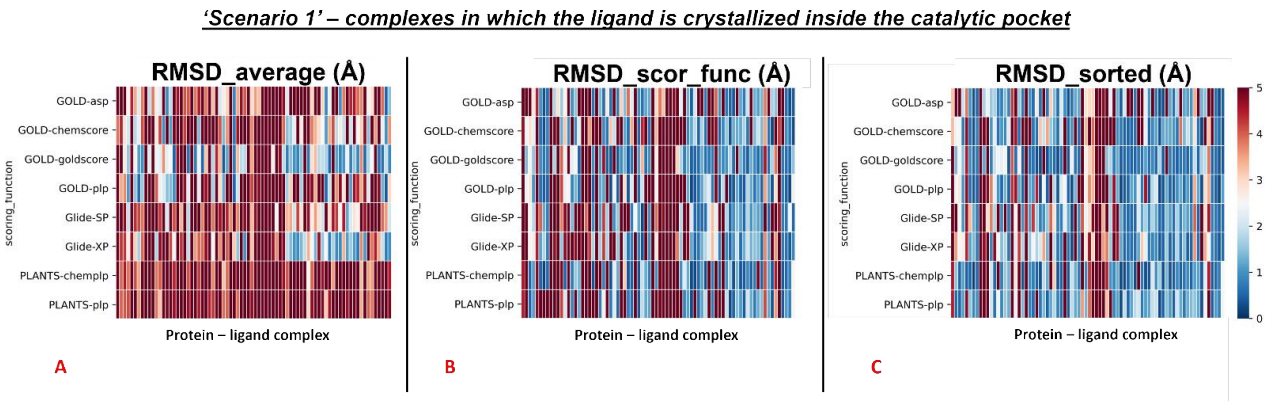


Figure 4. *Colormaps represent the results of the self-docking experiments just for the ligands crystallized inside the orthosteric pocket in the situation in which the crystallographic water molecules are not considered during the docking runs. The three plots illustrate respectively: (A) the results coming from the average of the RMSDs of all the poses for each docking run; (B) the results derived just from the RMSD between the crystallographic ligand coordinates and the pose classified as the best from the scoring function; (C) the results of the self-docking experiments if just the pose showing the best RMSD value between its coordinates and the crystallographic ones is retained. The x-axis lists all the different protein-ligand complexes, which are plotted against the different pairs docking program-scoring function used for this study, reported in the y-axis.*

Second, the same steps were done for the rest of the protein-ligand complexes, which are the ones in which the crystallographic ligand is located outside the orthosteric binding site. Also, in this case, the first passage involved the calculation of the average RMSD value of all the poses generated for these systems (“RMSD_average”), and its measure was 8.08Å. Then, the mean of the RMSD values coming from the poses which received the highest rank from the scoring functions was calculated (“RMSD_scor_func”) and is revealed to be 8.29Å. In the end, the average value of the lowest RMSDs of each run was computed (“RMSD_sorted”), and its measure has shown to be 6.08Å.

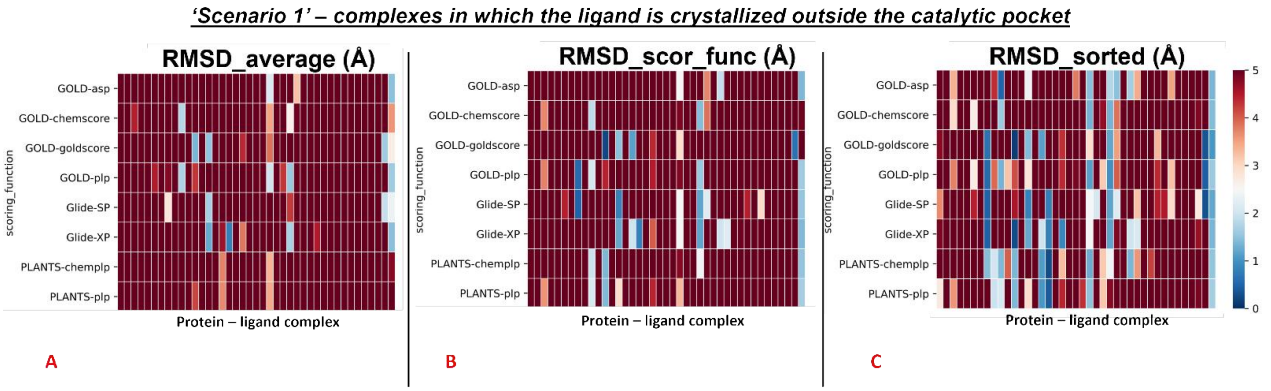


Figure 5. *Colormaps represent the results of the self-docking experiments just for the ligands crystallized outside the orthosteric pocket in the case in which the crystallographic water molecules are not considered during the docking runs. The three plots illustrate respectively: (A) the results coming from the average of the RMSDs of all the poses for each docking run; (B) the results derived just from the RMSD between the crystallographic ligand coordinates and the pose classified as the best from the scoring function; (C) the results of the self-docking experiments if just the pose showing the best RMSD value between its coordinates and the crystallographic ones is retained. The x-axis lists all the different protein-ligand complexes, which are plotted against the different pairs docking program-scoring function used for this study, reported in the y-axis.*

The results obtained for "scenario 1" are summarized in Table 1.

Results for "scenario 1" – no water molecules considered in docking calculations			
	RMSD_average (Å)	**RMSD_scor_func (Å)**	**RMSD_sorted (Å)**
All the 119 protein-ligand complexes	5.76	5.10	3.70
The 78 complexes with the ligand inside the catalytic pocket	4.54	3.43	2.45
The 41 complexes with the ligand outside the catalytic pocket	8.08	8.29	6.08

Table 1. *Table representing the self-docking results obtained for "scenario 1".*

2.2. "Scenario 2" – docking calculations considering the crystallographic water molecules

The outcomes of our molecular docking experiment for this section (which are reported in the Supplementary materials file "Selfdocking_scenario2.csv") are graphically represented with colormaps, which have been created with the same criteria listed in the previous paragraph. The results reported in the colormaps in Figures 6, 7, and 8 reveal the self-docking performance obtained on the different M^{pro} crystal structures in the case in which the crystallographic water molecules within 5 Å from the ligand were retained during the calculation. Also, in this case, it is easy to notice that the values result to be far better for the complexes in which the small molecule of interest is in the orthosteric binding site.

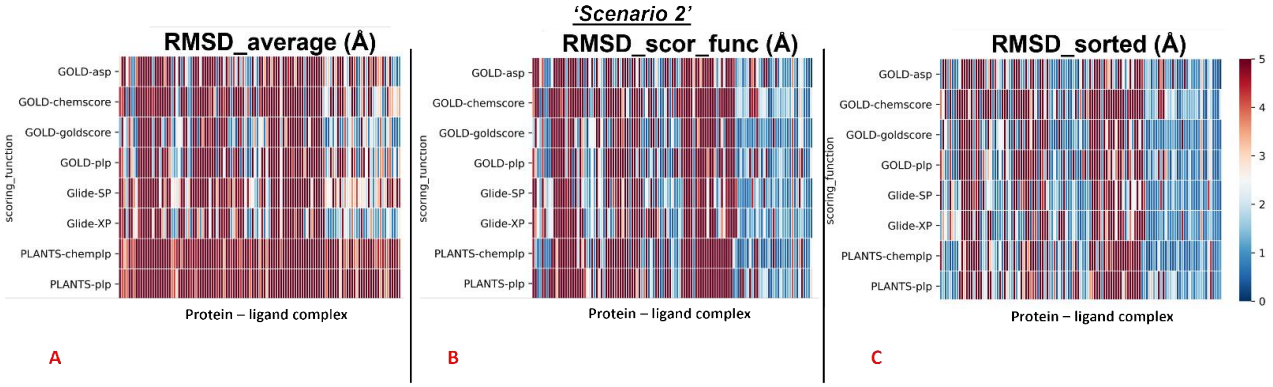


Figure 6. *Colormaps represent the results of the self-docking experiments in the case in which the crystallographic water molecules at 5 Å or nearer to the ligand itself are taken into account during the docking runs. The three plots illustrate respectively: (A) the results coming from the average of the RMSDs of all the poses for each docking run; (B) the results derived just from the RMSD between the crystallographic ligand coordinates and the pose classified as the best from the scoring function; (C) the results of the self-docking experiments if just the pose showing the best RMSD value between its coordinates and the crystallographic ones is retained. The x-axis lists all the different protein-ligand complexes, which are plotted against the different pairs docking program-scoring function used for this study, reported in the y-axis.*

Similar to the first scenario, we divided each colormap into two sets, one with the 78 proteins having the ligand located into the catalytic pocket, and the other including all the remaining cases (41 proteins). Considering all the protein-ligand complexes with all the different pairs docking program-scoring function, the mean values of all the RMSDs obtained ("RMSD_average") was 5.64Å, but focusing only on the mean of the RMSDs derived from the poses which were given the highest rank from the algorithms ("RMSD_scor_func"), the value resulted to be 4.83Å. Looking only at the best RMSDs for each docking run ("RMSD_sorted"), the average of the values

was 3.68 Å. As already done for "scenario 1", also in "scenario 2" the analysis was divided between the complexes in having the crystallographic ligand crystallized into the catalytic pocket, and for all other situations. We reported the colormaps which resulted from this evaluation, and those are represented in Figures 7 and 8. We started from the complexes in which the ligand is located inside the catalytic pocket in the crystal. For those systems, the mean of the RMSD values coming from all the poses("RMSD_average") resulted to be 4.22Å. Then, the average of the RMSDs derived from the scoring function highest-ranked poses in all the docking runs ("RMSD_scor_func") was computed, and its value was 3.11Å. In the end, also the average value between the lowest of the RMSDs in each docking run was calculated ("RMSD_sorted") and is revealed to be 2.26Å.

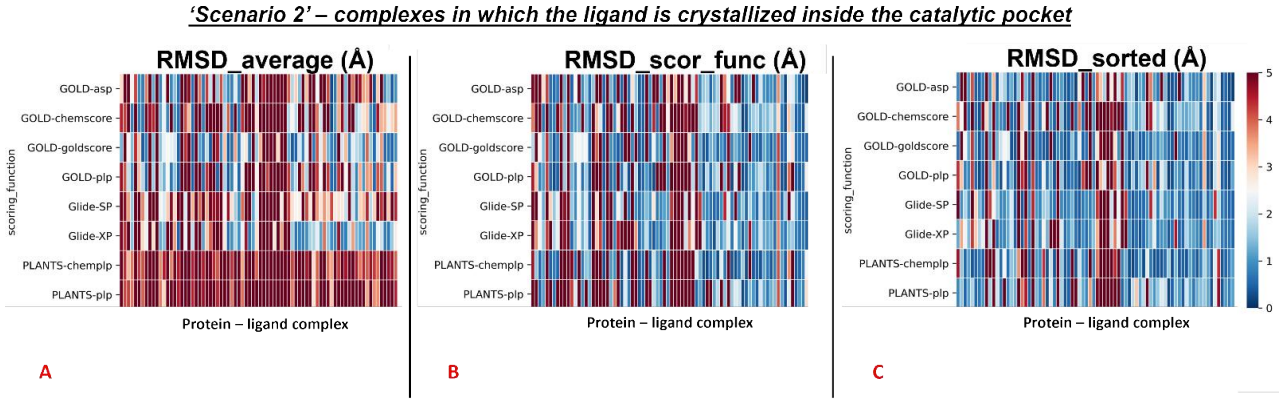


Figure 7. *Colormaps represent the results of the self-docking experiments just for the ligands crystallized inside the orthosteric pocket in the situation in which the crystallographic water molecules at 5 Å or nearer to the ligand itself are taken into account during the docking runs. The three plots illustrate respectively: (A) the results coming from the average of the RMSDs of all the poses for each docking run; (B) the results derived just from the RMSD between the crystallographic ligand coordinates and the once of the pose classified as the best from the scoring function; (C) the results of the self-docking experiments if just the pose showing the best RMSD value between its coordinates and the crystallographic ones are retained. The x-axis lists all the different protein-ligand complexes, which are plotted against the different pairs docking program-scoring function used for this study, reported in the y-axis.*

Second, we repeated the analysis for all the complexes in which the crystallographic ligand is located outside the orthosteric pocket. For these systems, the average of the RMSD coming from all the poses collected in the docking runs ("RMSD_average") was calculated to be 8.32Å. Next, we computed the mean of the RMSD values derived from the poses which received the highest score (from the scoring functions) in each run ("RMSD_scor_func"), and this value was 8.11Å.

Last, also the average value between the lowest of the RMSDs in each docking run was calculated ("RMSD_sorted"), giving 6.36 Å.

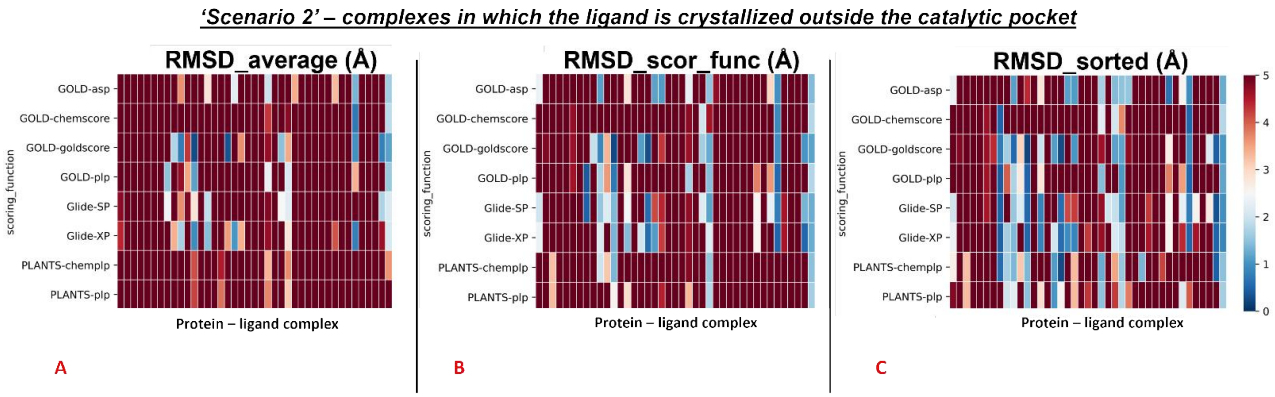


Figure 8. *Colormaps represent the results of the self-docking experiments only for the ligands crystallized outside the orthosteric pocket in the situation in which the crystallographic water molecules at 5 Å or nearer to the ligand itself are taken into account during the docking runs. The three plots illustrate respectively: (A) the results coming from the average of the RMSDs of all the poses for each docking run; (B) the results derived just from the RMSD between the crystallographic ligand coordinates and the once of the pose classified as the best from the scoring function; (C) the results of the self-docking experiments if just the pose showing the best RMSD value between its coordinates and the crystallographic ones are retained. The x-axis lists all the different protein-ligand complexes, which are plotted against the different pairs docking program-scoring function used for this study, reported in the y-axis.*

The results obtained for "scenario 1" are summarized in Table 2.

Results for "scenario 2" – water molecules 5 Å or nearer to the ligand considered in docking calculations			
	RMSD_average (Å)	**RMSD_scor_func (Å)**	**RMSD_sorted (Å)**
All the 119 protein-ligand complexes	5.64	4.83	3.68
The 78 complexes with the ligand inside the catalytic pocket	4.22	3.11	2.26
The 41 complexes with the ligand outside the catalytic pocket	8.32	8.11	6.36

Table 2. *Table representing the self-docking results obtained for "scenario 2".*

Just analyzing the numbers coming from the average values, is depictable how the performance of the docking programs dramatically increases when the ligand is docked inside the catalytic pocket rather than on the surface of the protein, in line with the fact that the molecules have a limitation in the conformation that they can explore into a binding site. Together with this, the ligands can exploit their accessible surface area to interact with the protein more efficiently, following the principle of "complementarity" [27,28].

2.3. Solvent exposure analysis

The results of the docking calculations were then analyzed in light of the data coming from the solvent exposure analysis. For each docking program-scoring function pair, the best RMSDs given by the docking calculation were evaluated against the solvent exposure of the ligand in its crystallographic pose. The results were reported in different plots, one for each couple docking program-scoring function, also in this case dividing the graphs in respect to the "scenario" from which the docking result was coming. To give an example, we reported in this article the plots for the pair GOLD-goldscore for each of these cases (Figures 9 and 10).

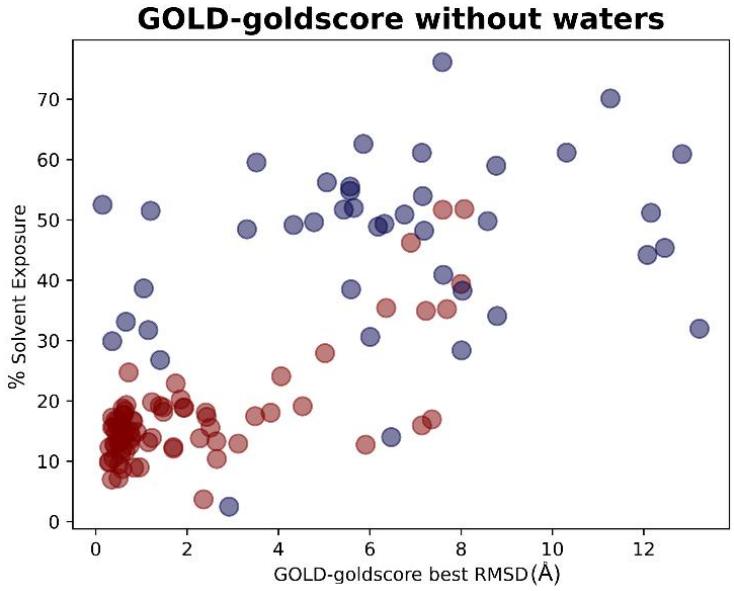


Figure 9. *Scatter plots showing the different distribution of the RMSD values between the coordinates of the best pose from the GOLD-goldscore docking experiment in respect to the solvent exposure of the corresponding crystallographic ligands. The red dots represent the values having the ligand crystallized inside the catalytic pocket while the blue dots represent the ligands crystallized on the other side of* M^{pro}*. As can be noticed, the molecules showing the*

best values of RMSD are in most cases located inside the orthosteric pocket and characterized by low solvent exposure. This plot depicts part of the results of "scenario 1", and so the crystallographic water molecules are not considered in the docking runs of which the outcomes are here represented.

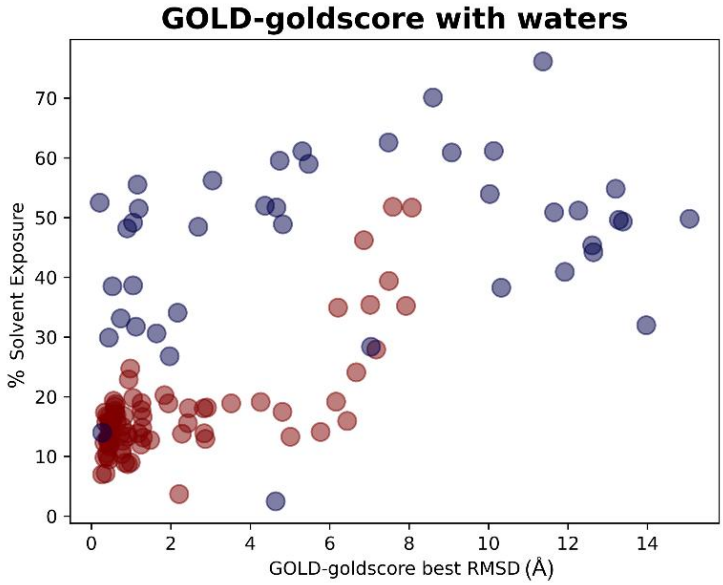


Figure 10. *Scatter plots showing the different distribution of the RMSD values between the coordinates of the best pose from the GOLD-goldscore docking experiment in respect to the solvent exposure of the corresponding crystallographic ligands. The red dots represent the ligands that are originally crystallized inside the catalytic pocket, while the blue dots represent the ligands crystallized in the other parts of M^{pro}. As can be noticed, the molecules showing the best values of RMSD are in most cases located inside the orthosteric pocket and characterized by low solvent exposure. This plot depicts part of the results of "scenario 2", which means that the crystallographic water molecules at 5 Å or nearer to the ligand are also considered in the docking runs of which the outcomes are here represented.*

The plots arising from all other docking program-scoring function pairs, both in "scenario 1" and "scenario 2", are reported in the supplementary material (Figure S1). From these graphs, it is easily depictable how the best RMSDs values tended to derive from protein-ligand complexes in which the solvent exposure of the ligand is low, and most of the time this means that the small molecule is crystallized in the binding pocket (indicated with the red dots in the plots). There are some cases in which the mean RMSD values were suboptimal also for this kind of ligands, and this can be due to several reasons. In some situations, of which the complexes 5REH, 5RE9, 5RGK (represented in Figure 11), and 7AVD are an example, the solvent exposure was tendentially higher in respect to the other orthosteric ligands, while in other cases the increase in RMSD can be attributable to the small dimensions of the ligand itself, which made harder for the docking

algorithms to reproduce the reference pose in a pocket of such considerable volume (the complexes 5R82 and 5RG0 are an example for this)[31].

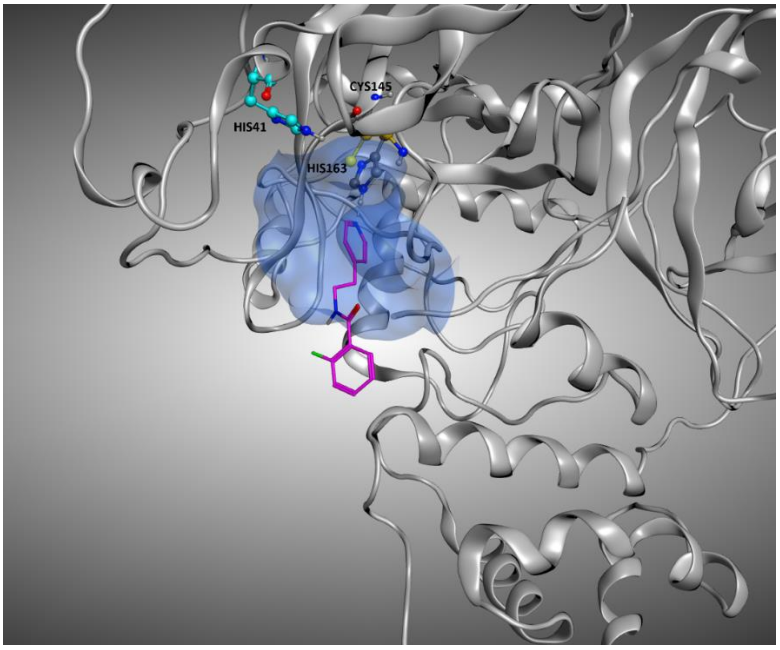


Figure 11. *Representation of the crystallographic complex conformation of 5RGK, one of the protein-ligand complexes in which the crystallographic ligand is located inside the orthosteric binding site, but the docking calculation results in high RMSD values. This is mainly due to the high level of solvent exposure that characterizes this ligand, which locates just a small portion of its structure inside the pocket, leaving the rest in an outer zone. The ligand is represented with stick representation (C-atom are colored in magenta), the catalytic dyad (Cys145 and His41) is highlighted, as well as the His163, the binding site residue interacting with the ligand. To give a better representation, the surface of the protein in a 5 Å radius from the ligand is represented and colored in blue*

On the other hand, there are also some cases in which the best RMSD given by the protocol was pretty low even if the crystallographic ligand is not placed inside the orthosteric pocket. This is the case, for example, of 7LFP (the crystallographic pose is reported in Figure 12); the ligand is placed at the interface between the monomers, and so its solvent exposure and RMSDs values were low even if is marked to be “outside the catalytic pocket”. A similar situation is observed on 5RF0, where the ligand, even if not located into the orthosteric pocket, is neither situated in the peripheral part of the protease.

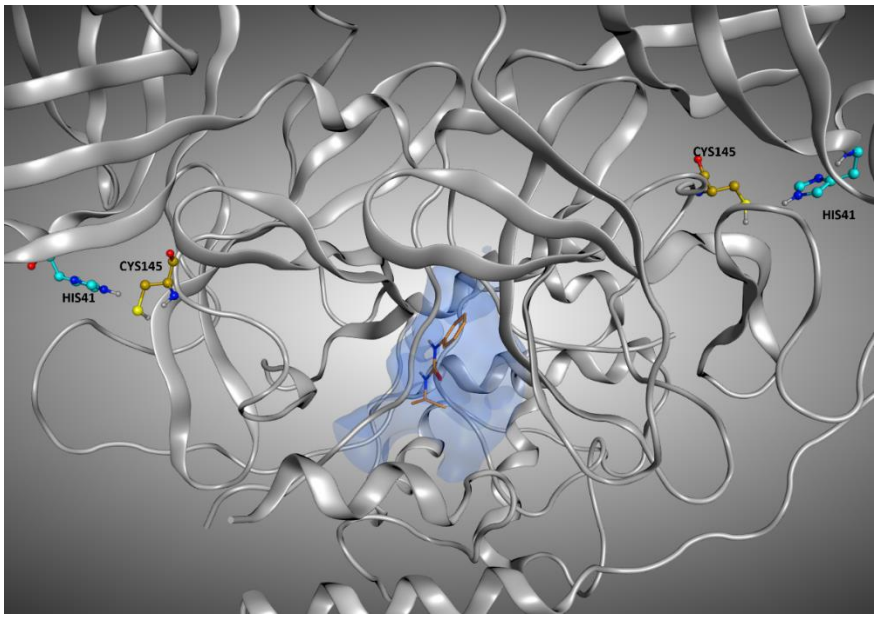


Figure 12. *Representation of the crystallographic pose of 7LFP, which is one of the protein-ligand complexes in which, even if the crystallographic ligand is located outside the orthosteric binding site, the RMSD values between the original coordinates and the ones given from the docking runs are considerably low. The reason for this can be found in the very low solvent exposure of this ligand, which is located in the interface between the monomers, and so is shielded by them. In the Figure, the ligand is represented in orange and the catalytic dyad (Cys145 and His41) of both monomers is highlighted. To give a better representation, the surface of the protein in a 5 Å radius from the ligand is represented and colored in blue.*

2.4. Molecular dynamics simulations

For each of the 119 crystallographic complexes, 3 different molecular dynamics simulations (MD) of 5 ns each were collected, to examine the behavior of the ligands in a dynamic environment. The trajectories were wrapped, aligned to the first frame and the root-mean-square fluctuation (RMSF) of the ligand, as well as the RMSD between its crystallographic and final coordinates ("RMSD_final"), were calculated for every single experiment. For each protein, the values coming from the average of the RMSFs and "RMSD_final" derived from the replicas were considered. Considering all the simulations collected, the average of all the ligand RMSF values was calculated to be 5.28 Å, while the average of the RMSD values between the coordinates of the crystallographic conformation of the ligand and the ones coming from the last frame of the trajectory ("RMSD_final") was of 8.89 Å.

As already done for the docking results analysis, we firstly focused on the complexes in which the crystallographic ligand is originally located inside the orthosteric pocket. For these systems, the

average of all the RMSFs coming from the simulations was 2.19 Å. The mean value of the RMSDs of the ligands in the last frame of each trajectory ("RMSD_final") was instead calculated to be 4.43 Å.

Second, we concentrated on the systems in which the crystallographic position of the ligand (and so its initial location) is outside the catalytic pocket. For these systems, the average value of all the ligand RMSFs during the trajectories was calculated to be 11.15Å. Then, the RMSD value between the final coordinates of the ligands and their crystallographic ones ("RMSD_final") were considered. The average of these values, for all the trajectories collected for these complexes, was 17.66Å. The output files of the molecular dynamics simulation geometric analysis are available in the supplementary material ("MD_data.csv").

As already done for the docking experiments, also for MD results the average values of RMSF and "RMSD_final" were plotted against the percentage of solvent exposure of the crystallographic conformation of the ligand, and the plots obtained have been reported in Figures 13 and 14.

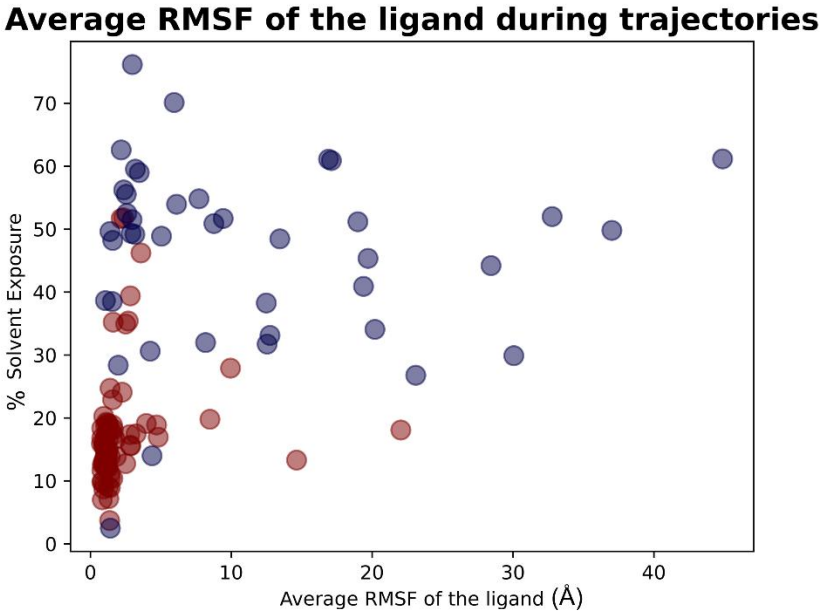


Figure 13. *Scatter plots showing the different distribution of the mean RMSF values between the coordinates of the M^{pro} ligands compared to crystallographic ones after the molecular dynamics simulations in respect to the solvent exposure of the corresponding crystallographic ligands. The red dots represent the ligands that are originally crystallized inside the catalytic pocket, while the blue dots represent the ligands crystallized in the other parts of M^{pro}. As can be noticed, the molecules showing the best values of RMSF after the analysis of the trajectories are mainly located inside the catalytic pocket and characterized by a low solvent exposure of the original crystallographic pose.*

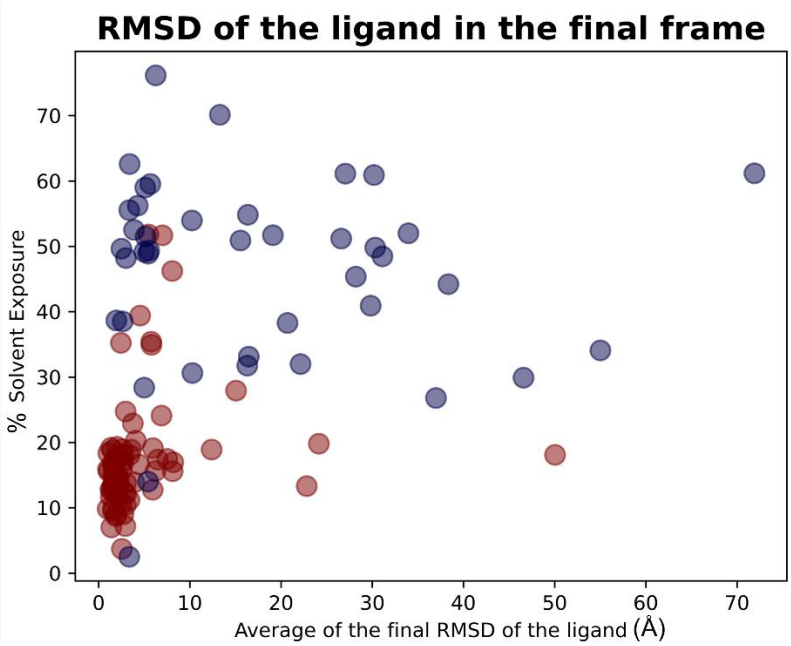


Figure 14. Scatter plots showing the different distribution of the mean RMSD values between the final coordinates of the M^{pro} ligands compared to crystallographic ones after the molecular dynamics simulations in respect to the solvent exposure of the corresponding crystallographic ligands. The red dots represent the ligands that are originally crystallized inside the catalytic pocket, while the blue dots represent the ligands crystallized in the other parts of M^{pro}. As can be noticed, the molecules showing the best values of RMSF after the analysis of the trajectories are mainly located inside the catalytic pocket and characterized by a low solvent exposure of the original crystallographic pose.

As expected, the complexes in which the ligand is crystallized in the orthosteric site (marked with the red dots in the scatter plot) tended to fluctuate much less than the ligands which are complexed in the external parts of the protease (represented with the blue dots in the graphs). As depicted, MD analysis confirms that the best values in terms of RMSF and "RMSD_final", which are correlated to a more energetically stable situation for the protein-ligand complex, come from the systems in which the crystallographic ligand is localized inside the catalytic pocket and are characterized by a low percentage of solvent exposure. These outcomes further support the already mentioned hypothesis about the correlation between the improvement of the docking performances in the case in which the binding site is a pocket rather than a surface.

The overall results obtained with molecular dynamics simulations are summarized in Table 3. A graphical representation of the molecular dynamics simulations is reported in the Supplementary

material (“Video_S1.mp4”). In this video, the ligands crystallized into the catalytic pocket were colored in magenta, while the other ligands were colored in cyan.

Results of the molecular dynamics simulations		
	RMSD_final (Å)	**RMSF_average (Å)**
All the 119 protein-ligand complexes	8.98	5.28
The 78 complexes with the ligand inside the catalytic pocket	4.43	2.19
The 41 complexes with the ligand outside the catalytic pocket	17.66	11.15

Table 3. *Table representing the results of the molecular dynamics experiments.*

3. Materials and Methods

3.1. Software overview

The molecular modeling operations were executed with the Molecular Operating Environment (MOE) suite (version 2019.01)[32]. The molecular docking calculations have been carried out with CCDC GOLD (version 2020), Schrodinger Glide (from the Schrödinger suite 2021.3), and PLANTS. The solvent exposure calculation was performed with a series of SVL commands (exploiting “moebatch” of the MOE suite) implemented into a python script. The systems for molecular dynamics simulations were prepared using tleap[33] and VMD[34]. The simulations were then carried out with ACEMD3[35] (version 3.3.0), a licensed program based upon OpenMM[36] (version 7.4.0). The modeling and docking calculations have been performed on a 12 CPU (Intel Xeon E5-1620 3.50 GHz) Linux Workstation, while the MD simulations were carried out on a GPUs-cluster based composed of 20 NVIDIA GPUs.

3.2. Structure preparation for docking calculations

The different crystal structures of M^{pro} were collected from the Protein Data Bank[37]. Among these, the proteins which did not present any ligand, or which were complexed with a covalent ligand, were excluded. This way, only the non-covalent protein-ligand complexes were retained,

and the complete list of all the 119 complexes used is available in the Supplementary material (Table S1). The structures were grouped into a database and were prepared with MOE "QuickPrep" tool. With this tool, each complex was properly prepared to recreate the small missing loops in the structure, assigning the proper conformation to the residues with alternate orientation (based on occupancy) and adding the hydrogens to the system (this last passage was performed with the MOE "Protonate 3D" tool). The hydrogens added this way were then minimized using the AMBER10:EHT force field implemented in MOE[38].

After these preliminary but crucial steps, each complex was manually examined and treated to eliminate every molecule except the crystallographic waters and the main ligand. Each complex was then independently saved.

3.3. Docking calculations

For each of the complexes prepared, the crystallographic ligand was separated from the protein and self-docked into its binding site. For each docking program, all the scoring functions available were used for separate runs, and each run 5 poses were collected for the ligand. GOLD supports 4 different scoring functions: goldscore, chemscore, asp, and plp; Glide supports two main functions for docking, which are Glide-SP and Glide-XP, while PLANTS implements plp and chemplp.

For each docking program-scoring function couple, the docking calculation has been carried out in two different scenarios: one in which the crystallographic water molecules were not considered (which we will refer to as "scenario 1") and one in which also the water molecules 5 Å or nearer from the ligand atoms were taken into account into the computation (which we will refer to as "scenario 2").

When all the docking calculations were executed, the ligand root-mean-square deviation (RMSD) between the coordinates of each one of the poses and the crystallographic conformations were computed. The data of major interest were: the RMSD in respect to the pose which is marked with the highest score by the program (RMSD_scor_func), the lowest RMSD of the docking run (RMSD_sorted), and the average of the RMSDs of all the poses generated (RMSD_average). The

output files of the self-docking experiments executed are available in the supplementary material ("Selfdocking_scenario1.csv" and "Selfdocking_scenario2.csv").

3.4. Solvent exposure calculation

For each M^{pro} complex, the solvent exposure of the main crystallographic ligand was calculated with an SVL script based on MOE "moebatch". The output of such calculation was the percentage of the ligand surface which is exposed to the solvent in the protein-ligand crystallographic complex. A table comprising all the percentages obtained is present in the Supplementary material (Table S2).

3.5. Molecular dynamics simulations set up and execution

All the protein-ligand M^{pro} systems were independently prepared for molecular dynamics simulations. The program tleap was used for the creation of the simulation box, which was set to be cubic and characterized by a 15 Å padding. The solvation model used was the explicit TIP3P, and the neutralization of the system was performed adding Na^+ and Cl^- ions until the salt concentration inside the box reached the value of 0.154 M.

The systems then underwent a two-passage equilibration. In the first one, both protein and ligand atoms were subjected to a harmonic position restrain of 5 kcal/mol. The length of this step was set to 0.1 ns and the ensemble used was the canonical one (NVT). During the second equilibration step, the ensemble was set to NPT (isothermal-isobaric), the length was 0.5 ns and the harmonic restrains (always 5 kcal/mol) were applied both on the ligand and on the alpha-carbons of the protein backbone.

After these preliminary steps, 3 replicas of 5 ns each were collected for each system, the ensemble was again the NVT one and no restraints were applied. At the end of the simulations, the average root-mean-square fluctuation (RMSF) of the ligand during the trajectory, as well as the RMSD between

crystallographic coordinates of the ligand and the ones coming from the last frame of the trajectory were collected.

4. Conclusion

In this study, we have evaluated the performance of three orthogonal docking algorithms in reproducing the crystallographic pose of several ligands, located in different parts of the same target, which in our case was the Sars-COV 2 main protease (M^{pro}). Our analysis revealed how, even if the docking programs used operate in different ways to give the final conformations to the user, all of them perform much better in the case in which the ligands are located in a binding pocket rather than crystallized outside of it. Specifically, we reported that their performance tends to decrease with the increment of the exposure of the crystallographic pose to the solvent. This was confirmed both from the experiments executed without considering the crystallographic water molecules in the docking calculations and from the ones taking into account the waters 5 Å or nearer to the ligand. Molecular dynamics simulations further give credit to our study, demonstrating how the less-fluctuating ligands (and so the most stable) through the trajectories were the once crystallized inside the orthosteric binding site of M^{pro}.

References

[1] I. D. Kuntz, J. M. Blaney, S. J. Oatley, R. Langridge, and T. E. Ferrin, “A geometric approach to macromolecule-ligand interactions,” *J. Mol. Biol.*, vol. 161, no. 2, pp. 269–288, Oct. 1982, doi: 10.1016/0022-2836(82)90153-X.

[2] X.-Y. Meng, H.-X. Zhang, M. Mezei, and M. Cui, “Molecular Docking: A Powerful Approach for Structure-Based Drug Discovery,” *Curr. Comput. Aided-Drug Des.*, vol. 7, no. 2, pp. 146–157, Jun. 2011, doi: 10.2174/157340911795677602.

[3] T. Lengauer and M. Rarey, “Computational methods for biomolecular docking,” *Curr. Opin. Struct. Biol.*, vol. 6, no. 3, pp. 402–406, Jun. 1996, doi: 10.1016/S0959-440X(96)80061-3.

[4] D. B. Kitchen, H. Decornez, J. R. Furr, and J. Bajorath, “Docking and scoring in virtual screening for drug discovery: methods and applications,” *Nat. Rev. Drug Discov.*, vol. 3, no. 11, pp. 935–949, Nov. 2004, doi: 10.1038/nrd1549.

[5] G. Jones, P. Willett, R. C. Glen, A. R. Leach, and R. Taylor, “Development and validation of a genetic algorithm for flexible docking 1 1Edited by F. E. Cohen,” *J. Mol. Biol.*, vol. 267, no. 3, Apr. 1997, doi: 10.1006/jmbi.1996.0897.

[6] T. A. Halgren *et al.*, “Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening,” *J. Med. Chem.*, 2004, doi: 10.1021/jm030644s.

[7] G. M. Morris *et al.*, “AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility,” *J. Comput. Chem.*, vol. 30, no. 16, Dec. 2009, doi: 10.1002/jcc.21256.

[8] O. Trott and A. J. Olson, “AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading,” *J. Comput. Chem.*, p. NA-NA, 2009, doi: 10.1002/jcc.21334.

[9] O. Korb, T. Stützle, and T. E. Exner, “PLANTS: Application of Ant Colony Optimization to Structure-Based Drug Design,” 2006.

[10] D. Yusuf, A. M. Davis, G. J. Kleywegt, and S. Schmitt, “An Alternative Method for the Evaluation of Docking Performance: RSR vs RMSD,” *J. Chem. Inf. Model.*, vol. 48, no. 7, pp. 1411–1422, Jul. 2008, doi: 10.1021/ci800084x.

[11] E. D. Boittier, Y. Y. Tang, M. E. Buckley, Z. P. Schuurs, D. J. Richard, and N. S. Gandhi, “Assessing Molecular Docking Tools to Guide Targeted Drug Discovery of CD38 Inhibitors,” *Int. J. Mol. Sci.*, vol. 21, no. 15, p. 5183, Jul. 2020, doi: 10.3390/ijms21155183.

[12] Z. Wang *et al.*, “Comprehensive evaluation of ten docking programs on a diverse set of protein–ligand complexes: the prediction accuracy of sampling power and scoring power,” *Phys. Chem. Chem. Phys.*, vol. 18, no. 18, pp. 12964–12975, 2016, doi: 10.1039/C6CP01555G.

[13] U. D. Ramirez, F. Myachina, L. Stith, and E. K. Jaffe, “Docking to Large Allosteric Binding Sites on Protein Surfaces,” 2010, pp. 481–488.

[14] C. Jacquemard, M. N. Drwal, J. Desaphy, and E. Kellenberger, “Binding mode information improves fragment docking,” *J. Cheminform.*, vol. 11, no. 1, p. 24, Dec. 2019, doi: 10.1186/s13321-019-0346-7.

[15] Y. Fan, K. Zhao, Z.-L. Shi, and P. Zhou, “Bat Coronaviruses in China,” *Viruses*, vol. 11, no. 3, p. 210, Mar. 2019, doi: 10.3390/v11030210.

[16] M. Lotfi, M. R. Hamblin, and N. Rezaei, “COVID-19: Transmission, prevention, and potential therapeutic opportunities,” *Clin. Chim. Acta*, vol. 508, pp. 254–266, Sep. 2020, doi: 10.1016/j.cca.2020.05.044.

[17] Latest access: 8th november 2021, “https://covid19.who.int /.” .

[18] Z. Jin *et al.*, “Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors,” *Nature*, vol. 582, no. 7811, pp. 289–293, Jun. 2020, doi: 10.1038/s41586-020-2223-y.

[19] C.-H. Zhang *et al.*, “Potent Noncovalent Inhibitors of the Main Protease of SARS-CoV-2 from Molecular Sculpting of the Drug Perampanel Guided by Free Energy Perturbation Calculations,” *ACS Cent. Sci.*, vol. 7, no. 3, pp. 467–475, Mar. 2021, doi: 10.1021/acscentsci.1c00039.

[20] D. R. Owen *et al.*, “An oral SARS-CoV-2 M ^{pro} inhibitor clinical candidate for the treatment of COVID-19,” *Science (80-.).*, Nov. 2021, doi: 10.1126/science.abl4784.

[21] M. Pavan, G. Bolcato, D. Bassani, M. Sturlese, and S. Moro, “Supervised Molecular Dynamics (SuMD) Insights into the mechanism of action of SARS-CoV-2 main protease inhibitor PF-07321332,” *J. Enzyme Inhib. Med. Chem.*, vol. 36, no. 1, pp. 1646–1650, Jan. 2021, doi: 10.1080/14756366.2021.1954919.

[22] M. Rudrapal *et al.*, “Identification of bioactive molecules from Triphala (Ayurvedic herbal formulation) as potential inhibitors of SARS-CoV-2 main protease (Mpro) through computational investigations,” *J. King Saud Univ. - Sci.*, vol. 34, no. 3, p. 101826, Apr. 2022, doi: 10.1016/j.jksus.2022.101826.

[23] V. Di Sarno *et al.*, “Identification of a dual acting SARS-CoV-2 proteases inhibitor through in silico design and step-by-step biological characterization,” *Eur. J. Med. Chem.*, vol. 226, p. 113863, Dec. 2021, doi: 10.1016/j.ejmech.2021.113863.

[24] A. Cuzzolin, M. Sturlese, I. Malvacio, A. Ciancetta, and S. Moro, “DockBench: An Integrated Informatic Platform Bridging the Gap between the Robust Validation of Docking Protocols and Virtual Screening Simulations,” *Molecules*, vol. 20, no. 6, pp. 9977–9993, May 2015, doi: 10.3390/molecules20069977.

[25] G. Bolcato, A. Cuzzolin, M. Bissaro, S. Moro, and M. Sturlese, “Can We Still Trust Docking Results? An Extension of the Applicability of DockBench on PDBbind Database,” *Int. J. Mol. Sci.*, vol. 20, no. 14, Jul. 2019, doi: 10.3390/ijms20143558.

[26] S. Zev, K. Raz, R. Schwartz, R. Tarabeh, P. K. Gupta, and D. T. Major, “Benchmarking the Ability of Common Docking Programs to Correctly Reproduce and Score Binding Modes in SARS-CoV-2 Protease Mpro,” *J. Chem. Inf. Model.*, vol. 61, no. 6, pp. 2957–2966, Jun. 2021, doi:

10.1021/acs.jcim.1c00263.

[27] G. Bolcato, M. Bissaro, M. Sturlese, and S. Moro, “Comparing Fragment Binding Poses Prediction Using HSP90 as a Key Study: When Bound Water Makes the Difference,” *Molecules*, vol. 25, no. 20, p. 4651, Oct. 2020, doi: 10.3390/molecules25204651.

[28] G. Bolcato *et al.*, “A Computational Workflow for the Identification of Novel Fragments Acting as Inhibitors of the Activity of Protein Kinase CK1δ,” *Int. J. Mol. Sci.*, vol. 22, no. 18, p. 9741, Sep. 2021, doi: 10.3390/ijms22189741.

[29] Y. Yan and S.-Y. Huang, “Pushing the accuracy limit of shape complementarity for protein-protein docking,” *BMC Bioinformatics*, vol. 20, no. S25, p. 696, Dec. 2019, doi: 10.1186/s12859-019-3270-y.

[30] H. A. Gabb, R. M. Jackson, and M. J. E. Sternberg, “Modelling protein docking using shape complementarity, electrostatics and biochemical information 1 1Edited by J. Thornton,” *J. Mol. Biol.*, vol. 272, no. 1, pp. 106–120, Sep. 1997, doi: 10.1006/jmbi.1997.1203.

[31] M. L. Verdonk, I. Giangreco, R. J. Hall, O. Korb, P. N. Mortenson, and C. W. Murray, “Docking Performance of Fragments and Druglike Compounds,” *J. Med. Chem.*, vol. 54, no. 15, pp. 5422–5431, Aug. 2011, doi: 10.1021/jm200558u.

[32] “Molecular Operating Environment (MOE), 2019.01; Chemical Computing Group ULC, 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2021.” .

[33] D.A. Case et al., “Amber 2021, University of California, San Francisco.” 2021.

[34] W. Humphrey, A. Dalke, and K. Schulten, “VMD: Visual molecular dynamics,” *J. Mol. Graph.*, vol. 14, no. 1, pp. 33–38, Feb. 1996, doi: 10.1016/0263-7855(96)00018-5.

[35] M. J. Harvey, G. Giupponi, and G. De Fabritiis, “ACEMD: Accelerating Biomolecular Dynamics in the Microsecond Time Scale,” *J. Chem. Theory Comput.*, vol. 5, no. 6, Jun. 2009, doi: 10.1021/ct9000685.

[36] P. Eastman *et al.*, “OpenMM 7: Rapid development of high performance algorithms for molecular dynamics,” *PLOS Comput. Biol.*, vol. 13, no. 7, p. e1005659, Jul. 2017, doi: 10.1371/journal.pcbi.1005659.

[37] H. M. Berman, “The Protein Data Bank,” *Nucleic Acids Res.*, vol. 28, no. 1, pp. 235–242, Jan. 2000, doi: 10.1093/nar/28.1.235.

[38] D.A. Case et al., “Amber 10, University of California, San Francisco.” .

Ribose and Non-Ribose A_{2A} Adenosine Receptor Agonists: Do They Share the Same Receptor Recognition Mechanism?

Giovanni Bolcato, Matteo Pavan, **Davide Bassani**, Mattia Sturlese and Stefano Moro

Bolcato, G., Pavan, M., Bassani, D., Sturlese, M., & Moro, S. (2022). Ribose and Non-Ribose A2A Adenosine Receptor Agonists: Do They Share the Same Receptor Recognition Mechanism? *Biomedicines 2022, Vol. 10, Page 515*, *10*(2), 515. https://doi.org/10.3390/BIOMEDICINES10020515

Abstract

Adenosine receptors have been a promising class of targets for the development of new therapies for several diseases. In recent years a renewed interest in this field has risen, thanks to the implementation of a novel class of agonists that lack the ribose moiety, once considered essential for the agonistic profile. Recently an X-ray crystal structure of A_{2A} Adenosine Receptor has been solved providing insights about the receptor activation from this novel class of agonists. Starting from this structural information, we have performed supervised molecular dynamics (SuMD) simulations to investigate the binding pathway of a non-nucleoside adenosine receptor agonist, as well as one of three classic agonists. Furthermore, we analyzed the possible role of water molecules in receptor activation.

1. Introduction

Adenosine is the endogenous agonist of a group of class A G protein-coupled receptors (GPCRs) named adenosine receptors (AR); four receptors belong to this family: A_1, A_{2A}, A_{2B}, and A_3. A_1 and A_3 are generally coupled to $G_{\alpha i}$ protein (so they inhibit adenylate cyclase enzyme upon activation), A_{2A} and A_{2B} are coupled to $G_{\alpha s}$ protein (and therefore stimulate the production of cAMP upon activation)[1].

AR are targets of interest for the treatment of several diseases[2]: Parkinson's disease[3][4], asthma[5], pain treatment[6], several cancer types[7], and cardiovascular diseases[8]. Despite this broad range of potential therapeutic applications, only two AR ligands have been approved: the A_{2A} antagonist Istradefylline, approved for the treatment of Parkinson's disease, and the A_{2A} agonist Regadenoson, used as a coronary vasodilator. One of the main problems in the translation of AR

ligands into therapeutic agents is the presence of unacceptable side effects due to the lack of selectivity of the drug candidates among different AR subtypes as well as off-target effects[9].

Progress was made in the field of AR agonists as therapeutic agents[10] with the publication by Bayer of some patents regarding non-nucleoside AR agonists[11]. While this novel class of AR agonists presents several advantages over classic adenosine-derived ligands (easier synthesis, improved pharmacokinetics, and oral bioavailability), the AR activation for this class of compounds has been difficult to understand since they lack the ribose moiety which was considered essential for the agonistic profile of AR ligands[12]. Some modifications on this moiety are tolerated and can improve both metabolic stability and potency, but often these alterations on the ribose unit lead to inactive compounds or switch the ligand activity toward an antagonistic profile[13].

To gain some insights on the structural basis of AR activation several site-directed mutagenesis data have been collected over the years[14]. Interestingly it was proven that the non-nucleoside A_{2A} agonist LUF5834 is sensitive to mutagenesis experiments in a different way compared to classic adenosine-like AR agonists like CGS21680[15]. In particular, the agonistic profile of LUF5834 is not affected when Ser-277 and Thr-88 are mutated in alanine (note that the enumeration, as well as in all this work, refers to A_{2A}). These two residues are essential for the agonistic activity of classic AR agonists along with His-278.

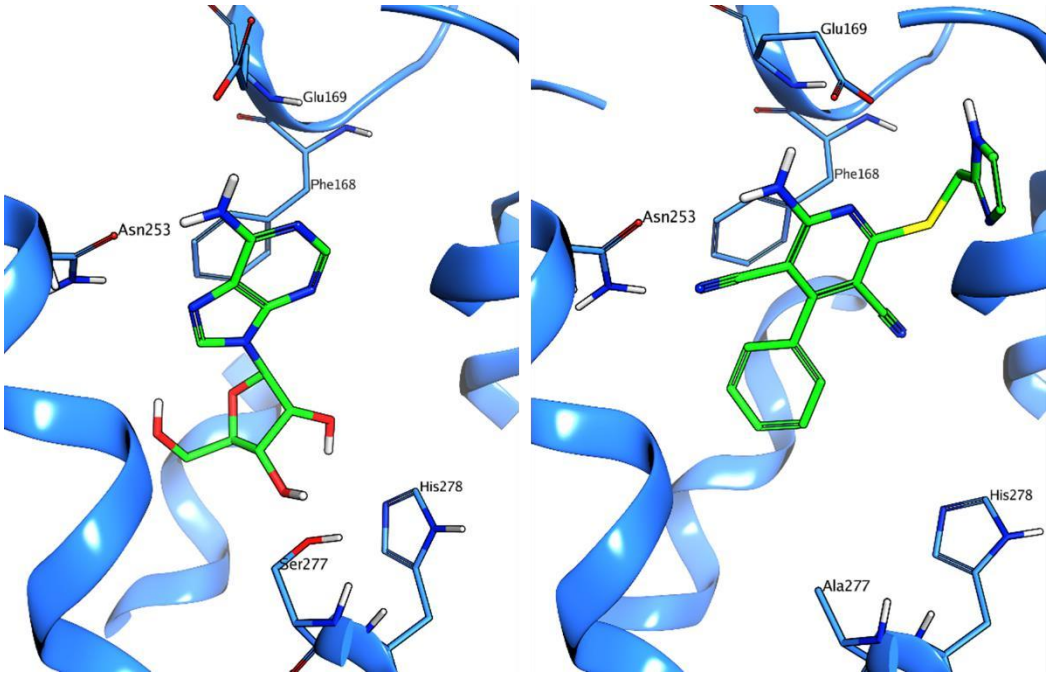


Figure 1. *Binding modes of Adenosine (left) and of LUF5833 (right) as observed in X-ray crystal structures (PDB code 2YDO and 7ARO respectively). Please note that the AR used to obtain the crystal structure with LUF5833 presents some thermostabilizing mutations including Ser-277. Other AR agonists, like CGS21680 and NECA, also interact with His-250 and Thr-88. The binding mode of these two ligands can be found in Figure 2.*

Fundamental progress in the comprehension of the structural basis of the agonistic action of non-nucleoside agonists has been made recently with the obtainment of the X-ray crystal structure of A_{2A} AR in complex with a close analog of LUF5834, LUF5833[16]. Interestingly the ligand does not form any hydrogen bonds with the abovementioned residues that are considered essential for the activation of A_{2A} AR (Thr-88, Ser-277, His-278).

In Figure 1, a comparison of the binding mode (as observed in X-ray crystal structures) that the endogenous agonist adenosine and of LUF5833 adopt in the orthosteric binding site of A_{2A} AR is reported.

Starting from this structural information, in the present work we have investigated the recognition process of LUF5833 and different classic adenosine-like agonists: CGS21680, NECA, and adenosine itself. The study has been carried out using supervised molecular dynamics simulations (SuMD) in order to gain structural information beyond the observed final bound state. SuMD is indeed a molecular dynamics-based approach that allows the sampling of events involving infrequent particle collision such as protein-ligand binding, without applying any energetic bias to the system.

The comparison between the binding trajectories collected for the two different classes of AR agonists reveals a different recognition pathway. Moreover, a detailed analysis of the behavior of water molecules during the binding event provides some insights into the possible role of the solvent molecules in the activation of the A_{2A} adenosine receptor.

2. Materials and Methods

2.1. System Setup

The three-dimensional structure of the protein-ligand complexes examined in this work (PDB codes: 2YDO, 2YDV, 4UG2, and 7ARO) was retrieved from the Protein Data Bank (PDB) and prepared for subsequent calculations using various tools provided by the Molecular Operating Environment (MOE) suite, version 2019.01 [17]. Residues with alternate conformation were assigned to the one with the highest occupancy. Missing hydrogen atoms were added to the system with the Protonate3D tool, assigning each titratable residue to the most probable protonation state at pH = 7.4. Crystallographic water molecules, ions, and other molecules present in the crystallization buffer were then removed, and the ligand was moved away from the binding site into the bulk, at a distance of at least 30 Å from the nearest receptor atom (higher than the cutoff chosen for electrostatic interaction computation).

The system preparation for Supervised Molecular Dynamics (SuMD) simulations was carried out with Visualize Molecular Dynamics (VMD[18]) version 1.9.3. At first, the protein-ligand system was explicitly solvated in a cubic TIP3P[19] water box, ensuring a distance of 15 Å between the box borders and any protein atom. Then, the system charge was neutralized by the addition of sodium and chlorine ions until a physiological concentration (0.154 M) was reached. Finally, the receptor was embedded in a lipid bilayer consisting of phosphatidylcholine (POPC) units.

From a methodological point of view, one main limitation of the SuMD technique, as is the case for traditional molecular dynamics (MD), is the fact that simulations are carried out assuming fixed protonation states. The prediction of the protonation state of titratable residues relies on a static structure (the crystal complex, which is the starting point for the simulations) and can sometimes be imprecise in those cases where the protein is flexible[20] or the residues are highly buried[21]. Furthermore, the co-existence of protonated and deprotonated states and dynamical processes coupled to a change in protonation states cannot be directly studied if the protonation states are fixed.

A second limitation is represented by the fact that the lipid constitution of the phospholipid membrane does not include the presence of cholesterol, which could exert some form of allosteric modulation on AR[22].

2.2. Molecular Dynamics

All simulations were carried out using the ACEMD[23] molecular dynamics engine. The system was described using parameters from the CHARMM36[24] force field (protein, lipids, ions, and water molecules), while ligand parameters were retrieved from Paramchem[25], a web front-end for the CgenFF[26] force field. If the parameters associated with specific dihedral angles of ligands presented high penalties, these have been parametrized using FFParam[27]. A QM scansion of the dihedral angle has been performed using MP2 level of theory with 6-31G** basis set, then the QM profile has then been fitted to retrieve the new force field parameters.

The simulation protocol consisted of a four-stage equilibration phase, followed by a productive SuMD simulation phase. For both equilibration and productive simulations, the integration timestep was set to 2 fs, the temperature was set to 310 K through a Langevin thermostat (friction coefficient = 0.1 ps^{-1}), the M-SHAKE algorithm was employed to constrain the length of bonds involving hydrogen atoms and the particle-mesh Ewald (PME[28]) was exploited to compute electrostatic interactions (grid length = 1 Å). Finally, a 9.0 Å cutoff was applied to long-term interactions.

2.3. Equilibration phase

Before equilibration MD simulations, 1500 steps of energy minimization using the conjugate-gradient method were performed, to remove clashes and bad contacts within the system. The first three equilibration MD simulations were carried out in the isothermal-isobaric ensemble (NPT), maintaining the system pressure fixed at 1 atm through the Berendsen barostat[29], while the fourth and final one was performed in the isothermal ensemble (NVT).

The first equilibration stage consisted of a 5 ns simulation with 1 Kcal mol^{-1} $Å^{-2}$ harmonic positional constraints applied on each receptor, ligand, and membrane atom. The second equilibration stage consisted of a 10 ns simulation with the same constraints applied only on each protein, ligand, and phosphorus atom. The third equilibration stage consisted of a 5 ns simulation with the same constraints applied only on the protein alpha carbons and on ligand atoms. Finally, a 10 ns equilibration MD simulation was performed without any constraints applied to the system.

2.4. Supervised Molecular Dynamics simulations

SuMD[30] is an enhanced sampling MD method that allows investigating molecular recognition processes at an atomistic level of detail in the nanosecond timescale without any energetic bias. The SuMD code is written in Python 2.7 and mainly exploits the Numpy and ProDy[31] packages to perform geometric supervision over a series of short classic MD trajectories (defined as "SuMD steps") carried out with the ACEMD engine. As reported in the original publication, each suMD step lasts 600 ps.

During each SuMD step, the distance between the center of mass (i.e., the hypothetical point where the entire mass of an object is assumed to be concentrated) of both the ligand and the binding site is monitored and collected at 5 evenly spaced time intervals. At the end of each step, these data are fitted in a straight line, which is then processed by a tabù-like algorithm: if the line slope is negative (indicating that the ligand is approaching the binding site), the step is considered productive and retained for the generation of the final MD trajectory, while the final state of this simulation is set as the initial state for the successive step. On the contrary, in the case where the slope is positive (indicating that the ligand is not approaching the binding site), the step is considered not productive and is discarded: in this case, the step is repeated reassigning the velocities through the Langevin thermostat. This process continues until the distance between the two centers of mass gets below

5 Å: from that point on, the supervision is turned off and the simulation proceeds as a classic MD simulation for other 30 SuMD steps.

2.5. Trajectory analysis

A per-residue energetic analysis was performed using an in-house developed Python script.

At first, the MDAnalysis[32][33] Python package is exploited to parse each MD trajectory and compute the number of contacts between the ligand and each protein residue, using a cutoff distance of 4.5 Å.

Afterward, the interaction energy (defined as the sum of the electrostatic and van der Waals contribution) is computed between the ligand and each one of the top 25 most contacted residues alongside each MD trajectory using the NAMD Energy Plugin[34] for VMD (version 1.4).

Finally, a heatmap is generated exploiting the Seaborn Python package: on the horizontal axis the simulation time in nanoseconds is reported, while on the vertical axis the residue name and index are reported for each residue considered for this analysis. The interaction energy is then plotted onto the heatmap using a colormap which ranges from red (indicating positive energy values, *i.e.,* a repulsive interaction) to blue (indicating negative energy values, *i.e.,* an attractive interaction). The first and the third quartile with regards to the distribution of interaction energy values are used as mask values for the heatmap generation.

To inspect the peculiar hydrodynamic profile of ligand LUF5833, the trajectory was analyzed with AquaMMapS[35], an in-house developed tool that allows investigating the behavior of water molecules within a receptor, based on their persistency across an MD trajectory. For this purpose, the simulation box is discretized in a voxel grid and the occupancy value for each cell is calculated as the ratio between the number of frames in which a water molecule occupies that cell and the total number of frames.

3. Results and Discussion

For all the four ligands a SuMD trajectory where the crystallographic observed binding mode is well reproduced has been collected. In Figure 2 the crystallographic binding mode and the final pose obtained with SuMD for the four ligands under examination are reported: as it can be seen the experimentally observed binding mode is well reproduced.

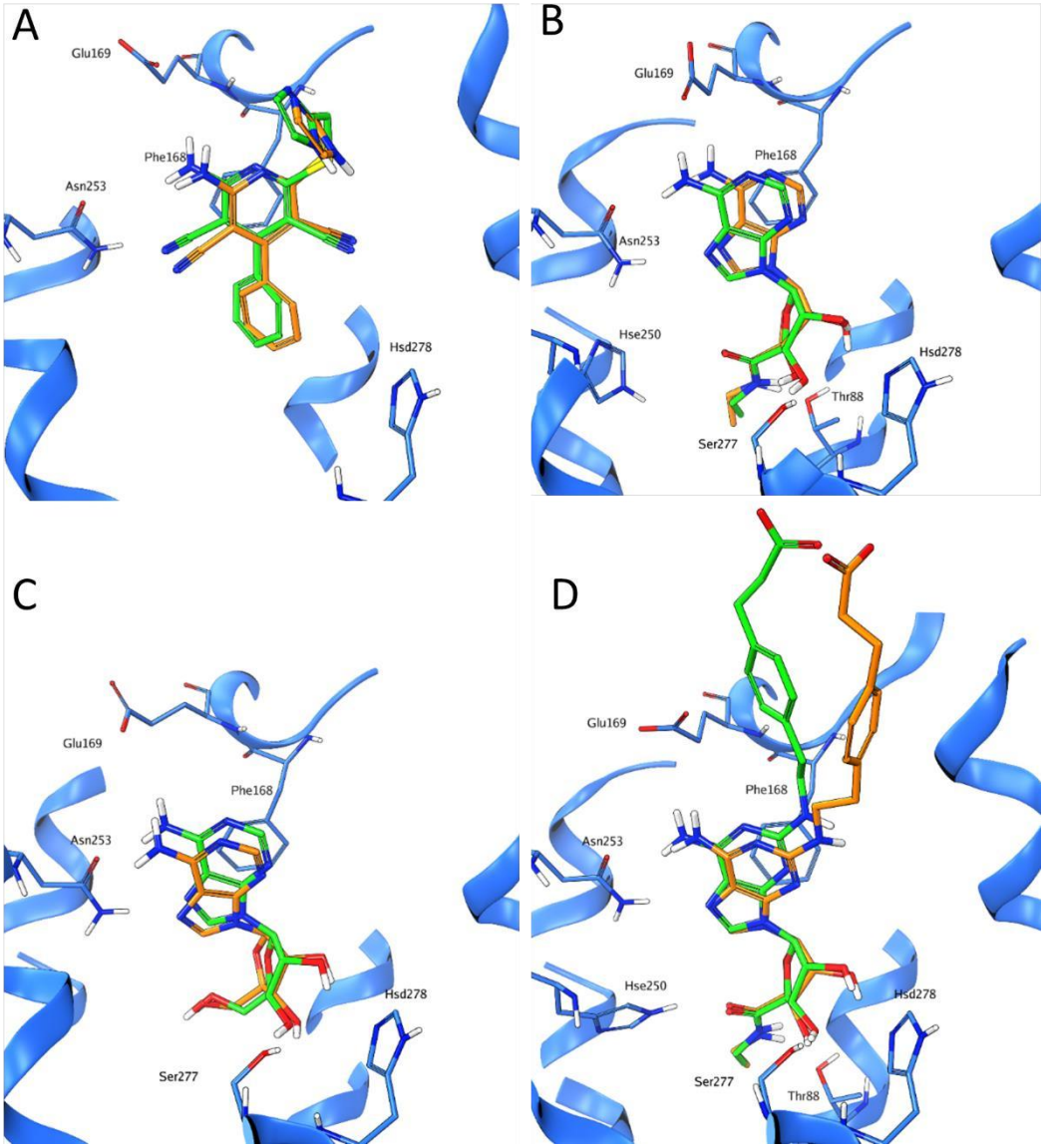


Figure 2. *The crystallographic binding mode (green) and the final pose obtained using SuMD (orange) for the four agonists used in this study (A: LUF5833; B: NECA, C: Adenosine; D: CGS21680).*

Detailed analysis on each trajectory has been performed to understand the recognition process for the four agonists (Figures 3-6). This analysis consists of a per-residue decomposition of the interaction energy between the ligand and the protein, during the binding event.

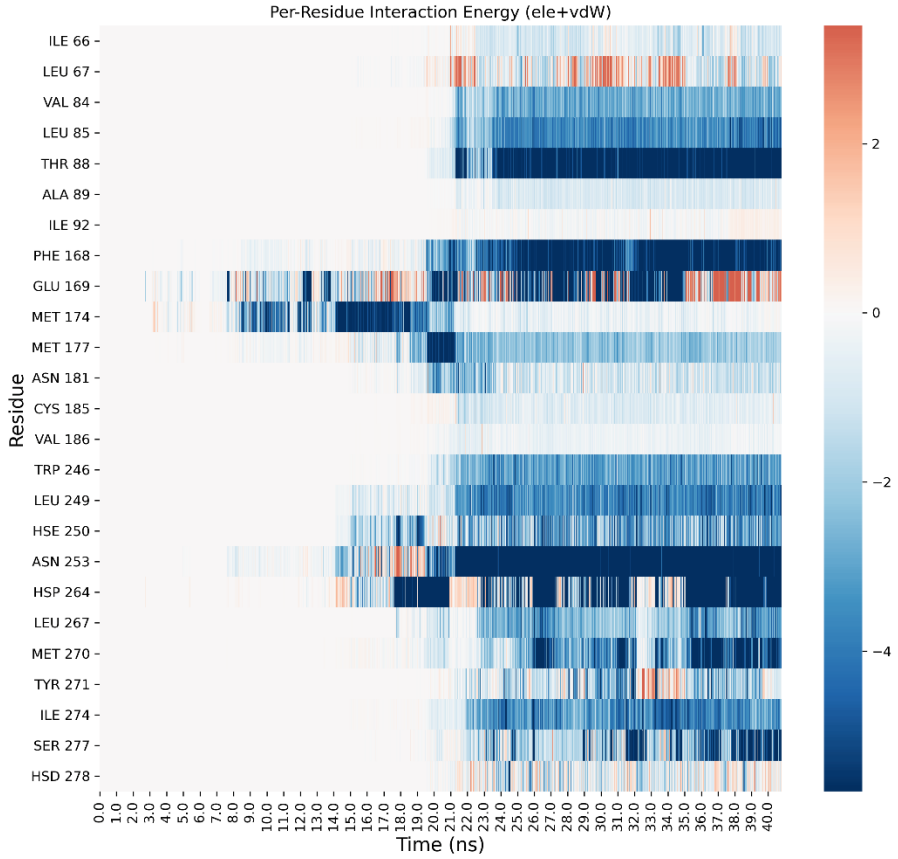


Figure 3. *Per-residue energetic analysis of the SuMD trajectory of CGS21680.*

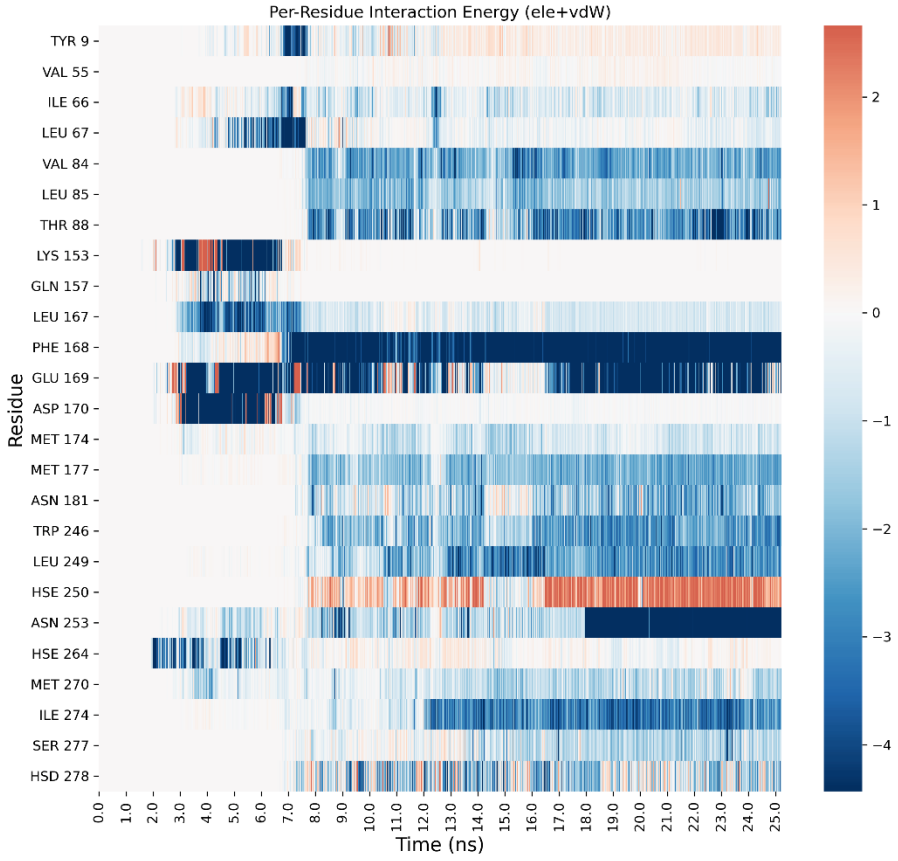


Figure 4. *Per-residue energetic analysis of the SuMD trajectory of Adenosine.*

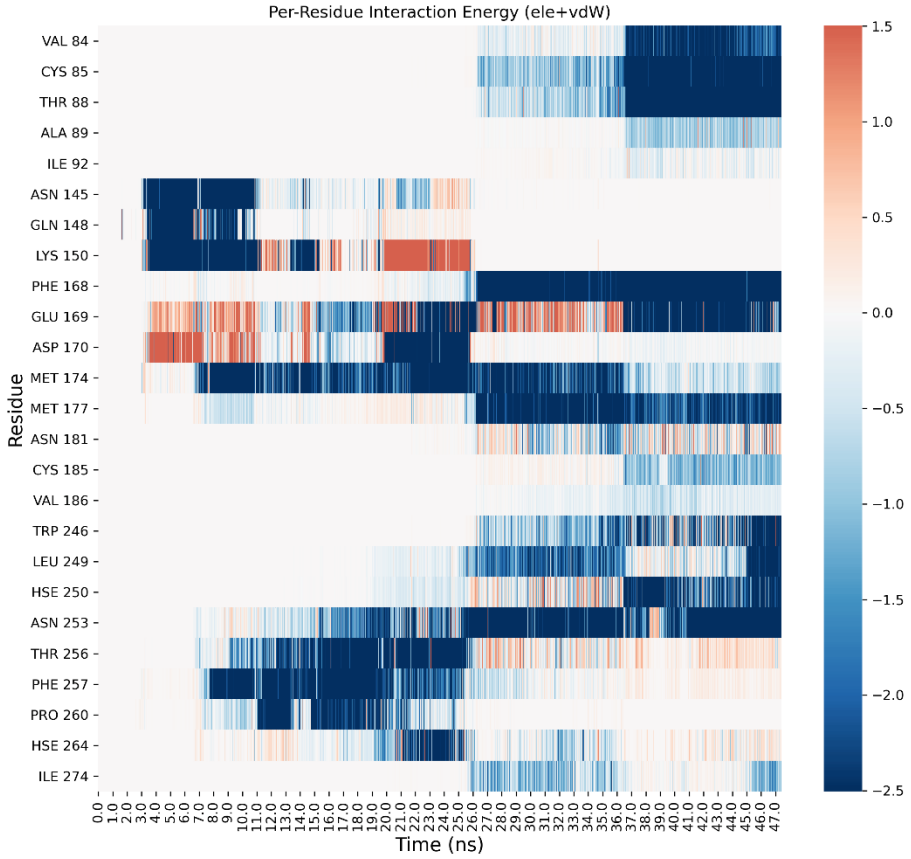


Figure 5. *Per-residue energetic analysis of the SuMD trajectory of NECA.*

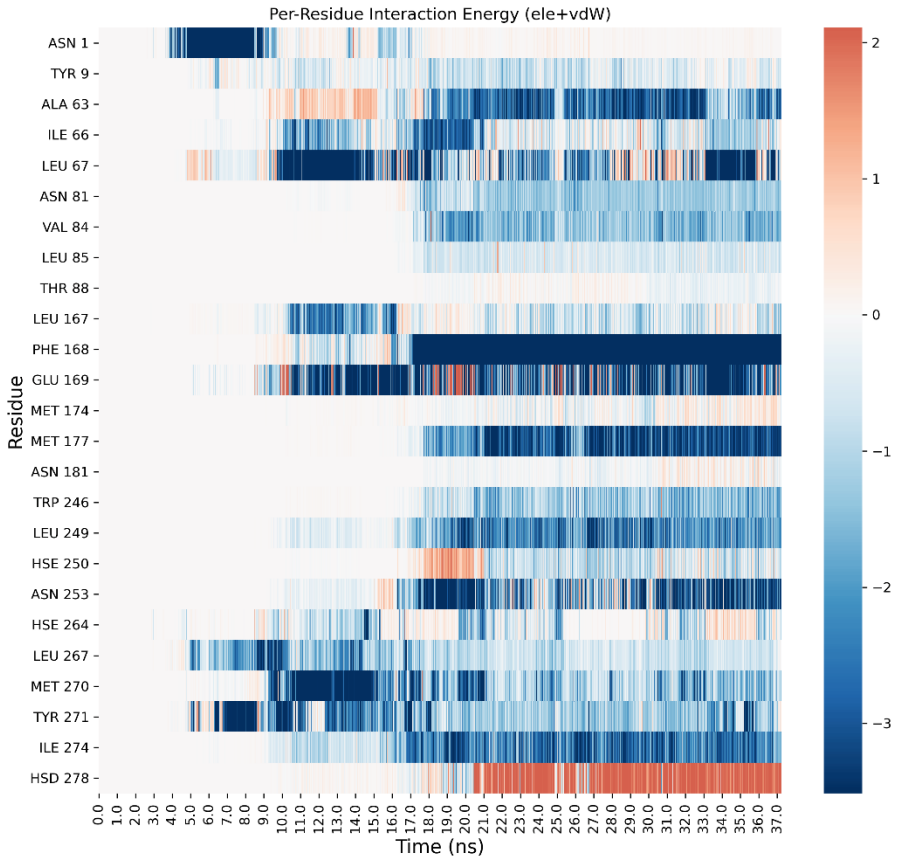


Figure 6. *Per-residue energetic analysis of the SuMD trajectory of LUF5833.*

The binding pathway for the four ligands as well as the most contacted regions of the protein can be visualized in Figure 7 and the videos collected in the Supplementary Information.

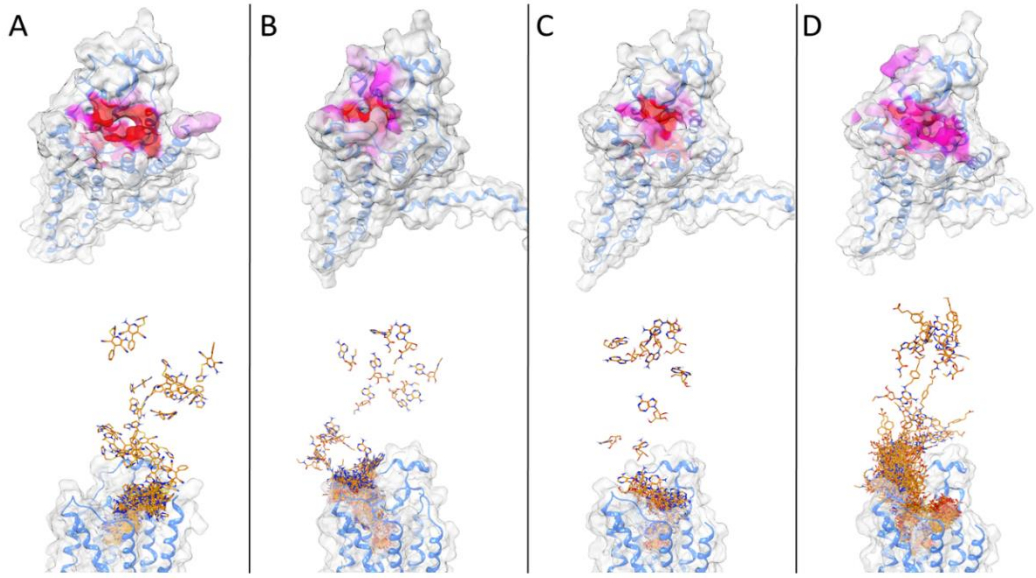


Figure 7. *In the upper part of the Figure the protein surface is colored according to the number of contacts with the ligand during the trajectory (scale white to red, from less contacted to more contacted residues). In the lower part the SuMD trajectory is displayed superposing each frame. A: LUF5833; B: NECA; C: Adenosine; D: CGS21680.*

The trajectory of LUF5833 has been prolonged for 25 ns at the end of the SuMD simulation. This prolonged trajectory has been analyzed using AquaMMapS (see Materials and Methods) to gain information on the possible role of the solvent in the activation mechanism of AR by non-nucleoside agonists.

For comparison also the first part (before the ligand reaches the orthosteric site) of the SuMD trajectory of LUF5833 has been analyzed using AquaMMapS. This analysis can provide some additional information on the solvent behavior in the apo form of the receptor. The results of these analyses are reported in Figure 8.

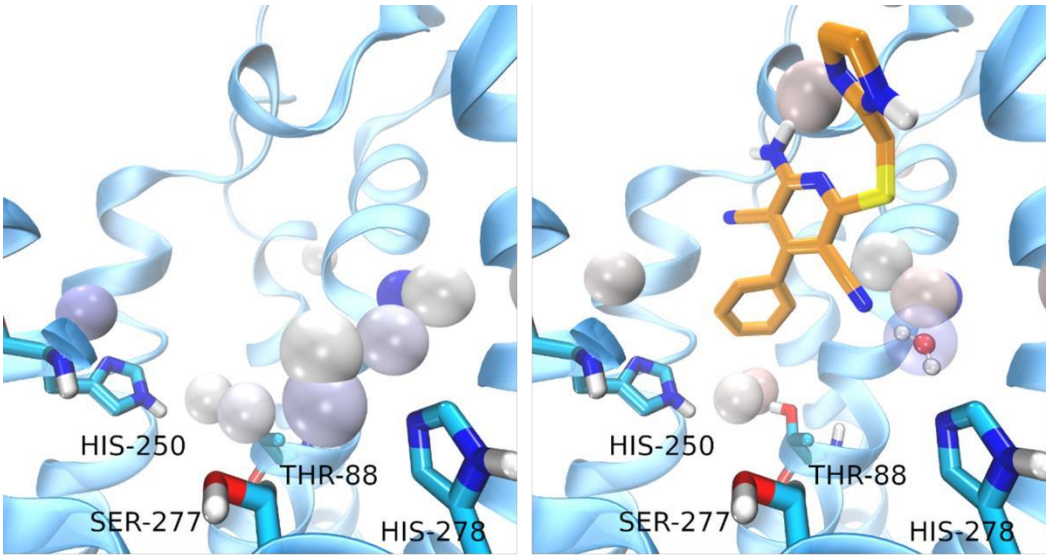


Figure 8. *Results of the AquaMMapS analysis for the "APO" trajectory on the left and for the prolonged SuMD trajectory of LUF5833. The cells where the water molecules have an occupancy value higher than 25% are displayed as spheres coloured according to the occupancy value (from white to blue).*

As it can be observed, the three classic ribose-containing agonists approach the protein taking contacts in a region that includes residues of the extracellular loop (ECL) 2 and 3 and transmembrane helical segments (TMs) 4, 5, and 6. This meta-binding site has already been described in our previous works[36][37]. LUF5833 instead approaches the receptor from the other side, making contacts in the region between the extracellular portion of TMs 1, 2, and 7. This suggests a different binding pathway for the two classes of AR agonists. Trajectory analysis correctly highlights the pivotal role that is played by Phe-168 (which is involved in a π-stacking with both types of agonists), according to mutagenesis studies which flagged this residue as fundamental for ligand binding[15]. Moreover, Asn-253, which establishes a double hydrogen bond with the adenine moiety of ribose agonists and a single hydrogen bond through one of the two nitrile groups of LUF5833, is also marked as an important residue for the recognition of both classes of agonists, according to mutagenesis data that illustrate how an N253A mutation would be detrimental for the activity of both ribose and non-ribose agonists[15]. Aside from these common interaction features that regard the adenine or "pseudo-adenine" portion of the molecule, the main difference in the recognition pattern of these two classes of agonists is related to the role of Ser-277 and Thr-88: as highlighted by our trajectory analysis, neither of these two residues establishes a direct interaction with LUF5833, in agreement with mutagenesis data which shows that mutation of these two residues negatively impact ribose agonists but have no effect on the affinity of non-ribose one[15].

Regarding the solvent behavior in the orthosteric site, it is interesting to note that the water molecules in the apo form of the receptor seem to adopt an interactive pattern that mimics the one

observed for agonists ligands. Indeed, key residues for the activation of the receptor, like Thr-88, His-250, Ser-277, and His-278 are well solvated and stabilize water molecules through hydrogen bonds.

It is tempting to argue that this observation (the organization of solvent molecules in a way that mimics agonists interactions) can provide a possible explanation for the concept of receptor basal activity, defined as the activation of the receptor in the absence of the ligand.

In detail, it seems that the stable water molecule interacting with Ser-277 is displaced upon LUF5833 binding while the water molecule interacting with His-278 is further stabilized by the cyano group in position 3. This water molecule is displayed in Figure 8. So, while it is true that LUF5833 does not interact directly with any key residues for the receptor activation, at least one of these interactions (the one with His-278) is still present and is mediated by a stable water molecule. Interestingly also the interaction between Adenosine and His-250 is mediated by a bridging water molecule, while NECA and CGS21680 interact with this residue using their amide tail.

Concerning water-bridged interactions, AquaMMapS analysis illustrates how LUF5833 seems to stabilize two water molecules that form a hydrogen bond bridge between His-250 and Thr-88 (Figure 8), playing a similar role to the amide tail of both NECA and CGS. Mutation of both residues has a detrimental role on ribose agonists' affinity, coherently with their direct interaction with the ribose moiety[38][39]. This could indicate that, while not interacting with His-250 and Thr-88, non-ribose agonists such as LUF5833 could stabilize a water molecule network that mimics the same interaction pattern of ribose agonists. The hydrophobic pocket which houses these stable water molecules is completed by Leu-85: this residue was determined to have a big impact on the affinity of ribose agonists such as CGS but has a smaller effect on the affinity of non-ribose agonists such as LUF5833[15]. This could be explained by the fact that this residue interacts directly with the ligand in the case of CGS, while in the case of LUF5833 its main involvement seems to be in the definition of a “hydrophobic” trap for these two water molecules that mimic the interaction pattern of ribose agonists. Notably, this bound water network extends also towards Asn-253: based on mutagenesis studies which show that, in the case of non-ribose agonists, the reduction of potency is mainly related to efficiency rather than on binding affinity, it is also tempting to speculate that this water network stabilized by LUF5833 is somehow involved in receptor activation, thereby validating the role of LUF5833 as a partial agonist.

Altogether, our SuMD simulations provide an overview of the mechanistic details regarding the recognition process between AR and their agonists, shedding light upon differences in the binding

event between nucleoside and non-nucleoside ones. Despite the useful information that can be gathered from our simulations, some AR-specific features cannot be captured by the SuMD technique, thereby impairing a clear and complete depiction of the agonist mechanism. Firstly, our simulations consider the interaction between one single ligand molecule and an individual receptor in a defined lipidic and ionic environment: despite being a sufficiently accurate approximation of reality for the evaluation of geometric properties related to the binding event, these boundary conditions cannot take into account the complex network of interactions of AR within a cellular environment, including the ones with themselves, other GPCRs and a plethora of ancillary factors[40], which leads to surprising pharmacological properties[41][42]. Secondly, a key aspect of AR agonist signaling is portrayed by the ligand residence time, which has been flagged as a more efficient predictor of “in vivo” functional efficacy than binding affinity[43]. Although the evaluation of this aspect of agonist signaling was beyond the scope of this scientific work, it is important to underline that the association process is only the first part of a more complex and intricate story.

References

[1] Müller, C. E. & Jacobson, K. A. Recent developments in adenosine receptor ligands and their potential as novel drugs. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1808**, 1290–1308 (2011).

[2] Chen, J. F., Eltzschig, H. K. & Fredholm, B. B. Adenosine receptors as drug targets-what are the challenges? *Nature Reviews Drug Discovery* (2013) doi:10.1038/nrd3955.

[3] Navarro, G., Borroto-Escuela, D. O., Fuxe, K. & Franco, R. Purinergic signaling in Parkinson's disease. Relevance for treatment. *Neuropharmacology* **104**, 161–168 (2016).

[4] Shah, U. & Hodgson, R. Recent progress in the discovery of adenosine A2A receptor antagonists for the treatment of Parkinson's disease. *Current Opinion in Drug Discovery and Development* Preprint at (2010).

[5] Gao, Z.-G. G. & Jacobson, K. A. Purinergic Signaling in Mast Cell Degranulation and Asthma. *Frontiers in Pharmacology* **8**, (2017).

[6] Vincenzi, F., Pasquini, S., Borea, P. A. & Varani, K. Targeting Adenosine Receptors: A Potential Pharmacological Avenue for Acute and Chronic Pain. *International Journal of Molecular Sciences* **21**, 8710 (2020).

[7] Fishman, P. *et al.* Adenosine Receptors and Cancer. in 399–441 (2009). doi:10.1007/978-3-540-89615-9_14.

[8] Mustafa, S. J., Morrison, R. R., Teng, B. & Pelleg, A. Adenosine Receptors and the Heart: Role in Regulation of Coronary Blood Flow and Cardiac Electrophysiology. in 161–188 (2009). doi:10.1007/978-3-540-89615-9_6.

[9] Schmidt, J. & Ferk, P. Safety issues of compounds acting on adenosinergic signalling. *Journal of Pharmacy and Pharmacology* **69**, 790–806 (2017).

[10] Jacobson, K. A., Tosh, D. K., Jain, S. & Gao, Z.-G. Historical and Current Adenosine Receptor Agonists in Preclinical and Clinical Development. *Frontiers in Cellular Neuroscience* **13**, (2019).

[11] Albrecht-Küpper, B. E., Leineweber, K. & Nell, P. G. Partial adenosine A1 receptor agonists for cardiovascular therapies. *Purinergic Signalling* **8**, 91–99 (2012).

[12] Soudijn, W., Wijngaarden, I. & Ijzerman, A. Medicinal Chemistry of Adenosine A1 Receptor Ligands. *Current Topics in Medicinal Chemistry* **3**, 355–367 (2003).

[13] Cristalli, G., Lambertucci, C., Marucci, G., Volpini, R. & Ben, D. A2A Adenosine Receptor and its Modulators: Overview on a Druggable GPCR and on Structure-Activity Relationship Analysis and Binding Requirements of Agonists and Antagonists. *Current Pharmaceutical Design* **14**, 1525–1552 (2008).

[14] Jespers, W. *et al.* Structural Mapping of Adenosine Receptor Mutations: Ligand Binding and Signaling Mechanisms. *Trends in Pharmacological Sciences* **39**, 75–89 (2018).

[15] Lane, J. R. *et al.* A Novel Nonribose Agonist, LUF5834, Engages Residues That Are Distinct from Those of Adenosine-Like Ligands to Activate the Adenosine A 2a Receptor. *Molecular Pharmacology* **81**, 475–487 (2012).

[16] Amelia, T. *et al.* Crystal Structure and Subsequent Ligand Design of a Nonriboside Partial Agonist Bound to the Adenosine A 2A Receptor. *Journal of Medicinal Chemistry* **64**, 3827–3842 (2021).

[17] Molecular Operating Environment (MOE), 2019.01; Chemical Computing Group ULC, 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2021. https://www.chemcomp.com/Research-Citing_MOE.htm.

[18] Humphrey, W., Dalke, A. & Schulten, K. Sartorius products. *Journal of molecular graphics* **14**, 33–38 (1996).

[19] Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J Chem Phys* **79**, 926–935 (1983).

[20] Wallace, J. A. & Shen, J. K. Predicting pKa values with continuous constant pH molecular dynamics. *Methods in enzymology* **466**, 455–475 (2009).

[21] Wallace, J. A. *et al.* Toward accurate prediction of pKa values for internal protein residues: The importance of conformational relaxation and desolvation energy. *Proteins: Structure, Function, and Bioinformatics* **79**, 3364–3373 (2011).

[22] Huang, S. K. *et al.* Allosteric modulation of the adenosine A2A receptor by cholesterol. *eLife* **11**, (2022).

[23] Harvey, M. J., Giupponi, G. & de Fabritiis, G. ACEMD: Accelerating biomolecular dynamics in the microsecond time scale. *J Chem Theory Comput* **5**, 1632–1639 (2009).

[24] Brooks, B. R. *et al.* No Title. **30**, 1545–1614 (2009).

[25] CGenFF Home. https://cgenff.umaryland.edu/.

[26] Vanommeslaeghe, K. *et al.* CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J Comput Chem* **31**, 671–690 (2010).

[27] Kumar, A., Yoluk, O. & MacKerell, A. D. FFParam: Standalone package for CHARMM additive and Drude polarizable force field parametrization of small molecules. *Journal of Computational Chemistry* **41**, 958–970 (2020).

[28] Essmann, U. *et al.* A smooth particle mesh Ewald method. *J Chem Phys* **103**, 8577 (1998).

[29] Berendsen, H. J. C., van der Spoel, D. & van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications* **91**, (1995).

[30] Sabbadin, D. & Moro, S. Supervised molecular dynamics (SuMD) as a helpful tool to depict GPCR-ligand recognition pathway in a nanosecond time scale. *J Chem Inf Model* **54**, 372–376 (2014).

[31] Bakan, A., Meireles, L. M. & Bahar, I. ProDy: Protein dynamics inferred from theory and experiments. *Bioinformatics* **27**, 1575–1577 (2011).

[32] Michaud-Agrawal, N., Denning, E. J., Woolf, T. B. & Beckstein, O. MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *J Comput Chem* **32**, 2319–2327 (2011).

[33] Gowers, R. J. *et al.* MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. *Proceedings of the 15th Python in Science Conference* 98–105 (2016) doi:10.25080/MAJORA-629E541A-00E.

[34] Phillips, J. C. *et al.* Scalable molecular dynamics on CPU and GPU architectures with NAMD. *J Chem Phys* **153**, 044130 (2020).

[35] Cuzzolin, A., Deganutti, G., Salmaso, V., Sturlese, M. & Moro, S. AquaMMapS: An Alternative Tool to Monitor the Role of Water Molecules During Protein–Ligand Association. *ChemMedChem* **13**, 522–531 (2018).

[36] Sabbadin, D., Ciancetta, A., Deganutti, G., Cuzzolin, A. & Moro, S. Exploring the recognition pathway at the human A 2A adenosine receptor of the endogenous agonist adenosine using supervised molecular dynamics simulations. *MedChemComm* **6**, 1081–1085 (2015).

[37] Deganutti, G., Welihinda, A. & Moro, S. Comparison of the Human A 2A Adenosine Receptor Recognition by Adenosine and Inosine: New Insight from Supervised Molecular Dynamics Simulations. *ChemMedChem* **12**, 1319–1326 (2017).

[38] Jiang, Q. *et al.* Hydrophilic side chains in the third and seventh transmembrane helical domains of human A2A adenosine receptors are required for ligand recognition. *Molecular Pharmacology* **50**, (1996).

[39] Kim, J., Wess, J., van Rhee, A. M., Schöneberg, T. & Jacobson, K. A. Site-directed Mutagenesis Identifies Residues Involved in Ligand Recognition in the Human A2a Adenosine Receptor. *Journal of Biological Chemistry* **270**, 13987–13997 (1995).

[40] Fredholm, B. B., Ijzerman, A. P., Jacobson, K. A., Klotz, K. N. & Linden, J. International Union of Pharmacology. XXV. Nomenclature and classification of adenosine receptors. *Pharmacological Reviews* **53**, 527–552 (2001).

[41] Lopes, L. V., Cunha, R. A. & Ribeiro, J. A. Cross talk between A(1) and A(2A) adenosine receptors in the hippocampus and cortex of young adult and old rats. *Journal of neurophysiology* **82**, 3196–3203 (1999).

[42] Cunha, R. A., Correia-de-Sá, P., Sebastião, A. M. & Ribeiro, J. A. Preferential activation of excitatory adenosine receptors at rat hippocampal and neuromuscular synapses by adenosine formed from released adenine nucleotides. *British journal of pharmacology* **119**, 253–260 (1996).

[43] Guo, D., Mulder-Krieger, T., IJzerman, A. P. & Heitman, L. H. Functional efficacy of adenosine A2A receptor agonists is positively correlated to their receptor residence time. *British Journal of Pharmacology* **166**, 1846–1859 (2012).

Omicron variant of SARS-CoV-2 virus: *in silico* evaluation of the possible impact on people affected by diabetes mellitus

Davide Bassani, Eugenio Ragazzi, Annunziata Lapolla, Giovanni Sartore, Stefano Moro

D. Bassani, E. Ragazzi, A. Lapolla, G. Sartore, and S. Moro, "Omicron Variant of SARS-CoV-2 Virus: In Silico Evaluation of the Possible Impact on People Affected by Diabetes Mellitus," *Front. Endocrinol. (Lausanne).*, vol. 13, Mar. 2022, doi: 10.3389/fendo.2022.847993.

Abstract

The Omicron variant of SARS-CoV-2 (Spike mutant B.1.1.529) carrying more than 30- point mutations in its structure, of which 15 are localized in the receptor-binding domain (RBD), allows to hypothesize a relevant change in interactivity with ACE2. In previous reports we hypothesized that the worse outcome of the COVID-19 disease in diabetes mellitus condition could be related to the non-enzymatic glycation of ACE2 receptor and an in silico evaluation led to the demonstration that the number of interactions is decreased in comparison to the unmodified model, possibly shifting the virus attack through different, multiple alternative entry routes. Given the evidenced features of this variant, we aimed to investigate with a computational approach the characteristics of Omicron SARS-CoV-2 with respect to its binding to human ACE-2 receptor, in a particular population, namely people affected by diabetes mellitus, at risk for unfavorable outcomes of the COVID-19. The computational analysis, considering the case in which all the lysine residues in the system are subjected to non-enzymatic glycation, confirmed that lysine glycation causes a general loss of interactivity between wild-type (WT)-Spike-RBD and ACE2. In the Omicron variant, Lys417 mutates into an asparagine, preventing the possible non-enzymatic glycation of this residue. Therefore, if non-enzymatic glycation seemed to cause a shift in the way in which the virus enters the cell from the ACE2-mediated mechanism to other pathways, in the case of the Omicron variant the ACE2-mediated approach of the virus seems to remain an important event to take into account. Indeed, interaction profile analysis, together with molecular mechanics–generalized Born surface area (MM-GBSA) calculations, suggests that the Omicron-Spike-RBD maintains a higher affinity for ACE2 subsequently to non-enzymatic glycation with respect to WT-Spike-RBD. The finding of the present computational study may suggest a different clinical relevance of the Omicron variant for the diabetes mellitus field, also in the possible direction of a lower severity of the disease.

1. Introduction

The unprecedented occurrence of a set of several genetic mutations in the Omicron variant of the SARS-CoV-2 virus has drawn particular attention within scientists and media. Omicron-Spike mutant B.1.1.529 carries more than 30-point mutations in its structure, of which 15 are localized in the receptor-binding domain (RBD) [1]. Significant changes have occurred, also with respect to the Delta-Spike variant (Figure 1). Moreover, 11 of these mutations are located at the contact surface with ACE2, allowing us to hypothesize that Omicron-Spike could have a relevant change in interactivity with ACE2. Some of these mutations, such as Q498R and N501Y, have already proven to lead to an increased affinity with ACE2 in respect to WT-Spike [2]. The Technical Advisory Group on SARS-CoV-2 Virus Evolution (TAG-VE) has advised WHO that the Omicron variant should be designated as a variant of concern (VOC), due to the epidemiological parameters initially reported in South Africa, and now spreading around the world [3]. This is the fifth VOC to be reported since the beginning of the pandemic [4]; following the experience with the previously reported variants associated with new worsening of the pandemic, a great concern has arisen whether a relevant change in transmissibility and binding affinity is to be expected with the new Omicron variant. Moreover, most recent data suggest that mutations occurring at RBD influence negatively the activities of neutralizing antibodies induced by vaccines, or administered as monoclonal antibody therapy [5]. However, preliminary observations from South Africa suggest that the SARS-CoV-2 Omicron variant is linked to a reduced risk of severe disease when compared to the Delta variant [6,7]. Since the first times of the pandemic, increased morbidity has been evidenced in people affected by diabetes mellitus [8][9][10][11], and the reason for this phenomenon is at present still debated. In a previous report, we hypothesized that the worse outcome of the COVID-19 disease in diabetes mellitus condition could be related to the non-enzymatic glycation of ACE2 receptor, due to hyperglycemic environment, triggering a higher interaction with virus Spike protein [12]. However, a specific in silico evaluation of the interaction between ACE2 receptor and SARS-CoV-2 Spike protein under different conditions of nonenzymatic glycation has led to the demonstration that the number of interactions is decreased in comparison to the unmodified model [13], possibly shifting the virus attack through different, multiple alternative entry routes. In particular, the interaction of SARS-CoV-2 with human cells has been suggested also be mediated by Transmembrane Protease Serine-2 (TMPRSS2) activity [14][15], as well as by other receptor pathways, such as Neuropilin-1 (NRP1) [14][15][16], dipeptidyl peptidase 4

(DPP4) also known as cluster of differentiation 26 (CD26) [14][15][16][17], the transmembrane glycoprotein CD147 (basigin 2) [14], and glucose-regulated protein 78 (GRP78) [18]. However, the role of these alternative routes of virus entry in diabetes is still a matter of debate and investigation [13]. In view of the evidenced features of this variant, we aimed to investigate with a computational approach the characteristics of Omicron SARS-CoV-2 with respect to its binding to human ACE-2 receptor, in a particular population, namely people affected by diabetes mellitus, at risk for unfavorable outcomes of the COVID-19. The in silico analysis, after the recognition of the main interactions occurring between virus and ACE2 receptor, was directed to the specific evaluation of the impact to the affinity of the system, induced by a hyperglycemic environment, conditioning non-enzymatic glycation at lysine residues of both Omicron-Spike-RBD and ACE2 receptor.

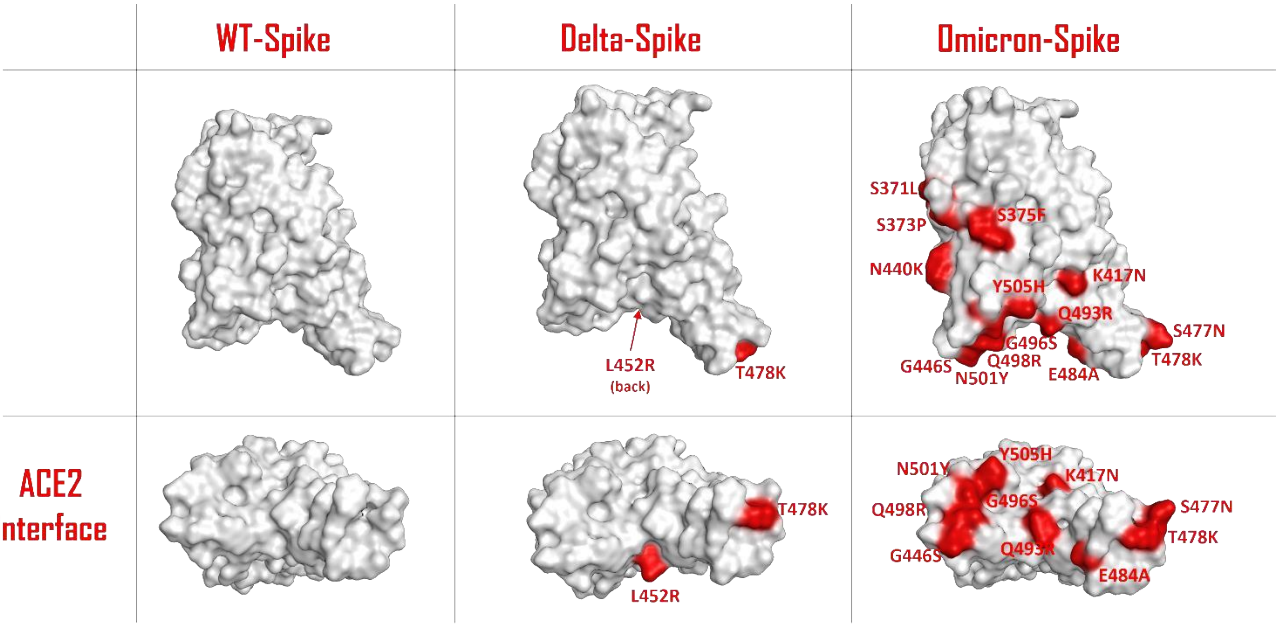


Figure 1. *Schematic representation of the SARS-COV-2 Spike protein receptor-binding domain (RBD) surface for, respectively, the wild-type (WT), the Delta variant and the Omicron variant. The residues involved in mutations are colored in red and are also labelled. The upper panels represent a lateral view of the Spike-RBDs considered, while the lower panels highlight the protein surface facing the ACE2 interaction interface for each situation examined.*

2. Materials and Methods

The computational analysis presented in this study was conducted starting from two different experimental structures, which were both downloaded from the Protein Data Bank (PDB) [19]. The first of these systems represents the wild-type (WT) form of SARS-CoV-2 Spike RBD complexed with ACE2 receptor (PDB code: 6M0J; method: X-ray diffraction; resolution: 2.45 Å) [20], while the second involves

the SARS-CoV-2 Spike-RDB Omicron variant forming a complex with ACE2 (PDB code: 7T9L; method: cryo-electron microscopy (Cryo-EM); resolution: 2.66 Å) [21]. The proteins were prepared for molecular modeling with the "Structure Preparation" tool implemented in the Molecular Operating Environment (MOE) suite [22]. The missing hydrogen atoms were added exploiting the MOE "Protonate 3D" tool, which assigns to each residue the most probable protonation state at the chosen pH, which in our case was set at the value of 7.4. These hydrogen atoms were then minimized using the AMBER10:EHT [23] force field implemented in MOE. To create the glycated forms of both WT-Spike-RBD/ACE2 and Omicron-Spike-RBD/ACE2 systems, the lysine residues of the prepared complexes 6M0J and 7T9L were manually changed. This operation involved the addition of a cyclic amino sugar moiety, which results from the Maillard reaction between the lysine amino acid and D-glucose, followed by the Amadori rearrangement [24] (as depicted in Supplementary Figure 1). This step allowed us to obtain the four systems of interest for our study, considering various conditions of non-enzymatic glycation (here labeled as "glyco"). These complexes are WT-Spike-RBD with ACE2, WT-Spike-RBD-glyco with ACE2-glyco, Omicron-Spike-RBD with ACE2, and Omicron-Spike-RBD-glyco with ACE2-glyco. Once the different systems were obtained, the glycated lysine residues were energetically minimized under the AMBER10:EHT [23] force field implemented in MOE. This passage was essential in order to allow the newly introduced sugar groups to be properly orientated in the environment. After these preliminary steps, the interactions between Spike-RBD and ACE2 were analyzed both visually and with the "GetContacts" tool [25]. This last program can extrapolate and classify all the interactions between the biological entities of a system. In our case, the specific contacts between Spike-RBD and ACE2 were evaluated for all the cases. To give a better view of the macroscopic changes brought by the mutation and the glycation of the systems, the electrostatic surface of both ACE2 and Spike-RBD were also calculated for all the situations considered (as depicted in Figure 2). To further inspect the effect of glycation on the interaction between Spike-RBD and ACE2 receptor for both WT and Omicron variants, molecular mechanics–generalized Born surface area (MM-GBSA) calculations were executed in the complexes using the Schrödinger Prime application [26]. This method consists of the estimation of the binding free energy between two entities in an environment, exploiting an implicit solvation model. In our case, the Surface Generalized Born Model and Variable Dielectric (VSGB) solvation model [27] was used, and the force field in which the complexes were evaluated was the OPLS4 force field [28] implemented in Prime.

3. Results

A visual representation of the complexes created considering various conditions of non-enzymatic glycation on the SARSCoV-2 Spike RBD/ACE2 receptor system is reported in Supplementary Figures 2–5. A brief report of the number of the interactions (expressed in terms of the number of pairs of residues that are in contact between the two proteins) in each system is reported in Table 1.

Interaction type	WT-Spike-RBD on ACE2	WT-Spike-RBD-glyco on ACE2-glyco	Omicron-Spike-RBD on ACE2	Omicron-Spike-RBD-glyco on ACE2-glyco
Salt bridge interactions	1	0	2	1
Hydrogen bonds	9	10	11	10
van der Waals interactions	30	21	32	37

Table 1. *Table representing the numerosity of the interactions (expressed in terms of the number of residues pairs that are in contact between the two proteins) extrapolated from the four systems considered in the study presented (WT-Spike-RBD on ACE2, WT-Spike-RBD-glyco on ACE2-glyco, Omicron-Spike-RBD on ACE2, and Omicron-Spike-RBD-glyco on ACE2-glyco). The "GetContacts" tool was exploited in order to calculate all the contacts between Spike and ACE2 in each of the scenarios considered.*

The interaction pattern occurring in the native WT-SpikeRBD/ACE2 system (Supplementary Figure 2) is characterized by polar bonds (comprising both salt bridges and hydrogen bonds) as well as non-polar van der Waals interactions. The WTSpike-RBD-glycated/ACE2-glycated system (Supplementary Figure 3), mimicking the maximum level of glycation attainable in a hyperglycemic condition in the case of native SARS-CoV-2, is characterized by a reduction in the number of interactions, mainly of a non-polar entity. The effect of glycation on the new Omicron SARS-CoV-2 variant (compare Supplementary Figure 4 and Supplementary Figure 5), produces a diminished number of bonds as well, occurring in virus/receptor interaction. As depicted from the results obtained, while for the WT-Spike-RBD/ACE2 situation the reduction in the number of non-polar interactions due to glycation is not linked to an overall change in the number of polar contacts, for the Omicron variant, the small decrease in the number of hydrogen bonds subsequent to the non-enzymatic glycation of lysine amino acids seems to be compensated by an increase in non-polar interactions.

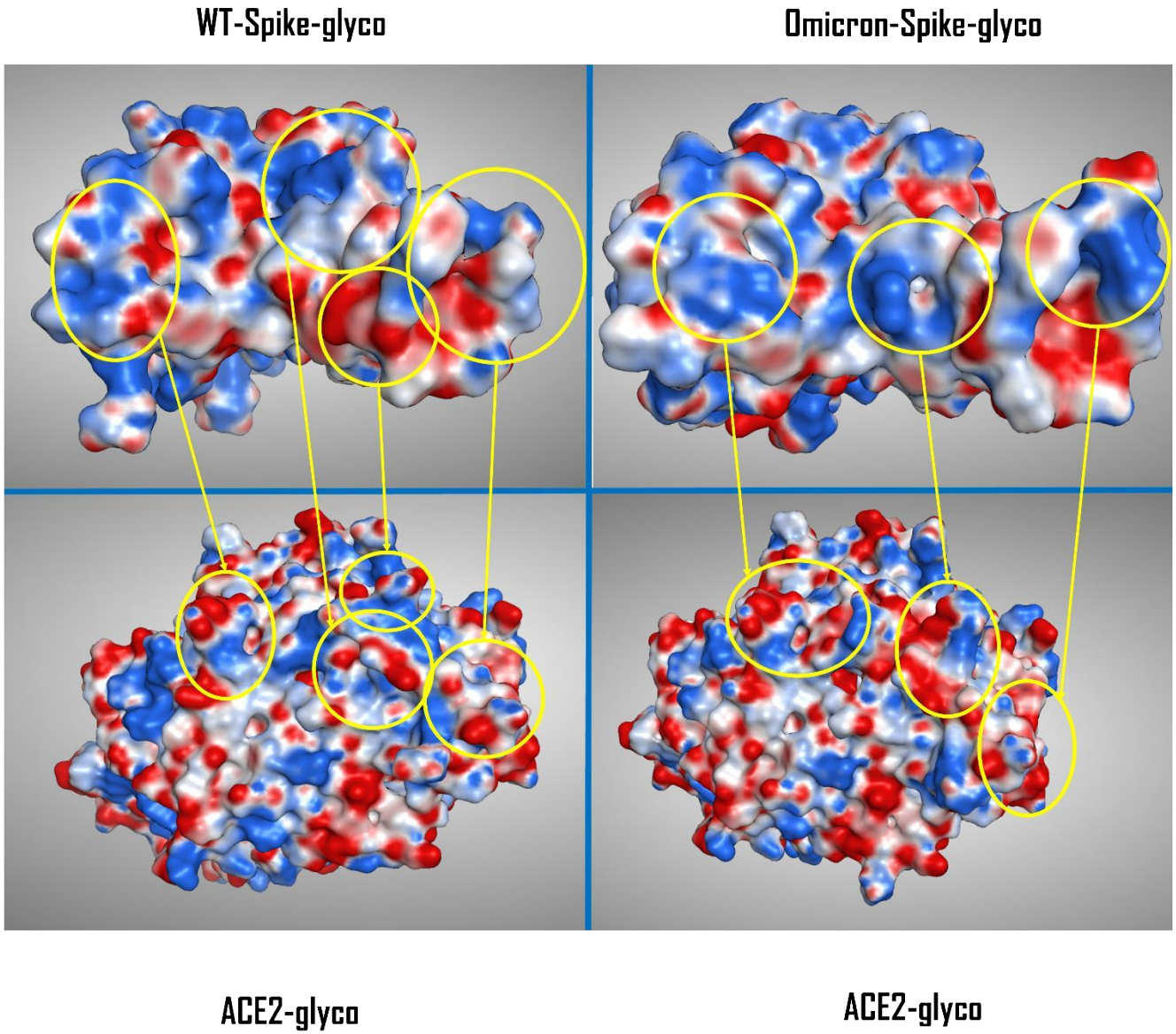


Figure 2. *Comparison between the electrostatic contact surface between the WT-Spike-RBD-glyco/ACE2-glyco system (upper and lower left) and the Omicron-Spike-RBD-glyco/ACE2-glyco complex (upper and lower right). The areas of highest variation in electrostatic distribution between glycated WT and Omicron Spike-RBDs are circled in yellow and are then connected with the corresponding ACE2-glyco interacting surface portion. The Spike-RBD/ACE2 backbone conformation for the WT variant comes from the crystal with PDB code 6M0J, while for the Omicron variant the data come from the Cryo-EM structure with PDB code 7T9L. WT, wild type; RBD, receptor-binding domain; PDB, Protein Data Bank; Cryo-EM, cryo-electron microscopy*

Consequently, we can hypothesize that, if the decrement in non-polar contacts between WT-Spike-RBD/ACE2 and WT-Spike-RBD-glyco/ACE2-glyco systems is not balanced by a significant strengthening of the polar component of the interaction, for the Omicron variant, where the number of polar bonds is higher with respect to the WT in both native and glycated forms, the interaction between the two

glycated proteins could be more efficiently preserved. This is also supported by the MM-GBSA calculation results, reported in Table 2, which highlight the loss in interaction energy that both the WT-Spike-RBD/ACE2 and the Omicron-Spike-RBD/ ACE2 complexes experienced after glycation. The data coming from this analysis confirm that Omicron-Spike-RDB is able to form with ACE2 a more stable complex in respect to the WT-Spike-RBD (with a difference in the binding free energy of about 11 kcal/mol). Moreover, the binding free energies obtained demonstrate that glycation causes a general loss in interactivity towards ACE2 for both for WT-Spike-RBD and Omicron Spike-RBD, and that this drop is much more significant for the WT variant.

Interaction type	WT-Spike-RBD on ACE2	WT-Spike-RBD-glyco on ACE2-glyco	Omicron-Spike-RBD on ACE2	Omicron-Spike-RBD-glyco on ACE2-glyco
MM-GBSA total free energy of binding (kcal/mol)	-114.37	-89.29	-125.59	-111.83

Table 2. *Table reporting the free energy of binding (expressed in kcal/mol) obtained with the MM-GBSA calculations executed on the different complexes considered in this study (WT-Spike-RBD/ACE2, WT-Spike-RBD-glyco/ACE2-glyco, Omicron-Spike-RBD/ACE2, and Omicron-Spike-RBD-glyco/ACE2-glyco).*

To allow a more comprehensive view of the overall changes caused by the Omicron mutation and the glycation of the systems, the electrostatic surface of both ACE2 and Spike-RBD were calculated for all the situations considered, and the differences between the electrostatic distribution were highlighted (Supplementary Figure 6 and Figure 2). As illustrated by these pictures, it is interesting to notice the linkage between the change in the surface features of the proteins and the locations in which glycation takes place. To represent this parallelism, the glycation sites were highlighted on the Spike-RBD surface, and the results are reported in Figure 3. As depicted, the only Spike-RBD lysine residue directly in contact with ACE2 is Lys417 in the native form. Although in the Omicron variant this lysine mutates into an arginine, Thr478 mutates into lysine (T478K), forming another glycation site on the ACE2 interface.

4. Discussion

The computational analysis allowed us to evaluate the changes in the interaction pattern between Spike-RBD and ACE2 receptor considering both mutation (focusing on the Omicron variant) and the presence of a hyperglycemic environment. The modified affinity between these two biological entities, which seems higher for Omicron-Spike-RBD to WT-Spike-RBD, has been reported [21][29][30]. In a

hyperglycemic environment, typical of diabetes, we considered the case in which all the lysine residues in the system are subjected to non-enzymatic glycation. The behavior of the proteins will so change with respect to the position of the lysine amino acids. A recent work of our group highlighted the possibility that diabetic patients in which this glycation process takes place are less prone to be infected by SARS-CoV-2 in an ACE2-dependent way [13]. Indeed, our analysis here confirms the result and suggests that lysine glycation causes a general loss of interactivity between WT-Spike-RBD and ACE2.

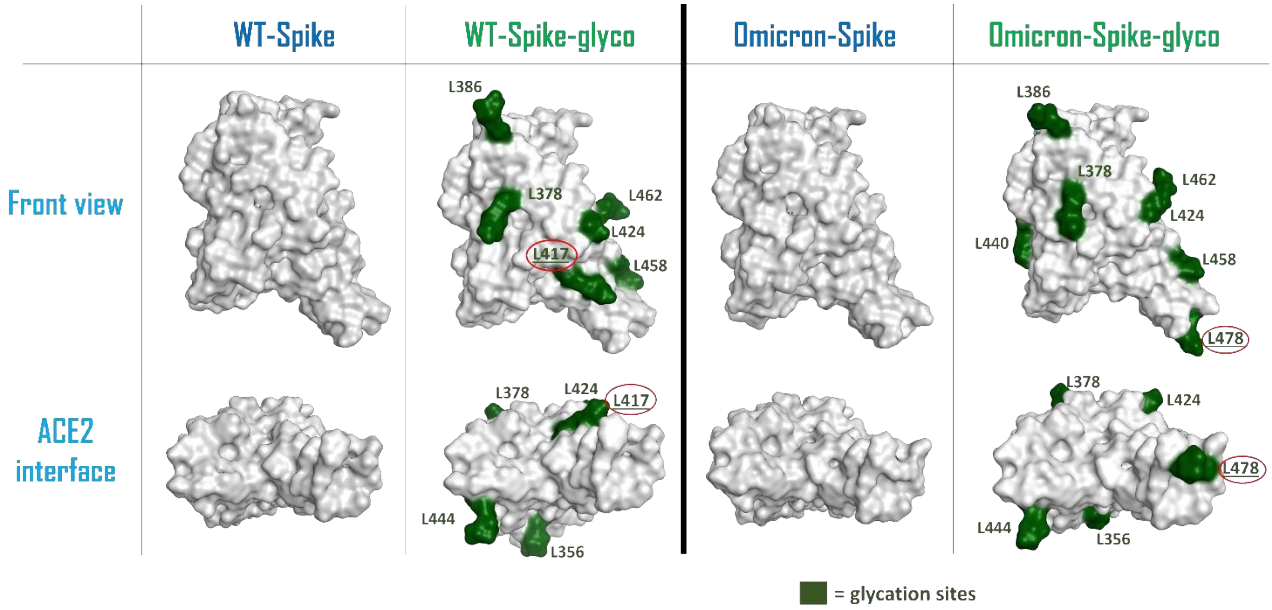


Figure 3. *Schematic representation of the surfaces of the possible forms of Spike-RBD considered in this computational study. From the left, the systems considered are: wild-type (WT) Spike-RBD, the lysine-glycated form of wild-type (WT) Spike-RBD, the Omicron variant Spike-RBD and the glycated form of Omicron variant Spike-RBD. For the glycated systems, the surface of lysine amino acids has been colored in dark green and has also been labeled. The lysine residues which are located on the contact interface with ACE2 (Lys417 for the WT-Spike-RBD-glyco and Lys478 for the Omicron-Spike-RBD-glyco) have their label underlined and circled in red. The upper figures represent a lateral view of the Spike-RBDs considered, while the lower highlight the protein surface facing the ACE2 interaction interface for each situation examined.*

One of the main WT-Spike-RBD residues involved in the interaction with ACE2 is Lys417, which is the lysine amino acid that is most exposed to the ACE2 interface when Spike-RBD approaches the host cells. Indeed, glycation at this residue could lead to a weakening of the contacts between the two entities, mainly due to the loss of the strong salt bridge between the positive-charged nitrogen of Spike-RBD Lys417 and the negative-charged oxygen of ACE2 Asp30. Moreover, the increased hindrance of the glycated moiety at the contact interface can contribute to a general decrease of the interaction strength between Spike-RBD and ACE2 residues at the interface. In the Omicron variant, Lys417 mutates into an

asparagine, preventing the possible non-enzymatic glycation of this residue. This variation, together with several mutations on Spike-RBD that were linked to an increased affinity for ACE2 (such as S477N, Q498R, and N501Y) [31], is predicted to contribute to an increase in the interactivity between Omicron-Spike-RBD and ACE2 in a hyperglycemic environment. This behavior is shown by Supplementary Figures 2–5, in which the number of contacts between Spike-RBD and ACE2 (derived from the calculation executed with the "GetContacts" tool) is reported for the different systems considered. Table 1 summarizes the overall number of residues pairs in contact between each kind of Spike RBD/ACE2 molecular complex. First, as mentioned, both interaction analysis and MM-GBSA calculations highlight the higher affinity that Omicron exerts for ACE2 with respect to the WT variant. The data coming from the MM-GBSA calculation show that the reduction in the number of contacts between WT Spike-RBD and ACE2 subsequent to glycation causes an increase in the binding free energy which is more significant than the one experienced by Omicron-Spike-RBD. The glycation, as expected, causes in both WT and Omicron variants a reduction in interaction strength with ACE2 receptor, and this can be attributed to different molecular reasons. For WT-Spike-RBD, the glycation causes mainly a loss in non-polar interaction, which could be also associated with a loss in the efficiency of hydrogen-bonding, leading to an overall noticeable increase in binding free energy. In the case of Omicron-Spike-RBD, where the loss in the number of polar contacts is compensated by an increase in non-polar interaction, and the number of non-polar contacts not only does not decrease, but also increases, the loss in binding free energy seems to be fully attributable to a drop in the efficiency of hydrogen bonding. Indeed, the increased hindrance created by the glycated residues present on the Omicron-Spike-RBD-glyco/ACE2-glyco interface (e.g., ACE2 Lys353) could destabilize the web of polar contacts that, even if the number is retained, decrease in strength. In any case, the data coming both from the analysis of the number of interactions and from MMGBSA calculations highlight higher stability of the Omicron Spike-RBD-glyco/ACE2-glyco complex over the WT-Spike-RBD-glyco/ACE2-glyco system. The results of our analysis allow us to hypothesize that the affinity between the viral protein Spike and the human receptor ACE2 is higher for the Omicron variant in respect to the WT both in native conditions and also in the case of non-enzymatic glycation, typical of the hyperglycemic environment. The results of the computational analysis here conducted allow us to hypothesize that, if non-enzymatic glycation seemed to cause a shift in the way the virus enters the cell from the ACE2-mediated mechanism to other pathways, in the case of the Omicron variant the ACE2-mediated approach of the virus seems to remain an important event

to take into account. Indeed, both our number and our MM-GBSA analysis suggest that, even if a loss in interactivity is noticeable between the Omicron-Spike-RBD/ACE2 and the Omicron-Spike-RBD-glyco/ACE2-glyco, these systems are able to maintain a higher affinity with each other in respect to, respectively, the WT-Spike RBD/ACE2 and WT-Spike-RBD-glyco/ACE2-glyco complexes. The finding of the present computational study may suggest several consequences of potential clinical relevance for the diabetes mellitus field. Non-enzymatic glycation is a well-documented phenomenon occurring in diabetes, and the fast kinetics of the event has been proven. Atanasova et al. [32] demonstrated that non-enzymatic glycation of protein amino groups (Maillard reaction) can occur at high glucose concentrations very quickly, already after few minutes. Therefore, a non-enzymatic glycation of ACE2 receptor in target tissues exposed to virus entry is very likely a fast spreading-out process, and the virus Spike protein can be glycated as well very quickly when high glucose concentrations occur. If on one side the glycation might prevent the ACE2-Spike interaction in diabetes mellitus, as previously suggested [13], on the other side the mutations of the Omicron-variant Spike protein, leading to the loss of a lysine glycation site, may be responsible for a relatively augmented binding affinity to ACE2 (compare with Figure 3). Therefore, the risk of a more pronounced virus binding to ACE-2 receptor may occur in patients with diabetes, possibly conditioning a higher susceptibility to SARS-CoV-2 infection, as indicated in our first hypothesis [12]. Moreover, in patients with decompensated diabetes, the greater affinity of Omicron Spike for ACE2 could on the one hand enhance the contact with the virus and therefore infection rate, but on the other hand, may reduce the possibility of entry through alternative routes and thus modify the overall course and severity of the disease. The new suggested roles of a novel furin cleavage site, exploited by SARS-CoV-2 to become fully active [33], could represent a possible escape mechanism used by the virus to produce an infection that, in the case of diabetes mellitus, may assume a pernicious evolution. The possible occurrence of alternative ways of virus entry in people affected by diabetes, as recently suggested [13], makes the new Omicron virus mutation of peculiar interest, not only in the general population, but also particularly in diabetes and prediabetes areas. However, the present results seem to point towards a lower severity of the disease with the Omicron variant also for patients with diabetes mellitus, supporting very recently published data [7], which show that diabetes did not appear to be a co-morbidity factor influencing disease severity, but only age. The severity of COVID-19 in diabetes remains a clinical query, also because the overall course of the infectious disease is influenced by a pre-existing reactivity, characterized by an increased pro-inflammatory profile [34],

also linked to an excess of adipose tissue, which is associated with augmented lymphocyte activation and cytokine production [35]. Whether this new variant may influence the real-world clinical evolution of COVID-19 in people affected by diabetes mellitus is still a matter of speculation. As suggested by the comment of Karim and Karin [4], we await knowledge of how this new variant will develop.

References

[1] Stanford University, “Coronavirus Antiviral & Resistance, SARS-CoV-2 Variants, DATABASE.”

[2] J. Zahradník *et al.*, “SARS-CoV-2 variant prediction and antiviral drug design are enabled by RBD in vitro evolution,” *Nat. Microbiol.*, vol. 6, no. 9, pp. 1188–1198, Sep. 2021, doi: 10.1038/s41564-021-00954-4.

[3] World Health Organization, “Classification of Omicron (B.1.1.529): SARSCoV-2 Variant of Concern. Statement, 26 November 2021.”

[4] S. S. A. Karim and Q. A. Karim, “Omicron SARS-CoV-2 variant: a new chapter in the COVID-19 pandemic,” *Lancet*, vol. 398, no. 10317, pp. 2126–2128, Dec. 2021, doi: 10.1016/S0140-6736(21)02758-6.

[5] L. Liu *et al.*, “Striking antibody evasion manifested by the Omicron variant of SARS-CoV-2,” *Nature*, vol. 602, no. 7898, pp. 676–681, Feb. 2022, doi: 10.1038/s41586-021-04388-0.

[6] B. Christie, “Covid-19: Early studies give hope omicron is milder than other variants,” *BMJ*, p. n3144, Dec. 2021, doi: 10.1136/bmj.n3144.

[7] N. Wolter *et al.*, “Early assessment of the clinical severity of the SARS-CoV-2 omicron variant in South Africa: a data linkage study,” *Lancet*, vol. 399, no. 10323, pp. 437–446, Jan. 2022, doi: 10.1016/S0140-6736(22)00017-4.

[8] M. Norouzi *et al.*, “Type-2 Diabetes as a Risk Factor for Severe COVID-19 Infection,” *Microorganisms*, vol. 9, no. 6, p. 1211, Jun. 2021, doi: 10.3390/microorganisms9061211.

[9] L. Zhu *et al.*, “Association of Blood Glucose Control and Outcomes in Patients with COVID-19 and Pre-existing Type 2 Diabetes,” *Cell Metab.*, vol. 31, no. 6, pp. 1068-1077.e3, Jun. 2020, doi: 10.1016/j.cmet.2020.04.021.

[10] Z. Wu, Y. Tang, and Q. Cheng, “Diabetes increases the mortality of patients with COVID-19: a meta-analysis,” *Acta Diabetol.*, vol. 58, no. 2, pp. 139–144, Feb. 2021, doi: 10.1007/s00592-020-01546-0.

[11] S. R. Varikasuvu, N. Dutt, B. Thangappazham, and S. Varshney, “Diabetes and COVID-19: A pooled analysis related to disease severity and mortality,” *Prim. Care Diabetes*, vol. 15, no. 1, pp. 24–27, Feb. 2021, doi: 10.1016/j.pcd.2020.08.015.

[12] G. Sartore, E. Ragazzi, L. Faccin, and A. Lapolla, “A role of glycation and methylation for SARS-CoV-2 infection in diabetes?,” *Med. Hypotheses*, vol. 144, Nov. 2020, doi: 10.1016/j.mehy.2020.110247.

[13] G. Sartore, D. Bassani, E. Ragazzi, P. Traldi, A. Lapolla, and S. Moro, “In silico evaluation of the interaction between ACE2 and SARS-CoV-2 Spike protein in a hyperglycemic environment,” *Sci. Rep.*, vol. 11, no. 1, p. 22860, Dec. 2021, doi: 10.1038/s41598-021-02297-w.

[14] S. F. Masre, N. F. Jufri, F. W. Ibrahim, and S. H. Abdul Raub, “Classical and alternative receptors for SARS-CoV-2 therapeutic strategy,” *Rev. Med. Virol.*, vol. 31, no. 5, pp. 1–9, Sep. 2021, doi: 10.1002/rmv.2207.

[15] P. Katopodis *et al.*, “COVID-19 and SARS-CoV-2 host cell entry mediators: Expression profiling of TMRSS4 in health and disease,” *Int. J. Mol. Med.*, vol. 47, no. 4, Feb. 2021, doi: 10.3892/ijmm.2021.4897.

[16] I. Kyrou, H. S. Randeva, D. A. Spandidos, and E. Karteris, “Not only ACE2—the quest for additional host cell mediators of SARS-CoV-2 infection: Neuropilin-1 (NRP1) as a novel SARS-CoV-2 host cell entry mediator implicated in COVID-19,” *Signal Transduct. Target. Ther.*, vol. 6, no. 1, Dec. 2021, doi: 10.1038/s41392-020-00460-9.

[17] A. A. Raha *et al.*, “Investigation of CD26, a potential SARS-CoV-2 receptor, as a biomarker of age and

pathology," *Biosci. Rep.*, vol. 40, no. 12, Dec. 2020, doi: 10.1042/BSR20203092.

[18] M. V. Mahtab Shahriari Felordi, Arash Memarnejadian, Mustapha Najimi, "Is There any Alternative Receptor for SARS-CoV-2?," *Cell*, vol. 23, no. 2, pp. 247–250, 2021.

[19] H. M. Berman, "The Protein Data Bank," *Nucleic Acids Res.*, vol. 28, no. 1, pp. 235–242, Jan. 2000, doi: 10.1093/nar/28.1.235.

[20] J. Lan *et al.*, "Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor," *Nature*, vol. 581, no. 7807, pp. 215–220, May 2020, doi: 10.1038/s41586-020-2180-5.

[21] D. Mannar *et al.*, "SARS-CoV-2 Omicron variant: Antibody evasion and cryo-EM structure of spike protein–ACE2 complex," *Science (80-.).*, Jan. 2022, doi: 10.1126/science.abn7760.

[22] "Molecular Operating Environment (MOE), 2019.01; Chemical Computing Group ULC, 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2021." .

[23] D.A. Case et al., "Amber 10, University of California, San Francisco." .

[24] Brito-Arias M, "The Chemistry of Oxidative Stress and Glycoxidation As Risk Factors for Developing Degenerative Disease," *Ann Rev Resear 4(5) 555648.* , 2019.

[25] "GetContacts - https://getcontacts.github.io/." .

[26] M. P. Jacobson *et al.*, "A hierarchical approach to all-atom protein loop prediction," *Proteins Struct. Funct. Bioinforma.*, vol. 55, no. 2, Mar. 2004, doi: 10.1002/prot.10613.

[27] J. Li, R. Abel, K. Zhu, Y. Cao, S. Zhao, and R. A. Friesner, "The VSGB 2.0 model: A next generation energy model for high resolution protein structure modeling," *Proteins Struct. Funct. Bioinforma.*, vol. 79, no. 10, pp. 2794–2812, Oct. 2011, doi: 10.1002/prot.23106.

[28] C. Lu *et al.*, "OPLS4: Improving Force Field Accuracy on Challenging Regimes of Chemical Space," *J. Chem. Theory Comput.*, vol. 17, no. 7, pp. 4291–4300, Jul. 2021, doi: 10.1021/acs.jctc.1c00302.

[29] S. R. H. . S. P. . et al. Dejnirattisai W., "Reduced neutralisation of SARS-COV-2 Omicron-B.1.1.529 variant by postimmunisation serum ," *medRxiv 2021.12.10.21267534; doi https//doi.org/10.1101/2021.12.10.21267534*.

[30] M. Shah and H. G. Woo, "Omicron: A Heavily Mutated SARS-CoV-2 Variant Exhibits Stronger Binding to ACE2 and Potently Escapes Approved COVID-19 Therapeutic Antibodies," *Front. Immunol.*, vol. 12, Jan. 2022, doi: 10.3389/fimmu.2021.830527.

[31] T. N. Starr *et al.*, "Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding," *Cell*, vol. 182, no. 5, pp. 1295-1310.e20, Sep. 2020, doi: 10.1016/j.cell.2020.08.012.

[32] M. Atanasova, E. Konova, T. Betova, and S. Baydanoff, "Non-Enzymatic Glycation of Human Fibrillin-1," *Gerontology*, vol. 55, no. 1, pp. 73–81, 2009, doi: 10.1159/000157436.

[33] Y. Zhang *et al.*, "A second functional furin site in the SARS-CoV-2 spike protein," *Emerg. Microbes Infect.*, vol. 11, no. 1, pp. 182–194, Dec. 2022, doi: 10.1080/22221751.2021.2014284.

[34] I. Bonyek-Silva *et al.*, "LTB4-Driven Inflammation and Increased Expression of ALOX5 / ACE2 During Severe COVID-19 in Individuals With Diabetes," *Diabetes*, vol. 70, no. 9, pp. 2120–2130, Sep. 2021, doi: 10.2337/db20-1260.

[35] S. Tsalamandris *et al.*, "The Role of Inflammation in Diabetes: Current Concepts and Future

Perspectives," *Eur. Cardiol. Rev.*, vol. 14, no. 1, Apr. 2019, doi: 10.15420/ecr.2018.33.1.

Sodium or Not Sodium: Should Its Presence Affect the Accuracy of Pose Prediction in Docking GPCR Antagonists?

Davide Bassani, Matteo Pavan, Mattia Sturlese and Stefano Moro

D. Bassani, M. Pavan, M. Sturlese, and S. Moro, "Sodium or Not Sodium: Should Its Presence Affect the Accuracy of Pose Prediction in Docking GPCR Antagonists?," Pharmaceuticals, vol. 15, no. 3, p. 346, Mar. 2022, doi: 10.3390/ph15030346.

Abstract

The function of the allosteric sodium ion in stabilizing the inactive form of GPCRs has been extensively described in the past decades. Its presence has been reported to be essential for the binding of antagonist molecules in the orthosteric site of these very important therapeutical targets. Among the GPCR–antagonist crystal structures available, in most cases, the sodium ion could not be experimentally resolved, obliging computational scientists using GPCRs as targets for virtual screening to ask: "Should the sodium ion affect the accuracy of pose prediction in docking GPCR antagonists?" In the present study, we examined the performance of three orthogonal docking programs in the self-docking of GPCR antagonists to try to answer this question. The results of the present work highlight that if the sodium ion is resolved in the crystal structure used as the target, it should also be taken into account during the docking calculations. If the crystallographic studies were not able to resolve the sodium ion then no advantage would be obtained if this is manually inserted in the virtual target. The outcomes of the present analysis are useful for researchers exploiting molecular docking-based virtual screening to efficiently identify novel GPCR antagonists.

1. Introduction

G protein-coupled receptors (GPCRs) represent one of the most important protein superfamilies encoded by the human genome. The members of this protein superfamily (more than 800 entities [1]) have been proven to perform a great variety of biological functions in the organism. Among these, very remarkable are the regulation of senses (e.g., smell, taste, gustatory), the regulation of the nervous and immune systems, homeostasis modulation, pain control, and mood balancing [2]. Indeed, it becomes clear why GPCRs are one of the most interesting protein superfamilies for drug discovery, with more than 160 validated drug targets among them [3]. The fact that encourages the scientific community in putting efforts into the research about GPCRs is their huge therapeutic

potential. At the present date, about 35% of the FDA-approved drugs are directed towards a GPCR [4,5], and more than 300 molecules are currently in clinical trials, with near one-fifth targeting a novel GPCR protein [4]. These data make clear that the drug discovery research in this field is very active, and much about this superfamily of proteins has yet to be understood.

GPCRs are cataloged into six classes based on their sequence and function similarities: class A (rhodopsin-like receptors), class B (known as "secretin family"), class C (metabotropic glutamate receptors), class D (fungal mating pheromone receptors), class E (cyclic adenosine monophosphate (cAMP) receptors), and class F (frizzled and smoothened receptors) [6]. All GPCRs share a similar organization in their three-dimensional structure; they are membrane protein receptors constituted of a transmembrane domain formed by seven α-helices (7TM domain), which are linked by three extracellular and three intracellular loops (three ECLs and three ICLs, respectively). The N-terminal (N-ter) domain is located in the extracellular side, while the C-terminal (C-ter) is found intracellularly. The functions of GPCRs are strongly dependent on their conformation and on the changes of this confirmation in time. They exist in an equilibrium between an active and an inactive state [7], and this balance can be shifted upon ligand binding. Indeed, three main families of GPCR ligands have been reported: agonists, antagonists, and inverse agonists. The first group of binders shifts the equilibrium towards the active arrangement of the receptor, while inverse agonists exert the opposite effect, increasing the conformational inactive population and decreasing the GPCR basal activity. Antagonists simply bind to the receptor and prevent the binding of other ligands, without affecting the conformational balance of the GPCR [8].

In drug discovery campaigns aimed to find new molecular entities for GPCR binding, several techniques are used to select, prioritize, and optimize the most promising compounds. Computational tools have acquired a very important role in the latest decades for drug design and discovery, strongly reducing both the time and money required to obtain new drug candidates and elucidating the most important features required to achieve a desired therapeutic effect. The approach chosen from computational medicinal chemistry to reach these ambitious goals depends on the data available about the target of interest. The presence of a three-dimensional structure of the drug target implies the possibility to exploit a structure-based drug design (SBDD) procedure, while its absence prompts the prioritization of the ligand-based drug design (LBDD) techniques. SBDD has proven to be very successful through the pharmaceutical history [9,10], with several campaigns leading to approved drugs or the repositioning of existing drugs on different targets. The most applied technique belonging to the SBDD family is surely "molecular docking" [11].

Molecular docking is a computational approach that aims to find the conformation in which a molecule binds to its recognition site, forming a stable complex [12]. Specifically, in the case of drug discovery, the main goal is to elucidate how a ligand (which could be a small molecule, a peptide, or a macromolecule) binds to a biological target of interest (usually a protein or a nucleic acid). Docking algorithms are composed of two main parts: a conformational search algorithm and a scoring function. The first aims to search through the conformational space of the ligand, while the second has the goal of ranking the conformations obtained based on their eligibility for target binding. This fitness evaluation is based on several factors, taking into account different geometrical and energetical parameters. Molecular docking has been successfully applied multiple times for virtual screening (VS) aimed at GPCR drug discovery, both in academic and industrial environments [13]. In these specific cases, attention must be paid to obtain reasonable results from the VS, tuning the docking experiment with respect to both the specific target and the family of ligands considered. A recent study demonstrated that the results of molecular docking on adenosine receptor A_{2A} change if the sodium ion stabilized in the transmembrane domain is considered or not during the calculations [14]. Specifically, that work highlighted a concordance between the computational data and the literature regarding A_{2A} receptor modulation, showing that docking algorithms tend to more efficiently reproduce antagonists' crystallographic binding modes when the sodium ion also is taken into consideration during the calculations. Indeed, the sodium ion has been reported to be present in the middle of the 7TM region of the receptor in several structures of class A GPCRs, helping stabilize the inactive conformation. The sodium ion, together with its solvation sphere, has been demonstrated to negatively modulate the binding of agonists, without influencing the binding of antagonists [15].

To date, the GPCR group which has prevailed for importance for drug discovery is class A (known as rhodopsin-like receptors), mainly for their centrality in the diseases in which they are involved, as well as for the abundance of resolved structures [16]. These proteins are divided into 19 subfamilies (A1–A19) based on phylogenetic analysis [17], including some receptors which have already become very famous in the pharmaceutical world, such as opioid, adrenergic, histaminergic, cannabinoid, and adenosine receptors.

Our evaluation starts from the already cited work of Margiotta et al. [14] to explore the influence of the allosteric sodium ion when molecular docking experiments for the diverse class A GPCR antagonists are performed. Indeed, we evaluated the performance of three different and orthogonal docking algorithms (GOLD, Glide, and PLANTS) in reproducing the ligand crystallographic pose of

protein–ligand complexes involving an antagonist bound to a class A GPCR. We extended the study to the class A subfamilies of which some antagonist–protein experimental structure is available, also taking into consideration the eventually present complexes involving a reverse agonist bound to the orthosteric binding site.

2. Results and Discussion

The complete results of the docking runs are reported in the Supplementary Materials (files "Self-docking_without_sodium.csv" and "Selfdocking_with_sodium.csv"), while a brief per-protocol report is here described by Tables 1 and 2. A graphical representation of the outcomes of the docking runs is also reported using a colormap representation in Figures 1 and 2. In these plots, the colorimetric scale delineating the RMSD values starts from 0 Å, corresponding to a docking pose perfectly superposable to the crystallographic one (maximum docking performance, represented by the dark blue color), and reaches values of 4 Å or higher (minimum docking performance, all represented by the dark red color), which stands for a very suboptimal overlay between the coordinates of the pose produced and the ones of the crystallographic conformation. The results have been reported using three different metrics: "RMSD_average", which represents the mean RMSD of all the poses obtained; "RMSD_scor_func", which is the average value of the RMSDs obtained by the poses which were top-ranked by the scoring functions in each docking run; and "RMSD_sorted", which represents the mean value of the RMSDs obtained from the poses with the lowest RMSD value in each docking calculation. As mentioned, the analysis of the results has also been executed on each docking program–scoring function pair exploited in the study (Tables 1 and 2). Moreover, to better inspect the effect of the sodium ion in the docking simulations, the analysis has also been applied to separate the group of proteins in which the sodium ion is present in the crystallographic structure considered (26 systems) from the other entries (92 complexes). The per-protocol inspections of these last results are reported in the Supplementary Materials (Tables S2–S5).

Self-Docking Results—Na^+ and H_2O Not Considered			
	RMSD_average (Å)	**RMSD_scor_func (Å)**	**RMSD_sorted (Å)**
GOLD-goldscore	3.60	2.83	1.86
GOLD-chemscore	4.45	3.25	2.46
GOLD-asp	3.87	2.91	2.14
GOLD-plp	4.60	3.48	2.56
Glide-SP	4.16	2.57	1.73
Glide-XP	2.67	2.46	1.89
$PLANTS_{CHEMPLP}$	4.96	2.12	1.35
$PLANTS_{PLP}$	5.18	2.58	1.54
All the molecular docking experiments	**4.19**	**2.78**	**1.94**

Table 1. *Table showing the results of the self-docking calculations executed without considering the sodium ion.*

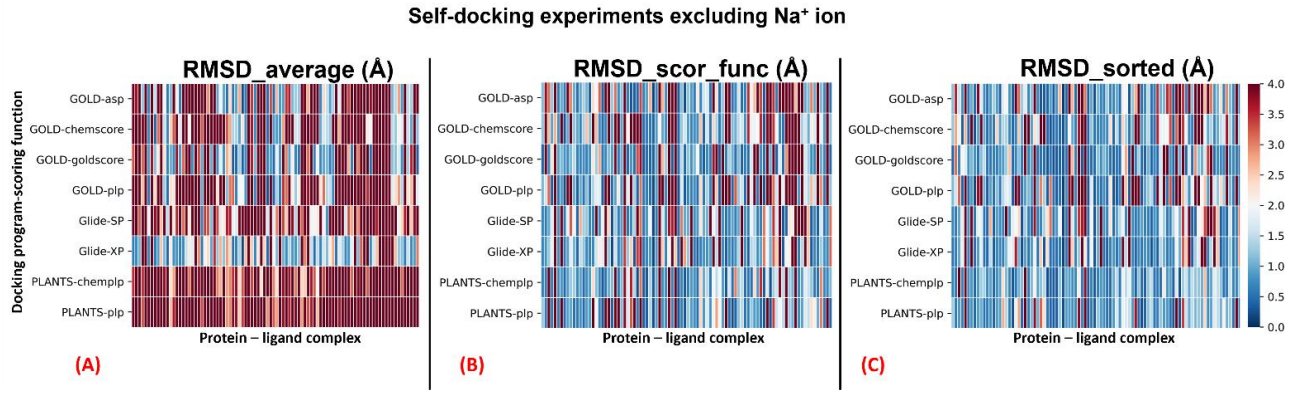


Figure 1. *Colormaps show the results of the self-docking calculations not considering the sodium ion within the GPCR receptor 7TM region of the 118 complexes examined in this study. The three plots depict respectively: (**A**) the outcomes derived from the average of the RMSDs of all the poses for each docking run ("RMSD_average"); (**B**) the results obtained just from the RMSD between the crystallographic ligand coordinates and the best-ranked pose from the scoring function ("RMSD_scor_func"); (**C**) the results of the self-docking experiments if just the pose showing the best RMSD value between its coordinates and the crystallographic ones are considered ("RMSD_sorted"). The x-axis enumerates all the different GPCR–antagonist complexes, which are plotted against the different docking program–scoring function pairs used for our study, reported on the y-axis.*

Self-Docking Results—Na^+ and H_2O Placed at 4 Å or Nearer to It Both Considered			
	RMSD_average (Å)	**RMSD_scor_func (Å)**	**RMSD_sorted (Å)**
GOLD-goldscore	4.07	3.93	2.33
GOLD-chemscore	4.53	3.90	2.82
GOLD-asp	4.13	3.25	2.40
GOLD-plp	4.52	3.51	2.56
Glide-SP	4.40	2.55	1.67
Glide-XP	2.81	2.62	1.90
$PLANTS_{CHEMPLP}$	5.16	2.60	1.49
$PLANTS_{PLP}$	5.14	2.76	1.62
All the molecular docking experiments	**4.34**	**3.14**	**2.10**

Table 2. *Table showing the results of the self-docking calculations executed considering the sodium ion and the water molecules surrounding it.*

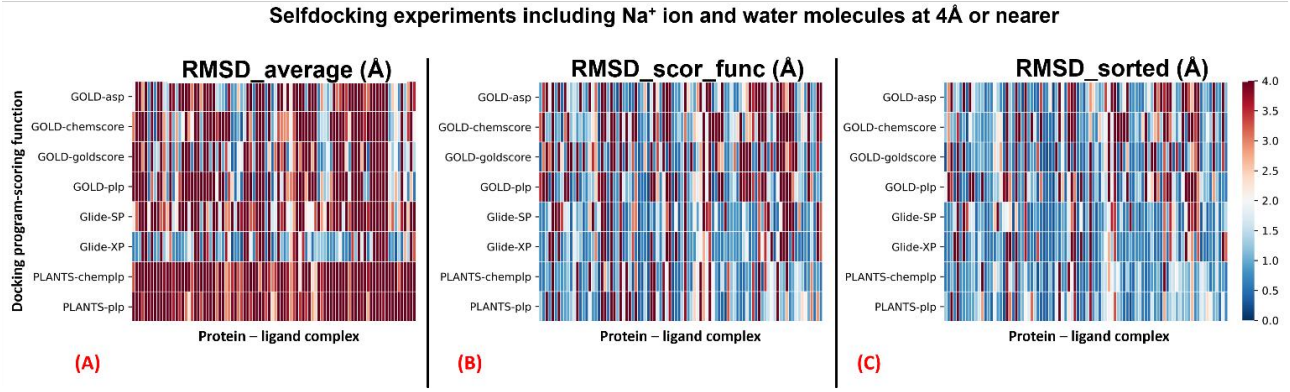


Figure 2. *Colormaps show the results of the self-docking calculations executed considering the sodium ion and the water molecules at 4 Å or nearer to it within the GPCR receptor 7TM region of the 118 complexes examined in this study. The three plots depict respectively: (**A**) the outcomes derived from the average of the RMSDs of all the poses for each docking run ("RMSD_average"); (**B**) the results obtained just from the RMSD between the crystallographic ligand coordinates and the best-ranked pose from the scoring function ("RMSD_scor_func"); (**C**) the results of the self-docking experiments if just the pose showing the best RMSD value between its coordinates and the crystallographic ones are considered ("RMSD_sorted"). The x-axis enumerates all the different GPCR–antagonist complexes, which are plotted against the different docking program–scoring function pairs used for our study, reported on the y-axis.*

The outcomes of our experiment highlight how all the algorithms used show an overall good performance in GPCR–antagonist self-docking. Among the others, the pairs "Glide-SP", "$PLANTS_{CHEMPLP}$", and "$PLANTS_{PLP}$" were always able to produce an "RMSD_sorted" value of less than 2 Å with respect to the crystallographic coordinates. Even if the scoring functions allowed obtaining

reasonable RMSD values (as observable from the “RMSD_scor_func” columns in Tables 1 and 2), the solutions are given by them rarely corresponded to the ones with the lowest RMSD. As expected, an ant colony optimization algorithm such as PLANTS tends to produce poses with less three-dimensional conservation compared to a genetic algorithm such as GOLD or a systematic method such as Glide, and this is evidenced by the higher values of “RMSD_average” given by both docking program–scoring function pairs involving PLANTS. On the other hand, the higher variability in the poses produced could be the reason for the fact that PLANTS can obtain solutions with very low RMSD, as demonstrated by the “RMSD_sorted” results, which are far below 2 Å in all the cases reported in this study (also when the complexes are separated based on the presence of the sodium ion in the original PDB structure, as depicted in Supplementary Materials, Tables S2–S5). GOLD and Glide both performed remarkably, with “goldscore” giving the best results among the scoring functions implemented for GOLD in all the metrics used for the analysis (exception made for the “RMSD_scor_func” value when considering the sodium ion and the water molecules at 4 Å or nearer to it in the calculations). Comparing “Glide-SP” and “Glide-XP” outcomes, even if the first can obtain lower “RMSD_sorted” values, is important to notice that the XP protocol is the overall best performing when considering the “RMSD_average”, always giving a value below 3 Å for this parameter. The choice between the two for GPCR antagonist virtual screening (VS) should so be based on the specific case examined. Indeed, “Glide-SP” would be more beneficial in the VS of a GPCR antagonist with already known scaffold and properties (eventually coming from “focused libraries”), while “Glide-XP” would be more effective when a library with molecules characterized by higher diversity is taken into account. When considering the use of “Glide-XP” instead of “Glide-SP” for large VS of high-diversity entities for GPCR antagonism, the medicinal chemists should always consider the higher computational times required for the XP function (passing from the 10 s/compound of “Glide-SP” to about 2 min/compound of “Glide-XP”, as reported on the developer’s page [18]).

A graphical representation of the comparison between the performance of the algorithms when the sodium ions are considered or not is reported in Figure 3, while two analog diagrams are reported in Figure 4 (based on Tables S2–S5, which can be examined in the Supplementary Materials) to give a more immediate visualization of the outcomes divided based on whether the sodium ion is present in the original crystallographic complexes.

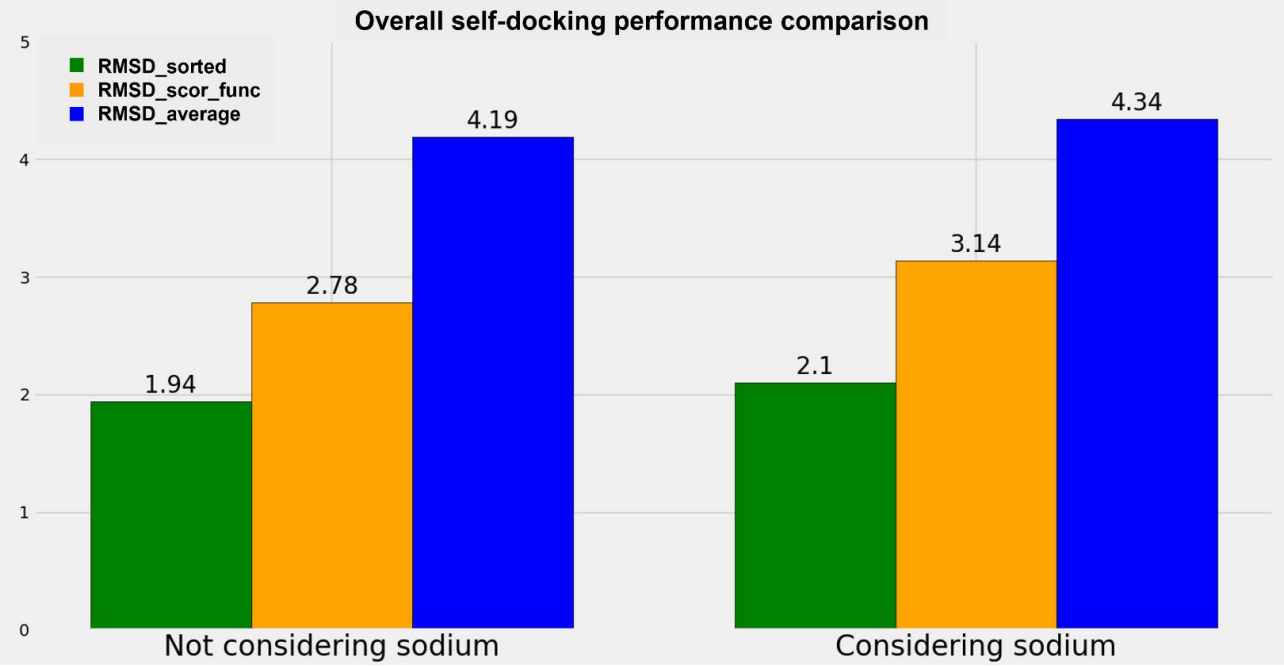


Figure 3. *Plot representing the comparison of the overall performance of the different docking algorithms implemented in this study when the sodium ion is not considered during the calculation (on the **left**) and when both the sodium ion and the crystal water molecules at 4 Å or nearer to it are included (on the **right**). The metrics used for the comparison are the "RMSD_average", the "RMSD_scor_func", and the "RMSD_sorted" values already described in the present study.*

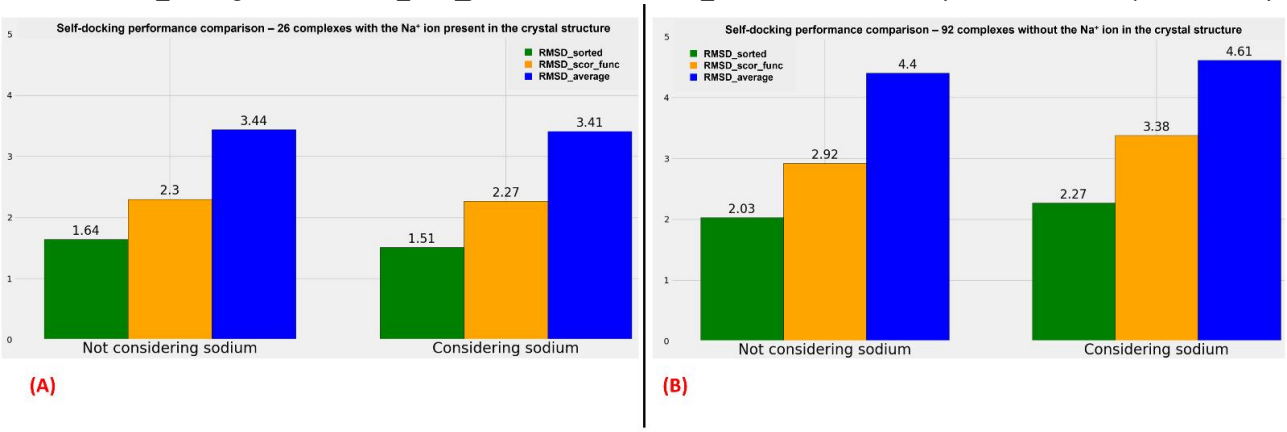


Figure 4. *Panel (**A**) Plot representing the comparison of the overall performance of the different docking algorithms implemented in this study when the sodium ion is not considered during the calculation (on the **left**) and when both the sodium ion and the crystal water molecules at 4 Å or nearer to it are included (on the **right**), focusing just on the 26 GPCR–antagonist complexes in which the sodium ion is already present in the crystal structure. Panel (**B**) Graphical representation of the comparison of the overall performance of the different docking algorithms used in this study when the sodium ion is not considered during the calculation (on the **left**) and when both the sodium ion and the crystal water molecules at 4 Å or nearer to it are included (on the **right**), focusing just on the 92 GPCR–antagonist complexes in which the sodium ion is not present on the original crystal structure.*

The results obtained show a small decrease in all the metrics used when the sodium ion is not considered in the docking runs. Specifically, the decrement in "RMSD_average" is 3.46%, the reduction in "RMSD_scor_func" is 11.47% (this higher value has to be attributed to the scoring functions), and the diminution in "RMSD_sorted" is 7.62%. Considering the decreases in the order of magnitude of the RMSD of the docking results (which is around the very promising value of 2 Å for the best pose

obtained and around 3 Å for the best solution given by the scoring functions), we can conclude that no big difference in the docking performance regarding a GPCR–antagonist system is achieved if the sodium ion is taken into account during the calculation.

The metrics used for the comparisons are the “RMSD_average”, the “RMSD_scor_func”, and the “RMSD_sorted” values already described in the present study. The overall figure is useful to compare the performance of the docking algorithms when the sodium is present in the original crystal structure and when it is not.

With an examination of the data coming from Tables S2–S5, of which the comparison of the overall results is plotted in Figure 4, we can see an analog trend of the outcomes when sodium is considered or not during the docking runs. It is interesting to notice that when the sodium is already present in the crystal structure, the RMSD values obtained from the docking poses tend to be more promising, but this has to be weighed on the fact that, in that case, the exact position of the sodium is known, and so the possible error coming from the manual placing of this alkaline ion in the 7TM region is removed. Moreover, a comparison should be made very carefully when data coming from only 26 complexes (the ones having the sodium crystallographically resolved) are juxtaposed to the ones derived from a larger set of 92 structures (the complexes in which the sodium ion is missing in the crystal structure).

On the contrary, important information is obtainable if the comparison is limited between the two groups of proteins. Indeed, as shown in Figure 4 (as well as Tables S2–S5), if the sodium ions are already present in the GPCR–antagonist crystallographic complex, no relevant difference can be noticed between the results coming from the docking runs in which the alkaline ion is considered and the ones derived from the calculation in which also sodium and the water molecules surrounding it are taken into account. Furthermore, the analysis shows a very slight decrease in the RMSD values when the crystallographic sodium and the water molecules at 4 Å or nearer to it are retained during the calculations.

On the other side, more important changes in the RMSD metrics used are highlightable comparing the outcomes of the docking runs for the complexes in which the sodium ion is not present in the crystal structure. In this case, all “RMSD_average”, “RMSD_scor_func”, and “RMSD_sorted” values show an increase if the alkaline ion with its surrounding water molecules is inserted in the complex and considered during the calculation.

The results of our study show that when performing molecular docking experiments on GPCR antagonists, the sodium ion present in the allosteric 7TM pocket should be considered during the

calculation only if it is already present in the crystal structure used as the protein target. If the GPCR on which the research is based does not have antagonist-bound crystallographic structures in which the sodium ion is present, any advantage will be obtained if it is manually placed in its allosteric pocket, and so the execution of the docking calculations without this alkaline ion should be considered. A possible reason for this behavior could be related to the fact that the small benefit coming from taking into account the sodium ion when performing the virtual screening would be demolished by the inevitable error coming from the manual placing of this ion in its allosteric pocket. We also assert that this type of uncertainty would not be canceled even if advanced computational approaches would be used for sodium placement, because of the errors that these techniques inexorably bring with them.

The importance of the allosteric sodium ion for the binding of antagonists to class A GPCRs has been extensively described in the literature [15,19]. Moreover, as observable from Figure 5, the side chains of the amino acids located in the allosteric sodium binding site are conservatively orientated towards the alkaline ion location even if all the structures represented do not have the sodium ion present in the crystallographic complex, showing that this alkaline ion has to be present in its site to guarantee the antagonist activity.

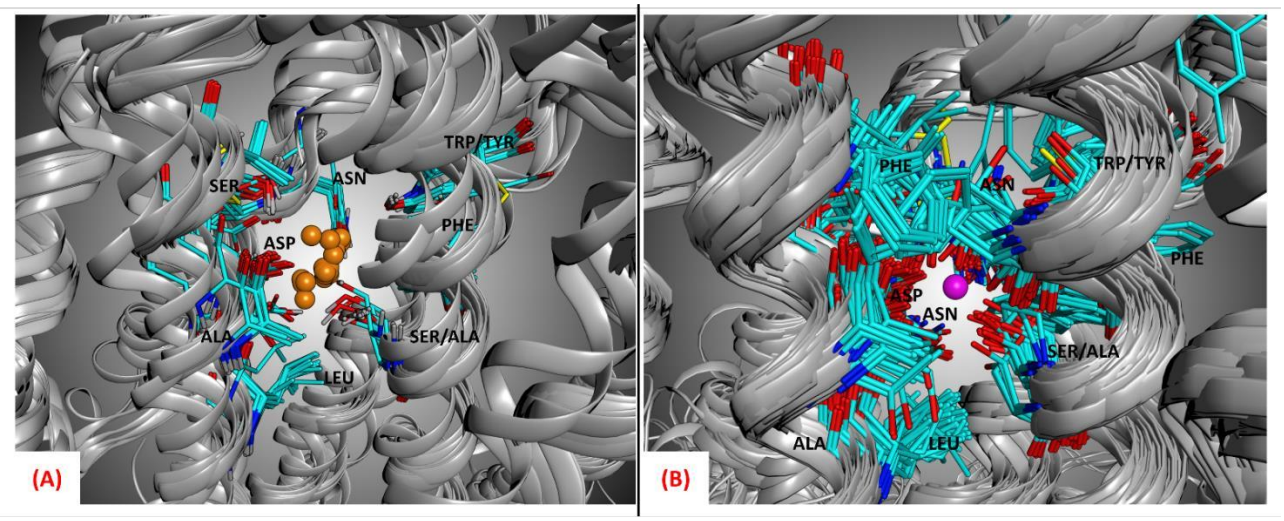


Figure 5. *(**A**) Representation of the allosteric sodium binding site of all the 26 GPCR–antagonist complexes considered in this study which had the sodium ion present in their crystallographic structures. As depicted, the position of the sodium (the orange atoms in the image) is well conserved, as are the type and orientation of the side chains of the amino acids surrounding it, which help stabilize the alkaline ion. (**B**) Representation of the allosteric sodium binding site of all the 92 GPCR–antagonist complexes considered in this study in which the sodium ion is not present in the crystallographic structures. As can be seen, the potential position of the sodium (the purple atom in the image, which comes from the crystal upon which all proteins have been superposed, PDB code: 5IU4) is well conserved, as are the type and orientation of the side chains of the amino acids surrounding it, which help to stabilize the alkaline ion in its 7TM allosteric pocket.*

Molecular docking techniques are known for not being able to distinguish agonism from antagonism. Indeed, this family of computational approaches has the goal of highlighting the potential

binders for a target, but their results cannot be related to a specific type of outcomes that this binding will have on the target itself (which has to be evaluated by the medicinal chemistry experts, based on their expertise and the communication with other professionals of the pharmaceutical world). These limitations of the technique of molecular docking may be the reason for the very low difference between the results coming from the cases in which sodium is considered or not in the calculations.

3. Materials and Methods

For each of the 19 GPCR class A subfamilies, the crystal structures available in the Protein Data Bank [20] (PDB, latest access 15 January 2022) were inspected. Each entry with a human GPCR protein complexed with a small molecule orthosteric antagonist crystallized together was selected to build the starting database of our study. If multiple crystals of a protein bound to the same ligand existed, only the highest resolution crystal with the sodium ion present was selected. To give a more comprehensive panoramic of the role of allosteric sodium in GPCR binding, the structures with an inverse agonist were also considered for this study (e.g., 6K1Q, 7F83, 7B6W, 7BVQ). In the end, 118 protein–ligand complexes involving a GPCR and a small molecule antagonist were obtained (a comprehensive list is reported in the Supplementary Materials, Table S1).

The 118 complexes were downloaded from the PDB and imported into Molecular Operating Environment (MOE) suite [21], the main molecular modeling program that we used in this work. Each system was then prepared with a protocol involving the tools included in the MOE package. First, the “Structure Preparation” program was used to rebuild the small missing loops in the structures and to adequately select the orientation of alternate crystallographic residues based on occupancy. Then, the most proper protonation state for each amino acid was determined with the “Protonate 3D” tool, setting 7.4 as the pH value for the environment. Subsequently, the added hydrogen atoms were minimized with the AMBER10:EHT [22] force field implemented in MOE. Finally, each non-protein, non-ligand, and non-sodium molecule was deleted from the systems, with an exception made for the water molecules solvating the sodium ion (we used 4 Å as the cut-off radius), when present.

The systems were then separated based on whether they had the sodium ion crystallized in their original PDB structure. Among all the complexes downloaded, 26 already had the sodium present in the crystal, while 92 did not (the distinction is highlighted in the Supplementary Materials, Table S1). All the systems in which the sodium was not present were properly treated, inserting the

sodium ion with its solvation water molecules. The position of the sodium and the water molecules was chosen according to the PDB crystal 5IU4, the complex with the best resolution, R-value, and R-free value balance among all the entries considered. This choice was also supported by the fact that when superposing all the 7TM regions of the protein–ligand systems with the sodium crystallized, the position of this alkaline ion is very conservative, as observable from Figure 5 (the average RMSD between the coordinates of the sodium ions and the sodium ion of the reference structure 5IU4 was calculated to be 0.75 Å). It is important to mention that the crystal 5IU4 was not considered for the docking calculations because, even if its resolution is optimal, it is significantly mutated in the 7TM region. Another ADORA2 crystal structure bearing the same ligand (ZMA), 6LPJ, shows a very similar resolution (1.80 Å versus the 1.72 Å of 5IU4) and does not show mutations in the 7TM domain.

Our self-docking approach consisted of the separation of each ligand from its crystallographic GPCR structure, its preparation, and its molecular docking inside the orthosteric binding site with three different orthogonal programs, namely GOLD [23] (based on a genetic algorithm, developed and licensed by CCDC), Glide [24] (a systematic docking program developed and distributed by Schrödinger), and PLANTS [25] (an ant colony optimization algorithm developed by the University of Tübingen). For each of the programs, all the scoring functions supported were used. Specifically, GOLD was used in four different parallel runs, applying the scoring functions “goldscore”, “chemscore”, “asp”, and “plp”. The two Glide calculations for each ligand were executed first with “Glide-SP” and then with “Glide-XP”, while the docking runs with PLANTS exploited the scoring functions “PLANTS$_{CHEMPLP}$” and “PLANTS$_{PLP}$”. For each program–scoring function pair, five poses were produced for each ligand, and each of those was compared with the crystallographic pose to calculate the root-mean-square deviation (RMSD) between the coordinates of the two conformations.

This whole procedure was executed twice, first setting the docking programs to not consider the sodium ion and the water molecules solvating it and then setting the programs to take into account both sodium and the water molecules placed at 4 Å or nearer to the alkaline ion.

4. Conclusions

In the present study, we examined the effect of considering the allosteric sodium ion when molecular docking approaches are applied to GPCR antagonists. To accomplish our task, we collected 118 GPCR–antagonist complexes, both with and without the sodium ion present in the crystallographic structure. For the systems in which this alkaline ion was not present, a manual insertion of the sodium and its surrounding water molecules was executed based on superposition with a very high

resolution structure (PDB: 5IU4), after having established that the position of the ion is very conservative in the GPCR–antagonist crystals. Then, we executed self-docking experiments of the orthosteric GPCR ligands with three orthogonal docking programs (GOLD, Glide, and PLANTS) both considering and not considering the sodium ion and its surrounding water molecules. What emerged from the present work is the finding that the performance of the docking programs (enucleated in three different metrics, “RMSD_average”, “RMSD_sorted”, and “RMSD_scor_func”) does not significantly change between the two cited scenarios. Going deeper into the analysis of the results, we highlighted that a small increment in the docking programs’ performance is observable if the sodium ion is kept during the docking runs just for those crystal structures in which the alkaline ion was resolved, while for the other complexes the trend is the opposite, favoring the solution of not considering sodium during the docking calculations. The outcomes of the present work are helpful to increase the knowledge about the performance of docking programs when applied to research about GPCR antagonists, and we are confident that the pharmaceutical experts that are putting effort into this fascinating field will benefit from our work.

References

[1] Gacasan, S.B.; Baker, D.L.; Parrill, A.L. G protein-coupled receptors: The evolution of structural insight. *AIMS Biophys.* **2017**, *4*, 491–527. https://doi.org/10.3934/biophy.2017.3.491.

[2] Heng, B.C.; Aubel, D.; Fussenegger, M. An overview of the diverse roles of G-protein coupled receptors (GPCRs) in the pathophysiology of various human diseases. *Biotechnol. Adv.* **2013**, *31*, 1676–1694. https://doi.org/10.1016/j.biotechadv.2013.08.017.

[3] Yang, D.; Zhou, Q.; Labroska, V.; Qin, S.; Darbalaei, S.; Wu, Y.; Yuliantie, E.; Xie, L.; Tao, H.; Cheng, J.; et al. G protein-coupled receptors: Structure- and function-based drug discovery. *Signal Transduct. Target. Ther.* **2021**, *6*, 7. https://doi.org/10.1038/s41392-020-00435-w.

[4] Hauser, A.S.; Attwood, M.M.; Rask-Andersen, M.; Schiöth, H.B.; Gloriam, D.E. Trends in GPCR drug discovery: New agents, targets and indications. *Nat. Rev. Drug Discov.* **2017**, *16*, 829–842. https://doi.org/10.1038/nrd.2017.178.

[5] Insel, P.A.; Sriram, K.; Gorr, M.W.; Wiley, S.Z.; Michkov, A.; Salmerón, C.; Chinn, A.M. GPCRomics: An Approach to Discover GPCR Drug Targets. *Trends Pharmacol. Sci.* **2019**, *40*, 378–387. https://doi.org/10.1016/j.tips.2019.04.001.

[6] Lee, Y.; Basith, S.; Choi, S. Recent Advances in Structure-Based Drug Design Targeting Class A G Protein-Coupled Receptors Utilizing Crystal Structures and Computational Simulations. *J. Med. Chem.* **2018**, *61*, 1–46. https://doi.org/10.1021/acs.jmedchem.6b01453.

[7] Weis, W.I.; Kobilka, B.K. The Molecular Basis of G Protein–Coupled Receptor Activation. *Annu. Rev. Biochem.* **2018**, *87*, 897–919. https://doi.org/10.1146/annurev-biochem-060614-033910.

[8] Syrovatkina, V.; Alegre, K.O.; Dey, R.; Huang, X.-Y. Regulation, Signaling, and Physiological Functions of G-Proteins. *J. Mol. Biol.* **2016**, *428*, 3850–3868. https://doi.org/10.1016/j.jmb.2016.08.002.

[9] Meng, X.-Y.; Zhang, H.-X.; Mezei, M.; Cui, M. Molecular Docking: A Powerful Approach for Structure-Based Drug Discovery. *Curr. Comput. Aided-Drug Des.* **2011**, *7*, 146–157. https://doi.org/10.2174/157340911795677602.

[10] Anderson, A.C. The Process of Structure-Based Drug Design. *Chem. Biol.* **2003**, *10*, 787–797. https://doi.org/10.1016/j.chembiol.2003.09.002.

[11] Wang, G.; Zhu, W. Molecular docking for drug discovery and development: A widely used approach but far from perfect. *Future Med. Chem.* **2016**, *8*, 1707–1710, https://doi.org/10.4155/fmc-2016-0143.

[12] Lengauer, T.; Rarey, M. Computational methods for biomolecular docking. *Curr. Opin. Struct. Biol.* **1996**, *6*, 402–406. https://doi.org/10.1016/S0959-440X(96)80061-3.

[13] Ballante, F.; Kooistra, A.J.; Kampen, S.; de Graaf, C.; Carlsson, J. Structure-Based Virtual Screening for Ligands of G Protein–Coupled Receptors: What Can Molecular Docking Do for You? *Pharmacol. Rev.* **2021**, *73*, 1698–1736. https://doi.org/10.1124/pharmrev.120.000246.

[14] Margiotta, E.; Deganutti, G.; Moro, S. Could the presence of sodium ion influence the accuracy and precision of the ligand-posing in the human A2A adenosine receptor orthosteric binding site using a molecular docking approach? Insights from Dockbench. *J. Comput. Aided. Mol. Des.* **2018**, *32*, 1337–1346. https://doi.org/10.1007/s10822-018-0174-2.

[15] Katritch, V.; Fenalti, G.; Abola, E.E.; Roth, B.L.; Cherezov, V.; Stevens, R.C. Allosteric sodium in class A GPCR signaling. *Trends Biochem. Sci.* **2014**, *39*, 233–244. https://doi.org/10.1016/j.tibs.2014.03.002.

[16] Basith, S.; Cui, M.; Macalino, S.J.Y.; Park, J.; Clavio, N.A.B.; Kang, S.; Choi, S. Exploring G Protein-Coupled Receptors (GPCRs) Ligand Space via Cheminformatics Approaches: Impact on Rational Drug Design. *Front. Pharmacol.* **2018**, *9*, 128. https://doi.org/10.3389/fphar.2018.00128.

[17] Joost, P.; Methner, A. Phylogenetic analysis of 277 human G-protein-coupled receptors as a tool for the prediction of orphan receptor ligands. *Genome Biol.* **2002**, *3*, research0063.1. https://doi.org/10.1186/gb-2002-3-11-research0063.

[18] Available online: https://www.schrodinger.com/science-articles/docking-and-scoring (accessed on 18th January 2022).

[19] Gentry, P.R.; Sexton, P.M.; Christopoulos, A. Novel Allosteric Modulators of G Protein-coupled Receptors. *J. Biol. Chem.* **2015**, *290*, 19478–19488. https://doi.org/10.1074/jbc.R115.662759.

[20] Berman, H.M. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242. https://doi.org/10.1093/nar/28.1.235.

[21] *Molecular Operating Environment (MOE), 2019.01*; Chemical Computing Group ULC, 1010 Sherbooke St. West, Suite #910: Montreal, QC, Canada, 2021.

[22] Case, D.A; et al. *Amber 10*; University of California: San Francisco, CA, USA.

[23] Jones, G.; Willett, P.; Glen, R.C.; Leach, A.R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking 1 1Edited by F. E. Cohen. *J. Mol. Biol.* **1997**, *267*, 727–748. https://doi.org/10.1006/jmbi.1996.0897.

[24] Friesner, R.A.; Banks, J.L.; Murphy, R.B.; Halgren, T.A.; Klicic, J.J.; Mainz, D.T.; Repasky, M.P.; Knoll, E.H.; Shelley, M.; Perry, J.K.; et al. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* **2004**, *47*, 1739–1749. https://doi.org/10.1021/jm0306430.

[25] Korb, O.; Stützle, T.; Exner, T.E. PLANTS: Application of Ant Colony Optimization to Structure-Based Drug Design. *Lect. Notes Comput. Sci.* **2006**, *4150*, 247–258.

Bat coronaviruses related to SARS-CoV-2: what about their 3CL Proteases (M^{pro})?

Matteo Pavan, **Davide Bassani**, Mattia Sturlese, and Stefano Moro

Matteo Pavan, Davide Bassani, Mattia Sturlese & Stefano Moro (2022) Bat coronaviruses related to SARS-CoV-2: what about their 3CL proteases (MPro)?, Journal of Enzyme Inhibition and Medicinal Chemistry, 37:1, 1077-1082, DOI: 10.1080/14756366.2022.2062336

Abstract

Despite a huge effort by the scientific community to determine the animal reservoir of SARS-CoV-2, which led to the identification of several SARS-CoV-2-related viruses both in bats and in pangolins, the origin of SARS-CoV-2 is still not clear. Recently, Temmam et al. reported the discovery of bat coronaviruses with a high degree of genome similarity with SARS-CoV-2, especially concerning the RBDs of the S protein, which mediates the capability of such viruses to enter and therefore infect human cells through a hACE2-dependent pathway. These viruses, especially the one named BANAL-236, showed a higher affinity for the hACE2 compared to the original strain of SARS-CoV-2. In the present work, we analyze the similarities and differences between the 3CL protease (main protease, M^{pro}) of these newly reported viruses and SARS-CoV-2, discussing their relevance relative to the efficacy of existing therapeutic approaches against COVID-19, particularly concerning the recently approved orally available Paxlovid, and the development of future ones.

1. Introduction

Since its outbreak in December 2019, the COVID-19 pandemic has caused to date the death of almost 6 million people all around the world[1][2]. This worldwide-spread disease is caused by a betacoronavirus known as SARS-CoV-2, which infects the respiratory system of the host organism compromising its health status[3]. The symptoms of this illness range from the ones typical of influenza (cough, fever, and headache) to very serious complications such as breathing difficulty, pneumonia, and hypoxia, eventually leading to respiratory failure and death[4]. The high transmissibility of the SARS-CoV-2 virus allowed its fast diffusion all around the world, rapidly attracting the interest of experts in the medical, biological, and pharmaceutical environments, who

have extensively worked and are still putting relevant efforts into the elaboration of proper solutions to fight this pathogen.

The first approach to finding viable therapeutic options was the so-called “drug-repurposing”, i.e. the use of drugs that are already marketed for the treatment of different diseases to cure COVID-19 patients. Concerning this, particular attention was directed towards HIV protease inhibitors such as Kaletra (therapeutic combination of Lopinavir and Ritonavir)[5] and antimalarial drug Plaquenil (commercial name of hydroxychloroquine)[6]. Unfortunately, despite the promising premises (especially from a timescale perspective[7]), this approach was unsuccessful, with investigated drugs showing little to no efficacy in randomized clinical trials[8].

Parallel to the first approach, a considerable amount of labor by both the industry and academia has been spent on developing tools that prevent the detrimental effect of the pathology and has resulted in the approval by the Food and Drug Administration (FDA) of several vaccines[9]. These therapeutic entities can be divided into three different classes[10]: the first one is composed of the inactivated virus vaccines, such as Chinese CoronaVac and the Russian CoviVac, the second family is formed by the ones based on adenovirus vectors, like Vaxzevria, Sputnik V, and the Janssen COVID-19 vaccine, while the third and final family consists of the mRNA-based ones such as the Pfizer-BioNTech “Comirnaty” and the Moderna “Spikevax”.

While vaccines based on inactivated viruses have given poor results, several studies have proven the efficacy of vaccination campaigns with the other two classes of vaccines (especially m-RNA ones) all around the world[11][12]. Despite the success of said vaccines, the SARS-CoV-2 Spike protein is often subjected to immune system-escaping mutations which lead to the development of new viral variants[13], obliging the vaccines to be periodically updated to maintain their efficacy.

The high variability of the Spike protein among different coronavirus strains, which threatens the efficacy of already approved vaccines in the long period, led the scientific community to join forces to identify effective treatments for ongoing infections and to prevent future pandemic waves. Regarding this, a remarkable example is portrayed by the COVID Moonshot consortium, a collaborative project that involved scientists from all over the world in an attempt to design and develop an orally available drug against COVID-19[14][15]. COVID Moonshot aside, the great amount of knowledge accumulated on the target since the SARS-CoV epidemic in 2002/2003 rapidly resulted in the approval of the first COVID-19 specific treatments.

The first drug to be approved was Remdesivir, a polymerase inhibitor that was initially designed against the Ebola Virus and has then been repositioned against COVID-19. This drug, unfortunately, has an unfavorable pharmacokinetic profile, which limits its administration to the intravenous route in a hospital setting[16][17]. Tocilizumab, an interleukin-6 receptor monoclonal antibody originally developed to cure rheumatoid arthritis, obtained the emergency use authorization (EUA) for the treatment of COVID-19 in the United States in June 2021[18]. The oral RNA-polymerase inhibitor Favipiravir has also been approved for marketing in countries such as Japan, China, India, Saudi Arabia, and the United Arab Emirates, but is still under examination from the FDA[19].

An important milestone has been achieved at the end of 2021 with the FDA approval of the therapy based on the SARS-CoV-2 main protease (M^{pro}) inhibitor Nirmatrelvir (also known as PF-07321332) in combination with Ritonavir, sold under the commercial name "Paxlovid" (which is available also in Europe since the end of January 2022)[20]. Thanks to its efficient, reversibly covalent inhibition of M^{pro}, the Nirmatrelvir-based therapy demonstrated to lower of 89% the risk of severe complications after COVID-19 infection in symptomatic, unvaccinated, non-hospitalized adults[21].

A recent scientific work by Temmam et al. reported the discovery of a high level of sequence similarity between the SARS-CoV-2 genome and that of other coronavirus species infecting cave bats living in North Laos[22], raising serious concerns about the potential threat to public health that these coronaviruses could portray. Despite giving an in-depth analysis on the similarities and differences between the S protein of these newly reported viruses, no consideration is reported in the original work about their main proteases. For this reason, in order to assess the impact that these bat coronaviruses could have on public health, we performed a computational analysis to shed light on similarities and differences between the main protease of SARS-CoV-2 and that of these newly discovered bat coronaviruses, discussing the role that these alterations could have on the efficacy of existing therapies (Paxlovid, in particular) and the development of future ones.

2. Materials and Methods

The genome sequence for SARS-CoV-2, BANAL-52, BANAL-103, BANAL-236, and RaTG13 was obtained through GenBank. Table 1 reports the accession codes for each of the considered genomes. The protein sequence associated with the 3CL protease (main protease, M^{pro}), was extracted, aligned using the appropriate tool from MOE 2019.01[23], and used for the generation

of the correspondent homology model (except for SARS-CoV-2, for which several crystal structures are available in the Protein Data Bank).

The structure of SARS-CoV-2 main protease in its unliganded state was retrieved from the Protein Data Bank (PDB ID: 6Y2E[24]) and prepared using MOE 2019.01. At first, the functional dimer was restored applying the symmetric crystallographic transformation to each asymmetric unit. Secondly, residues with fractional occupancy values were assigned to the most probable state. Then, missing hydrogen atoms were added to the system according to the most probable protonation state at pH 7.4 for each titratable residue exploiting the "Protonate 3D" tool. Afterward, hydrogen atoms coordinates were energy minimized according to the AMBER10: EHT force field until a gradient of 0.1 Kcal mol^{-1} $Å^{-2}$ was reached. Finally, crystallographic water molecules were removed.

Four different homology models were generated exploiting the "Homology Model" tool, one for each bat coronavirus considered in the present work. The sequences used for the generation of homology models are reported in Table 1, while the structure 6Y2E, prepared as described before, was used as a template for the model generation.

Organism	Isolate	Accession Code	Product	Protein ID	Residues
SARS-CoV-2	"Wuhan-Hu-1"	NC_045512.2	ORF1ab polyprotein	YP_009724389.1	S3264-Q3569
Bat coronavirus	"BANAL-20-52/Laos/2020"	MZ937000.1	ORF1ab polyprotein	UAY13216.1	S3255-Q3560
Bat coronavirus	"BANAL-20-103/Laos/2020"	MZ937001.1	ORF1ab polyprotein	UAY13228.1	S3256-Q3561
Bat coronavirus	"BANAL-20-236/Laos/2020"	MZ937003.2	ORF1ab polyprotein	UAY13252.1	S3256-Q3561
Bat coronavirus RaTG13	"RaTG13"	MN996532.2	ORF1ab polyprotein	QHR63299.2	S3263-Q3568

Table 1. *This table reports the protein sequences used in this work and their origin.*

3. Results

In order to compare similarities and dissimilarities between the SARS-CoV-2 M^{pro} and correspondent proteases in the most closely related bat coronaviruses, four different homology models (one for each different virus considered in this work) were generated, as reported in Materials and Methods. Due to the high degree of sequence identity (99,7% for BANAL-52, BANAL-103, and BANAL-236, 99,4% for RaTG13) between considered bat coronaviruses and SARS-CoV-2 M^{pro}, homology modeling is expected to give a representative result, very closely related to the experimental data. As illustrated by Figure 1, there are only two differences in the primary sequences of considered viruses. These small changes to the amino acid sequences led to the generation of homology models that are practically superimposable to the reference structure (6Y2E), as is also depicted in Figure 1. Figure 2, instead, reports a comparison between the four homology models and SARS-CoV-2 M^{pro} from structure 6Y2E, mapping the differences between various proteases onto their three-dimensional structure.

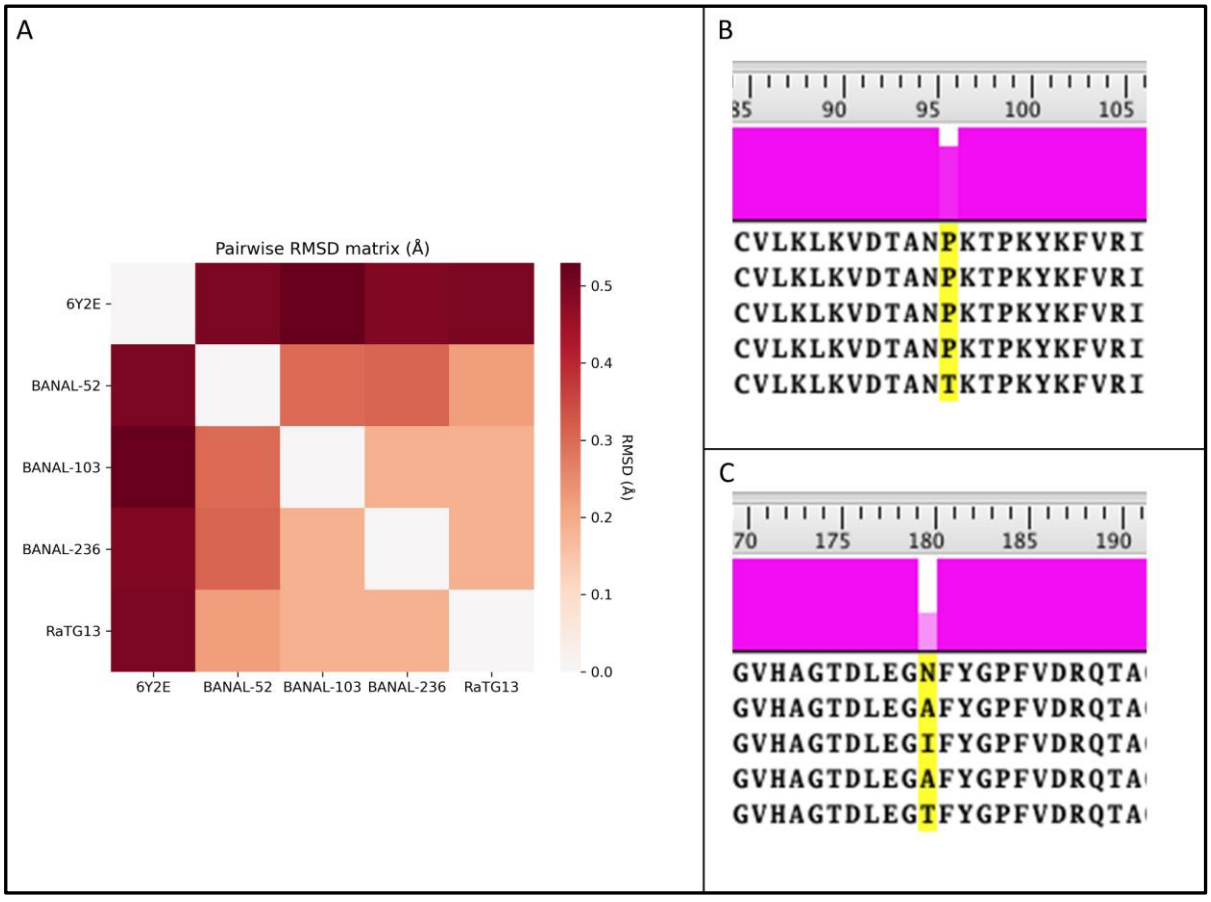


Figure 1. *Comparison between SARS-CoV-2 3CL protease (M^{pro}) from crystal structure 6Y2E (blue) and homology models of M^{pro} from four different bat coronaviruses, reported in Table 1. In Panel A, the pairwise RMSD matrix derived from the superposition of each bat coronavirus homology model to the template structure 6Y2E is reported. Panel B and C summarize the differences in the primary sequence between SARS-CoV-2 and bat coronaviruses M^{pro}.*

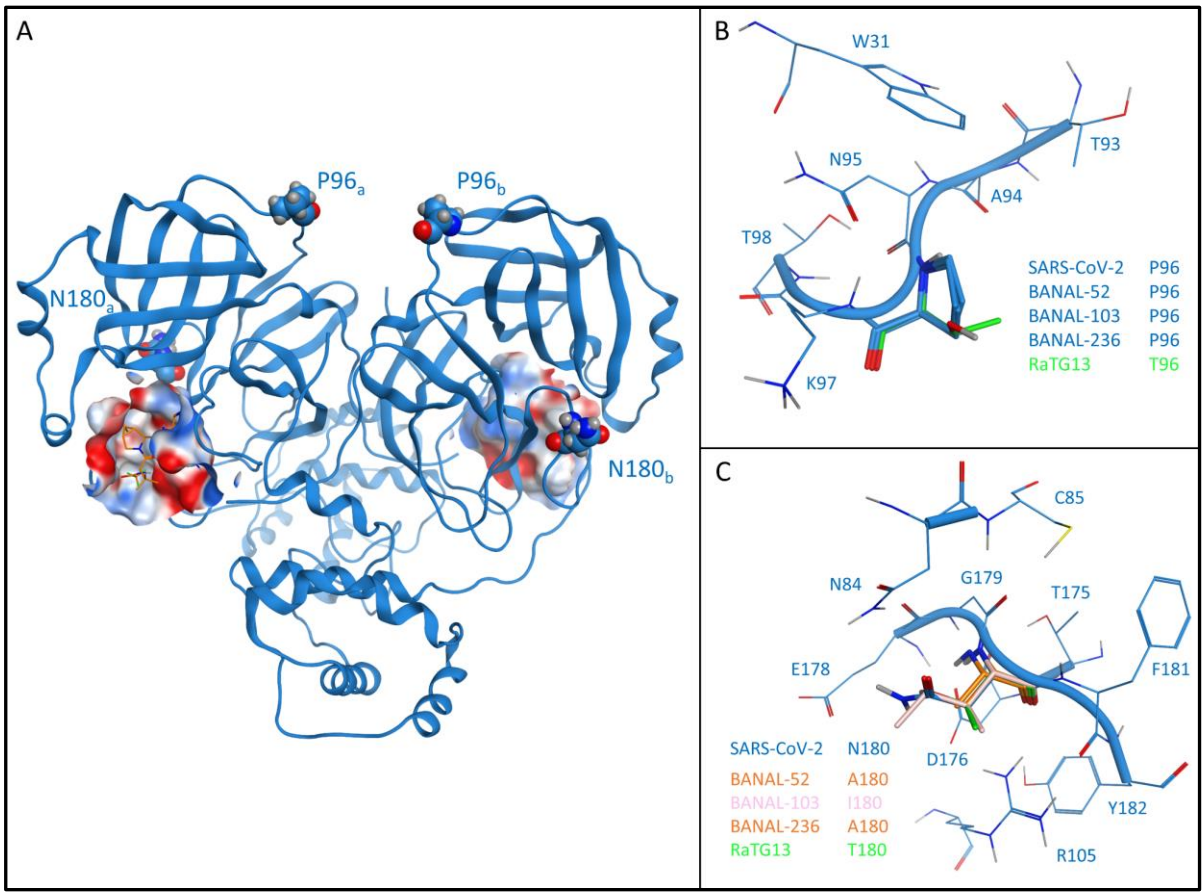


Figure 2. *Comparison between SARS-CoV-2 3CL protease (M^{pro}) from crystal structure 6Y2E (blue) and homology models of M^{pro} from four different bat coronaviruses, reported in Table 1. Panel A reports the structure of SARS-CoV-2 M^{pro} (PDB ID: 6Y2E) in its free form. The protein is depicted in blue ribbons, while mutated residues (namely, P96 and N180) in comparison with bat coronaviruses are highlighted and depicted as CPK models. For visual reference, Nirmatrelvir (also known as PF-07321332, commercial name Paxlovid) from structure 7RFS is also shown in the picture, alongside the binding site surface colored according to electrostatic properties. Panel B highlights the comparison between residue 96 of SARS-CoV-2 M^{pro} and homology models of bat coronaviruses M^{pro}. Panel C reports a comparison between residue 180 of SARS-CoV-2 M^{pro} and homology models of bat coronaviruses M^{pro}.*

The first difference is related to residue 96, which in the case of SARS-CoV-2 is a proline. This residue is conserved in each BANAL coronavirus reported by Temmam et al. but is not conserved in RaTG13, which was previously considered to be the most closely related bat coronavirus to SARS-CoV-2. Instead of a proline, RaTG13 presents a threonine residue at the 96 position, which is expected to increase both the flexibility and the hydrophilicity of the surroundings.

The second structural alteration is referred to residue 180, which in the case of SARS-CoV-2 is an asparagine. In this case, there is a higher variability between different coronavirus strains, with each BANAL virus presenting a hydrophobic residue (alanine, in the case of BANAL-52 and BANAL-236, isoleucine in the case of BANAL-103), while RaTG13 once again differentiate from both BANAL viruses and SARS-CoV-2 presenting a hydrophilic threonine residue.

4. Discussion

The comparison between the crystal structure of SARS-CoV-2 and homology models of bat coronaviruses M^{pro} showed that there are two main structural differences, both of which do not involve the catalytic site.

In the native SARS-CoV-2 structure, Phe96 is involved in a series of hydrophobic contacts with the side chain of Trp31, Thr93, and Lys97 through its pyrrolidine core. In the case of RaTG13, the only bat coronavirus that presents an alteration at this position, the presence of a threonine causes a reduction of possible hydrophobic contacts with the surrounding aminoacids but does not cause the loss of any crucial interaction for structural integrity. Moreover, this residue is located in a solvent-exposed flexible loop region that connects between two beta-sheets, a further indication that this substitution should not compromise the structural integrity of the protease.

Concerning the second structural alteration, in the native SARS-CoV-2 structure Asn180 is involved in a double interaction with the sidechain of two charged residues, namely Asp176 and Arg105. Both of these interactions happen with the backbone of Asn180 and do not involve its sidechain, which is stretched towards the solvent. Intriguingly, in this case, the newly discovered bat coronaviruses all present a hydrophobic residue at position 180: in all these cases, no loss of native interaction happens, coherently with the fact that they do not involve the sidechain of residue 180 and only occur through its backbone. Once again, RaTG13 is the most diverse one, being the only analyzed bat coronavirus that presents a polar aminoacid (a threonine) at this position. As previously mentioned, the sidechain of residue 180 is not involved in any structurally relevant interaction, and therefore the presence of a hydroxyethyl sidechain does not give a particular advantage to this virus strain. Furthermore, as is the case for Pro96, this structural modification is also located in a solvent-exposed, non-structured loop region, indicating that no critical harm to the protease integrity should be provoked by this alteration.

Altogether, our structural analysis reveals that neither of these two structural differences between SARS-CoV-2 and bat coronaviruses M^{pro} should determine any relevant structural alteration of the main protease. Notably, this observation is in agreement with a recent article that characterized the effect of each possible M^{pro} mutation on its functionality: both Pro96 and Asn180 are marked as highly tolerant to mutations[25].

Concerning the implications of these two mutations on the efficacy of M^{pro} inhibitors, several elements point to the conclusion that neither mutation should have a relevant effect. As can be

seen from Figure 2, which gives a depiction of the localization of these two mutations relative to the position of the catalytic site (which is also the binding site of most protease inhibitors, including PF-07321332, the active principle of Paxlovid) shows that both these mutations are not directly linked to the catalytic site, indicating that the binding cleft that harbors PF-07321332 should not be altered. Moreover, as thoroughly assessed in a previous scientific work from our laboratory, neither of these two residues is in any way involved in the recognition process of PF-07321332[26], complementing the structural information provided by crystal structures 7VH8[27], 7RFS, and 7RFW which clearly show how none of this two residues contributes to the interaction with PF-07321332 in the final bound state.

The fact that the SARS-CoV-2 main protease and the one from closely related bat coronaviruses are very similar and practically identical at the catalytic site supports the idea that targeting this protease is still a viable therapeutic option not only for the present but also for the prevention of future pandemic waves.

To date, several studies have contributed to thoroughly characterizing the nature of the shallow and solvent-exposed catalytic site of the SARS-CoV-2[28], which has proven to be readily investigable with both time-dependent and time-independent structure based-approaches such as molecular docking[29] and molecular dynamics[30], leading to the development of compounds with affinities in the low nanomolar range[31][32].

All these factors, combined with the fact that striking 3D structure similarities exists also with other coronaviral M^{pro} such as the one from Porcine transmissible Gastroenteritis virus (TGEV)[33], Human coronavirus strain 229E (HCoV)[34], Infectious bronchitis virus (IBV)[24] and MERS-CoV[35], validate the pursue of novel M^{pro} inhibitors that could increase the pool of available treatment for COVID-19 and also for future coronavirus-related diseases, acting as pan-coronaviral drugs.

5. Conclusions

Recently, a scientific work by Temmam et al. reported the discovery of bat coronaviruses closely related to SARS-CoV-2 that can infect human cells. This scientific work raised the attention of both the scientific community and the general audience to the possible threat to public health that these newly discovered coronaviruses could represent. Despite a thorough characterization of Spike protein of these bat coronaviruses, no information was given in the original work about their main

proteases, which is considered the main target for the development of COVID-19 specific active principles.

In the present scientific work, we performed a computational analysis to characterize structural similarities and differences between the main proteases of SARS-CoV-2 and closely related bat coronaviruses. A comparison between the crystal structure of SARS-CoV-2 M^{pro} and homology models of bat coronavirus M^{pro} shows that two main differences exist, involving the mutation of Pro96 and Asn180. None of these structural alterations are predicted to have an impact on the protease structural integrity, functionality, or affinity for existing inhibitors (especially the recently approved orally available Paxlovid), nor towards the development of novel protease inhibitors. Furthermore, the high degree of structural conservation among main proteases from different coronaviruses suggests that Mpro is not only a valid target for the treatment of COVID-19, but that the knowledge acquired on this target could be useful in the identification and development of pan-coronaviral drugs that can cure different diseases and prevent future pandemic waves.

References

[1] Guarner, J. Three Emerging Coronaviruses in Two Decades: The Story of SARS, MERS, and Now COVID-19. *American Journal of Clinical Pathology* vol. 153 420–421 Preprint at https://doi.org/10.1093/ajcp/aqaa029 (2020).

[2] COVID Live - Coronavirus Statistics - Worldometer. https://www.worldometers.info/coronavirus/.

[3] Zhou, F. *et al.* Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *The Lancet* **395**, 1054–1062 (2020).

[4] https://www.cdc.gov/coronavirus/2019-ncov/hcp/clinical-guidance-management-patients.html. Centers for Disease Control and Prevention, 'Interim Clinical Guidance for Management of Patients with Confirmed Coronavirus Disease (COVID-19)'.

[5] Bolcato, G., Bissaro, M., Pavan, M., Sturlese, M. & Moro, S. Targeting the coronavirus SARS-CoV-2: computational insights into the mechanism of action of the protease inhibitors lopinavir, ritonavir and nelfinavir. *Sci Rep* **10**, 20927 (2020).

[6] Gautret, P. *et al.* Hydroxychloroquine and azithromycin as a treatment of COVID-19: results of an open-label non-randomized clinical trial. *International Journal of Antimicrobial Agents* **56**, (2020).

[7] Mani, D., Wadhwani, A. & Krishnamurthy, P. T. Drug Repurposing in Antiviral Research: A Current Scenario. *Journal of Young Pharmacists* **11**, 117–121 (2019).

[8] Viveiros Rosa, S. G. & Santos, W. C. Clinical trials on drug repositioning for COVID-19 treatment. *Revista Panamericana de Salud Publica/Pan American Journal of Public Health* **44**, e40 (2020).

[9] Fiolet, T., Kherabi, Y., MacDonald, C. J., Ghosn, J. & Peiffer-Smadja, N. Comparing COVID-19 vaccines for their characteristics, efficacy and effectiveness against SARS-CoV-2 and variants of concern: a narrative review. *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases* **28**, 202–221 (2022).

[10] Pollard, A. J. & Bijker, E. M. A guide to vaccinology: from basic principles to new developments. *Nature Reviews Immunology* **21**, 83–100 (2021).

[11] Moghadas, S. M. *et al.* The Impact of Vaccination on Coronavirus Disease 2019 (COVID-19) Outbreaks in the United States. *Clinical Infectious Diseases* **73**, 2257–2264 (2021).

[12] Rinott, E., Youngster, I. & Lewis, Y. E. Reduction in COVID-19 Patients Requiring Mechanical Ventilation Following Implementation of a National COVID-19 Vaccination Program — Israel, December 2020–February 2021. *MMWR. Morbidity and Mortality Weekly Report* **70**, 326–328 (2021).

[13] Harvey, W. T. *et al.* SARS-CoV-2 variants, spike mutations and immune escape. *Nature Reviews Microbiology* **19**, 409–424 (2021).

[14] Consortium, T. C. M. *et al.* Open Science Discovery of Oral Non-Covalent SARS-CoV-2 Main Protease Inhibitor Therapeutics. *bioRxiv* 2020.10.29.339317 (2022) doi:10.1101/2020.10.29.339317.

[15] Morris, A. *et al.* Discovery of SARS-CoV-2 main protease inhibitors using a synthesis-directed de novo design model. *Chemical communications (Cambridge, England)* **57**, 5909–5912 (2021).

[16] Kokic, G. *et al.* Mechanism of SARS-CoV-2 polymerase stalling by remdesivir. *Nature Communications* **12**, 1–7 (2021).

[17] Beigel, J. H. *et al.* Remdesivir for the Treatment of Covid-19 — Final Report. *New England Journal of Medicine* **383**, 1813–1826 (2020).

[18] Xu, X. *et al.* Effective treatment of severe COVID-19 patients with tocilizumab. *Proceedings of the National Academy of Sciences* **117**, 10970–10975 (2020).

[19] Manabe, T., Kambayashi, D., Akatsu, H. & Kudo, K. Favipiravir for the treatment of patients with COVID-19: a systematic review and meta-analysis. *BMC Infectious Diseases* **21**, 1–13 (2021).

[20] Owen, D. R. *et al.* An oral SARS-CoV-2 M pro inhibitor clinical candidate for the treatment of COVID-19. *Science (1979)* **374**, 1586–1593 (2021).

[21] Hammond, J. *et al.* Oral Nirmatrelvir for High-Risk, Nonhospitalized Adults with Covid-19. *New England Journal of Medicine* (2022) doi:10.1056/NEJMoa2118542.

[22] Temmam, S. *et al.* Bat coronaviruses related to SARS-CoV-2 and infectious for human cells. *Nature* (2022) doi:10.1038/s41586-022-04532-4.

[23] Molecular Operating Environment (MOE), 2019.01; Chemical Computing Group ULC, 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2021. https://www.chemcomp.com/Research-Citing_MOE.htm.

[24] Xue, X. *et al.* Production of authentic SARS-CoV M(pro) with enhanced activity: application as a novel tag-cleavage endopeptidase for protein overproduction. *Journal of molecular biology* **366**, 965–975 (2007).

[25] Flynn, J. M. *et al.* Comprehensive fitness landscape of SARS-CoV-2 M^{pro} reveals insights into viral resistance mechanisms. *bioRxiv* 2022.01.26.477860 (2022) doi:10.1101/2022.01.26.477860.

[26] Pavan, M., Bolcato, G., Bassani, D., Sturlese, M. & Moro, S. Supervised Molecular Dynamics (SuMD) Insights into the mechanism of action of SARS-CoV-2 main protease inhibitor PF-07321332. *J Enzyme Inhib Med Chem* **36**, 1646–1650 (2021).

[27] Zhao, Y. *et al.* Crystal structure of SARS-CoV-2 main protease in complex with protease inhibitor PF-07321332. *Protein & Cell* (2021) doi:10.1007/s13238-021-00883-2.

[28] Fornasier, E. *et al.* A new inactive conformation of SARS-CoV-2 main protease. *Acta Crystallographica Section D Structural Biology* **78**, (2022).

[29] Bassani, D., Pavan, M., Bolcato, G., Sturlese, M. & Moro, S. Re-Exploring the Ability of Common Docking Programs to Correctly Reproduce the Binding Modes of Non-Covalent Inhibitors of SARS-CoV-2 Protease Mpro. *Pharmaceuticals* **15**, 180 (2022).

[30] Bissaro, M. *et al.* Inspecting the mechanism of fragment hit binding on SARS-CoV-2 Mpro by using supervised molecular dynamics (SuMD) simulations. *ChemMedChem* (2021) doi:10.1002/cmdc.202100156.

[31] Zhang, C. H. *et al.* Potent Noncovalent Inhibitors of the Main Protease of SARS-CoV-2 from Molecular Sculpting of the Drug Perampanel Guided by Free Energy Perturbation Calculations. *ACS Central Science* **7**, 467–475 (2021).

[32] Luttens, A. *et al.* Ultralarge Virtual Screening Identifies SARS-CoV-2 Main Protease Inhibitors with Broad-Spectrum Activity against Coronaviruses. *J Am Chem Soc* **144**, 2905–2920 (2022).

[33] Anand, K. *et al.* Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra alpha-helical domain. *The EMBO journal* **21**, 3213–3224 (2002).

[34] Anand, K., Ziebuhr, J., Wadhwani, P., Mesters, J. R. & Hilgenfeld, R. Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs. *Science (New York, N.Y.)* **300**, 1763–1767 (2003).

[35] Ho, B. L. *et al.* Critical Assessment of the Important Residues Involved in the Dimerization and Catalysis of MERS Coronavirus Main Protease. *PloS one* **10**, (2015).

The Multifaceted Role of GPCRs in Amyotrophic Lateral Sclerosis: A New Therapeutic Perspective?

Davide Bassani, Matteo Pavan, Stephanie Federico, Giampiero Spalluto, Mattia Sturlese and Stefano Moro

D. Bassani, M. Pavan, S. Federico, G. Spalluto, M. Sturlese, and S. Moro, "The Multifaceted Role of GPCRs in Amyotrophic Lateral Sclerosis: A New Therapeutic Perspective?," Int. J. Mol. Sci., vol. 23, no. 9, p. 4504, Apr. 2022, doi: 10.3390/ijms23094504.

Abstract

Amyotrophic lateral sclerosis (ALS) is a degenerating disease involving the motor neurons, which causes a progressive loss of movement ability, usually leading to death within 2 to 5 years from the diagnosis. Much effort has been put into research for an effective therapy for its eradication, but still, no cure is available. The only two drugs approved for this pathology, Riluzole and Edaravone, are only able to slow down the inevitable disease progression. As assessed in the literature, drug targets such as protein kinases have already been extensively examined as potential drug targets for ALS, with some molecules already in clinical trials. Here, we focus on the involvement of another very important and studied class of biological entities, G protein-coupled receptors (GPCRs), in the onset and progression of ALS. This work aims to give an overview of what has been already discovered on the topic, providing useful information and insights that can be used by scientists all around the world who are putting efforts into the fight against this very important neurodegenerating disease.

1. Introduction

Amyotrophic Lateral Sclerosis (ALS, also referred to as "motor neuron disease") indicates a clinical situation in which the motor neurons of patients undergo a progressive loss in their function and number[1]. This type of neuronal cell, whose cell body is localized in the motor cortex, the brainstem, and the spinal cord, is responsible for the innervation and the control of muscle fibers, essential for voluntary muscle contraction [2]. Their loss has very important consequences on the patient's life, firstly impairing the ability to chew and walk, then to speak and to move, until even the ability to breath is affected, leading, after 2–5 years, to death due to respiratory failure [3]. ALS can be classified into two main types, "sporadic ALS" (the great majority of all cases), which has no

known cause and typically has its onset between the ages of 58 and 63 years, and "familial ALS" (about 5–10% of cases), which is linked to genetic factors, and has its onset between the ages of 47 and 52 years [4].

In both scenarios, the pathology starts with the manifestation of muscle weakness and atrophy, with methods and timing very variable based on the patient and on the parts of the motor neurons that are affected first [5]. Indeed, a classification of the onset of the pathology can be made with reference to the site of its onset. For two-thirds of patients, the limb muscles are affected first ("spinal ALS"), with manifestations mainly in the distal muscles of the dominant hand for the upper limb and in the hamstrings for the lower limb. For the greater part of the remaining patients, the bulbar muscles represent the onset site ("bulbar ALS"), and in this case, dysphagia and chewing problems represent the first manifestations of the pathology [6].

The next steps of the disease involve the progressive spreading of the neurodegeneration process to the unaffected motor neurons, causing an increasing worsening in the patient's daily life, making activities such as eating and walking continuously more difficult and leading to their complete loss. The final and worst clinical scenario has its onset when the respiratory function is significantly affected, progressively increasing the risk of respiratory failure, which is the main cause of death due to ALS [7].

Even if much effort has been made among both academic and industrial scientific groups, no cure has yet emerged for ALS. Riluzole [8] and the recently FDA-approved drug Edaravone [9] (both represented in Figure 1) constitute the only two small molecules used for the ALS treatment, and only succeed in slowing down the disease's progression [10].

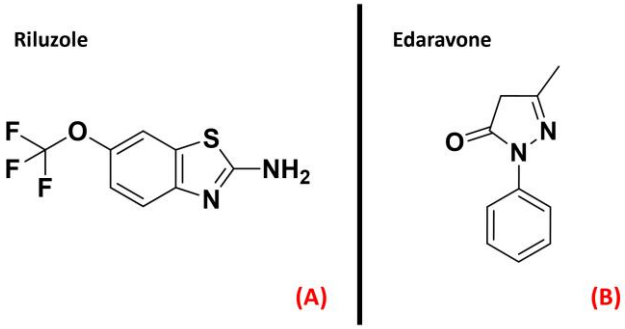


Figure 1. *The chemical structures of Riluzole (**A**) and Edaravone (**B**), which are the only two small-molecule drugs approved currently for ALS treatment.*

Such a neurodegenerating process, affecting 1.75–3 people per 100,000[11], has, in the great majority of cases, no known cause [12], making it even harder to design a therapy for this disease. From

a biochemical point of view, the hallmark of ALS is considered to be the presence of inclusion bodies in the cytoplasm of motor neurons. These aggregates are formed by the TAR DNA-binding protein 43 (TDP-43) [13], a protein involved in several important physiological functions such as DNA repair, splicing, and transcriptional regulation. Even if its main localization site is the nucleus, processes such as its hyperphosphorylation or the mutation of its gene (TARDBP) lead to its aggregation in the cytoplasm [14]. This mislocalization directly causes the dysregulation of several cellular events related to RNA metabolism, DNA replication, and oxidative stress management, leading to the loss of the motor neurons affected [15]. Other molecular targets that have been demonstrated to be important for ALS onset and progression are superoxide dismutase (SOD1) [16] and DNA/RNA-binding protein FUS/TLS (FUused in Sarcoma/Translocated in LipoSarcoma, also called "FUS") [17], which appear to be mutated in the patients.

The knowledge that hyperphosphorylation of TDP-43 is one of the main processes leading to its aggregation has led the scientific community to devote some effort to identifying the protein kinases responsible for such processes, in order to find proper inhibitors for such species [18]. Recent work by Guo et al. goes deep in the examination of the involvement of kinases in ALS progression, enucleating species such as CK1, ERK, GSK3 β, and JAK3 as promising targets for the treatment of this neurodegenerating disease [19]. Another article by Palomo et al. gives an exhaustive panoramic view of the protein kinase inhibitors currently in clinical trials for ALS treatment [20]. Riluzole has proven to increase life expectancy by about 2–3 months [21], and even if its main target still remains the NMDA receptor, recent work by Bissaro et al. suggested that this mechanism could be due to its action on the delta isoform of CK1 [22].

Many molecular candidates (both new chemical entities and compounds coming from repurposing strategies) are nowadays in clinical trials for ALS [23], acting on different biological pathways, with the common aim being to restore the neuronal health status in the affected patients, possibly trying to go in the direction to find a proper cure for this pathology [24]. A comprehensive list of the potential small-molecule drugs now being evaluated by the FDA in clinical phases is reported in Table 1.

Molecule	Target/Mechanism	Developer	Clinical Phase
Ibudilast	Macrophage migration inhibitory factor inhibitor	MediciNova	Phase II/III
Prosetin	Mitogen-activated protein kinase inhibitor	ProJenX	Phase I

Sotuletinib	Macrophage colony-stimulating factor receptor antagonist	Novartis	Phase II
EPI 589	NAD(P)H dehydrogenase modulator	PTC Therapeutics	Phase II
DNL 343	Eukaryotic initiation factor 2b stimulant	Denali Therapeutics Inc	Phase I
Celecoxib/ciprofloxacin	Cyclo-oxygenase 2 inhibitors/DNA gyrase inhibitors	NeuroSense Therapeutics	Phase I
Fingolimod	Apoptosis stimulant and immunosuppressant	ALS Therapy Development Institute	Phase II
Trehalose	Autophagy stimulant and protein aggregation inhibitor	Massachusetts General Hospital	Phase II/III
Sodium cromoglicate	Glial cell modulator and mast cell stabilizer	AZTherapies	Phase II
Dexpramipexole	Antioxidant and apoptosis inhibitor	Knopp Biosciences	Phase II
Masitinib	Tyrosine kinase inhibitor	AB Science	Phase III
NP 001	Macrophage modulator	Neuvivo	Phase II
Fasudil	Rho-associated kinase inhibitor and vasodilatator	Woolsey Pharmaceuticals	Phase II
Levosimendan	Calcium-sensitising phosphodiesterase inhibitor and potassium channel agonist	Orion	Phase III
Apilimoddimesylate	Interleukin 12 inhibitor and interleukin 23 inhibitor	AI Therapeutics	Phase II
Verdiperstat	Peroxidase inhibitor	Biohaven Pharmaceuticals	Phase II/III
Pridopidine	Sigma-1 receptor agonist	Massachusetts General Hospital, Prilenia Therapeutics	Phase II/III
Triheptanoin	Triglyceride replacement agent	Ultragenyx Pharmaceutical	Phase I/II
Reldesemtiv	Troponin stimulant	Cytokinetics	Phase III
BIIB 100	Exportin-1 protein inhibitor	Biogen	Phase I
AGX 201	Histamine receptor modulator	AgoneX Biopharmaceuticals	Phase I
Ranolazine extended release	Sodium channel antagonist	Gilead Sciences	Phase II
GDC 0134	Mitogen-activated protein kinase 12 inhibitor	Genentech	Phase I
NPT520 34	Phosphatidylinositol 3 kinase modulator	Neuropore Therapies	Phase I

Table 1. *Table reporting the different small molecules currently in FDA clinical trials for ALS treatment (updated 13 April 2022).*

Even if some effort has been directed toward trying to highlight the role of protein kinases in ALS progression, this has not been recently or extensively done with respect to G-protein coupled

receptors, biological actors which have been demonstrated to be detrimental to neuronal and physiological conditions. It is important to remember that ALS is a non-cell-autonomous disease, which means that the neuronal damage characterizing the pathology is caused by aberrant processes also happening outside the neurons themselves. Indeed, ALS progression has been demonstrated to be strongly related to glial cell dysregulation (mainly microglia and astrocytes) [25]. GPCRs are very widely expressed proteins in the human organism [26], and so a beneficial effect could also be obtained by targeting extraneuronal receptors, which could trigger biological processes that, in the overall scenario, could mitigate if not reverse the disease progression.

GPCRs are membrane receptors and constitute one of the main protein families encoded by human genes, with more than 800 members already identified [27], divided into six different classes (identified alphabetically with letters from "A" to "F") based on their similarities in sequence and function. They all share a common architecture formed of a seven-α-helix transmembrane domain (usually referred to as "7-TM"), an extracellular N-terminal domain, and an intracellular C-terminal domain. These proteins exert their roles by coupling with an intracellular messenger called "heterotrimeric G protein", which is formed by α,β, and γ subunits, and interacts with different intracellular partners based on its type. The α subunit is displaced from the βγ-complex upon GPCR–ligand binding, and its fate depends on its Gα family belonging. Indeed, activated $G_{i/0}$ proteins inhibit adenylyl cyclase (AC), reducing the production of the second messenger cyclic adenosine monophosphate (cAMP); G_s, conversely, activates adenylyl cyclase, and $G_q\alpha$ subunits activate phospholipase C (PLC), leading to an increase in Ca^{2+} influx in the cytoplasm [28]. The physiological roles of GPCRs include homeostasis modulation, mood balancing, immune system regulation, neuronal plasticity, and many more [29]. The goal of the present work is to give a panoramic view of the GPCRs which have been linked to ALS onset and progression, presenting what has already been done to modulate their action, and highlighting new potential therapeutical scenarios. Table 2 summarizes the outcomes of our study, listing the GPCR targets that will be discussed and highlighting the new possible paths that can be taken in order to exploit their therapeutic potential for ALS.

Receptor/Receptor Family	Cellular Expression	Potential for ALS Treatment	References
Adenosine receptors	Circulatory, immune, respiratory, and nervous systems	Ambiguous	[32–39]
Purinergic receptors P2Y	Almost all human tissues	Antagonism	[44–49]
Chemokine receptors	Predominantly on leukocytes surface	CXCR3, CXCR4, and CCR2 Antagonism	[53–56]
Angiotensin II receptors	Adrenal cortex, kidneys, vascular and cardiac muscles, nervous system	AT_1 Antagonism	[64–66]
Dopamine receptors	Arteries, heart, kidneys, CNS	D2R Agonism	[71–78]
Serotonin receptors	Almost all human tissues	Ambiguous	[84–88]
GPR17 receptor	CNS, kidneys, heart	Antagonism	[95–102]
Adrenergic receptor β_2	GI tract, respiratory system, blood vessels, pancreas, nervous system	Agonism	[106,107]
Histamine receptors	GI tract, circulatory, immune, and nervous systems.	Ambiguous	[111–114]
Cannabinoid receptors	CNS and immune system	CB_2 agonism	[119–122]
Prostaglandin E_2 receptor	GI tract, kidneys, reproductive, skeletal, immune, and nervous systems.	Ambiguous	[136–138]
Vasoactive intestinal peptide receptors	Almost all human tissues	Agonism	[145–149]
Metabotropic glutamate receptors	Nervous system	mGluR I antagonism/ mGluR II and mGluR III agonism	[152–156]

Table 2. *Table summarizing the evidence about the therapeutic potential of the GPCRs examined in this article for ALS treatment.*

Our work will be beneficial for all of the scientists who are dedicating their knowledge and efforts to the eradication of ALS.

2. GPCRs Involved in ALS

2.1. Puringergic Receptors P2Y and Adenosine Receptor A_{2A}AR

Purinergic receptors are a peculiar class of membrane receptors, sensitive to a wide series of purinergic ligands such as ATP, ADP, UTP, UDP, UDP-glucose, and adenosine. The kind of molecules interacting with them defines their classification into one of the three subfamilies forming this class. The first group, called "P1 receptors", is formed of GPCRs activated upon adenosine binding (and for these reasons are also known as "adenosine receptors"), while "P2Y receptors" are GPCRs that can bind to ATP, UDP, and their diphosphate analogs ADP and UDP (with the addition of UDP-glucose). The last subfamily, named "P2X receptors", are ligand-gated ion channels exclusively sensitive to ATP [30]. Being the P2X family not formed by GPCRs, our evaluations will focus on the first two families of receptors.

The "P1" subfamily, more commonly referred to as adenosine receptors (ARs), is a group of purinergic class A GPCRs divided into four subtypes, A_1AR, A_{2A}AR, A_{2B}AR, and A_3AR, each involved in many different physiological processes. While the functions of A_{2B}AR and A_{3A}R are mainly related to the circulatory, immune and respiratory systems, the A_1AR and A_{2A}AR proteins are importantly present in the central nervous system (CNS) [31]. Moreover, the A_1AR and the A_{2A}AR receptors have been demonstrated to play a crucial role in neuroprotection, neuronal survival, and neuroinflammation [32]. A study from Vincenzi et al. reported an upregulation of A_{2A}AR receptors in the lymphocytes of people affected by ALS [33], while Yoshida et al. measured adenosine levels in the cerebrospinal fluid of ALS patients, finding out that these were significantly higher with respect to the control subjects [34]. Unexpectedly, treatment with the A_{2A}AR antagonist caffeine (Figure 2, panel A), which is usually referred to as a protective agent against Alzheimer's Disease (AD) and Parkinson's Disease (PD), was demonstrated to shorten the survival of ALS-affected $SOD1^{G93A}$ mice, a well-known experimental model for ALS (Potenza et al.) [35], even if some explanation for this phenomenon can be provided by the non-selectivity of caffeine [36]. Indeed, Ng et al. showed that suppression of A_{2A}AR signaling delays the progression of ALS in the same $SOD1^{G93A}$ mouse model [37]. This is in accordance with the study of Mojsilovic-Petrovic et al., which demonstrated that A_{2A}AR inhibitors (such as the non-selective enprofylline (Figure 2, panel B), and the A_{2A}AR-selective KW-6002 (Figure 2, panel C), also called Istradefylline) protect motor neurons from toxic insult, highlighting the beneficial effects of such activity for ALS patients [38].

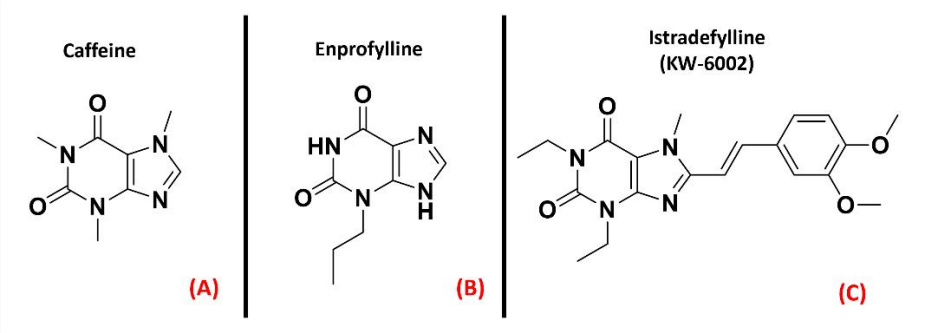


Figure 2. *The chemical structures of the non-selective adenosine receptors antagonists caffeine (**A**) and Enprofylline (**B**), and the selective A_{2A}AR antagonist Istradefylline (**C**).*

Despite all this evidence, a study from Liu et al. showed that A_{2A}AR activation, suppressing AMPK activation, suppressed TDP-43 mislocalization [39]. The multifactorial nature of ALS makes it very difficult to define sharply whether agonism or antagonism of the A_{2A}AR receptor has the best risk/benefit ratio, but the literature clearly defines this GPCR (represented in Figure 3) as one of the promising targets for ALS treatment.

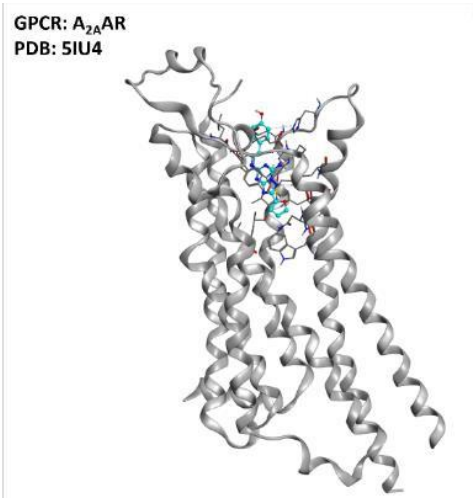


Figure 3. *Representation of the structure of the A_{2A}AR receptor (sourced from the Protein Data Bank [40], PDB code: 5IU4 [41], method: X-ray diffraction, resolution: 1.72 Å). The image was created and rendered with the Molecular Operating Environment (MOE) suite [42].*

The second subfamily of purinergic receptors, "P2Y", is present in a great variety of human tissues, but their main biological roles are identifiable in blood clotting, vasodilatation, and immune response [43]. The P2Y family comprises eight different isoforms, among which $P2Y_{12}$ has gained the interest of the scientific community for its role in neuroinflammation, as addressed by Morillas et al. [44] and Amadio et al. [45].

Jacobson et al. recently highlighted the proinflammatory effect of P2Y agonists, reporting that antagonizing this class of GPCRs could be considered a way of treating inflammatory conditions [46].

Even if inhibitors of the $P2Y_{12}$ isoform are already marketed as antiplatelet drugs (e.g., Clopidogrel, Prasugrel, Ticagrelor, all represented in Figure 4), Jacobson et al. pointed out that there is a lack of selective and versatile P2Y ligands for each subtype, meaning that the drug discovery process is still very active in this specific field.

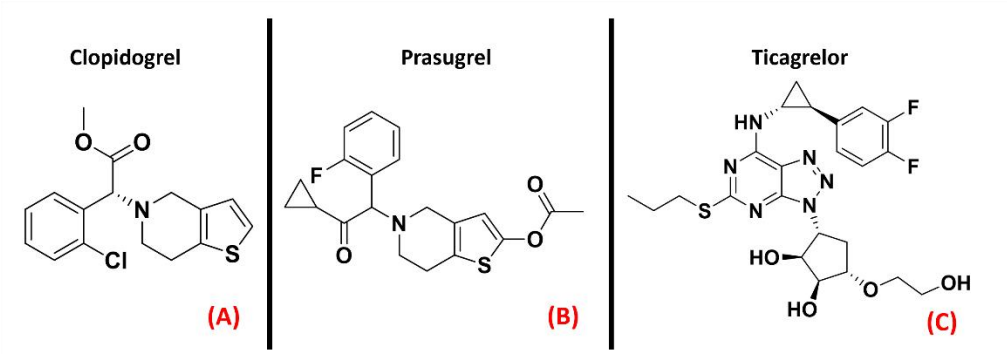


Figure 4. *The chemical structures of the selective and irreversible $P2Y_{12}$ receptor antagonists Clopidogrel (**A**) and Prasugrel (**B**). Ticagrelor, a selective, reversible, allosteric $P2Y_{12}$ receptor antagonist, is also reported (**C**).*

Specifically, D'Ambrosi et al. reported an upregulation of $P2Y_6$ receptors in the microglia SOD1 mutant models of ALS, also remembering that this phenomenon is associated with brain damage [47]. Moreover, $P2Y_{12}$ (represented in Figure 5) is upregulated in spinal cord microglia upon nerve injury, as pointed out by Kobayashi et al. [48]. Converging information is provided by a study from Moore et al. [49], further confirming $P2Y_{12}$ as a potential target for modulating neuroinflammation and neuronal damage. The data currently available help in suggesting the practical possibility of ALS regulation through purinergic receptor modulation, and this will be realizable as soon as proper inhibitors can be designed, potentially avoiding the non-desired antiplatelet effect of these molecules.

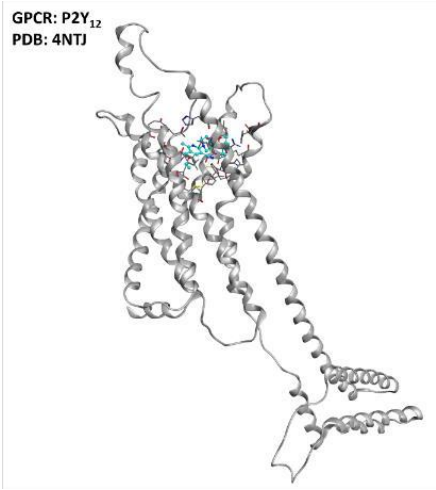


Figure 5. *Representation of the structure of the $P2Y_{12}$ receptor (sourced from the Protein Data Bank, PDB code: 4NTJ [50], method: X-ray diffraction, resolution: 2.62 Å). The image was created and rendered with MOE.*

2.2. Chemokine Receptors

Chemokine receptors constitute a group of about 20 classes of GPCRs found mainly on the surface of leukocytes, which respond to specific ligands to control chemotaxis. The ligands for these proteins are called chemokines and are a peculiar kind of cytokine used for inducing a directional movement of certain types of cells, such as epithelial and immune ones. The chemokines, as well as the receptors they act on, can be divided into four families, namely CC (e.g., chemokine CCL4), CXC (e.g., chemokine CXCL8, also known as IL-8), XC (e.g., XCL1), and CX3C (of which the only member today is CX3CL1, also called neurotactin) [51]. The activation of chemokine receptors leads to Ca^{2+} influx and cell mobilization [52]. Several of these receptors are important in the progression of motor neuron damage. La Cognata et al. highlighted an upregulation of CXCR2 in both sporadic ALS patients and $SOD1^{G93A}$ mice, showing that treating the mouse models with the CXCR2 allosteric inhibitor Reparixin (Figure 6, panel B), the neuromuscular function of the subjects was improved [53]. Another interesting paper published by Rabinovich-Nikitin et al. highlighted the benefits in terms of lifespan and motor function obtained on $SOD1^{G93A}$ mouse models through the administration of the CXCR4 antagonist AMD3100 (also known as "Plerixafor", Figure 6, panel A) [54].

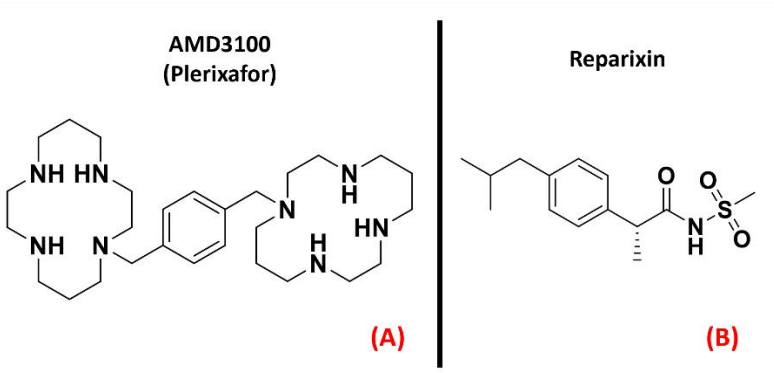


Figure 6. *The chemical structures of the CXCR4 antagonist AMD3100 (also known as "Plerixafor", (**A**) and the inhibitor of CXCR1 and CXCR2 known as Reparixin (**B**).*

Several scientific works have reported an increase in circulating chemokines and cytokines in ALS patients, as recently detailed by Liu et al. [55], and the upregulation of chemokine receptors CXCR3, CXCR4, and CCR2 was also highlighted in the pathology of interest by Perner et al. [56], who also proposed CXCR3 and its ligands as possible therapeutic targets for ALS. These scientific works converge in addressing chemokine receptor modulation as a possibility for ALS treatment, focusing on the antagonism of certain isoforms. The three-dimensional structures of CXCR2, CXCR3, and CXCR4 are represented in Figure 7.

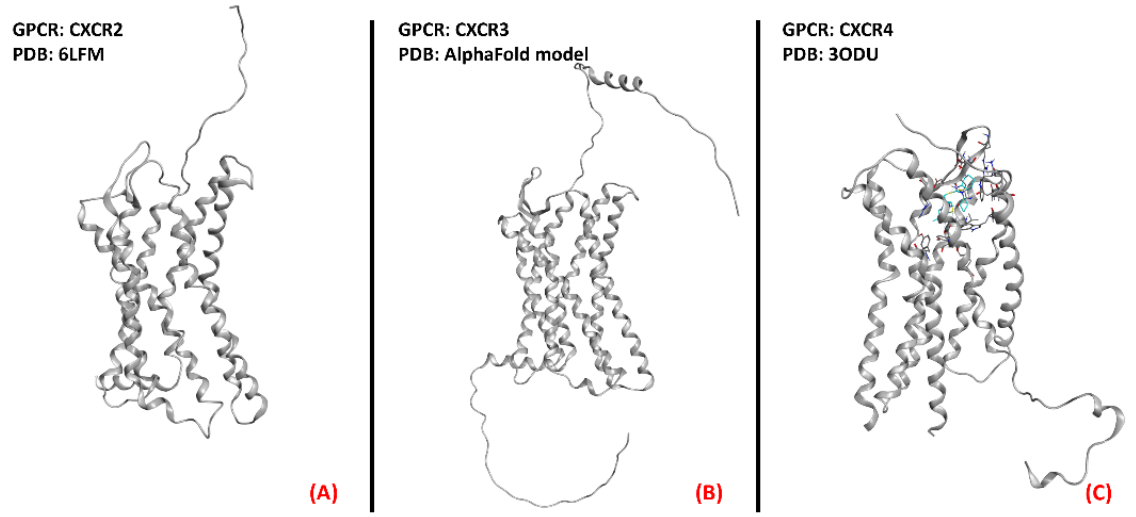


Figure 7. *Representation of the three-dimensional structures of the chemokine receptors that could be considered for ALS treatment. CXCR2* ***(A)*** *receptor (sourced from the Protein Data Bank, PDB code: 6LFM [57], method: cryo-EM, resolution: 2.50 Å), CXCR3 (****B****) (with no experimentally resolved structure available, the model from the AlphaFold [58] database is presented), and CXCR4* ***(C)*** *receptor (sourced from the Protein Data Bank, PDB code: 3ODU [59], method: X-ray diffraction, resolution: 2.50 Å). The images were created and rendered with MOE.*

2.3. Angiotensin II Receptors (ATRs)

Angiotensin II receptors (ATRs) are a group of GPCRs that have gained fame for their importance in the therapy of hypertension. Indeed, their main physiological role is related to the renin–angiotensin–aldosterone system, one of the main physiological pathways for blood pressure regulation and fluid and electrolyte balance [60]. Briefly, the peptide hormone angiotensinogen is secreted by the liver and cleaved by renin to form angiotensin I, which is then converted to angiotensin II by the angiotensin-converting enzyme (ACE), produced by the lungs. Angiotensin II acts on its receptors and modulates several processes, such as aldosterone secretion (in the adrenal glands), water and sodium retention (in the kidneys), sanguine pressure, and vasopressin production (in the CNS) [61]. To treat hypertension, many efforts have been directed towards the creation of drugs acting as ATRs inhibitors. The most famous drug family designed for this purpose is represented by the “Sartans”, which selectively bind to the first isoform of angiotensin receptors [62]. Indeed, ATRs can be divided into four isoforms, AT_1, AT_2, AT_3, and AT_4. While the latter two are still in the early stages of research, the first two isoforms have been more deeply characterized. AT_1 is mainly found in blood vessels, heart, kidney, brain, and adrenal cortex, mediating vasoconstrictive effects [60]. AT_2 receptors are more concentrated in the fetus and neonate, and their functions are more strongly related to neuronal development and excitability [63]. A study by Kawajiri et al. highlighted a reduction in angiotensin II levels in the CSF coming from ALS patients, reporting two opposite consequences: the reduction of protection and repair mediated by AT_2, on the one hand, and the reduction of oxidative

stress due to AT_1 on the other. Indeed, Kawajiri et al. hypothesized that angiotensin II could be downregulated in CSF of ALS patients as a protective reaction, avoiding excessive activation of AT_1 [64]. The benefits of AT_1 antagonism have also been underlined by Iwasaki et al., who reported evidence of the neurotrophic effects on spinal motor neurons of the drug Olmesartan (Figure 8, panel A), specifically referring to its potential application in ALS [65]. Furthermore, an article by Mammana et al. highlighted the AT_1 antagonism-mediated neuroprotective effects of Telmisartan (Figure 8, panel B), also outlining the decrease in neuronal injury and microglial activation caused by it [66].

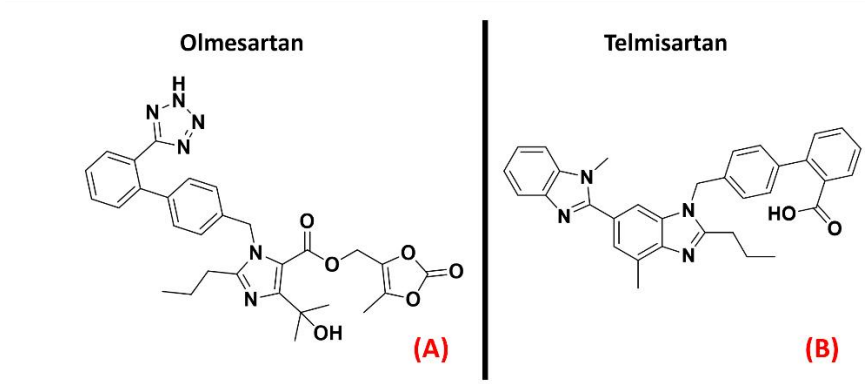


Figure 8. *The chemical structures of the AT_1 antagonists Olmesartan (**A**) and Telmisartan (**B**).*

Summing up the information obtainable from the literature, AT_1 inhibition could be examined as a potential new therapeutic method of fighting ALS conditions. Both AT_1 and AT_2 receptors are represented in Figure 9 below.

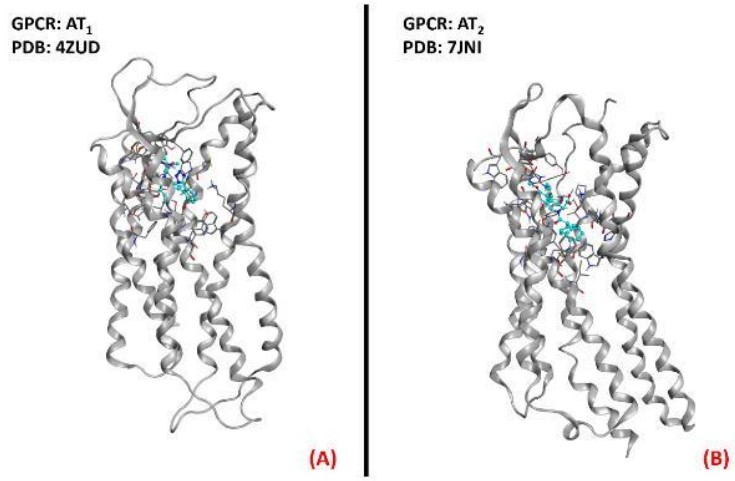


Figure 9. *Representation of the three-dimensional structures of the angiotensin II receptors that could be considered for ALS treatment. AT_1receptor **(A)** (sourced from the Protein Data Bank, PDB code: 4ZUD [67], method: X-ray diffraction, resolution: 2.80 Å) and AT_2receptor **(B)** (sourced from the Protein Data Bank, PDB code: 7JNI [68], method: X-ray diffraction, resolution: 3.00 Å). The images were created and rendered with MOE.*

2.4. Dopamine Receptors

Dopamine receptors are among the most important and widely studied G-protein coupled receptors, mainly for their important physiological roles in neurotransmission. This class of GPCRs is divided into five isoforms, which are separated into two classes. The first, also called the "D1-like family", comprises the D1R and D5R receptors, which are coupled to a G_s protein responsible for adenylyl cyclase activation upon binding. The second family, also known as the "D2-like" family, comprises the D2R, D3R, and D4R proteins, all coupled with a G_i protein with inhibitory activity on adenylyl cyclase. Dopamine receptors are localized in different peripheral parts of the organism, such as arteries, heart, and kidneys, but their activities much more determinant within the CNS. Indeed, dopamine is the main neurotransmitter involved in the reward system, and its signaling is of crucial importance for processes such as cognition, memory, and motor control [69]. Dysregulation of the dopaminergic system in the brain represents the main cause of very important diseases such as schizophrenia, attention deficit hyperactivity disorder (ADHD), and Parkinson's disease [70]. Several articles highlight a correlation between dopamine signaling and ALS development [71]. We have previously reported what was assessed by Liu et al. regarding the protective effect of the A_{2A} receptor on TDP-43 mislocalization [39]. A recent study by Lai et al. details how this beneficial activity can be blocked by D2R activation [72]. Despite this, D2R was also identified as important for the modulation of motor neuron excitability by Huang et al. [73]

Fujimori et al. showed how the treatment with Ropinirole (Figure 10, panel A), an agonist for receptors D2R, D3R, and D4R mainly used for Parkinson's disease, has neuroprotective effects in ALS models [74]. Additionally, D2R agonists such as Bromocriptine and Sumanirole (both represented in Figure 10, panels B and C, respectively) were tested by Huang et al., who reported that the final effect of such activity on ALS models was an increase in motor neuron survival [73]. Another agonist for the "D2-like family" of dopamine receptors is the R(+) enantiomer of the Parkinson's disease drug Pramipexole, known as Dexpramipexole (Figure 10, panel D).

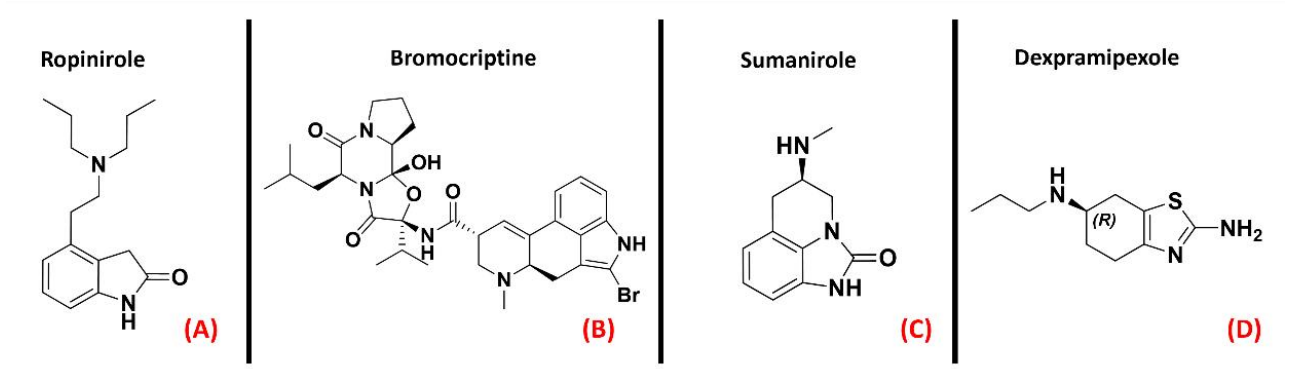


Figure 10. *The chemical structures of the dopamine receptor agonists Ropinirole (**A**), which has an affinity for D2R, D3R, and D4R, Bromocriptine (**B**), also non-selective with an affinity for D2R, D3R, and D4R, and Sumanirole (**C**), selective for D2R. (**D**) Chemical structure of Dexpramipexole (its neuroprotective effects are attributed to dopaminergic-independent activities).*

Even if Pramipexole is a powerful agonist of D2R, D3R, and D4R (all depicted in Figure 11) [75], its R(+) enantiomer has a very low affinity for dopamine receptors, so its neuroprotective effects have to be due to a non-dopaminergic action [76]. This molecule has been considered a promising candidate for ALS conditions [77]. After the phase III clinical trial, however, its development in Europe was discontinued [78].

D2R is still a very relevant target for Parkinson's disease treatment, but the presented literature concords in considering it also a protein of high therapeutic potential for treating ALS.

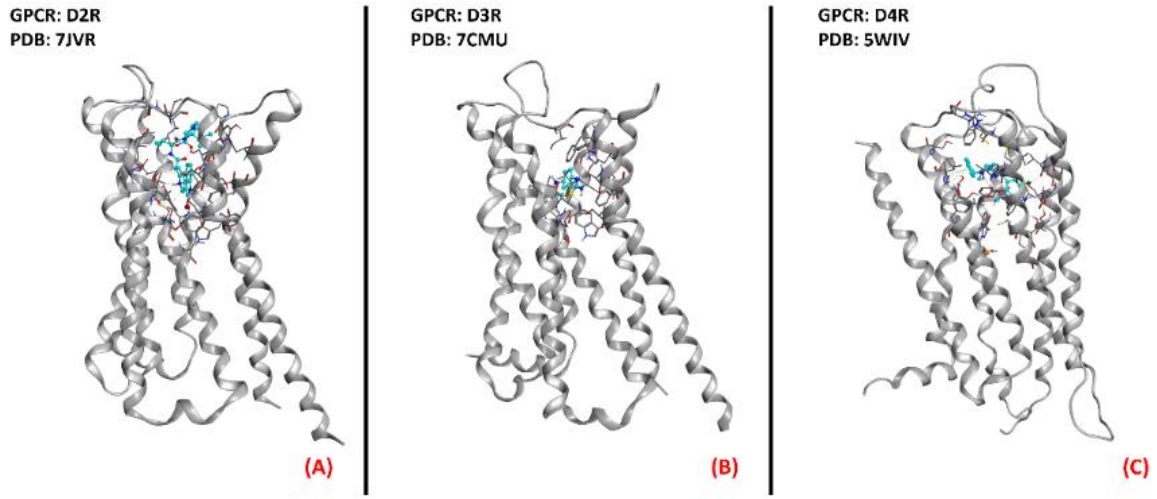


Figure 11. *Representation of the three-dimensional structures of the dopamine receptors that could be considered for ALS treatment. (**A**) D2R (sourced the Protein Data Bank, PDB code: 7JVR [79], method: cryo-EM, resolution: 2.80 Å), (**B**)D3R (sourced from the Protein Data Bank, PDB code: 7CMU [80], method: cryo-EM, resolution: 3.00 Å), and (**C**) D4R (sourced from the Protein Data Bank, PDB code: 5WIV [81], method: X-ray diffraction, resolution: 2.14 Å). All of the images were created and rendered with MOE.*

2.5. Serotonin (5-HT) Receptors

Serotonin (also called 5-hydroxytryptamine, or 5-HT) receptors represent one of the most populated subfamilies of class A GPCRs, consisting of 13 G-protein coupled receptor isoforms (5-HT_{1A}, 5-HT_{1B}, 5-HT_{1D}, 5-HT_{1E}, 5-HT_{1F}, 5-HT_{2A}, 5-HT_{2B}, 5-HT_{2C}, 5-HT_{4}, 5-HT_{5A}, 5-HT_{5B}, 5-HT_{6}, and 5-HT_{7}) distributed throughout the entire human organism, and a cation channel (5-HT_{3}), mainly involved in gastrointestinal motility [82]. It is also interesting to note that almost all of these isoforms are present in the CNS [83]. Concerning ALS, the serotonin receptor which has gained the greatest popularity is 5-HT_{2B}. An article from Oussini et al. reported that the activity of this biological entity could limit the degeneration of spinal cord mononuclear phagocytes, which is a process typical of neurodegenerative diseases. This article highlighted that the ablation of the 5-HT_{2B} gene resulted in an acceleration of ALS progression in mutant SOD1 mouse models. Indeed, they showed that the administration of a 5-HT_{2B} selective antagonist (SB204741, Figure 12, panel A) caused an important reduction in microglia viability, while treatment with the agonist BW723C86 (Figure 12, panel B) induced an increase in viability [84]. Another work by Dentel et al. reported that the spasticity associated with ALS progression could be strongly alleviated by the administration of inverse agonists of 5-$HT_{2B/C}$ such as SB206553 and Cyproheptadine (both depicted in Figure 12, panels C and D, respectively) [85]. A recent article by Arnoux et al., on the other hand, highlighted the lack of beneficial effects when ALS-affected $SOD1^{G86R}$ mutants were treated with the 5-HT_{2B} agonist BW723C86 [86]. The main factor that has always limited the development of 5-HT_{2B} agonists is their inherent cardiotoxicity, which can determine valvular heart disease [87]. Two randomized, double-blind, placebo-controlled multicenter studies (phase III) were conducted in 2004 by Meininger et al. to evaluate the potential benefits of Xaliproden (Figure 12, panel E), a 5-HT_{1A} receptor agonist with neuroprotective effects, in ALS patients.

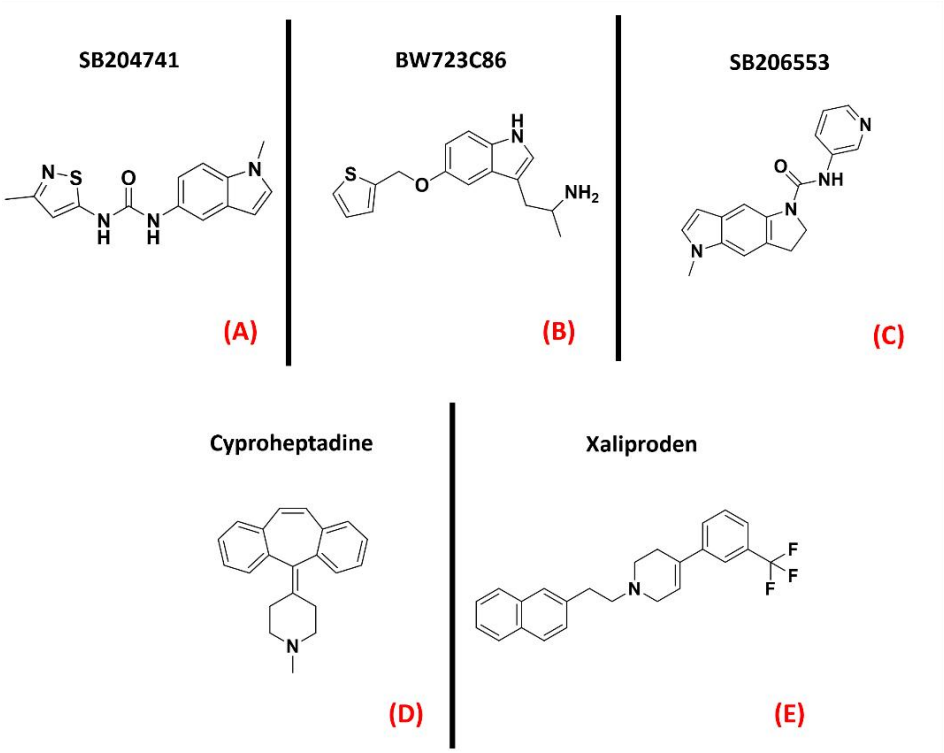


Figure 12. *The chemical structures of the serotonin receptor modulators are treated in this review. The selective 5-HT$_{2B}$ antagonist SB204741 (**A**), the 5-HT$_{2B}$ agonist BW723C86 (**B**). (**C**) Structure of the 5-HT$_{2B/C}$ mixed inverse agonist SB206553; (**D**) chemical structure of Cyproheptadine, a non-selective mixed serotonin receptor antagonist (which is an inverse agonist of 5-HT$_{2B}$). (**E**) Structure of the 5-HT$_{1A}$ receptor agonist Xaliproden.*

Despite the promising outcomes of the prior experiments [88], no effective slowing down in the progression of the pathology was evidenced. Even if some limitations have been encountered in serotoninergic modulation for motor neuron disease, the 5-HT receptors have been demonstrated to be targets of relevance in the ALS scenario. The three-dimensional structures of 5-HT$_{1A}$, 5-HT$_{2B}$, and 5-HT$_{2C}$ are represented in Figure 13.

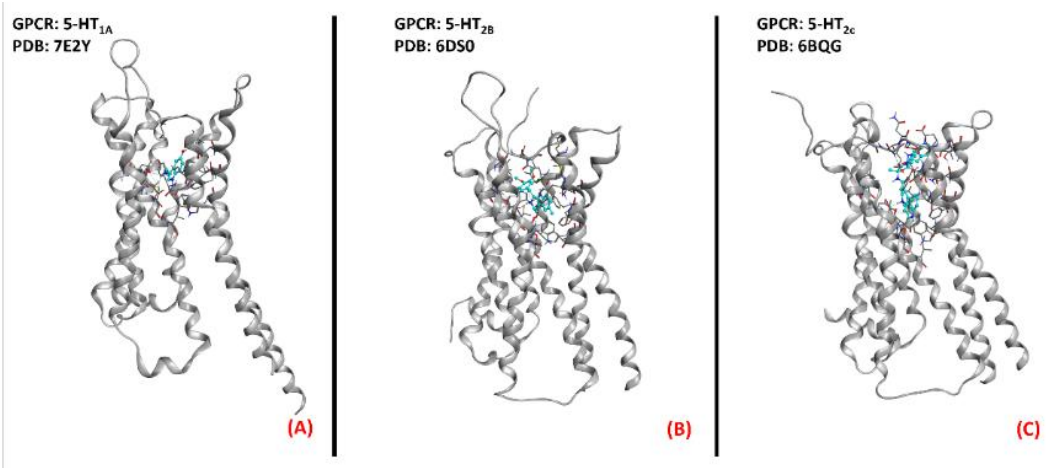


Figure 13. *Three-dimensional structures of the serotonin receptors that could be considered for ALS treatment. (**A**) 5-HT$_{1A}$ (sourced from the Protein Data Bank, PDB code: 7E2Y [89], method: cryo-EM,resolution: 3.00 Å), (**B**) 5-HT$_{2B}$ (sourced*

*from the Protein Data Bank, PDB code: 6DS0 [90], method: X-ray diffraction, resolution: 3.19 Å), and (**C**) 5-HT$_{2C}$ (sourced from the Protein Data Bank, PDB code: 6BQG [91], method: X-ray diffraction, resolution: 3.00 Å). All of the images were created and rendered with MOE.*

2.6. GPR17 Receptor

GPR17 (also known as "uracil nucleotide/cysteinyl leukotriene receptor") is a protein belonging to the 15th subfamily of class A GPCRs. One of its peculiarities is that its structure is phylogenetically related to both cysteinyl leukotriene (CysLT) receptors and to purinergic P2Y receptors [92]. This receptor is activated by uracil nucleotides such as UDP, UDP-glucose, and UDP-galactose, but is also sensitive to CysLTs, like Leukotriene D4 and C4 [93]. GPR17 is mainly expressed in the CNS (but also in kidneys, heart, and generally in organs that can experience ischemic damage), and a more pronounced presence of this protein has been highlighted in oligodendrocyte precursor cells (OPCs). Upregulation of GPR17 can be observed in neuronal cells surrounding an ischemic-injured area, making this protein a marker for cellular stress and death. It has been reported in the literature that in the case of a demyelinating event, GPR17 is involved in the remyelination process, but the mechanism of its involvement is still debated [94]. What is known is that GPR17 is deputed to accompany the OPCs in the early stages of their differentiation process, and so its downregulation is necessary for these cells to complete their maturation. As a result of this, overexpression of this protein leads to incomplete OPC development, impairing myelination and promoting inflammatory responses [95]. Moreover, GPR17 upregulation in neurons was linked to increased cell damage by Zhao et al., who also observed that its knockdown attenuated neuronal injury and microgliosis [96]. In the field of ALS, GPR17 was demonstrated by Bonfanti et al. to be upregulated in the spinal cord of SOD1^{G93A} mouse models [97]. As reported in a recent study by Raffaele et al., while the application of non-selective GPR17 antagonists such as HAMI3379 (Figure 14, panel A) or Montelukast (a marketed CysLT receptor inhibitor, which is represented in Figure 14, panel B) has been shown to improve remyelination processes (as also demonstrated by Merten et al. for the first of these two molecules [98]), the implementation of agonists has also been demonstrated to be beneficial in pushing OPCs to start differentiating [99]. Jin et al. asserted that the inhibition of GPR17 by Cangrelor (Figure 14, panel C) results in the amelioration of cognitive deficits through the inhibition of oxidative stress and neuroinflammation in Alzheimer's Disease mouse models [100].

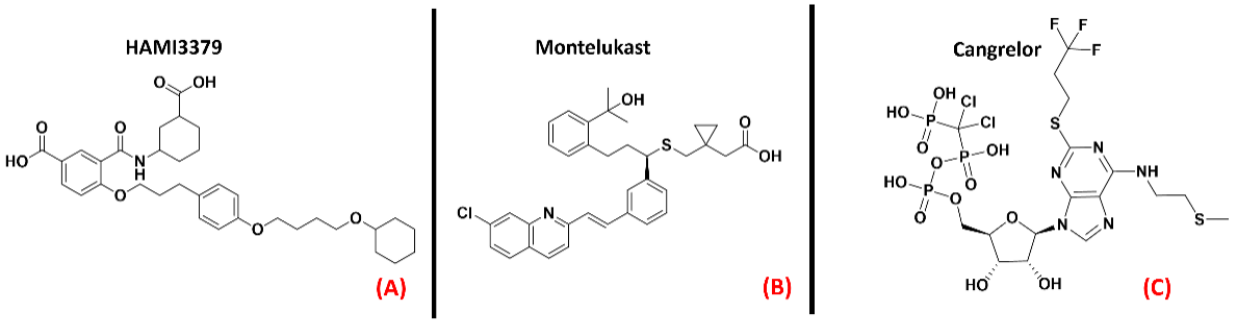


Figure 14. *The chemical structures of the GPR17 inhibitors. The non-selective inhibitors HAMI3397, Montelukast (sold as a CysLT receptors inhibitor for asthma), and Cangrelor (an antiplatelet drug, reversible inhibitor of $P2Y_{12}$ receptor), respectively (**A**–**C**).*

Marschallinger et al., in a recent paper, highlighted a restoration in cognitive function and a reduction in neuroinflammation in rats treated with Montelukast [101]. Another study by Burnstock et al. indicated that the in vivo knockdown of GPR17 markedly reduced brain damage [102].

The information available nowadays converges in indicating GPR17 (which three-dimensiona structure is provided in Figure 15) as a promising target for neuroinflammation and neurodegeneration diseases. Even if, at the present moment, these efforts are more focused on multiple sclerosis treatment, GPR17 regulation for ALS is also attracting increasing interest from the scientific community.

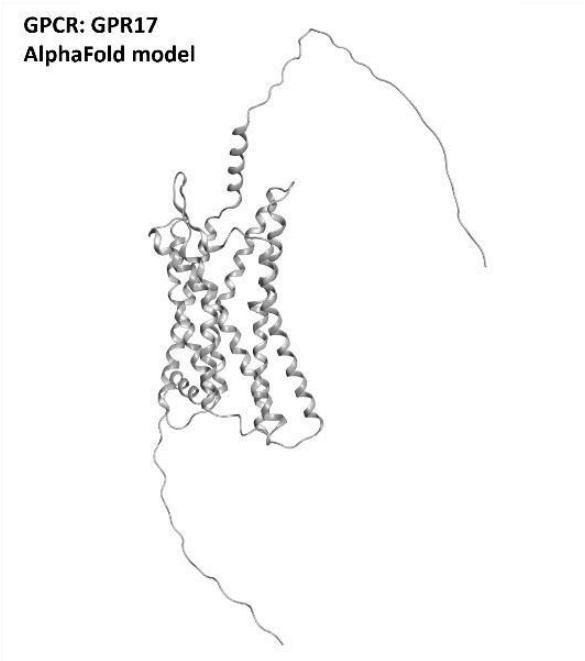


Figure 15. *Structure of the GPR17 receptors (with no experimentally resolved structure available, the model sourced from the AlphaFold database is presented). The image was created and rendered with MOE.*

2.7. Adrenergic Receptor β_2

Adrenergic receptors are part of the 17th subfamily of class A GPCRs and are divided into nine different isoforms (α_{1A}, α_{1B}, α_{1D}, α_{2A}, α_{2B}, α_{2C}, β_1, β_2, β_3), which are involved in very important physiological functions such as smooth muscle contraction and relaxation, heart muscle contraction (mainly receptors β_1 and β_2) [103], and glycogenolysis [104]. The research to find a link between adrenergic transmission and ALS has been focused on the β_2 isoforms of this GPCR. Historically, β_2 agonists represent one of the main solutions for asthma therapy [105], and drugs such as Salbutamol, Clenbuterol, and Formoterol (all depicted in Figure 16) are an example of this. A recent work by Bartus et al. highlighted the potentialities of β_2 agonists for ALS, reporting that the downstream effects of these molecules can be useful for protecting spinal cord neurons, both preserving and/or restoring their function [106].

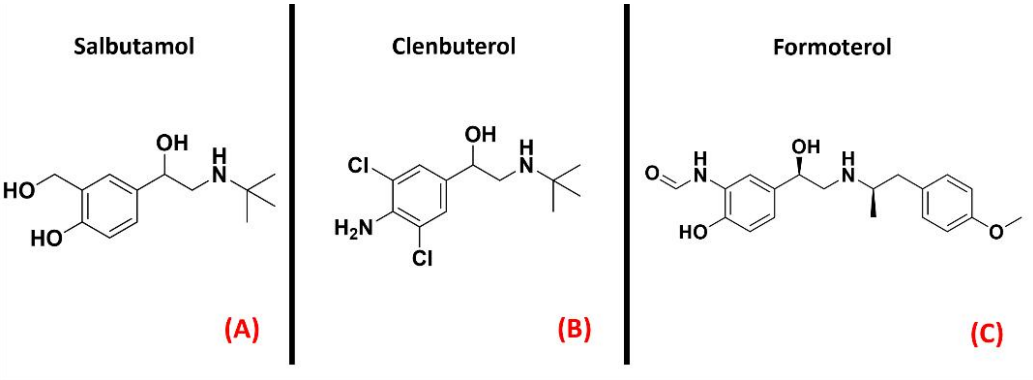


Figure 16. *The chemical structures of the adrenergic β_2 receptor agonists Salbutamol (**A**), Clenbuterol (**B**), and Formoterol (**C**).*

The pathways responsible for such outcomes presented by Bartus et al. are very biologically important for cell homeostasis, such as the cAMP/PKA/CREB pathway, the PI3K-Akt-mTOR pathway, and the PKA/SIRT1 pathway. Other than neuroprotection, other effects attributed by the authors to this class of ligands are the increase in muscle strength and the amelioration of mitochondrial function. Another study by Teng et al. reported a favorable effect of the β_2 agonist Clenbuterol on SOD1^{G93A} mice [107]. The outcomes of these studies open new possibilities in drug discovery for ALS, focusing special attention on adrenergic β_2 receptor modulation. A three-dimensional representation of the adrenergic β_2 receptor is provided in Figure 17.

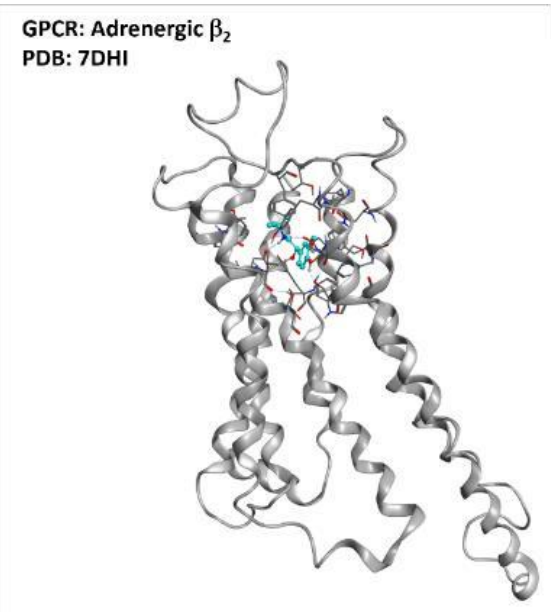


Figure 17. Representation of the structure of the adrenergic β_2 receptor (sourced from the Protein Data Bank, PDB code: 7DHI [108], method: cryo-EM, resolution: 3.26 Å). The image was created and rendered with MOE.

2.8. Histamine Receptors

Histamine receptors represent a group of class A GPCRs that has attracted a lot of interest in the pharmaceutical world in recent decades. This family is composed of four different members (H_1, H_2, H_3, and H_4), each with a specific localization in the organism. Their functions, vary from one isoform to another, ranging from vasoconstriction (H_1) to gastric acid secretion (H_2), to neurotransmitter release (H_3), to immunoregulation (mainly H_2 and H_4) [109]. For each histamine receptor, the research has mainly focused on the mechanism of antagonism, of which several marketed drugs are still now relevant examples (e.g., Cetirizine, Figure 18, panel A, for H_1 antagonism; or Famotidine, represented in Figure 18, panel B, for H_2 blockage) [110].

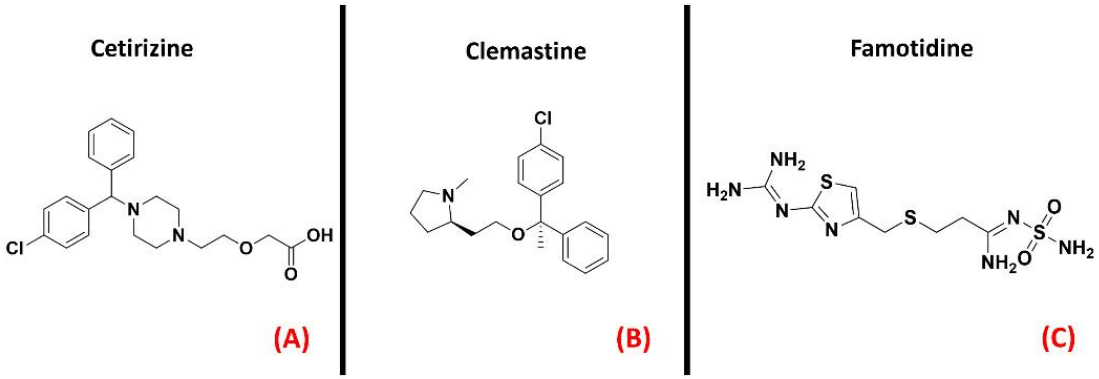


Figure 18. *The chemical structures of the H_1 receptor antagonists Cetirizine (**A**) and Clemastine (**B**). The H_2 receptor antagonist Famotidine is also reported (**C**).*

An article from Apolloni et al. reported the involvement of histaminergic signals in ALS progression, highlighting that histamine receptors are dysregulated in the cortex, spinal cord, and hypothalamus of $SOD1^{G93A}$ ALS-affected mice [111]. This study reported that histamine could counteract the pro-inflammatory phenotype of microglia, mainly through its H_1 and H_4 receptor isoforms. This would be mediated by both the reduction of NOX-2 and NF-kB expression and the increase in production of other species, such as IL-6 and IL-10. Another work by Volontè et al. highlighted the neuroprotective effects of histamine signaling in ALS, again giving higher relevance to H1R (represented in Figure 19, panel A) and H4R [112]. On the other hand, Zhang et al. described H_1 and H_4 receptors as being responsible for pro-inflammatory cytokine release in microglia, while H_2 and H_3 were considered to be the main actors of anti-inflammation in that environment (the H_2 receptor is depicted in its three-dimensional structure in Figure 19, panel B) [113]. Another article by Apolloni et al. reported an amelioration in ALS progression of $SOD1^{G93A}$ mice treated with the anti-histaminergic drug Clemastine [114]. Even if much more remains to be understood about the specific role of each histamine receptor isoform in ALS progression, what is certain is that this GPCR family has already proven to be a promising target for drug development against neuroinflammation and neurodegeneration.

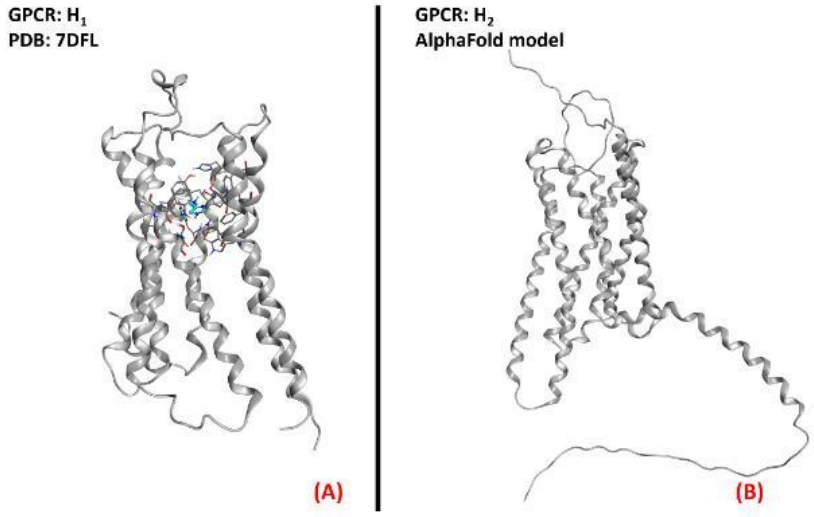


Figure 19. *Representation of the three-dimensional structures of the histamine receptors that could be considered for ALS treatment. (**A**) The H_1 receptor (sourced from the Protein Data Bank, PDB code: 7DFL [115], method: cryo-EM, resolution: 3.30 Å) and (**B**) the H_2 receptor (with no experimentally resolved structure available, the model sourced from the AlphaFold [116] database is presented). The images were created and rendered with MOE. The H3 and H4 receptors both lack experimentally resolved structures, but their AlphaFold models are available.*

2.9. Cannabinoid Receptors

A group of class A GPCRs that are of high interest at the present date is certainly the cannabinoid receptors. These biological entities are the main actors in the endocannabinoid system, and play

relevant roles in several physiological processes. Indeed, the first of its two main isoforms, called CB_1, is mainly located in both the central and peripheral nervous system, acting as a neurotransmitter release modulator in response to the binding of its agonists (mainly anandamide, but also 2-arachidonoylglycerol, both represented in Figure 20, panels A and B, respectively). In the majority of cases, CB_1 is coupled with $G_{i/o}$ protein, leading to adenylyl cyclase inhibition and consequent decrease in cAMP upon activation. The final effect of such an action is the reduction of neurotransmitter release in the synapse. On the other hand, the CB_2 receptor is mainly localized on the surface of the immune system cells. Its main agonist is 2-arachidonoylgycerol, binding of which leads to the inhibition of adenylyl cyclase through $G_{i/o}$ subunit action [117]. The final main effect is immunosuppression [118]. As reported by an article from Giacoppo and Mazzon, several studies have shown how the application of cannabinoid receptor agonists in $SOD1^{G93A}$ mouse models of ALS could be beneficial for the neuroprotective effects mediated by them [119]. Similarly, in 2019, Urbi et al. performed a meta-analysis on the studies regarding the application of cannabinoids in ALS murine models, highlighting the effective concordance in assessing that their application leads to a delay in disease progression [120]. A study by Shoemaker et al. highlighted that the increase in survival could be more addressable to the CB_2 isoform, showing that the administration of the CB_2 selective agonist AM-1241 (Figure 20, panel C) increased survival by 56% [121]. This molecule was also examined for cannabinoid-mediated ALS treatment by Kim et al., with similar results [122].

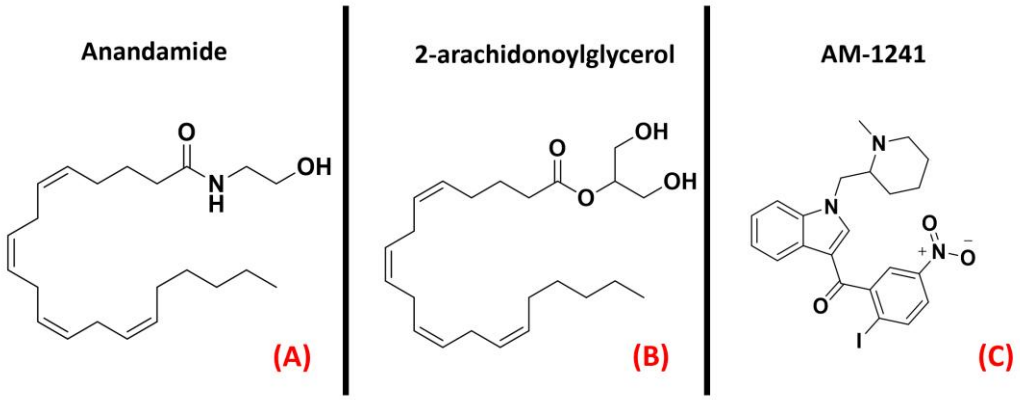


Figure 20. *The chemical structures of the endogenous CB_1 and CB_2 receptors agonists anandamide (**A**) and 2-arachidonoylglycerol (**B**). The selective CB_2 receptor agonist AM-1241 is also reported (**C**).*

In addition to this, Bilsland et al. reported that the knock-out of CB_1 receptors in $SOD1^{G39A}$ ALS-affected mice had no appreciable effect on disease onset [123], and regarding this, Shoemaker et al. reported that the activation of CB_1 could exacerbate disease progression [121]. The literature available today regarding the application of molecules acting on the endocannabinoid system for

ALS treatment converges in the possible evaluation of a therapy based on CB_2 selective agonists. A three-dimensional representation of both CB_1 and CB_2 receptors is provided in Figure 21.

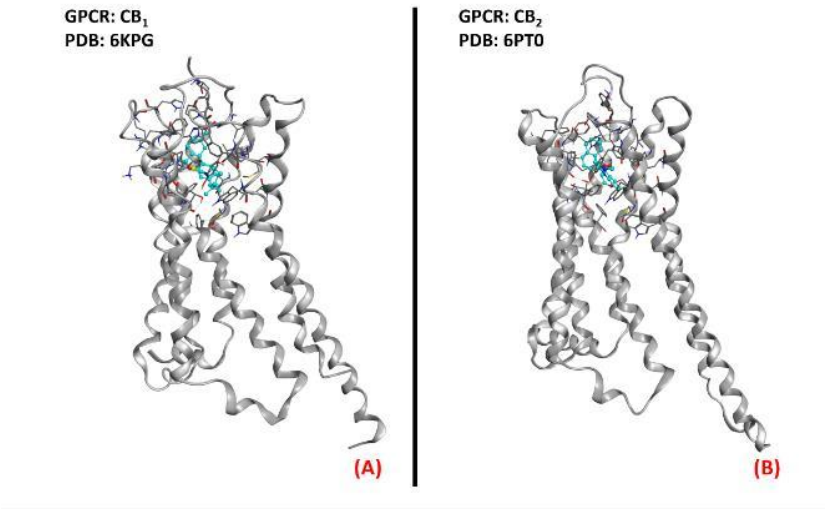


Figure 21. *The structures of (**A**) the CB_1 receptor (sourced from the Protein Data Bank, PDB code: 6KPG [124], method: cryo-EM, resolution: 3.00 Å) and (**B**) the CB_2 receptor (sourced from the Protein Data Bank, PDB code: 6PT0 [125], method: cryo-EM, resolution: 3.20 Å). The images were created and rendered with MOE.*

2.10. Prostaglandin E_2 Receptor (PGE2R)

Prostaglandin E_2 receptors (PGE_2) are a series of class A GPCRs that selectively bind to prostaglandin E_2 (also known as dinoprostone), an endogenous arachidonic acid derivative of high importance for several physiological functions. PGE_2 can be divided into four isoforms, named E_1, E_2, E_3, and E_4 (all represented in their three-dimensional structure in Figure 22). With the exception of the first, which stimulates phospholipase C if agonized, the other isoforms act on adenylyl cyclase and, specifically, the EP_2 and EP_4 isoforms (coupled with a G_s subunit) stimulate its function when agonized, while EP_3 inhibits AC through its action (being coupled to a $G_{i/o}$ subunit) [126]. EP_1 function has been correlated with hyperalgesia [127], immunoregulation [128], and colon cancer progression [129]. EP_2, which is active in the reproductive, visual, cardiovascular, skeletal, and nervous systems, has also been strictly related to tumor promotion, as highlighted in a 2018 study by Sun and Li [130]. Minor correlations with cancer have been reported for EP_3, which is also important for a large variety of functions, ranging from digestion [131] to blood pressure [132] and clotting [133], in addition to pain management [134]. The spectrum of systems in which the fourth isoform of PGE_2 receptors, EP_4, is involved is also very wide. Additionally, in this case, EP_4 has been reported to be hyper-

expressed in various types of cancer, mainly prostate cancer [135]. Talking about ALS onset and progression, Iłzecka found increased levels of PGE_2 in the cerebrospinal fluid of ALS patients, and therefore concluded that this mediator could play a role in disease progression, suggesting that its inhibition could be beneficial [136]. Additionally, Kosuge et al. highlighted the role of PGE_2 in the ROS generation pathway, focusing on its impact on ALS conditions. The same conclusion was reported in 2008 by Liang et al., who suggested EP_2 inhibition as a novel way to treat the neuroinflammation typical in ALS [137]. On the other hand, PGE_2 receptors were proposed to have an unexpected neuroprotective effect on motor neurons by Bilak et al., who reported that the neuroinflammatory process typical of ALS was mainly due to COX-2-mediated, prostaglandin-independent processes [138]. Taken together, all of these studies converge in evaluating PGE_2 receptors as interesting pharmacological targets for ALS, being strongly correlated with the significant neuroinflammation characterizing the pathology.

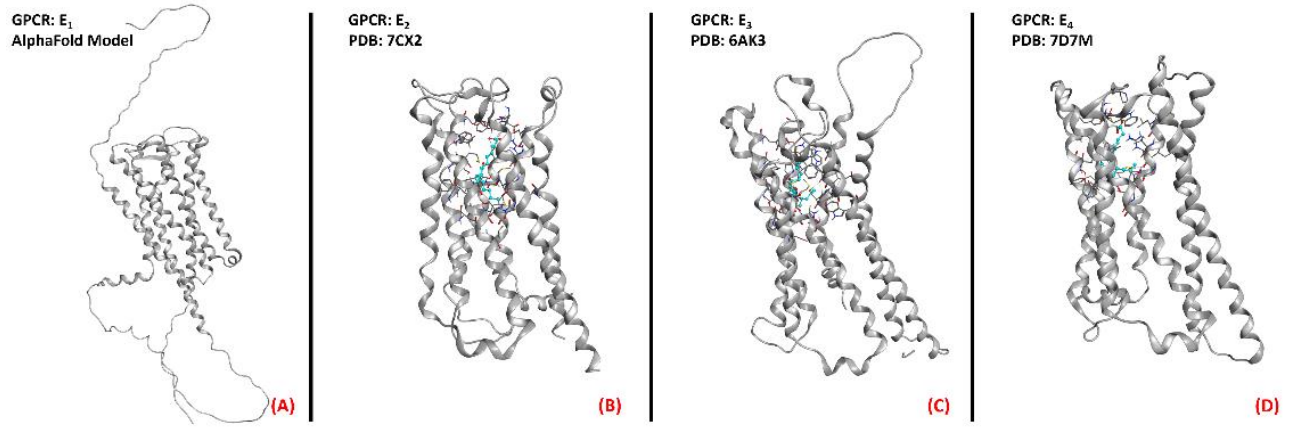


Figure 22. *Structure of the PGE_2 receptors (**A**) EP_1 (with no experimentally resolved structure available, the model sourced from the AlphaFold database is presented), (**B**) EP_2 (sourced from the Protein Data Bank, PDB code: 7CX2 [139], method: cryo-EM, resolution: 2.80 Å), (**C**) EP_3 (sourced from the Protein Data Bank, PDB code: 6AK3 [140], method: X-ray diffraction, resolution: 2.90 Å), and (**D**) EP_4 (sourced from the Protein Data Bank, PDB code: 7D7M [141], method: cryo-EM, resolution: 3.30 Å). All of the images were created and rendered with MOE.*

2.11. Vasoactive Intestinal Peptide Receptors

The receptors for the vasoactive intestinal peptide are part of the first subfamily of class B GPCRs. As part of this group of proteins, these receptors are responsive to signals mediated by the peptide hormone VIP (vasoactive intestinal polypeptide), formed by 28 amino acids and belonging to the glucagon/secretin superfamily [142]. After being produced by organs such as the gut, pancreas, and brain, VIP act in different physiological functions depending on the target tissue and the receptor isoform interacting with it. Indeed, two vasoactive intestinal polypeptide receptor isoforms are known, namely $VPAC_1$ and $VPAC_2$ (both represented in Figure 23). Both of these proteins are highly

expressed throughout the human body, from the smooth muscle of the GI tract and blood vessels to the reproductive system, lungs, spleen, and brain [143]. Their activity is mediated by a G_s protein, and involves the activation of adenylyl cyclase upon binding to VIP, consequently activating the protein kinase A (PKA) [144].

The implication of VIP in the CNS has been noticed when studying both the circadian rhythm and schizophrenia [143], but its importance as a potential target for ALS therapy is of more recent discovery. In 2008, Staines considered the possibility of studying vasoactive neuropeptides for degenerating pathologies such as MS and ALS [145], and a previous article by Iwasaki et al. specifically referred to the neurotrophic properties exerted by VIP in the degenerating diseases of motor neurons [146]. Solés-Tarrés et al. analyzed the neuroprotective effects of both VIP and PACAP (pituitary adenylate cyclase-activating polypeptide, a peptide hormone binding to both VPACs and PACAP receptors), also evaluating the synthetic derivatives available to mimic their action, with a special focus on the intrinsic pharmacokinetic problems of these species [147]. Waschek already identified both VIP and PACAP as promising targets for neuroinflammation in the CNS [148], as pointed out again in recent work by Martinez et al. [149].

The interaction with receptors of vasoactive peptides has been demonstrated to be a promising way to counteract the neuroinflammatory and degenerative effects of ALS, mainly through biological and/or chemical species mimicking the functions of the endogenous peptides. Further research on this topic will define the best way to accomplish this task.

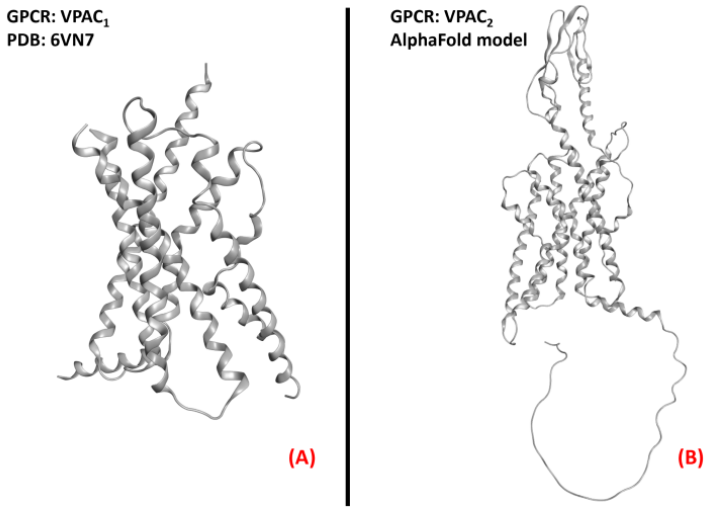


Figure 23. *Structure of the receptors (**A**) $VPAC_1$ (sourced from the Protein Data Bank, PDB code: 6VN7 [150], method: cryo-EM, resolution: 3.20 Å), and (**B**) $VPAC_2$ (with no experimentally resolved structure available, the model sourced from the AlphaFold database is presented). The images were created and rendered with MOE.*

2.12. Metabotropic Glutamate Receptors (mGluRs)

Metabotropic glutamate receptors (mGluRs) are GPCRs belonging to class C of this family of proteins. As the name suggests, these entities bind to the neurotransmitter glutamate, exerting different functions in both the central and peripheral nervous systems. They can be divided into three groups, with the first being composed of mGluR1 and mGluR5, predominantly postsynaptic, which, once activated, cause the stimulation of phospholipase C (PLC) through G_q-mediated signaling. Group II (formed by mGluRs 2 and 3) and III (of which mGluRs 4,6,7, and 8 are a part) receptors are mainly presynaptic and are all coupled with a $G_{i/0}$ subunit, which inhibits the activation of adenylyl cyclase, causing presynaptic inhibition. The functions of this family of proteins are majorly related to the nervous system, from the modulation of neurotransmission (e.g., gabaergic and dopaminergic) and of other proteins' signaling (e.g., NMDA receptors), to synaptic plasticity regulation [151]. This being said, it appears clear that the possibility of their involvement in ALS onset and progression is more than possible. Anneser et al. found a strong upregulation of mGluRs in the spinal cord with ALS, leading to the propagation of glial proliferation [152]. Hyperactivity of group I mGluRs has been correlated with neuroinflammation. Indeed, as demonstrated by Milanese et al., $SOD1^{G93A}$ ALS-affected mice with mGluR1 knockdown experience a reduction in microglia and astrocyte activation, decreasing mitochondrial damage and improving survival [153], and this phenomenon was also highlighted by Rossi et al. [154]

Anneser et al. showed the beneficial and protective effects of both agonism and antagonism of group I mGluRs for motor neuron disease, while less promising effects were derived from modulation of other mGluRs [155]. Crupi et al. recently pointed out that the beneficial therapeutic modulation of mGluRs is usually achieved through the reduction of the excitotoxicity drive via mGluR I inhibition or mGluR II and III agonism [156]. In conclusion, the literature supports the possibility of investing resources in the treatment of motor neuron diseases via mGluR modulation. A three-dimensional representation of mGluR1, mGluR2, and mGluR4 receptors is provided in Figure 24.

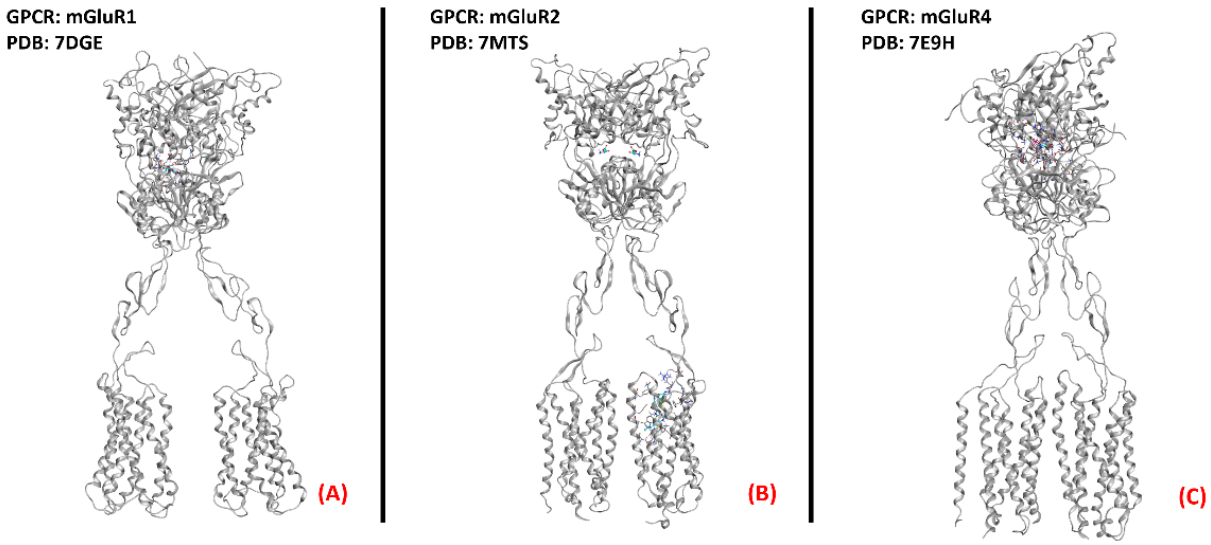


Figure 24. *One example of each group of the metabotropic glutamate receptors. (**A**) mGluR1, a member of the first group of mGluRs (sourced from the Protein Data Bank, PDB code: 7DGE [157], method: cryo-EM, resolution: 3.65 Å), (**B**) mGluR2 (owing to mGluRs group II, sourced from the Protein Data Bank, PDB code: 7MTS [158], method: cryo-EM, resolution: 3.20 Å), and (**C**) mGluR4, part of group III of the mGluRs (sourced from the Protein Data Bank, PDB code: 7E9H [159], method: cryo-EM, resolution: 4.00 Å). All of the images were created and rendered with MOE.*

3. Conclusions

In this review, we provided a panoramic view of the involvement of different G-protein-coupled receptors in the onset and progression of ALS, evaluating what has already been discovered on these biological entities, and highlighting what the next steps in research could be, always on the basis of the present literature on the topic. Our analysis shows that a GPCR-based therapy for ALS could be considered a practical possibility for the eradication of the ALS condition, and we encourage scientific groups all around the world in directing efforts towards this field.

References

[1]. J. Adams, M. Lee, and W. Peng, "Critical Review of Complementary and Alternative Medicine Use in Amyotrophic Lateral Sclerosis: Prevalence and Users' Profile, Decision-Making, Information Seeking, and Disclosure in the Face of a Lack of Efficacy," *Neurodegener. Dis.*, vol. 18, no. 4, pp. 225–232, 2018, doi: 10.1159/000492946..

[2]. Heckman, C.; Enoka, R.M. *Physiology of the Motor Neuron and the Motor Unit*; Levin, K., Chauvel, P., Eds.; Handbook of Clinical Neurophysiology; Elsevier: Amsterdam, The Netherlands, 2004; pp. 119–147. https://doi.org/10.1016/s1567-4231(04)04006-7.

[3]. Wijesekera, L.C.; Leigh, P.N. Amyotrophic lateral sclerosis. *Orphanet J. Rare Dis.* **2009**, *4*, 3. https://doi.org/10.1186/1750-1172-4-3.

[4]. Hardiman, O.; Al-Chalabi, A.; Chiò, A.; Corr, E.M.; Logroscino, G.; Robberecht, W.; Shaw, P.J.; Simmons, Z.; van den Berg, L.H. Amyotrophic lateral sclerosis. *Nat. Rev. Dis. Prim*. **2017**, *3*, 17071. https://doi.org/10.1016/S0140-6736(10)61156-7.

[5]. Ortí, J.D.L.R.; Armero, J.; Sanchis-Sanchis, C.; Sancho-Castillo, S.; Salazar, A.; Caplliure-Llopis, J.; Navarro-Illana, E.; Barrios, C.; Escribá-Alepuz, J.; Benlloch, M. Muscle Function Differences between Patients with Bulbar and Spinal Onset Amyotrophic Lateral Sclerosis. Does It Depend on Peripheral Glucose? *J. Clin. Med.* **2021**, *10*, 1582. https://doi.org/10.3390/jcm10081582.

[6]. van Es, M.A.; Hardiman, O.; Chio, A.; Al-Chalabi, A.; Pasterkamp, R.J.; Veldink, J.H.; van den Berg, L.H. Amyotrophic lateral sclerosis. *Lancet* **2017**, *390*, 2084–2098. https://doi.org/10.1016/s0140-6736(17)31287-4.

[7]. National Institute of Neurological Disorders and Stroke. Amyotrophic Lateral Sclerosis (ALS) Fact Sheet. Available online: www.ninds.nih.gov (accessed on 22nd March 2022).

[8]. Saitoh, Y.; Takahashi, Y. Riluzole for the treatment of amyotrophic lateral sclerosis. *Neurodegener. Dis. Manag.* **2020**, *10*, 343–355. https://doi.org/10.2217/nmt-2020-0033.

[9]. Breiner, A.; Zinman, L.; Bourque, P.R. Edaravone for amyotrophic lateral sclerosis: Barriers to access and lifeboat ethics. *Can. Med Assoc. J.* **2020**, *192*, E319–E320. https://doi.org/10.1503/cmaj.191236.

[10]. Jaiswal, M.K. Riluzole and edaravone: A tale of two amyotrophic lateral sclerosis drugs. *Med. Res. Rev.* **2018**, *39*, 733–748. https://doi.org/10.1002/med.21528.

[11]. Masrori, P.; van Damme, P. Amyotrophic lateral sclerosis: A clinical review. *Eur. J. Neurol.* **2020**, *27*, 1918–1929. https://doi.org/10.1111/ene.14393.

[12]. Boddy, S.L.; Giovannelli, I.; Sassani, M.; Cooper-Knock, J.; Snyder, M.P.; Segal, E.; Elinav, E.; Barker, L.A.; Shaw, P.J.; McDermott, C.J. The gut microbiome: A key player in the complexity of amyotrophic lateral sclerosis (ALS). *BMC Med.* **2021**, *19*, 13. https://doi.org/10.1186/s12916-020-01885-3.

[13]. Neumann, M.; Sampathu, D.M.; Kwong, L.K.; Truax, A.C.; Micsenyi, M.C.; Chou, T.T.; Bruce, J.; Schuck, T.; Grossman, M.; Clark, C.M.; et al. Ubiquitinated TDP-43 in Frontotemporal Lobar Degeneration and Amyotrophic Lateral Sclerosis. *Science* **2006**, *314*, 130–133. https://doi.org/10.1126/science.1134108.

[14]. Jo, M.; Lee, S.; Jeon, Y.-M.; Kim, S.; Kwon, Y.; Kim, H.-J. The role of TDP-43 propagation in neurodegenerative diseases: Integrating insights from clinical and experimental studies. *Exp. Mol. Med.* **2020**, *52*, 1652–1662. https://doi.org/10.1038/s12276-020-00513-7.

[15]. Suk, T.; Rousseaux, M.W.C. The role of TDP-43 mislocalization in amyotrophic lateral sclerosis. *Mol. Neurodegener.* **2020**, *15*, 45. https://doi.org/10.1186/s13024-020-00397-1.

[16]. Berdyński, M.; Miszta, P.; Safranow, K.; Andersen, P.M.; Morita, M.; Filipek, S.; Żekanowski, C.; Kuźma-Kozakiewicz, M. SOD1 mutations associated with amyotrophic lateral sclerosis analysis of variant severity. *Sci. Rep.* **2022**, *12*, 103. https://doi.org/10.1038/s41598-021-03891-8.

[17]. Ishigaki, S.; Sobue, G. Importance of Functional Loss of FUS in FTLD/ALS. *Front. Mol. Biosci.* **2018**, *5*, 44. https://doi.org/10.3389/fmolb.2018.00044.

[18]. Eck, R.J.; Kraemer, B.C.; Liachko, N.F. Regulation of TDP-43 phosphorylation in aging and disease. *GeroScience* **2021**, *43*, 1605–1614. https://doi.org/10.1007/s11357-021-00383-5.

[19]. Guo, W.; Vandoorne, T.; Steyaert, J.; Staats, K.A.; van den Bosch, L. The multifaceted role of kinases in amyotrophic lateral sclerosis: Genetic, pathological and therapeutic implications. *Brain* **2020**, *143*, 1651–1673. https://doi.org/10.1093/brain/awaa022.

[20]. Palomo, V.; Nozal, V.; Rojas-Prats, E.; Gil, C.; Martinez, A. Protein kinase inhibitors for amyotrophic lateral sclerosis therapy. *J. Cereb. Blood Flow Metab.* **2020**, *178*, 1316–1335. https://doi.org/10.1111/bph.15221.

[21]. Miller, R.G.; Mitchell, J.D.; Moore, D.H. Riluzole for amyotrophic lateral sclerosis (ALS)/motor neuron disease (MND). *Cochrane Database Syst. Rev.* **2012**, *65*, CD001447. https://doi.org/10.1002/14651858.cd001447.pub3.

[22]. Moro, S.; Bissaro, M. Rethinking to riluzole mechanism of action: The molecular link among protein kinase CK1δ activity, TDP-43 phosphorylation, and amyotrophic lateral sclerosis pharmacological treatment. *Neural Regen. Res.* **2019**, *14*, 2083–2085. https://doi.org/10.4103/1673-5374.262578.

[23]. Corcia, P.; Beltran, S.; Bakkouche, S.; Couratier, P. Therapeutic news in ALS. *Rev. Neurol.* **2021**, *177*, 544–549. https://doi.org/10.1016/j.neurol.2020.12.003.

[24]. Springer Nature, "AdisInsight—Small-molecules in clinical trials for ALS." https://adisinsight.springer.com/search.

[25]. Chen, H.; Kankel, M.W.; Su, S.C.; Han, S.W.S.; Ofengeim, D. Exploring the genetics and non-cell autonomous mechanisms underlying ALS/FTLD. *Cell Death Differ.* **2018**, *25*, 648–662. https://doi.org/10.1038/s41418-018-0060-4.

[26]. Insel, P.A.; Wilderman, A.; Zambon, A.C.; Snead, A.N.; Murray, F.; Aroonsakool, N.; McDonald, D.S.; Zhou, S.; McCann, T.; Zhang, L.; et al. G Protein–Coupled Receptor (GPCR) Expression in Native Cells: "Novel" endoGPCRs as Physiologic Regulators and Therapeutic Targets. *Mol. Pharmacol.* **2015**, *88*, 181–187. https://doi.org/10.1124/mol.115.098129.

[27]. Gacasan, S.B.; Baker, D.L.; Parrill, A.L. G protein-coupled receptors: The evolution of structural insight. *AIMS Biophys.* **2017**, *4*, 491–527. https://doi.org/10.3934/biophy.2017.3.491.

[28]. Lee, Y.; Basith, S.; Choi, S. Recent Advances in Structure-Based Drug Design Targeting Class A G Protein-Coupled Receptors Utilizing Crystal Structures and Computational Simulations. *J. Med. Chem.* **2017**, *61*, 1–46. https://doi.org/10.1021/acs.jmedchem.6b01453.

[29]. Heng, B.C.; Aubel, D.; Fussenegger, M. An overview of the diverse roles of G-protein coupled receptors (GPCRs) in the pathophysiology of various human diseases. *Biotechnol. Adv.* **2013**, *31*, 1676–1694. https://doi.org/10.1016/j.biotechadv.2013.08.017.

[30]. Zarrinmayeh, H.; Territo, P.R. Purinergic Receptors of the Central Nervous System: Biology, PET Ligands, and Their Applications. *Mol. Imaging* **2020**, *19*. https://doi.org/10.1177/1536012120927609.

[31]. Borea, P.A.; Gessi, S.; Merighi, S.; Vincenzi, F.; Varani, K. Pharmacology of Adenosine Receptors: The State of the Art. *Physiol. Rev.* **2018**, *98*, 1591–1625. https://doi.org/10.1152/physrev.00049.2017.

[32]. Gomes, C.; Kaster, M.; Tome, A.; Agostinho, P.; Cunha, R.A. Adenosine receptors and brain diseases: Neuroprotection and neurodegeneration. *Biochim. Biophys. Acta (BBA)-Biomembr.* **2011**, *1808*, 1380–1399. https://doi.org/10.1016/j.bbamem.2010.12.001.

[33]. Vincenzi, F.; Corciulo, C.; Targa, M.; Casetta, I.; Gentile, M.; Granieri, E.; Borea, P.A.; Popoli, P.; Varani, K. A2Aadenosine receptors are up-regulated in lymphocytes from amyotrophic lateral sclerosis patients. *Amyotroph. Lateral Scler. Front. Degener.* **2013**, *14*, 406–413. https://doi.org/10.3109/21678421.2013.793358.

[34]. Yoshida, Y.; Une, F.; Utatsu, Y.; Nomoto, M.; Furukawa, Y.; Maruyama, Y.; Machigashira, N.; Matsuzaki, T.; Osame, M. Adenosine and Neopterin Levels in Cerebrospinal Fluid of Patients with Neurological Disorders.. *Intern. Med.* **1999**, *38*, 133–139. https://doi.org/10.2169/internalmedicine.38.133.

[35]. Potenza, R.L.; Armida, M.; Ferrante, A.; Pèzzola, A.; Matteucci, A.; Puopolo, M.; Popoli, P. Effects of chronic caffeine intake in a mouse model of amyotrophic lateral sclerosis. *J. Neurosci. Res.* **2013**, *91*, 585–592. https://doi.org/10.1002/jnr.23185.

[36]. Mori, A.; Cross, B.; Uchida, S.; Walker, J.K.; Ristuccia, R. How Are Adenosine and Adenosine A_{2A} Receptors Involved in the Pathophysiology of Amyotrophic Lateral Sclerosis?. *Biomedicines* **2021**, *9*, 1027. https://doi.org/10.3390/biomedicines9081027.

[37]. Ng, S.K.; Higashimori, H.; Tolman, M.; Yang, Y. Suppression of adenosine 2a receptor (A2aR)-mediated adenosine signaling improves disease phenotypes in a mouse model of amyotrophic lateral sclerosis. *Exp. Neurol.* **2015**, *267*, 115–122. https://doi.org/10.1016/j.expneurol.2015.03.004.

[38]. Mojsilovic-Petrovic, J.; Jeong, G.-B.; Crocker, A.; Arneja, A.; David, S.; Russell, D.; Kalb, R.G. Protecting Motor Neurons from Toxic Insult by Antagonism of Adenosine A2a and Trk Receptors. *J. Neurosci.* **2006**, *26*, 9250–9263. https://doi.org/10.1523/jneurosci.1856-06.2006.

[39]. Liu, Y.-J.; Ju, T.; Chen, H.-M.; Jang, Y.-S.; Lee, L.-M.; Lai, H.-L.; Tai, H.-C.; Fang, J.-M.; Lin, Y.-L.; Tu, P.-H.; et al. Activation of AMP-activated protein kinase α1 mediates mislocalization of TDP-43 in amyotrophic lateral sclerosis. *Hum. Mol. Genet.* **2014**, *24*, 787–801. https://doi.org/10.1093/hmg/ddu497.

[40]. Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242. https://doi.org/10.1093/nar/28.1.235.

[41]. Segala, E.; Guo, D.; Cheng, R.K.Y.; Bortolato, A.; Deflorian, F.; Doré, A.S.; Errey, J.C.; Heitman, L.H.; Ijzerman, A.P.; Marshall, F.H.; et al. Controlling the Dissociation of Ligands from the Adenosine A_{2A} Receptor through Modulation of Salt Bridge Strength. *J. Med. Chem.* **2016**, *59*, 6470–6479. https://doi.org/10.1021/acs.jmedchem.6b00653.

[42]. Chemical Computing Group ULC. *Molecular Operating Environment (MOE)*; Chemical Computing Group ULC: Montreal, QC, Canada, 2021.

[43]. Burnstock, G. Purinergic Signalling: Therapeutic Developments. *Front. Pharmacol.* **2017**, *8*, 661. https://doi.org/10.3389/fphar.2017.00661.

[44]. Morillas, A.G.; Besson, V.; Lerouet, D. Microglia and Neuroinflammation: What Place for P2RY12?. *Int. J. Mol. Sci.* **2021**, *22*, 1636. https://doi.org/10.3390/ijms22041636.

[45]. Amadio, S.; Parisi, C.; Montilli, C.; Carrubba, A.S.; Apolloni, S.; Volonté, C. $P2Y_{12}$Receptor on the Verge of a Neuroinflammatory Breakdown. *Mediat. Inflamm.* **2014**, *2014*, 975849. https://doi.org/10.1155/2014/975849.

[46]. Jacobson, K.A.; Delicado, E.G.; Gachet, C.; Kennedy, C.; Von Kügelgen, I.; Li, B.; Miras-Portugal, M.T.; Novak, I.; Schöneberg, T.; Perez-Sen, R.; et al. Update of P2Y receptor pharmacology: IUPHAR Review 27. *J. Cereb. Blood Flow Metab.* **2020**, *177*, 2413–2433. https://doi.org/10.1111/bph.15005.

[47]. D'Ambrosi, N.; Finocchi, P.; Apolloni, S.; Cozzolino, M.; Ferri, A.; Padovano, V.; Pietrini, G.; Carrì, M.T.; Volonté, C. The Proinflammatory Action of Microglial P2 Receptors Is Enhanced in SOD1 Models for Amyotrophic Lateral Sclerosis. *J. Immunol.* **2009**, *183*, 4648–4656. https://doi.org/10.4049/jimmunol.0901212.

[48]. Kobayashi, K.; Yamanaka, H.; Fukuoka, T.; Dai, Y.; Obata, K.; Noguchi, K. P2Y12 Receptor Upregulation in Activated Microglia Is a Gateway of p38 Signaling and Neuropathic Pain. *J. Neurosci.* **2008**, *28*, 2892–2902. https://doi.org/10.1523/jneurosci.5589-07.2008.

[49]. Moore, C.S.; Ase, A.R.; Kinsara, A.; Rao, V.T.; Michell-Robinson, M.; Leong, S.Y.; Butovsky, O.; Ludwin, S.K.; Séguéla, P.; Bar-Or, A.; et al. P2Y12 expression and function in alternatively activated human microglia. *Neurol.-Neuroimmunol. Neuroinflammation* **2015**, *2*, e80. https://doi.org/10.1212/nxi.0000000000000080.

[50]. Zhang, K.; Zhang, J.; Gao, Z.-G.; Zhang, D.; Zhu, L.; Han, G.W.; Moss, S.M.; Paoletta, S.; Kiselev, E.; Lu, W.; et al. Structure of the human P2Y12 receptor in complex with an antithrombotic drug. *Nature* **2014**, *509*, 115–118. https://doi.org/10.1038/nature13083.

[51]. Hughes, C.E.; Nibbs, R.J.B. A guide to chemokines and their receptors. *FEBS J.* **2018**, *285*, 2944–2971. https://doi.org/10.1111/febs.14466.

[52]. Agle, K.A.; Vongsa, R.A.; Dwinell, M.B. Calcium Mobilization Triggered by the Chemokine CXCL12 Regulates Migration in Wounded Intestinal Epithelial Monolayers. *J. Biol. Chem.* **2010**, *285*, 16066–16075. https://doi.org/10.1074/jbc.m109.061416.

[53]. La Cognata, V.; Golini, E.; Iemmolo, R.; Balletta, S.; Morello, G.; De Rosa, C.; Villari, A.; Marinelli, S.; Vacca, V.; Bonaventura, G.; et al. CXCR2 increases in ALS cortical neurons and its inhibition prevents motor neuron degeneration in vitro and improves neuromuscular function in SOD1G93A mice. *Neurobiol. Dis.* **2021**, *160*, 105538. https://doi.org/10.1016/j.nbd.2021.105538.

[54]. Rabinovich-Nikitin, I.; Ezra, A.; Barbiro, B.; Rabinovich-Toidman, P.; Solomon, B. Chronic administration of AMD3100 increases survival and alleviates pathology in SOD1G93A mice model of ALS. *J. Neuroinflammation* **2016**, *13*, 123–123. https://doi.org/10.1186/s12974-016-0587-6.

[55]. Liu, E.; Karpf, L.; Bohl, D. Neuroinflammation in Amyotrophic Lateral Sclerosis and Frontotemporal Dementia and the Interest of Induced Pluripotent Stem Cells to Study Immune Cells Interactions With Neurons. *Front. Mol. Neurosci.* **2021**, *14*. https://doi.org/10.3389/fnmol.2021.767041.

[56]. Perner, C.; Perner, F.; Stubendorff, B.; Förster, M.; Witte, O.W.; Heidel, F.H.; Prell, T.; Grosskreutz, J. Dysregulation of chemokine receptor expression and function in leukocytes from ALS patients. *J. Neuroinflammation* **2018**, *15*, 99. https://doi.org/10.1186/s12974-018-1135-3.

[57]. Liu, K.; Wu, L.; Yuan, S.; Wu, M.; Xu, Y.; Sun, Q.; Li, S.; Zhao, S.; Hua, T.; Liu, Z.-J. Structural basis of CXC chemokine receptor 2 activation and signalling. *Nature* **2020**, *585*, 135–140. https://doi.org/10.1038/s41586-020-2492-5.

[58]. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589. https://doi.org/10.1038/s41586-021-03819-2.

[59]. Wu, B.; Chien, E.Y.T.; Mol, C.D.; Fenalti, G.; Liu, W.; Katritch, V.; Abagyan, R.; Brooun, A.; Wells, P.; Bi, F.C.; et al. Structures of the CXCR4 Chemokine GPCR with Small-Molecule and Cyclic Peptide Antagonists. *Science* **2010**, *330*, 1066–1071. https://doi.org/10.1126/science.1194396.

[60]. Singh, K.D.; Karnik, S.S. Angiotensin Receptors: Structure, Function, Signaling and Clinical Applications. *J. Cell Signal.* **2016**, *1*. https://doi.org/10.4172/jcs.1000111.

[61]. Ames, M.K.; Atkins, C.E.; Pitt, B. The renin-angiotensin-aldosterone system and its suppression. *J. Vet. Intern. Med.* **2018**, *33*, 363–382. https://doi.org/10.1111/jvim.15454.

[62]. Miura, S.-I.; Karnik, S.S.; Saku, K. Review: Angiotensin II type 1 receptor blockers: Class effects versus molecular effects. *J. Renin-Angiotensin-Aldosterone Syst.* **2010**, *12*, 1–7. https://doi.org/10.1177/1470320310370852.

[63]. Akishita, M.; Ito, M.; Lehtonen, J.Y.; Daviet, L.; Dzau, V.J.; Horiuchi, M. Expression of the AT2 receptor developmentally programs extracellular signal-regulated kinase activity and influences fetal vascular growth. *J. Clin. Investig.* **1999**, *103*, 63–71. https://doi.org/10.1172/jci5182.

[64]. Kawajiri, M.; Mogi, M.; Higaki, N.; Tateishi, T.; Ohyagi, Y.; Horiuchi, M.; Miki, T.; Kira, J.-I. Reduced angiotensin II levels in the cerebrospinal fluid of patients with amyotrophic lateral sclerosis. *Acta Neurol. Scand.* **2009**, *119*, 341–344. https://doi.org/10.1111/j.1600-0404.2008.01099.x.

[65]. Iwasaki, Y.; Ichikawa, Y.; Igarashi, O.; Kinoshita, M.; Ikeda, K. Trophic effect of Olmesartan, a novel AT1R antagonist, on spinal motor neuronsin vitroandin vivo. *Neurol. Res.* **2002**, *24*, 468–472. https://doi.org/10.1179/016164102101200357.

[66]. Mammana, S.; Fagone, P.; Cavalli, E.; Basile, M.S.; Petralia, M.C.; Nicoletti, F.; Bramanti, P.; Mazzon, E. The Role of Macrophages in Neuroinflammatory and Neurodegenerative Pathways of Alzheimer's Disease, Amyotrophic Lateral Sclerosis, and Multiple Sclerosis: Pathogenetic Cellular Effectors and Potential Therapeutic Targets. *Int. J. Mol. Sci.* **2018**, *19*, 831. https://doi.org/10.3390/ijms19030831.

[67]. Zhang, H.; Unal, H.; Desnoyer, R.; Han, G.W.; Patel, N.; Katritch, V.; Karnik, S.S.; Cherezov, V.; Stevens, R.C. Structural Basis for Ligand Recognition and Functional Selectivity at Angiotensin Receptor. *J. Biol. Chem.* **2015**, *290*, 29127–29139. https://doi.org/10.1074/jbc.m115.689000.

[68]. Perryman, R. Inhibition of the angiotensin II type 2 receptor AT2R is a novel therapeutic strategy for glioblastoma. Unpublished work, 2022.

[69]. Beaulieu, J.-M.; Gainetdinov, R.R. The Physiology, Signaling, and Pharmacology of Dopamine Receptors. *Pharmacol. Rev.* **2011**, *63*, 182–217. https://doi.org/10.1124/pr.110.002642.

[70]. Vallone, D.; Picetti, R.; Borrelli, E. Structure and function of dopamine receptors. *Neurosci. Biobehav. Rev.* **2000**, *24*, 125–132. https://doi.org/10.1016/s0149-7634(99)00063-9.

[71]. Borasio, G.D.; Linke, R.; Schwarz, J.; Schlamp, V.; Abel, A.; Mozley, P.D.; Tatsch, K. Dopaminergic deficit in amyotrophic lateral sclerosis assessed with [I-123] IPT single photon emission computed tomography. *J. Neurol. Neurosurg. Psychiatry* **1998**, *65*, 263–265. https://doi.org/10.1136/jnnp.65.2.263.

[72]. Lai, C.-Y.; Liu, Y.-J.; Lai, H.-L.; Chen, H.-M.; Kuo, H.-C.; Liao, Y.-P.; Chern, Y. The D2 Dopamine Receptor Interferes With the Protective Effect of the A2A Adenosine Receptor on TDP-43 Mislocalization in Experimental Models of Motor Neuron Degeneration. *Front. Neurosci.* **2018**, *12*, 187. https://doi.org/10.3389/fnins.2018.00187.

[73]. Huang, X.; Roet, K.C.; Zhang, L.; Brault, A.; Berg, A.P.; Jefferson, A.B.; Klug-McLeod, J.; Leach, K.L.; Vincent, F.; Yang, H.; et al. Human amyotrophic lateral sclerosis excitability phenotype screen: Target discovery and validation. *Cell Rep.* **2021**, *35*, 109224–109224. https://doi.org/10.1016/j.celrep.2021.109224.

[74]. Fujimori, K.; Ishikawa, M.; Otomo, A.; Atsuta, N.; Nakamura, R.; Akiyama, T.; Hadano, S.; Aoki, M.; Saya, H.; Sobue, G.; et al. Modeling sporadic ALS in iPSC-derived motor neurons identifies a potential therapeutic agent. *Nat. Med.* **2018**, *24*, 1579–1589. https://doi.org/10.1038/s41591-018-0140-5.

[75]. Dooley, M.; Markham, A. Pramipexole. *Drugs Aging* **1998**, *12*, 495–514. https://doi.org/10.2165/00002512-199812060-00007.

[76]. Gribkoff, V.K.; Bozik, M.E. KNS-760704 [(6R)-4,5,6,7-tetrahydro-N6-propyl-2, 6-benzothiazole-diamine dihydrochloride monohydrate] for the Treatment of Amyotrophic Lateral Sclerosis. *CNS Neurosci. Ther.* **2008**, *14*, 215–226. https://doi.org/10.1111/j.1755-5949.2008.00048.x.

[77]. Kingwell, K. Dexpramipexole shows promise for ALS in phase II trial. *Nat. Rev. Neurol.* **2011**, *8*, 4. https://doi.org/10.1038/nrneurol.2011.204.

[78]. Cudkowicz, M.E.; Berg, L.H.V.D.; Shefner, J.M.; Mitsumoto, H.; Mora, J.S.; Ludolph, A.; Hardiman, O.; E Bozik, M.; Ingersoll, E.W.; Archibald, D.; et al. Dexpramipexole versus placebo for patients with amyotrophic lateral sclerosis (EMPOWER): A randomised, double-blind, phase 3 trial. *Lancet Neurol.* **2013**, *12*, 1059–1067. https://doi.org/10.1016/s1474-4422(13)70221-7.

[79]. Zhuang, Y.; Xu, P.; Mao, C.; Wang, L.; Krumm, B.; Zhou, X.E.; Huang, S.; Liu, H.; Cheng, X.; Huang, X.-P.; et al. Structural insights into the human D1 and D2 dopamine receptor signaling complexes. *Cell* **2021**, *184*, 931–942.e18. https://doi.org/10.1016/j.cell.2021.01.027.

[80]. Xu, P.; Huang, S.; Mao, C.; Krumm, B.E.; Zhou, X.E.; Tan, Y.; Huang, X.-P.; Liu, Y.; Shen, D.-D.; Jiang, Y.; et al. Structures of the human dopamine D3 receptor-Gi complexes. *Mol. Cell* **2021**, *81*, 1147–1159.e4. https://doi.org/10.1016/j.molcel.2021.01.003.

[81]. Wang, S.; Wacker, D.; Levit, A.; Che, T.; Betz, R.M.; McCorvy, J.D.; Venkatakrishnan, A.J.; Huang, X.-P.; Dror, R.O.; Shoichet, B.K.; et al. D $_4$ dopamine receptor high-resolution structures enable the discovery of selective agonists. *Science* **2017**, *358*, 381–386. https://doi.org/10.1126/science.aan5468.

[82]. Nichols, D.E.; Nichols, C.D. Serotonin Receptors. *Chem. Rev.* **2008**, *108*, 1614–1641. https://doi.org/10.1021/cr078224o.

[83]. Frazer, A.; Hensler, J.G. Serotonin receptors. In *Basic Neurochemistry: Molecular, Cellular and Medical Aspects*, 6th ed.; Siegel, G.J., Agranoff, B.W., Albers, R.W. Eds.; Lippincott-Raven: Philadelphia, PA, USA, 1999.

[84]. El Oussini, H.; Bayer, H.; Scekic-Zahirovic, J.; Vercruysse, P.; Sinniger, J.; Dirrig-Grosch, S.; Dieterlé, S.; Echaniz-Laguna, A.; Larmet, Y.; Müller, K.; et al. Serotonin 2B receptor slows disease progression and prevents degeneration of spinal cord mononuclear phagocytes in amyotrophic lateral sclerosis. *Acta Neuropathol.* **2016**, *131*, 465–480. https://doi.org/10.1007/s00401-016-1534-4.

[85]. Dentel, C.; Palamiuc, L.; Henriques, A.; Lannes, B.; Spreux-Varoquaux, O.; Gutknecht, L.; René, F.; Echaniz-Laguna, A.; De Aguilar, J.-L.G.; Lesch, K.P.; et al. Degeneration of serotonergic neurons in amyotrophic lateral sclerosis: A link to spasticity. *Brain* **2013**, *136*, 483–493. https://doi.org/10.1093/brain/aws274.

[86]. Arnoux, A.; Ayme-Dietrich, E.; Dieterle, S.; Goy, M.-A.; Schann, S.; Frauli, M.; Monassier, L.; Dupuis, L. Evaluation of a 5-HT2B receptor agonist in a murine model of amyotrophic lateral sclerosis. *Sci. Rep.* **2021**, *11*, 23582. https://doi.org/10.1038/s41598-021-02900-0.

[87]. Elangbam, C.S.; Job, L.E.; Zadrozny, L.M.; Barton, J.C.; Yoon, L.W.; Gates, L.D.; Slocum, N. 5-Hydroxytryptamine (5HT)-induced valvulopathy: Compositional valvular alterations are associated with 5HT2B receptor and 5HT transporter transcript changes in Sprague-Dawley rats. *Exp. Toxicol. Pathol.* **2008**, *60*, 253–262. https://doi.org/10.1016/j.etp.2008.03.005.

[88]. Lacomblez, L.; Bensimon, G.; Douillet, P.; Doppler, V.; Salachas, F.; Meininger, V. Xaliproden in amyotrophic lateral sclerosis: Early clinical trials. *Amyotroph. Lateral Scler.* **2004**, *5*, 99–106. https://doi.org/10.1080/14660820410018973.

[89]. Xu, P.; Huang, S.; Zhang, H.; Mao, C.; Zhou, X.E.; Cheng, X.; Simon, I.A.; Shen, D.-D.; Yen, H.-Y.; Robinson, C.V.; et al. Structural insights into the lipid and ligand regulation of serotonin receptors. *Nature* **2021**, *592*, 469–473. https://doi.org/10.1038/s41586-021-03376-8.

[90]. McCorvy, J.D.; Wacker, D.; Wang, S.; Agegnehu, B.; Liu, J.; Lansu, K.; Tribo, A.R.; Olsen, R.H.J.; Che, T.; Jin, J.; et al. Structural determinants of 5-HT2B receptor activation and biased agonism. *Nat. Struct. Mol. Biol.* **2018**, *25*, 787–796. https://doi.org/10.1038/s41594-018-0116-7.

[91]. Peng, Y.; McCorvy, J.D.; Harpsøe, K.; Lansu, K.; Yuan, S.; Popov, P.; Qu, L.; Pu, M.; Che, T.; Nikolajsen, L.F.; et al. 5-HT2C Receptor Structures Reveal the Structural Basis of GPCR Polypharmacology. *Cell* **2018**, *172*, 719–730.e14. https://doi.org/10.1016/j.cell.2018.01.001.

[92]. Marucci, G.; Ben, D.D.; Lambertucci, C.; Navia, A.M.; Spinaci, A.; Volpini, R.; Buccioni, M. GPR17 receptor modulators and their therapeutic implications: Review of recent patents. *Expert Opin. Ther. Patents* **2019**, *29*, 85–95. https://doi.org/10.1080/13543776.2019.1568990.

[93]. Dziedzic, A.; Miller, E.; Saluk-Bijak, J.; Bijak, M. The GPR17 Receptor—A Promising Goal for Therapy and a Potential Marker of the Neurodegenerative Process in Multiple Sclerosis. *Int. J. Mol. Sci.* **2020**, *21*, 1852. https://doi.org/10.3390/ijms21051852.

[94]. Marucci, G.; Ben, D.D.; Lambertucci, C.; Santinelli, C.; Spinaci, A.; Thomas, A.; Volpini, R.; Buccioni, M. The G Protein-Coupled Receptor GPR17: Overview and Update. *ChemMedChem* **2016**, *11*, 2567–2574. https://doi.org/10.1002/cmdc.201600453.

[95]. Lecca, D.; Raffaele, S.; Abbracchio, M.P.; Fumagalli, M. Regulation and signaling of the GPR17 receptor in oligodendroglial cells. *Glia* **2020**, *68*, 1957–1967. https://doi.org/10.1002/glia.23807.

[96]. Zhao, B.; Zhao, C.; Zhang, X.; Huang, X.; Shi, W.; Fang, S.; Lu, Y.; Zhang, W.; Xia, Q.; Wei, E. The new P2Y-like receptor G protein-coupled receptor 17 mediates acute neuronal injury and late microgliosis after focal cerebral ischemia in rats. *Neuroscience* **2011**, *202*, 42–57. https://doi.org/10.1016/j.neuroscience.2011.11.066.

[97]. Bonfanti, E.; Bonifacino, T.; Raffaele, S.; Milanese, M.; Morgante, E.; Bonanno, G.; Abbracchio, M.P.; Fumagalli, M. Abnormal Upregulation of GPR17 Receptor Contributes to Oligodendrocyte Dysfunction in SOD1 G93A Mice. *Int. J. Mol. Sci.* **2020**, *21*, 2395. https://doi.org/10.3390/ijms21072395.

[98]. Merten, N.; Fischer, J.; Simon, K.; Zhang, L.; Schröder, R.; Peters, L.; Letombe, A.-G.; Hennen, S.; Schrage, R.; Bödefeld, T.; et al. Repurposing HAMI3379 to Block GPR17 and Promote Rodent and Human Oligodendrocyte Differentiation. *Cell Chem. Biol.* **2018**, *25*, 775–786.e5. https://doi.org/10.1016/j.chembiol.2018.03.012.

[99]. Raffaele, S.; Boccazzi, M.; Fumagalli, M. Oligodendrocyte Dysfunction in Amyotrophic Lateral Sclerosis: Mechanisms and Therapeutic Perspectives. *Cells* **2021**, *10*, 565. https://doi.org/10.3390/cells10030565.

[100]. Jin, S.; Wang, X.; Xiang, X.; Wu, Y.; Hu, J.; Li, Y.; Dong, Y.L.; Tan, Y.; Wu, X. Inhibition of GPR17 with cangrelor improves cognitive impairment and synaptic deficits induced by Aβ1–42 through Nrf2/HO-1 and NF-κB signaling pathway in mice. *Int. Immunopharmacol.* **2021**, *101*, 108335. https://doi.org/10.1016/j.intimp.2021.108335.

[101]. Marschallinger, J.; Schäffner, I.; Klein, B.; Gelfert, R.; Rivera, F.J.; Illes, S.; Grassner, L.; Janssen, M.; Rotheneichner, P.; Schmuckermair, C.; et al. Structural and functional rejuvenation of the aged brain by an approved anti-asthmatic drug. *Nat. Commun.* **2015**, *6*, 8466. https://doi.org/10.1038/ncomms9466.

[102]. Burnstock, G. An introduction to the roles of purinergic signalling in neurodegeneration, neuroprotection and neuroregeneration. *Neuropharmacology* **2016**, *104*, 4–17. https://doi.org/10.1016/j.neuropharm.2015.05.031.

[103]. Insel, P.A. Adrenergic Receptors—Evolving Concepts and Clinical Implications. *N. Engl. J. Med.* **1996**, *334*, 580–585. https://doi.org/10.1056/nejm199602293340907.

[104]. Subbarao, K.V.; Hertz, L. Effect of adrenergic agonists on glycogenolysis in primary cultures of astrocytes. *Brain Res.* **1990**, *536*, 220–226. https://doi.org/10.1016/0006-8993(90)90028-a.

[105]. Patel, M.; Shaw, D. A review of standard pharmacological therapy for adult asthma—Steps 1 to 5. *Chronic Respir. Dis.* **2015**, *12*, 165–176. https://doi.org/10.1177/1479972315573529.

[106]. Bartus, R.T.; Bétourné, A.; Basile, A.; Peterson, B.L.; Glass, J.; Boulis, N.M. β 2 -Adrenoceptor agonists as novel, safe and potentially effective therapies for Amyotrophic lateral sclerosis (ALS). *Neurobiol. Dis.* **2016**, *85*, 11–24. https://doi.org/10.1016/j.nbd.2015.10.006.

[107]. Teng, Y.D.; Choi, H.; Huang, W.; Onario, R.C.; Frontera, W.R.; Snyder, E.Y.; Sabharwal, S. Therapeutic effects of clenbuterol in a murine model of amyotrophic lateral sclerosis. *Neurosci. Lett.* **2006**, *397*, 155–158. https://doi.org/10.1016/j.neulet.2005.12.007.

[108]. Yang, F.; Ling, S.; Zhou, Y.; Zhang, Y.; Lv, P.; Liu, S.; Fang, W.; Sun, W.; A Hu, L.; Zhang, L.; et al. Different conformational responses of the β2-adrenergic receptor-Gs complex upon binding of the partial agonist salbutamol or the full agonist isoprenaline. *Natl. Sci. Rev.* **2020**, *8*, nwaa284. https://doi.org/10.1093/nsr/nwaa284.

[109]. Parsons, M.E.; Ganellin, C.R. Histamine and its receptors. *Br. J. Pharmacol.* **2006**, *147* (Suppl. 1), S127–S135. https://doi.org/10.1038/sj.bjp.0706440.

[110]. Tiligada, E.; Ennis, M. Histamine pharmacology: From Sir Henry Dale to the 21st century. *J. Cereb. Blood Flow Metab.* **2018**, *177*, 469–489. https://doi.org/10.1111/bph.14524.

[111]. Apolloni, S.; Fabbrizio, P.; Amadio, S.; Napoli, G.; Verdile, V.; Morello, G.; Iemmolo, R.; Aronica, E.; Cavallaro, S.; Volonté, C. Histamine Regulates the Inflammatory Profile of SOD1-G93A Microglia and the Histaminergic System Is Dysregulated in Amyotrophic Lateral Sclerosis. *Front. Immunol.* **2017**, *8*, 1689–1689. https://doi.org/10.3389/fimmu.2017.01689.

[112]. Volonté, C.; Apolloni, S.; Sabatelli, M. Histamine beyond its effects on allergy: Potential therapeutic benefits for the treatment of Amyotrophic Lateral Sclerosis (ALS). *Pharmacol. Ther.* **2019**, *202*, 120–131. https://doi.org/10.1016/j.pharmthera.2019.06.006.

[113]. Zhang, W.; Zhang, X.; Zhang, Y.; Qu, C.; Zhou, X.; Zhang, S. Histamine Induces Microglia Activation and the Release of Proinflammatory Mediators in Rat Brain Via H1R or H4R. *J. Neuroimmune Pharmacol.* **2019**, *15*, 280–291. https://doi.org/10.1007/s11481-019-09887-6.

[114]. Apolloni, S.; Fabbrizio, P.; Parisi, C.; Amadio, S.; Volonté, C. Clemastine Confers Neuroprotection and Induces an Anti-Inflammatory Phenotype in SOD1G93A Mouse Model of Amyotrophic Lateral Sclerosis. *Mol. Neurobiol.* **2014**, *53*, 518–531. https://doi.org/10.1007/s12035-014-9019-8.

[115]. Xia, R.; Wang, N.; Xu, Z.; Lu, Y.; Song, J.; Zhang, A.; Guo, C.; He, Y. Cryo-EM structure of the human histamine H1 receptor/Gq complex. *Nat. Commun.* **2021**, *12*, 2086. https://doi.org/10.1038/s41467-021-22427-2.

[116]. . DeepMind, "The AlphaFold Database." https://alphafold.ebi.ac.uk/ (accessed Mar. 22, 2022).

[117]. Zou, S.; Kumar, U. Cannabinoid Receptors and the Endocannabinoid System: Signaling and Function in the Central Nervous System. *Int. J. Mol. Sci.* **2018**, *19*, 833. https://doi.org/10.3390/ijms19030833.

[118]. Kaminski, N.E. Immune regulation by cannabinoid compounds through the inhibition of the cyclic AMP signaling cascade and altered gene expression. *Biochem. Pharmacol.* **1996**, *52*, 1133–1140. https://doi.org/10.1016/0006-2952(96)00480-7.

[119]. Mazzon, E.; Giacoppo, S. Can cannabinoids be a potential therapeutic tool in amyotrophic lateral sclerosis? *Neural Regen. Res.* **2016**, *11*, 1896–1899. https://doi.org/10.4103/1673-5374.197125.

[120]. Urbi, B.; Owusu, M.A.; Hughes, I.; Katz, M.; Broadley, S.; Sabet, A. Effects of cannabinoids in Amyotrophic Lateral Sclerosis (ALS) murine models: A systematic review and meta-analysis. *J. Neurochem.* **2018**, *149*, 284–297. https://doi.org/10.1111/jnc.14639.

[121]. Shoemaker, J.L.; Seely, K.A.; Reed, R.L.; Crow, J.P.; Prather, P.L. The CB2 cannabinoid agonist AM-1241 prolongs survival in a transgenic mouse model of amyotrophic lateral sclerosis when initiated at symptom onset. *J. Neurochem.* **2006**, *101*, 87–98. https://doi.org/10.1111/j.1471-4159.2006.04346.x.

[122]. Kim, K.; Moore, D.H.; Makriyannis, A.; Abood, M.E. AM1241, a cannabinoid CB2 receptor selective compound, delays disease progression in a mouse model of amyotrophic lateral sclerosis. *Eur. J. Pharmacol.* **2006**, *542*, 100–105. https://doi.org/10.1016/j.ejphar.2006.05.025.

[123]. Bilsland, L.G.; Dick, J.R.T.; Pryce, G.; Petrosino, S.; Di Marzo, V.; Baker, D.; Greensmith, L. Increasing cannabinoid levels by pharmacological and genetic manipulation delays disease progression in SOD1 mice. *FASEB J.* **2006**, *20*, 1003–1005. https://doi.org/10.1096/fj.05-4743fje.

[124]. Hua, T.; Li, X.; Wu, L.; Iliopoulos-Tsoutsouvas, C.; Wang, Y.; Wu, M.; Shen, L.; Brust, C.A.; Nikas, S.P.; Song, F.; et al. Activation and Signaling Mechanism Revealed by Cannabinoid Receptor-Gi Complex Structures. *Cell* **2020**, *180*, 655–665.e18. https://doi.org/10.1016/j.cell.2020.01.008.

[125]. Xing, C.; Zhuang, Y.; Xu, T.-H.; Feng, Z.; Zhou, X.E.; Chen, M.; Wang, L.; Meng, X.; Xue, Y.; Wang, J.; et al. Cryo-EM Structure of the Human Cannabinoid Receptor CB2-Gi Signaling Complex. *Cell* **2020**, *180*, 645–654.e13. https://doi.org/10.1016/j.cell.2020.01.007.

[126]. Reader, J.; Holt, D.; Fulton, A. Prostaglandin E2 EP receptors as therapeutic targets in breast cancer. *Cancer Metastasis Rev.* **2011**, *30*, 449–463. https://doi.org/10.1007/s10555-011-9303-2.

[127]. Johansson, T.; Narumiya, S.; Zeilhofer, H.U. Contribution of peripheral versus central EP1 prostaglandin receptors to inflammatory pain. *Neurosci. Lett.* **2011**, *495*, 98–101. https://doi.org/10.1016/j.neulet.2011.03.046.

[128]. Fedyk, E.R.; Phipps, R.P. Prostaglandin E2 receptors of the EP2 and EP4 subtypes regulate activation and differentiation of mouse B lymphocytes to IgE-secreting cells.. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 10978–10983. https://doi.org/10.1073/pnas.93.20.10978.

[129]. Watanabe, K.; Kawamori, T.; Nakatsugi, S.; Ohta, T.; Ohuchida, S.; Yamamoto, H.; Maruyama, T.; Kondo, K.; Ushikubi, F.; Narumiya, S.; et al. Role of the prostaglandin E receptor subtype EP1 in colon carcinogenesis. *Cancer Res.* **1999**, *59*, 5093–5096. PMID: 10537280

[130]. Sun, X.; Li, Q. Prostaglandin EP2 receptor: Novel therapeutic target for human cancers (Review). *Int. J. Mol. Med.* **2018**, *42*, 1203–1214. https://doi.org/10.3892/ijmm.2018.3744.

[131]. Takeuchi, K.; Kato, S.; Amagase, K. Prostaglandin EP Receptors Involved in Modulating Gastrointestinal Mucosal Integrity. *J. Pharmacol. Sci.* **2010**, *114*, 248–261. https://doi.org/10.1254/jphs.10r06cr.

[132]. Yang, T.; Du, Y. Distinct Roles of Central and Peripheral Prostaglandin E2 and EP Subtypes in Blood Pressure Regulation. *Am. J. Hypertens.* **2012**, *25*, 1042–1049. https://doi.org/10.1038/ajh.2012.67.

[133]. Friedman, E.A.; Ogletree, M.L.; Haddad, E.V.; Boutaud, O. Understanding the role of prostaglandin E2 in regulating human platelet activity in health and disease. *Thromb. Res.* **2015**, *136*, 493–503. https://doi.org/10.1016/j.thromres.2015.05.027.

[134]. Takasaki, I.; Nojima, H.; Shiraki, K.; Sugimoto, Y.; Ichikawa, A.; Ushikubi, F.; Narumiya, S.; Kuraishi, Y. Involvement of cyclooxygenase-2 and EP3 prostaglandin receptor in acute herpetic but not postherpetic pain in mice. *Neuropharmacology* **2005**, *49*, 283–292. https://doi.org/10.1016/j.neuropharm.2004.12.025.

[135]. Xu, S.; Zhou, W.; Ge, J.; Zhang, Z. Prostaglandin E2 receptor EP4 is involved in the cell growth and invasion of prostate cancer via the cAMP-PKA/PI3K-Akt signaling pathway. *Mol. Med. Rep.* **2018**, *17*, 4702–4712. https://doi.org/10.3892/mmr.2018.8415.

[136]. Iłżecka, J. Prostaglandin E2 is increased in amyotrophic lateral sclerosis patients. *Acta Neurol. Scand.* **2003**, *108*, 125–129. https://doi.org/10.1034/j.1600-0404.2003.00102.x.

[137]. Kosuge, Y.; Nango, H.; Kasai, H.; Yanagi, T.; Mawatari, T.; Nishiyama, K.; Miyagishi, H.; Ishige, K.; Ito, Y. Generation of Cellular Reactive Oxygen Species by Activation of the EP2 Receptor Contributes to Prostaglandin E2-Induced Cytotoxicity in Motor Neuron-Like NSC-34 Cells. *Oxidative Med. Cell. Longev.* **2020**, *2020*, 6101838. https://doi.org/10.1155/2020/6101838.

[138]. Bilak, M.; Wu, L.; Wang, Q.; Haughey, N.; Conant, K.; St. Hillaire, C.; Andreasson, K. PGE2 receptors rescue motor neurons in a model of amyotrophic lateral sclerosis. *Ann. Neurol.* **2004**, *56*, 240–248. https://doi.org/10.1002/ana.20179.

[139]. Qu, C.; Mao, C.; Xiao, P.; Shen, Q.; Zhong, Y.-N.; Yang, F.; Shen, D.-D.; Tao, X.; Zhang, H.; Yan, X.; et al. Ligand recognition, unconventional activation, and G protein coupling of the prostaglandin E 2 receptor EP2 subtype. *Sci. Adv.* **2021**, *7*, eabf1268. https://doi.org/10.1126/sciadv.abf1268.

[140]. Morimoto, K.; Suno, R.; Hotta, Y.; Yamashita, K.; Hirata, K.; Yamamoto, M.; Narumiya, S.; Iwata, S.; Kobayashi, T. Crystal structure of the endogenous agonist-bound prostanoid receptor EP3. *Nat. Chem. Biol.* **2018**, *15*, 8–10. https://doi.org/10.1038/s41589-018-0171-8.

[141]. Nojima, S.; Fujita, Y.; Kimura, K.T.; Nomura, N.; Suno, R.; Morimoto, K.; Yamamoto, M.; Noda, T.; Iwata, S.; Shigematsu, H.; et al. Cryo-EM Structure of the Prostaglandin E Receptor EP4 Coupled to G Protein. *Structure* **2020**, *29*, 252–260.e6. https://doi.org/10.1016/j.str.2020.11.007.

[142]. Umetsu, Y.; Tenno, T.; Goda, N.; Shirakawa, M.; Ikegami, T.; Hiroaki, H. Structural difference of vasoactive intestinal peptide in two distinct membrane-mimicking environments. *Biochim. Biophys. Acta (BBA)-Proteins Proteom.* **2011**, *1814*, 724–730. https://doi.org/10.1016/j.bbapap.2011.03.009.

[143]. Harmar, A.J.; Fahrenkrug, J.; Gozes, I.; Laburthe, M.; May, V.; Pisegna, J.R.; Vaudry, D.; Vaudry, H.; A Waschek, J.; I Said, S. Pharmacology and functions of receptors for vasoactive intestinal peptide and pituitary adenylate cyclase-activating polypeptide: IUPHAR Review 1. *J. Cereb. Blood Flow Metab.* **2012**, *166*, 4–17. https://doi.org/10.1111/j.1476-5381.2012.01871.x.

[144]. Couvineau, A.; Laburthe, M. VPAC receptors: Structure, molecular pharmacology and interaction with accessory proteins. *J. Cereb. Blood Flow Metab.* **2012**, *166*, 42–50. https://doi.org/10.1111/j.1476-5381.2011.01676.x.

[145]. Staines, D.R. Are multiple sclerosis and amyotrophic lateral sclerosis autoimmune disorders of endogenous vasoactive neuropeptides?. *Med. Hypotheses* **2008**, *70*, 413–418. https://doi.org/10.1016/j.mehy.2007.04.038.

[146]. Iwasaki, Y.; Ikeda, K.; Ichikawa, Y.; Igarashi, O. Vasoactive intestinal peptide influences neurite outgrowth in cultured rat spinal cord neurons. *Neurol. Res.* **2001**, *23*, 851–854. https://doi.org/10.1179/016164101101199298.

[147]. Solés-Tarrés, I.; Cabezas-Llobet, N.; Vaudry, D.; Xifró, X. Protective Effects of Pituitary Adenylate Cyclase-Activating Polypeptide and Vasoactive Intestinal Peptide Against Cognitive Decline in Neurodegenerative Diseases. *Front. Cell. Neurosci.* **2020**, *14*, 221. https://doi.org/10.3389/fncel.2020.00221.

[148]. Waschek, J. VIP and PACAP: Neuropeptide modulators of CNS inflammation, injury, and repair. *J. Cereb. Blood Flow Metab.* **2013**, *169*, 512–523. https://doi.org/10.1111/bph.12181.

[149]. Martínez, C.; Juarranz, Y.; Gutiérrez-Cañas, I.; Carrión, M.; Pérez-García, S.; Villanueva-Romero, R.; Castro, D.; Lamana, A.; Mellado, M.; González-Álvaro, I.; et al. A Clinical Approach for the Use of VIP Axis

in Inflammatory and Autoimmune Diseases. *Int. J. Mol. Sci.* **2019**, *21*, 65. https://doi.org/10.3390/ijms21010065.

[150]. Duan, J.; Shen, D.-D.; Zhou, X.E.; Bi, P.; Liu, Q.-F.; Tan, Y.-X.; Zhuang, Y.-W.; Zhang, H.-B.; Xu, P.-Y.; Huang, S.-J.; et al. Cryo-EM structure of an activated VIP1 receptor-G protein complex revealed by a NanoBiT tethering strategy. *Nat. Commun.* **2020**, *11*, 4121. https://doi.org/10.1038/s41467-020-17933-8.

[151]. Niswender, C.M.; Conn, P.J. Metabotropic Glutamate Receptors: Physiology, Pharmacology, and Disease. *Annu. Rev. Pharmacol. Toxicol.* **2010**, *50*, 295–322. https://doi.org/10.1146/annurev.pharmtox.011008.145533.

[152]. Anneser, J.M.H.; Chahli, C.; Ince, P.G.; Borasio, G.D.; Shaw, P. Glial Proliferation and Metabotropic Glutamate Receptor Expression in Amyotrophic Lateral Sclerosis. *J. Neuropathol. Exp. Neurol.* **2004**, *63*, 831–840. https://doi.org/10.1093/jnen/63.8.831.

[153]. Milanese, M.; Giribaldi, F.; Melone, M.; Bonifacino, T.; Musante, I.; Carminati, E.; Rossi, P.I.; Vergani, L.; Voci, A.; Conti, F.; et al. Knocking down metabotropic glutamate receptor 1 improves survival and disease progression in the SOD1G93A mouse model of amyotrophic lateral sclerosis. *Neurobiol. Dis.* **2014**, *64*, 48–59. https://doi.org/10.1016/j.nbd.2013.11.006.

[154]. Rossi, D.; Brambilla, L.; Valori, C.F.; Roncoroni, C.; Crugnola, A.; Yokota, T.; Bredesen, D.E.; Volterra, A. Focal degeneration of astrocytes in amyotrophic lateral sclerosis. *Cell Death Differ.* **2008**, *15*, 1691–1700. https://doi.org/10.1038/cdd.2008.99.

[155]. Anneser, J.; Chahli, C.; Borasio, G. Protective effect of metabotropic glutamate receptor inhibition on amyotrophic lateral sclerosis–cerebrospinal fluid toxicity in vitro. *Neuroscience* **2006**, *141*, 1879–1886. https://doi.org/10.1016/j.neuroscience.2006.05.044.

[156]. Crupi, R.; Impellizzeri, D.; Cuzzocrea, S. Role of Metabotropic Glutamate Receptors in Neurological Disorders. *Front. Mol. Neurosci.* **2019**, *12*, 20. https://doi.org/10.3389/fnmol.2019.00020.

[157]. Zhang, J.; Qu, L.; Wu, L.; Tang, X.; Luo, F.; Xu, W.; Xu, Y.; Liu, Z.-J.; Hua, T. Structural insights into the activation initiation of full-length mGlu1. *Protein Cell* **2020**, *12*, 662–667. https://doi.org/10.1007/s13238-020-00808-5.

[158]. Seven, A.B.; Barros-Álvarez, X.; de Lapeyrière, M.; Papasergi-Scott, M.M.; Robertson, M.J.; Zhang, C.; Nwokonko, R.M.; Gao, Y.; Meyerowitz, J.G.; Rocher, J.-P.; et al. G-protein activation by a metabotropic glutamate receptor. *Nature* **2021**, *595*, 450–454. https://doi.org/10.1038/s41586-021-03680-3.

[159]. Lin, S.; Han, S.; Cai, X.; Tan, Q.; Zhou, K.; Wang, D.; Wang, X.; Du, J.; Yi, C.; Chu, X.; et al. Structures of Gi-bound metabotropic glutamate receptors mGlu2 and mGlu4. *Nature* **2021**, *594*, 583–588. https://doi.org/10.1038/s41586-021-03495-2.

From the Wuhan-Hu-1 strain to the XD and XE variants: is targeting the SARS-CoV-2 Spike protein still a pharmaceutically relevant option against COVID-19?

Matteo Pavan, **Davide Bassani**, Mattia Sturlese, and Stefano Moro

Matteo Pavan, Davide Bassani, Mattia Sturlese & Stefano Moro (2022) From the Wuhan-Hu-1 strain to the XD and XE variants: is targeting the SARS-CoV-2 spike protein still a pharmaceutically relevant option against COVID-19?, Journal of Enzyme Inhibition and Medicinal Chemistry, 37:1, 1704-1714, DOI: 10.1080/14756366.2022.2081847

Abstract

Since the outbreak of the COVID-19 pandemic in December 2019, the SARS-CoV-2 genome has undergone several mutations, resulting in the rise of a multitude of viral variants, some of which have shown an increased pathogenic potential. The emergence of such variants has resulted in multiple pandemic waves, contributing to sustaining to date the number of infections, hospitalizations, and deaths despite the swift development of vaccines, since most of these mutations are concentrated on the Spike protein, a viral surface glycoprotein that is the main target for most vaccines. A milestone in the fight against the COVID-19 pandemic has been represented by the development of Paxlovid, the first orally available drug against COVID-19, which acts instead on a different viral target, i.e., the Main Protease (M^{pro}). In this article, we analyze the structural features of both the Spike protein and the Main Protease of the recently reported SARS-CoV-2 variant XE, as well as the closely related XD and XF ones, discussing their impact on the efficacy of existing treatments against COVID-19 and on the development of future ones.

1. Introduction

More than two years have now passed since the beginning of the COVID-19 pandemic, back in December 2019[1][2]. Caused by a betacoronavirus known as SARS-CoV-2 and characterized by para-influenzal symptoms such as fever, cough, and dyspnea, this worldwide-spread disease has resulted in the death of more than 6 million people around the world, becoming one of the deadliest illnesses in human history[3][4].

The SARS-CoV-2 virus was first identified in the Chinese city of Wuhan, where the pandemic was firstly spotted[5]. The genomic sequence of this virus (named Wuhan-Hu-1 from now on in the

article) is 80% identical to the one of the SARS-CoV virus[6][7], which was responsible for the Severe Acute Respiratory Syndrome (SARS) that stroke the South East of Asia in 2002/2003, causing the death of about 800 patients over 9000 cases (10 % death rate)[8][9]. The exact origin of the SARS-CoV-2 virus is still to this date unknown, however several pieces of evidence point out bat coronaviruses as closely related ancestors and to the pangolin as the intermediate host before the human spillover[10][11][12][13].

Soon after the original virus started spreading all over the world, several viral variants began to emerge[14][15], especially in the poorest countries where public health measures such as social distancing and wearing surgical masks in public places were difficult to implement[16][17][18]. Most of the genome mutations that characterized these variants were concentrated in the S gene[19], which encodes for the Spike protein, a surface glycoprotein that mediates the virus entry within the human cell through an interaction with the human ACE2 receptor[20]. Some of these mutations gathered the attention of the scientific community due to the selective advantage that they provided to the correspondent viral variants, regarding both the virus' ability to infect human cells and to escape the immune system response[21], gaining for these reasons the status of "variant of concern" (VOC).

The first SARS-CoV-2 variant to be labeled as VOC was the so-called Alpha variant (B.1.1.7). First identified in November 2020 in the Kent region of the United Kingdom and for this reason also known as the "English variant", B.1.1.7 was estimated to be 29% more transmissible than the original virus[22][23]. Despite being more transmissible than other circulating viral strains[24][25], and despite showing the first signs of reduced protection provided by vaccines, monoclonal antibodies, and convalescent sera[26][27][28], the indication from clinical studies showed that the vaccine coverage (especially in those who had already completed the vaccination cycle) was still able to contain the impact of this variant on the sanitary system[29].

Soon after the identification of the Alpha variant, a second VOC arose: the Delta variant (B.1.617.2), also known as the "Indian variant" due to being first detected in India in late 2020, rapidly overcame the Alpha variant becoming the dominant strain in the world, thanks to being 97% more transmissible than the original Wuhan virus[23]. The replacement of the less threatening Alpha variant with the Delta marked a significant change of pace in the pandemic trend, signing an increased burden for the health system[30] and posing for the first time a serious threat to the

protection provided by vaccines, convalescent sera, and monoclonal antibodies[31][32][33] due to its increased ability to evade the immune system response[34].

From November 2021 onwards, the Delta variant has been flanked by another VOC firstly identified in South Africa and defined as the Omicron variant (B.1.1.529)[35]. The rise of the Omicron variant, fuelled by a contemporary increase in transmissibility[36] and immune evasion[37], resulted in an unprecedented diffusion of the SARS-CoV-2 virus all over the world, being able to overcome even the protection provided by the full primary vaccination cycle and by most neutralizing antibodies used in therapy[38][39][40], thus inducing the introduction of a "booster dose" to bring the protective effect of vaccines back to adequate levels[41][42].

In the face of this increasingly troublesome variant landscape, characterized by a progressive reduction of the efficacy of existing therapeutic options against COVID-19, a light at the end of the tunnel is possibly represented by the development and release on the market of Paxlovid. This therapeutic combination between the active principle Nirmatrelvir (also known as PF-07321332) and the pharmacokinetic enhancer Ritonavir, represents the first orally available drug specifically designed against SARS-CoV-2 virus[43]. Instead of targeting the Spike protein, this peptidomimetic entity is designed to inhibit the SARS-CoV-2 Main Protease (M^{pro}) by covalently binding to Cysteine 145, one of the two components of the protease's catalytic diad[44]. Clinical studies showed a remarkable therapeutic efficacy of this novel treatment, with Paxlovid being able to lower by 89% the risk of severe complications associated with COVID-19 infection in symptomatic, non-vaccinated, and non-hospitalized adult patients[45].

Recently, three novel recombinant SARS-CoV-2 variants were identified in the United Kingdom: Xd, Xe, and Xf[46]. These variants are derived from the combination of the genomes of other major circulating variants, namely Delta, Omicron, and Omicron 2[46][47]. Among these three novel viral strains, particular worry is related to the XE variant, which derives from the recombination between two VOCs, Omicron and Omicron 2, and is supposed to be 13-20 % more transmissible than the Omicron 2 variant[46].

The rise of novel SARS-CoV-2 variants derived from the recombination of other threatening and heavily diffused ones poses a serious challenge in the fight against the COVID-19 pandemic, since it could contribute to render existing therapeutic options inefficient or practically useless. To evaluate the impact that these recently reported recombinant variants could have on the efficacy of existing vaccines and treatments (Paxlovid, in particular), we performed a computational analysis to shed

light on the key structural features that characterize both the Spike glycoprotein and the Main Protease of these novel viral strains. Moreover, we analyzed the structural evolution of these two viral proteins throughout the pandemic, discussing the impact that mutations found on these strains had and will have on the efficacy of existing therapeutic options against COVID-19 and the development of future ones.

2. Materials and Methods

The genome sequence for the SARS-CoV-2 virus and its variants, namely Delta, Omicron, XD, XE and XF, was obtained through GenBank[48]. Accession codes for each of the considered genomes are reported in Table 1. In the case of newly discovered variants XD, XE, XF, the sequence was chosen according to the one reported by the Nextclade project[49].

Organism	**Isolate**	**Accession Code**
SARS-CoV-2	"Wuhan-Hu-1"	NC_045512.2
SARS-CoV-2 "Delta"	"SARS-CoV-2/human/JPN/SARS-CoV-2"	OK091006.1
SARS-CoV-2 "Omicron"	"SARS-CoV-2/human/NLD/EMC-Omicron-1/2021"	OM287553.1
SARS-CoV-2 "XD"	"SARS-CoV-2/human/FRA/IHUCOVID-64762/2022"	OM990851.1
SARS-CoV-2 "XE"	/	OW018845.1
SARS-CoV-2 "XF"	/	OV940149.1

Table 1. *The genome sequences used in this work and their origin.*

All the basic molecular modeling operations have been executed with the Molecular Operating Environment (MOE) suite (version 2019.01)[50].

For what concerns the Spike protein, the approach chosen depended on the variant considered. For the wild-type (WT) Spike, the three-dimensional structure was retrieved from the Protein Data Bank (PDB code: 6ZDH[51], method: cryo-EM, resolution: 3.70 Å), as well as for the Delta (PDB code:

7W9E[52], method: cryo-EM, resolution: 3.10 Å), and the Omicron (PDB code: 7WPD[53], method: cryo-EM, 3.18 Å) variants. The cited structures were all subjected to the same preparation procedure for molecular modeling.

After being downloaded, the *Structure Preparation* tool implemented in MOE was applied in order to rebuild the missing loops in the structures, the proper protonation state was assigned to each amino acid with the MOE "*Protonate 3D"* application, and finally, the added hydrogens were minimized under the AMBER10:EHT[54] force field implemented in MOE. Since experimentally resolved structures for the XD, XE, and XF variants are not available in public databases, the models considered for our study were created starting from the WT SARS-CoV-2 Spike coming from the PDB code 6ZDH by manually mutating the residues, exploiting the MOE *"Protein builder"* tool, and subjecting each protein to the preparation procedure reported above. For the realization of the video reported in the Supplementary Materials, the program VMD 1.9.2[55] (Visual Molecular Dynamics) was used.

For what concerns M^{pro}, the protein sequences corresponding to the main protease were extracted from the whole genome sequence and aligned to the reference sequence (Wuhan-Hu-1) making use of the appropriate tool of MOE 2019.01. Subsequently, homology models for each variant were created making use of the "*Homology Model*" tool of MOE 2019.01, using the structure deposited in the Protein Data Bank with accession code 6Y2E (Crystal structure of the free enzyme of the SARS-CoV-2 main protease) as a template for model generation.

3. Results and Discussion

3.1 Structural analysis of Spike glycoprotein mutations found in SARS-CoV-2 XD, XE, and XF variants and their impact on hACE2 binding

The SARS-CoV-2 Spike protein (S) consists of a large biological entity formed by 1273 amino acids organized in different functional domains. The main role of the Spike protein is to mediate the virus entry into the host cell, with the principal and better-characterized mechanism being the pathway involving the binding to the human ACE2 receptor (hACE2)[56], a membrane-bound enzyme that is widely expressed in various districts of the human body (from the endothelial cells of the blood vessels to kidneys, liver, intestine, lungs[57], and cells of bronchial and nasal epithelia[58]).

The S protein, which works in a trimeric organization, is divided into two main subunits, S1 and S2. The second of these has very important roles in spike protein trimerization and in mediating the

virion entry into the host cell once the molecular contacts have been established. It is formed by relevant subdomains such as the fusion peptide (FP, residues 943-982, crucial for viral fusion to the host cell membrane), the transmembrane domain (TM, composed of 24 aminoacids and deputed both to the anchoring of S protein to the viral membrane and the maintenance of the Spike trimeric organization), and the cytoplasmatic fusion domain (CT, mediating virus-cell fusion).

The S1 subunit, instead, contains both the N-terminal and the C-terminal domains (NTD and CTD, respectively), which are involved in the binding to host cell receptors. Specifically, the CTD contains the receptor-binding domain (RBD, aminoacids 319-541), the region deputed to the binding with hACE2. This function is more precisely carried out by a particular RDB subdomain, called receptor-binding motif (RBM), which is formed by two beta-sheets (b5 and b6) composed of those residues which are in close contact with hACE2 (from 438 to 506[59])[60].

Looking at all the S proteins of the different SARS-CoV-2 relevant variants, the RBD contains the highest "single-point mutations/sequence length" ratio in all cases. Examining the different SARS-CoV-2 variants discovered up to date, the Spike protein is surely the viral entity that has mutated the most in the evolutionary process of the virus[61]. Its exposition on the viral surface and its crucial function in viral cell entry make this protein the eligible target for the host immune system[62].

The SARS-CoV-2 S protein has experienced several mutations in the past two years[63], as reported in Tables 1 and 2 for the variants considered in our study. As can be noticed, variants such as Delta (but also Alpha and Beta, not specifically treated in this article) showed few mutations in the overall viral genome, and Spike protein displayed never more than a tenth of single-point changes. The game-changing event was the advent of the Omicron variant, much different from its previous analogs, with 30-single nucleotide mutations involving the S protein only. Many of these, such as K417N, T478K, and P614G were inherited from the previous lineages (mainly Beta and Delta), but other mutations were completely new, such as G339D, G446S, or E484A.

Delta variant	**Omicron variant**	**XD variant**	**XE variant**	**XF variant**
T19R		T19R	T19R	
		A27S	A27S	
	A67V			A67V
	T95I	T95I		T95I
	G142D	G142D	G142D	
				Y145D
R158G		R158G		

	L212I	L212I		L212I
			V213G	
	G339D	G339D	G339D	G339D
	S371L	S371L	S371L	S371L
	S373P	S373P	S373P	S373P
	S375F	S375F	S375F	S375F
			T376A	
			D405N	
			R408S	
	K417N	K417N	K417N	K417N
	N440K	N440K	N440K	N440K
	G446S	G446S		G446S
L452R				
	S477N	S477N	S477N	S477N
T478K	T478K	T478K	T478K	T478K
	E484A	E484A	E484A	E484A
	Q493R	Q493R	Q493R	Q493R
	G496S	G496S		G496S
	Q498R	Q498R	Q498R	Q498R
	N501Y	N501Y	N501Y	N501Y
	Y505H	Y505H	Y505H	Y505H
	T547K	T547K		T547K
D614G	D614G	D614G	D614G	D614G
	H655Y	H655Y	H655Y	H655Y
	N679K	N679K	N679K	N679K
P681R	P681H	P681H	P681H	P681H
	N764K	N764K	N764K	N764K
	D796Y	D796Y	D796Y	D796Y
	N856K	N856K		N856K
	Q954H	Q954H	Q954H	Q954H
	N969K	N969K	N969K	N969K
	L981F	L981F		L981F

Table 2. *List of all the single-point mutations affecting the SARS-CoV-2 Spike protein for all the variants considered in our study (Delta, Omicron, XE, XD, and XF). The mutations have been aligned to give a better perspective of the ones which have been conserved through the evolutionary process. The mutations involving the RBD have been highlighted in green, while the ones involving the RBM are colored in cyan.*

Delta variant	Omicron variant	XD variant	XE variant	XF variant
			Δ24-26	
	Δ69-70			Δ69-70
				Δ142-144
	Δ143-145			
Δ156-157		Δ156-157		
	Δ211	Δ211		Δ211
	ins214EPE	ins214EPE		ins214EPE

Table 3. *List of all the insertions and deletions operated by SARS-CoV-2 Spike protein for all the variants considered in our study (Delta, Omicron, XD, XE, and XF). The mutations have been aligned to give a better perspective of the ones which have been conserved through the evolutionary process.*

Most of the mutations listed in Table 2 have been related to higher infectivity, mainly due to a consequent gain of affinity for hACE2 or improved shielding from the immune cells. As evidence of this, the most successful vaccination campaigns for COVID-19 always involved the forced recognition of Spike protein from the patients' immune cells[64].

Specifically, mutations such as S371L, K417N, and Q493R were related to a diminished binding to the anti-coronavirus monoclonal antibody Casivirimab, while mutations like N440K and G446S confer resistance towards the antibody Imdevimab[65]. The combination of Casivirimab and Imdevimab has been used to treat COVID-19 patients but has demonstrated to be ineffective against the Omicron variant[66]. Other mutations external to the RBD have been linked to different outcomes, such as increased viral replication (Δ69-70[67] and D614G[68]) and higher viral resistance (G339D and N440K[69]). In other scenarios, mutations have been reported to influence tropism of the S1/S2 cleavage, as in the cases of N679K and P681H[70].

The majority of the mutations highlighted up to date on the SARS-CoV-2 S protein impact the binding with hACE2, as in the cases of S477N, Q498R, and N501Y[71]. These last mutations, as can be seen from Table 2, have been conserved from all the variants following Omicron, assessing their importance for the viral evolutionary process.

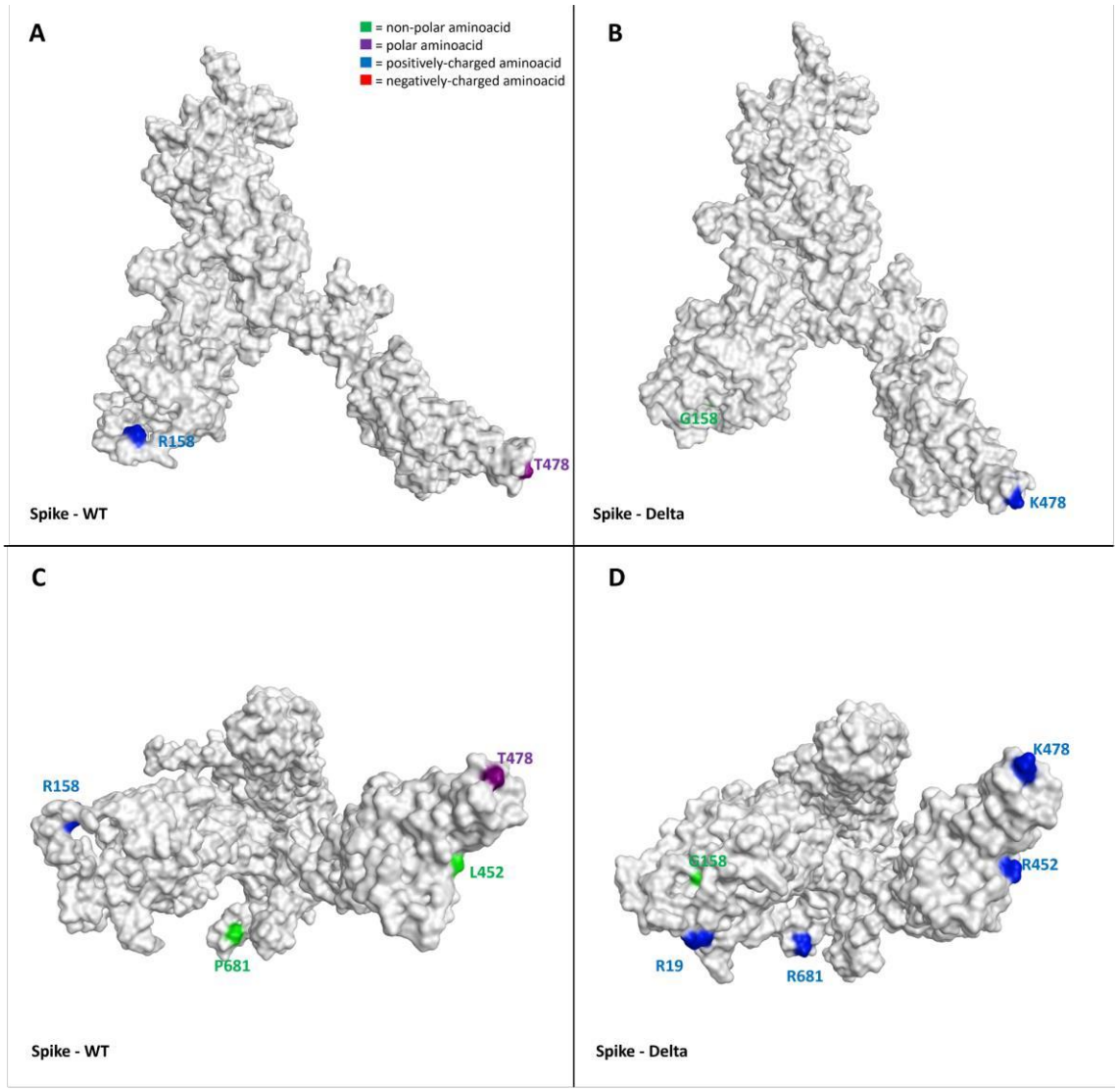


Figure 1. *Representation of structural differences between the WT (taken from the Protein Data Bank[72], PDB code: 6ZDH) and the Delta variant (retrieved from the PDB, code 7W9E[73]) of SARS-CoV-2 Spike protein. Panels A and B offer a front view of the comparison between the structures, while panels C and D shift the point of view to the bottom of the proteins. To give a clearer view of the mutations, only one monomer was considered to create the image. The aminoacids involved in mutations are labeled in the figure and are colored based on their kind, following the legend reported in the panel A. Specifically, Gly, Ala, Val, Leu, Ile, Pro, Cys, Met, Phe and Trp are considered non-polar aminoacids (green), Asp and Glu represent the negatively-charged aminoacids (red), and Lys, Arg, and His form the positively-charged aminoacid group (blue). Finally, Ser, Thr, Asn, Gln, and Tyr are all considered polar aminoacids (purple). All images were created and rendered using the Molecular Operating Environment (MOE) suite.*

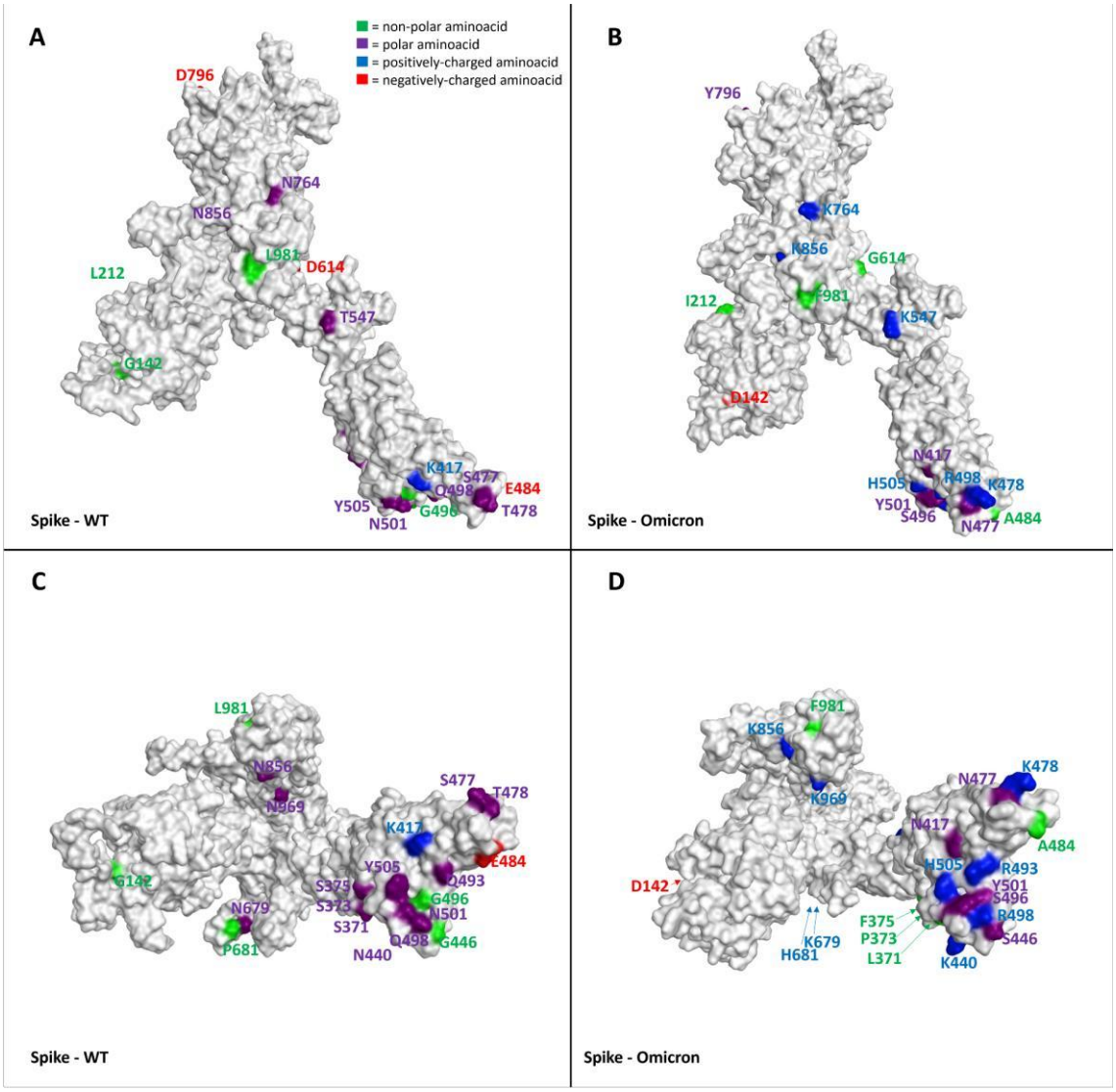


Figure 2. *Representation of the differences between the WT (taken from PDB code: 6ZDH) and the Omicron variant (retrieved from the PDB, code 7WPD) of SARS-CoV-2 Spike protein. Panels A and B offer a front view of the comparison between the structures, while panels C and D shift the point of view to the bottom of the proteins. To give a clearer view of the mutations, only one monomer was considered to create the image. The aminoacids involved in mutations are labeled in the figure and are colored based on their kind, following the legend reported in the panel A. Specifically, Gly, Ala, Val, Leu, Ile, Pro, Cys, Met, Phe and Trp are considered non-polar aminoacids (green), Asp and Glu represent the negatively-charged aminoacids (red), and Lys, Arg, and His form the positively-charged aminoacid group (blue). Finally, Ser, Thr, Asn, Gln, and Tyr are all considered polar aminoacids (purple). All images were created and rendered using the Molecular Operating Environment (MOE) suite.*

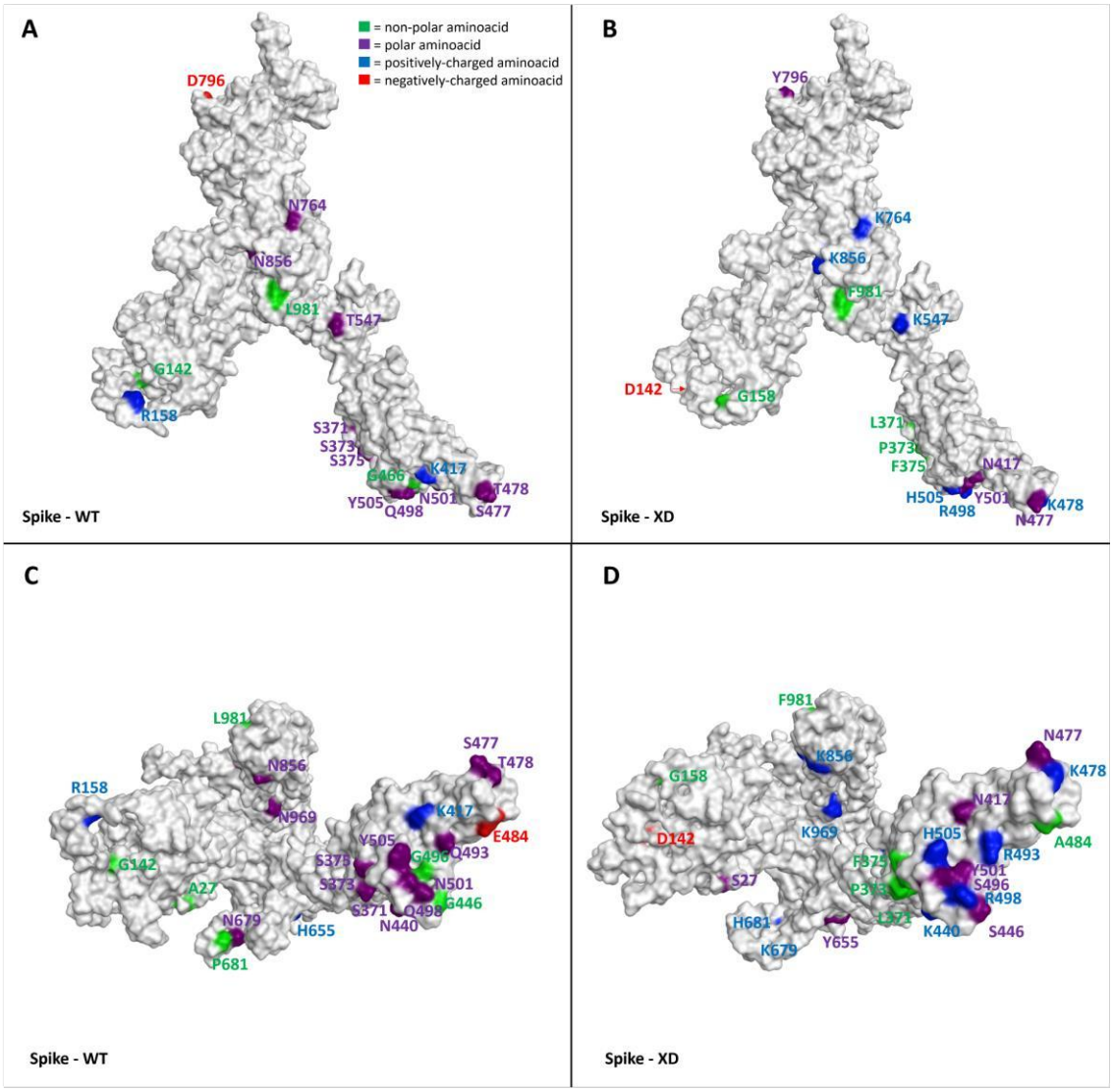


Figure 3. *Representation of the differences between the WT (taken from PDB code: 6ZDH) and the XD variant of SARS-CoV-2 Spike protein. Due to the lack of experimentally resolved structure of SARS-CoV-2 XD variant, the three-dimensional structure represented was obtained from the wild-type S protein coming from PDB code 6DZH, and then manually mutating the residues involved in the mutations (exploiting the MOE "Protein builder" tool). Panels A and B offer a front view of the comparison between the structures, while panels C and D shift the point of view to the bottom of the proteins. To give a clearer view of the mutations, only one monomer was considered to create the image. The aminoacids involved in mutations are labeled in the figure and are colored based on their kind, following the legend reported in the panel A. Specifically, Gly, Ala, Val, Leu, Ile, Pro, Cys, Met, Phe and Trp are considered non-polar aminoacids (green), Asp and Glu represent the negatively-charged aminoacids (red), and Lys, Arg, and His form the positively-charged aminoacid group (blue). Finally, Ser, Thr, Asn, Gln, and Tyr are all considered polar aminoacids (purple). All images were created and rendered using the Molecular Operating Environment (MOE) suite.*

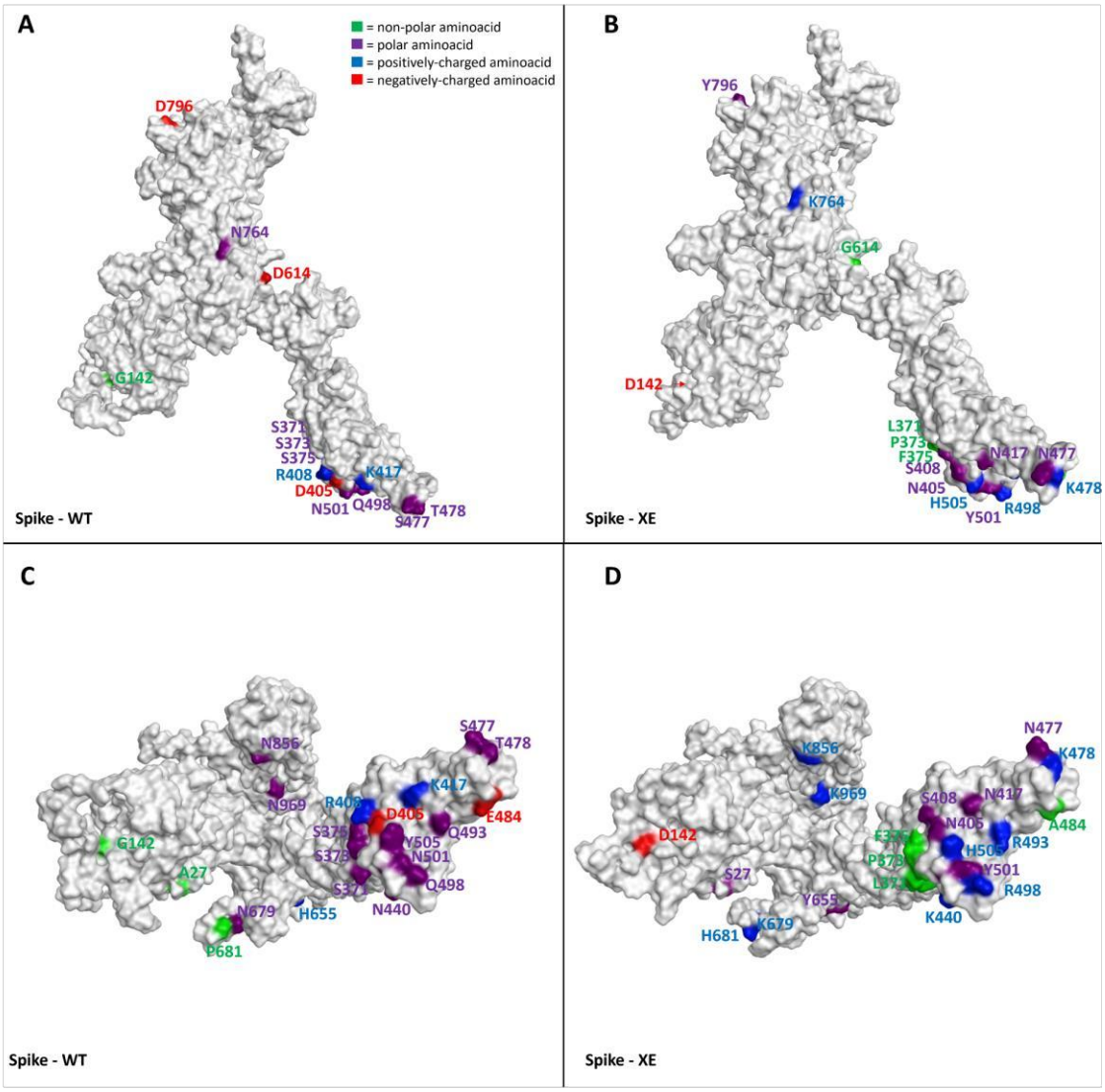


Figure 4. *Representation of the differences between the WT (taken from PDB code: 6ZDH) and the XE variant of SARS-CoV-2 Spike protein. Due to the lack of experimentally resolved structure of SARS-CoV-2 XE variant, the three-dimensional structure represented was obtained from the wild-type S protein coming from PDB code 6DZH, and then manually mutating the residues involved in the mutations (exploiting the MOE "Protein builder" tool). Panels A and B offer a front view of the comparison between the structures, while panels C and D shift the point of view to the bottom of the proteins. To give a clearer view of the mutations, only one monomer was considered to create the image. The aminoacids involved in mutations are labeled in the figure and are colored based on their kind, following the legend reported in the panel A. Specifically, Gly, Ala, Val, Leu, Ile, Pro, Cys, Met, Phe and Trp are considered non-polar aminoacids (green), Asp and Glu represent the negatively-charged aminoacids (red), and Lys, Arg, and His form the positively-charged aminoacid group (blue). Finally, Ser, Thr, Asn, Gln, and Tyr are all considered polar aminoacids (purple). All images were created and rendered using the Molecular Operating Environment (MOE) suite.*

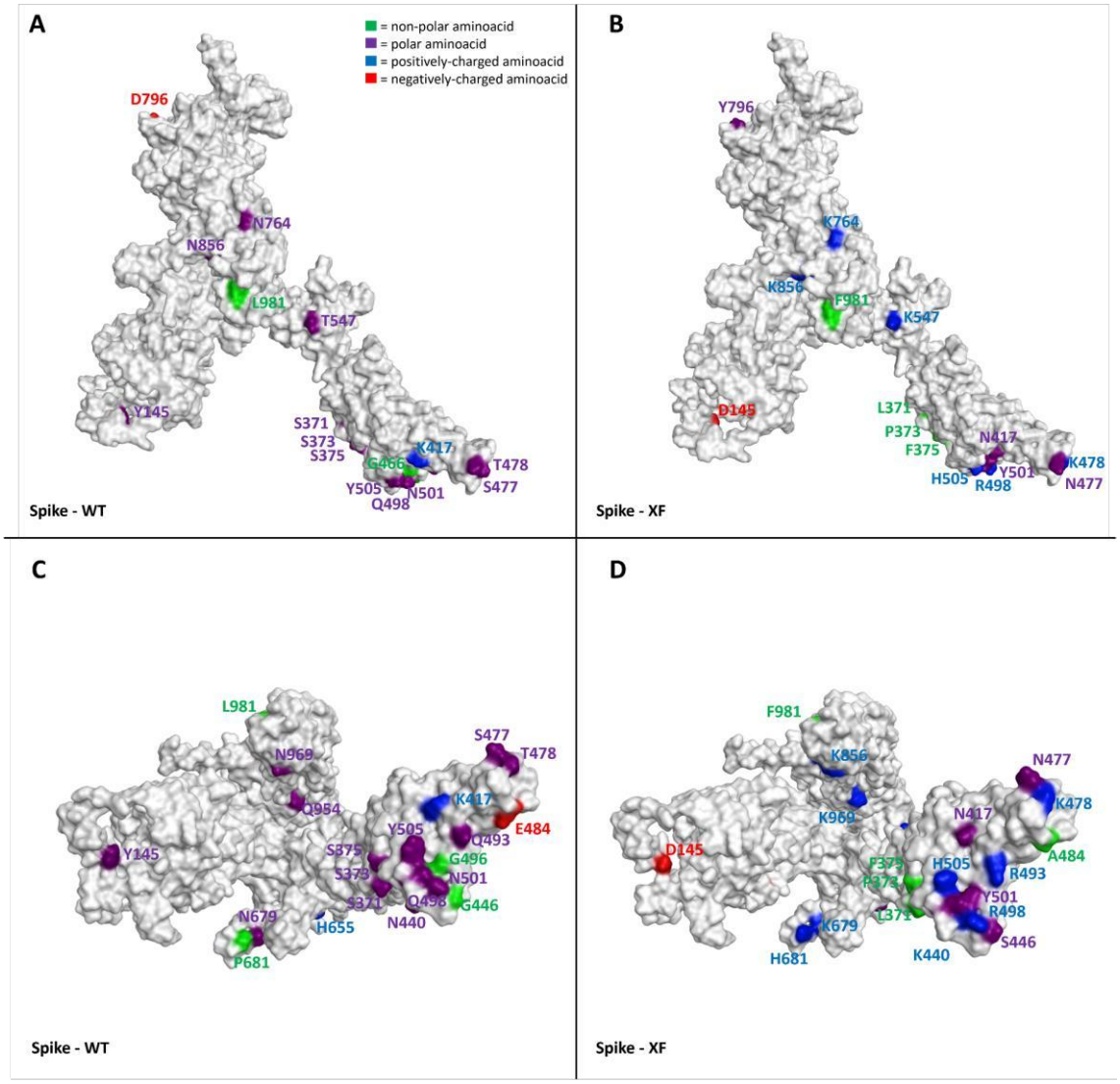


Figure 5. *Representation of the differences between the WT (taken from PDB code: 6ZDH) and the XF variant of SARS-CoV-2 Spike protein. Due to the lack of experimentally resolved structure of SARS-CoV-2 XF variant, the three-dimensional structure represented was obtained from the wild-type S protein coming from PDB code 6DZH, and then manually mutating the residues involved in the mutations (exploiting the MOE "Protein builder" tool). Panels A and B offer a front view of the comparison between the structures, while panels C and D shift the point of view to the bottom of the proteins. To give a clearer view of the mutations, only one monomer was considered to create the image. The aminoacids involved in mutations are labeled in the figure and are colored based on their kind, following the legend reported in the panel A. Specifically, Gly, Ala, Val, Leu, Ile, Pro, Cys, Met, Phe and Trp are considered non-polar aminoacids (green), Asp and Glu represent the negatively-charged aminoacids (red), and Lys, Arg, and His form the positively-charged aminoacid group (blue). Finally, Ser, Thr, Asn, Gln, and Tyr are all considered polar aminoacids (purple). All images were created and rendered using the Molecular Operating Environment (MOE) suite.*

As can be seen in Figures 1 to 5 and video.mp4 (Supplementary Materials), the highest "number of mutations/sequence length" ratio is owned by the RDB, as previously mentioned. Indeed, taking Omicron as an example, among the 30 mutations in the overall 1273-residues structure, 15 are located just in the 222 residues-sequence forming the RBD. The insertions and the deletions (summarized in Table 3), on the other hand, are located far outside the hACE2-binding domain in all

the variants examined, allowing us to assert that these mutations should not impact at all the host cell recognition process.

Interestingly, as depicted in the aforementioned figures, single-point mutations that are present in the RBD for all considered variants tend to progressively increase the positively-charged character of this protein region. Moreover, of all the changes operated by the evolutionary process of Spike protein, the very few mutations which transform a residue into a negatively-charged one (Asp or Glu) are always located away from the RBD (except for G339D, which is located in the posterior part of the RBD, away from the RBM that contacts hACE2). Indeed, in this region, the changes from polar aminoacids to positively charged ones are abundant (N440K, T478K, Q498R, Y505H), and there are also cases in which non-polar residues transform into polar ones (e.g., G446S and G496S, which are conserved in all examined post-Omicron variants except for the XE one). Another conserved structural feature across all post-Omicron variants is also the fact that E484, located in the RBM, mutates into an alanine, while a peculiar mutation exclusive to the XE variant is represented by the transformation of D405 into an asparagine. Taken together, all these pieces of evidence converge in assessing that an increase in the polar characteristics of the RBD (more specifically, the RBM), with particular relevance to an increase in the number of positively charged aminoacids, could be the mechanism adopted by SARS-CoV-2 to continuously increase its infectivity through an increase in the interaction with hACE2.

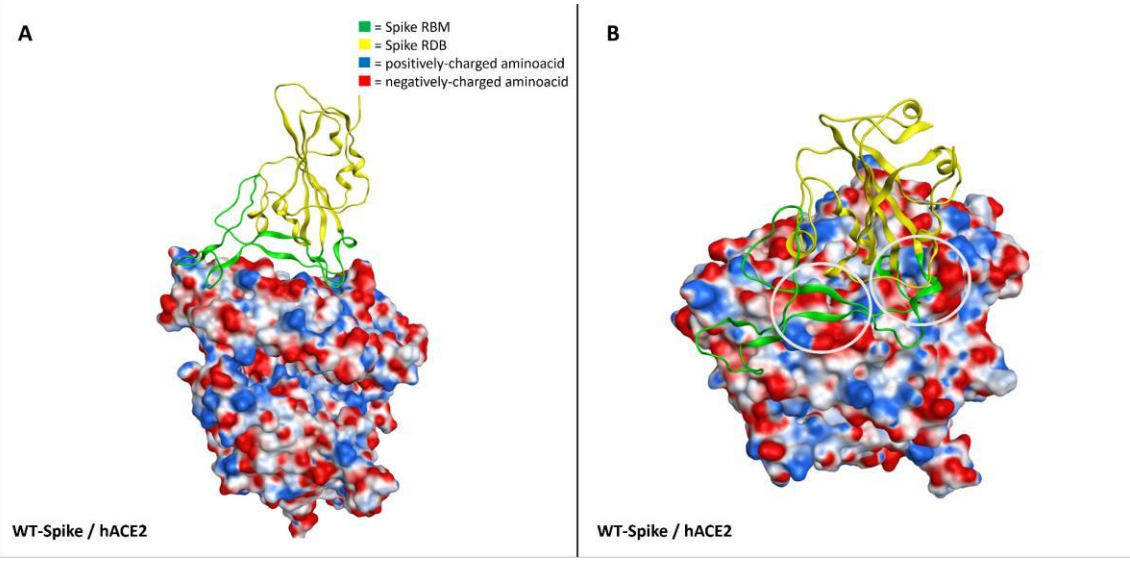


Figure 6. *Representation of the interaction between WT-Spike receptor-binding domain (RBD) and hACE2 (coming from PDB code: 6M0J[74]). The Spike RBD is colored in yellow, while the receptor-binding motif (RBM) is colored in green. The hACE2 surface is colored according to the electrostatic properties of underlying residues (blue, positively-charged regions, red, negatively-charged regions, white, neutral regions). Panel A offers a lateral view of the complex, while panel B focuses the attention on a top-lateral perspective. As can be seen from panel B, the hACE2 regions in contact*

with Spike-RBM are prevalently negatively-charged (red color): concerning this, for visualization purpose, the most extended negative regions at the Spike-hACE2 interface have also been highlighted with grey circles.

To further support this evidence, in Figure 6 we reported the electrostatic surface of hACE2 complexed with WT-Spike RBD highlighting the prevalence of negative charge on the surface facing Spike RBM (colored in green in the image). Coherently, the only mutation present in the RBD in which a positively-charged residue shifts into a polar one (K417N) has been reported to reduce the affinity with hACE2[75]. It is worth noting that other hACE2-independent entry routes for SARS-CoV-2 have been described in literature[76][77], but the lack of reliable structural information about the interaction with the target at the present date hampers and limits the possibility to analyze and discuss the impact that these mutations could have on them in a meaningful way. However, it cannot be excluded that this mutation pattern and other future Spike mutations could also impact these other entry pathways, contributing to making them more relevant for the ability of SARS-CoV-2 to infect human cells and increasing its overall infectivity.

3.2 Structural analysis of Main Protease mutations found in SARS-CoV-2 XD, XE and XF variants and their impact on the recognition of known inhibitors

The main protease M^{pro}, also known as 3C-like protease or $3CL^{pro}$, is a cysteine peptidase that is essential for the replication cycle of SARS-CoV-2[78][79]. Its catalytic activity revolves around the processing of two overlapping polyproteins, namely pp1a and pp1ab, which leads to the formation of 16 mature non-structural proteins (NSPs)[80]. Composed of 306 aminoacids, the SARS-CoV-2 M^{pro} shares 96% sequence identity and a highly conserved three-dimensional structure with the SARS-CoV M^{pro} (0.53 Å R.M.S.D. between PDB entries 6Y2E and 2BX4)[81][82]. Although a dynamic equilibrium between a monomeric and a dimeric form exists, only the dimer is responsible for the protease's enzymatic activity[83][84]. Each protomer composing the catalytically active dimer is composed of three different domains: the chymotrypsin-like β-barrel domains I (residues 1-99) and II (residues 100-182), which comprehend the substrate binding site and directly control the catalytic event, and the extra α-helical domain III (residues 198-306), which is connected to the remaining domains through a 16 residues loop and is involved in the dimerization process, thus playing an indirect role in the regulation of M^{pro} catalytic activity[83][85].

The catalytic site is a shallow, solvent-exposed cavity which is formed by several sub-pockets that are responsible for the recognition of various residues composing the substrate peptide

sequences[81][85]. Concerning this, particularly important is the conserved sequence Gln↓-Ser, where Gln↓- indicates the glutamine residue that precedes the cleavage site[86].

Despite its peculiar structural features, which makes it a difficult target to drug, rational Structure-Based approaches such as Molecular Docking[87] and Molecular Dynamics[88] have proven to be useful tools in the identification and characterization of M^{pro} small molecule inhibitors, leading to the discovery of both covalent and non-covalent lead compounds[89][90]. Further reinforcing the importance of the Main Protease as a key drug target against COVID-19, is the discovery and approval by regulatory agencies of Nirmatrelvir, the first drug specifically designed against SARS-CoV-2 to enter the market[43].

Due to its pivotal role for the virus replication cycle, the Main Protease is, on the contrary of Spike, particularly conserved in its primary sequence and its three-dimensional structural features among different viral strains. Taking a closer look at the main proteases from previously mentioned SARS-CoV-2 variants, only one out of 306 aminoacids is mutated compared to the reference sequence, precisely residue 132, which is a proline in the case of Wuhan-Hu-1, Delta and XD viral strains, while it is mutated to a histidine in the case of Omicron, XE and XF variants.

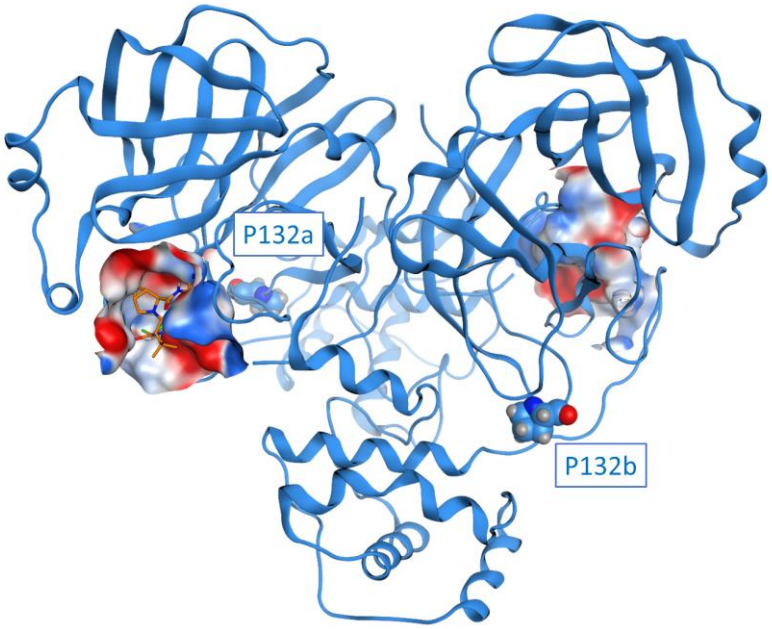


Figure 7. *The structure of SARS-CoV-2 M^{pro} (PDB ID: 6Y2E) in its free form. The protein is depicted in blue ribbons, while the mutated residue P132 in comparison with considered SARS-CoV-2 variants (Delta, Omicron, XD, XE and XF) is highlighted and depicted as a CPK model. For visual reference, Nirmatrelvir (also known as PF-07321332, commercial name Paxlovid) from structure 7RFS is also shown in the picture, alongside the binding site surface coloured according to electrostatic properties.*

As can be seen in Figure 7, this mutated residue is located outside the substrate binding site, specifically in a turn that precedes the sequence leading to the oxyanion loop (residues 138-145), which is a vital part of the catalytic machinery that is responsible for the processing of substrate peptides[81]. Although the position of such mutation could suggest a possible destabilization of the catalytic site related to an alteration of the enzymatic activity of the protease, a visual inspection of the surroundings of residue 132 suggests that this mutation should not affect in any way the stability of the threedimensional structure of the protease, thereby not harming its ability to correctly process the substrate.

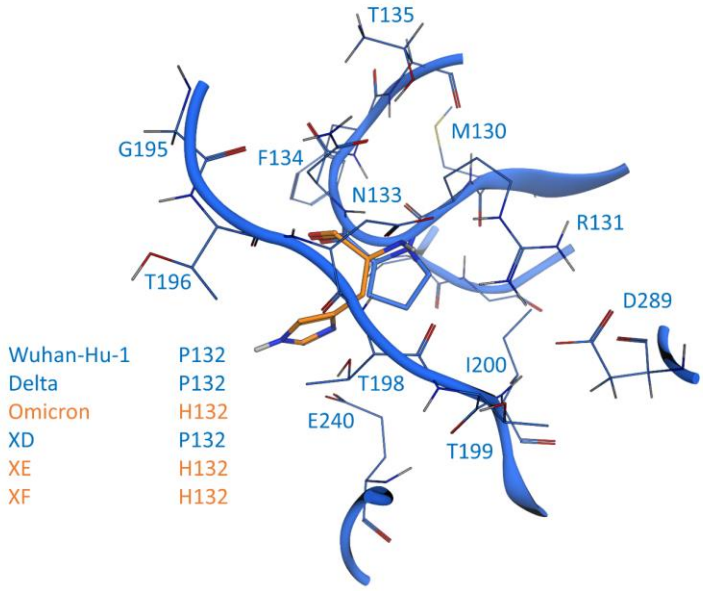


Figure 8. *Comparison between SARS-CoV-2 3CL protease (M^{pro}) from crystal structure 6Y2E (blue) and homology models of M^{pro} from five different SARS-CoV-2 variants, reported in Table 1: focus is on residue 132 (either a proline or a histidine) of SARS-CoV-2 M^{pro} and homology models of Delta, Omicron, XD, XE and XF variants M^{pro}.*

As can be seen in Figure 8, indeed, the proline residue is not involved in any intermolecular interaction relevant for the structural stability of the protease, suggesting that its only role could be limited to a joint between more relevant residues such as R131, which mediates several interactions through its sidechain guanidium group (specifically, a salt bridge with both D289 and D197, and a hydrogen bond with the backbone of T198) and its backbone (a hydrogen bond between its backbone amide proton and the amide carbonyl oxygen of T135 and another one between its carbonyl oxygen and the amide proton of F134), and N133, which is itself involved in a network of intermolecular interactions with both its backbone (hydrogen bond between its amide proton and the carboxyl oxygen of D197) and its sidechain (the amide proton donates to the carbonyl oxygen

of G195 while the carbonyl oxygen receives from the hydroxyl group of T135). These structural insights are confirmed also by a functional screening performed by Flynn et al., which showed that mutations at this position, especially the P132H found in these viral variants, are generally well tolerated, while mutations of both R131 and N133 drastically reduce or even abolish the catalytic activity of the protease[91].

Concerning the relevance of this mutation for the efficacy of existing therapeutic treatments and the development of future ones, a recent study from Greasley et al. reported the crystal structure of Nirmatrelvir in complex with the main protease from three different viral variants that presented a mutation on M^{pro92}. The analyzed mutations included the P132H, which characterizes both the Omicron SARS-CoV-2 variant and the recently found XE and XF. Greasley and collaborators established that the P132H mutation does not affect the affinity of Nirmatrelvir for the main protease catalytic site, thereby indicating the same data would be extendable also to XE and XF variants considered that they share the same P132H mutation as the Omicron variant.

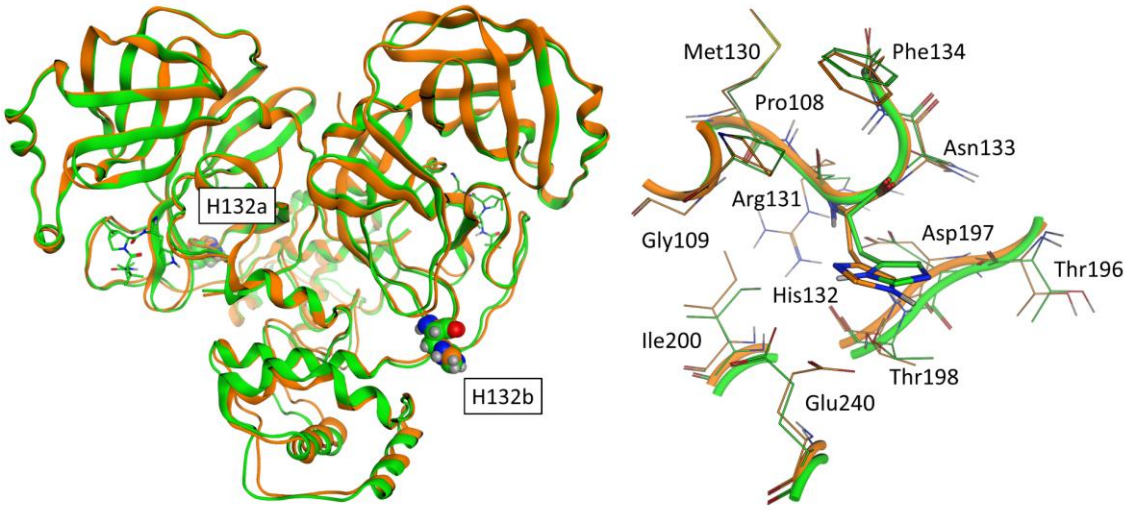


Figure 9. *This panel reports the structural superposition between the crystal structure of SARS-CoV-2 Omicron variant M^{pro} in complex with Nirmatrelvir (PDB ID: 7TLL, green) and the homology model of the XE/XF variant based on the unliganded state of the original SARS-CoV-2 virus (orange). In panel A, the whole protease structure is shown in ribbon: for visual reference, both Nirmatrelvir and H132 are reported in licorice and CPK models respectively. In panel B, a focused view of residue 132 and its surrounding residues is reported.*

As can be seen in Figure 9, despite the fact that our homology model of the XE/XF variant is based on the structure 6Y2E, which represents the SARS-CoV-2 main protease in its free/unliganded form, there is an almost perfect structural superposition between our homology model and the

experimentally resolved structure of the complex between the M^{pro} from the Omicron variant and Nirmatrelvir (PDB ID: 7TLL), as is also quantitatively assessed by the 0.67 Å R.M.S.D. between the two structures after optimal superimposition of the backbone. The congruence between our structural prediction and the experimental data supports the idea that the overall fold of the main protease is conserved across several variants and that the structural effect that residue mutations could have on the effectiveness of main protease inhibitor could be accurately predicted through the combination of computational techniques such as homology modelling, molecular docking and molecular dynamics. Moreover, based on available structural information, the high degree of structural similarity between the main proteases is not only shared by variants of the SARS-CoV-2 virus but also by other coronaviruses such as bat coronavirus[13], the Porcine transmissible Gastroenteritis virus (TGEV)[83], Human coronavirus strain 229E (HCoV)[85], Infectious bronchitis virus (IBV)[93] and MERS-CoV[94], thereby validating the pursue of novel M^{pro} inhibitors that could act as pan-coronaviral drugs and help preventing future coronavirus associated pandemics.

4. Conclusions

The recent emergence of novel recombinant SARS-CoV-2 variants, namely XD, XE and XF, poses a serious threat to the efficacy of existing therapeutic options against COVID-19. In the face of a continuous evolution of the SARS-CoV-2 genome under an evolutionary pressure opposed by the development of vaccines and by the natural immunity induced by infections, the more recent viral variants have increased both their infectiveness and their ability to escape the immune response. The structural analysis reported in this article depicts a scenario where the Spike protein, which is responsible for the ability of the virus to infect human cells by interaction with the hACE2 receptor, is the viral entity that is accumulating the highest number of mutations in an attempt to increase its affinity towards the hACE2 and decrease the one towards antibodies, while the main protease M^{pro}, a key enzyme for the virus replication cycle, is still practically identical to the wild-type virus. The difference behaviour of these two proteins in response to SARS-CoV-2 genome evolution could be vital not only for the development of efficient therapies against COVID-19 but, considering the striking structural similarities between the main protease from different viruses, also in the development of pan-coronaviral drugs that could prevent the development of future coronavirus-associated pandemics.

References

[1] Lu, H., Stratton, C. W. & Tang, Y. W. Outbreak of pneumonia of unknown etiology in Wuhan, China: The mystery and the miracle. *Journal of Medical Virology* **92**, 401 (2020).

[2] Heymann, D. L. & Shindo, N. COVID-19: what is next for public health? *The Lancet* vol. 395 542–545 Preprint at https://doi.org/10.1016/S0140-6736(20)30374-3 (2020).

[3] COVID Live - Coronavirus Statistics - Worldometer. https://www.worldometers.info/coronavirus/.

[4] Guarner, J. Three Emerging Coronaviruses in Two Decades: The Story of SARS, MERS, and Now COVID-19. *American Journal of Clinical Pathology* vol. 153 420–421 Preprint at https://doi.org/10.1093/ajcp/aqaa029 (2020).

[5] Zhou, F. *et al.* Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *The Lancet* **395**, 1054–1062 (2020).

[6] Bolcato, G., Bissaro, M., Pavan, M., Sturlese, M. & Moro, S. Targeting the coronavirus SARS-CoV-2: computational insights into the mechanism of action of the protease inhibitors lopinavir, ritonavir and nelfinavir. *Sci Rep* **10**, 20927 (2020).

[7] Gorbalenya, A. E. *et al.* The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nature Microbiology 2020 5:4* **5**, 536–544 (2020).

[8] Ksiazek, T. G. *et al.* A Novel Coronavirus Associated with Severe Acute Respiratory Syndrome. *New England Journal of Medicine* **348**, 1953–1966 (2003).

[9] Peiris, J. S. M. *et al.* Clinical progression and viral load in a community outbreak of coronavirus-associated SARS pneumonia: a prospective study. *The Lancet* **361**, 1767–1772 (2003).

[10] Li, W. *et al.* Bats are natural reservoirs of SARS-like coronaviruses. *Science* **310**, 676–679 (2005).

[11] Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C. & Garry, R. F. The proximal origin of SARS-CoV-2. *Nature Medicine 2020 26:4* **26**, 450–452 (2020).

[12] Temmam, S. *et al.* Bat coronaviruses related to SARS-CoV-2 and infectious for human cells. *Nature* (2022) doi:10.1038/s41586-022-04532-4.

[13] Pavan, M., Bassani, D., Sturlese, M. & Moro, S. Bat coronaviruses related to SARS-CoV-2: what about their 3CL proteases (MPro)? *https://doi.org/10.1080/14756366.2022.2062336* **37**, 1077–1082 (2022).

[14] Forster, P., Forster, L., Renfrew, C. & Forster, M. Phylogenetic network analysis of SARS-CoV-2 genomes. *Proceedings of the National Academy of Sciences of the United States of America* **117**, 9241–9243 (2020).

[15] Harvey, W. T. *et al.* SARS-CoV-2 variants, spike mutations and immune escape. *Nature Reviews Microbiology* **19**, 409–424 (2021).

[16] Chu, D. K. *et al.* Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis. *The Lancet* **395**, 1973–1987 (2020).

[17] Hellewell, J. *et al.* Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. *The Lancet Global Health* **8**, e488–e496 (2020).

[18] Oran, D. P. & Topol, E. J. Prevalence of Asymptomatic SARS-CoV-2 Infection. *https://doi.org/10.7326/M20-3012* **173**, 362–368 (2020).

[19] Garcia-Beltran, W. F. *et al.* Multiple SARS-CoV-2 variants escape neutralization by vaccine-induced humoral immunity. *Cell* **184**, 2372-2383.e9 (2021).

[20] Hoffmann, M. *et al.* SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* **181**, 271-280.e8 (2020).

[21] Li, Q. *et al.* The Impact of Mutations in SARS-CoV-2 Spike on Viral Infectivity and Antigenicity. *Cell* **182**, 1284-1294.e9 (2020).

[22] Frampton, D. *et al.* Genomic characteristics and clinical effect of the emergent SARS-CoV-2 B.1.1.7 lineage in London, UK: a whole-genome sequencing and hospital-based cohort study. *The Lancet. Infectious Diseases* **21**, 1246 (2021).

[23] Campbell, F. *et al.* Increased transmissibility and global spread of SARSCoV- 2 variants of concern as at June 2021. *Eurosurveillance* **26**, 1–6 (2021).

[24] Davies, N. G. *et al.* Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science* **372**, (2021).

[25] Volz, E. *et al.* Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. *Nature* **593**, 266–269 (2021).

[26] Collier, D. A. *et al.* Sensitivity of SARS-CoV-2 B.1.1.7 to mRNA vaccine-elicited antibodies. *Nature* **593**, 136–141 (2021).

[27] Chen, R. E. *et al.* Resistance of SARS-CoV-2 variants to neutralization by monoclonal and serum-derived polyclonal antibodies. *Nature Medicine* **27**, 717–726 (2021).

[28] Wang, P. *et al.* Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7. *Nature* **593**, 130–135 (2021).

[29] Haas, E. J. *et al.* Impact and effectiveness of mRNA BNT162b2 vaccine against SARS-CoV-2 infections and COVID-19 cases, hospitalisations, and deaths following a nationwide vaccination campaign in Israel: an observational study using national surveillance data. *The Lancet* **397**, 1819–1829 (2021).

[30] Twohig, K. A. *et al.* Hospital admission and emergency care attendance risk for SARS-CoV-2 delta (B.1.617.2) compared with alpha (B.1.1.7) variants of concern: a cohort study. *The Lancet Infectious Diseases* **22**, 35–42 (2022).

[31] Lopez Bernal, J. *et al.* Effectiveness of Covid-19 Vaccines against the B.1.617.2 (Delta) Variant. *New England Journal of Medicine* **385**, 585–594 (2021).

[32] Planas, D. *et al.* Reduced sensitivity of SARS-CoV-2 variant Delta to antibody neutralization. *Nature* **596**, 276–280 (2021).

[33] Liu, C. *et al.* Reduced neutralization of SARS-CoV-2 B.1.617 by vaccine and convalescent serum. *Cell* **184**, 4220-4236.e13 (2021).

[34] Mlcochova, P. *et al.* SARS-CoV-2 B.1.617.2 Delta variant replication and immune evasion. *Nature* (2021) doi:10.1038/S41586-021-03944-Y.

[35] Gao, S. J., Guo, H. & Luo, G. Omicron variant (B.1.1.529) of SARS-CoV-2, a global urgent public health alert! *Journal of Medical Virology* **94**, 1255–1256 (2022).

[36] Araf, Y. *et al.* Omicron variant of SARS-CoV-2: Genomics, transmissibility, and responses to current COVID-19 vaccines. *Journal of Medical Virology* **94**, 1825–1832 (2022).

[37] Liu, L. *et al.* Striking antibody evasion manifested by the Omicron variant of SARS-CoV-2. *Nature* **602**, 676–681 (2022).

[38] Dejnirattisai, W. *et al.* Reduced neutralisation of SARS-CoV-2 omicron B.1.1.529 variant by post-immunisation serum. *The Lancet* **399**, 234–236 (2022).

[39] Cao, Y. *et al.* Omicron escapes the majority of existing SARS-CoV-2 neutralizing antibodies. *Nature* **602**, 657–663 (2022).

[40] Hoffmann, M. *et al.* The Omicron variant is highly resistant against antibody-mediated neutralization: Implications for control of the COVID-19 pandemic. *Cell* **185**, 447-456.e11 (2022).

[41] Nemet, I. *et al.* Third BNT162b2 Vaccination Neutralization of SARS-CoV-2 Omicron Infection. *New England Journal of Medicine* **386**, 492–494 (2022).

[42] Garcia-Beltran, W. F. *et al.* mRNA-based COVID-19 vaccine boosters induce neutralizing immunity against SARS-CoV-2 Omicron variant. *Cell* **185**, 457-466.e4 (2022).

[43] Owen, D. R. *et al.* An oral SARS-CoV-2 M pro inhibitor clinical candidate for the treatment of COVID-19. *Science (1979)* **374**, 1586–1593 (2021).

[44] Pavan, M., Bolcato, G., Bassani, D., Sturlese, M. & Moro, S. Supervised Molecular Dynamics (SuMD) Insights into the mechanism of action of SARS-CoV-2 main protease inhibitor PF-07321332. *J Enzyme Inhib Med Chem* **36**, 1646–1650 (2021).

[45] Hammond, J. *et al.* Oral Nirmatrelvir for High-Risk, Nonhospitalized Adults with Covid-19. *New England Journal of Medicine* (2022) doi:10.1056/NEJMoa2118542.

[46] Investigation of SARS-CoV-2 variants: technical briefings - GOV.UK. https://www.gov.uk/government/publications/investigation-of-sars-cov-2-variants-technical-briefings.

[47] Lacek, K. A. *et al.* Identification of a Novel SARS-CoV-2 Delta-Omicron Recombinant Virus in the United States. *bioRxiv* 2022.03.19.484981 (2022) doi:10.1101/2022.03.19.484981.

[48] Aksamentov, I., Roemer, C., Hodcroft, E. & Neher, R. Nextclade: clade assignment, mutation calling and quality control for viral genomes. *Journal of Open Source Software* **6**, 3773 (2021).

[49] Aksamentov, I., Roemer, C., Hodcroft, E. & Neher, R. Nextclade: clade assignment, mutation calling and quality control for viral genomes. *Journal of Open Source Software* **6**, 3773 (2021).

[50] Molecular Operating Environment (MOE), 2019.01; Chemical Computing Group ULC, 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2021. https://www.chemcomp.com/Research-Citing_MOE.htm.

[51] Zhou, D. *et al.* Structural basis for the neutralization of SARS-CoV-2 by an antibody from a convalescent patient. *Nature Structural & Molecular Biology* **27**, 950–958 (2020).

[52] Wang, Y. *et al.* Structural basis for SARS-CoV-2 Delta variant recognition of ACE2 receptor and broadly neutralizing antibodies. *Nature Communications* **13**, 871 (2022).

[53] Yin, W. *et al.* Structures of the Omicron spike trimer with ACE2 and an anti-Omicron antibody. *Science* **375**, 1048–1053 (2022).

[54] D.A. Case et al. Amber 10, University of California, San Francisco.

[55] Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *J Mol Graph* **14**, 33–38 (1996).

[56] Huang, Y., Yang, C., Xu, X., Xu, W. & Liu, S. Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19. *Acta Pharmacologica Sinica* **41**, 1141–1149 (2020).

[57] Li, M.-Y., Li, L., Zhang, Y. & Wang, X.-S. Expression of the SARS-CoV-2 cell receptor gene ACE2 in a wide variety of human tissues. *Infectious Diseases of Poverty* **9**, 45 (2020).

[58] Jackson, C. B., Farzan, M., Chen, B. & Choe, H. Mechanisms of SARS-CoV-2 entry into cells. *Nature Reviews Molecular Cell Biology* **23**, 3–20 (2022).

[59] Lan, J. *et al.* Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* **581**, 215–220 (2020).

[60] Zhu, C. *et al.* Molecular biology of the SARs-CoV-2 spike protein: A review of current knowledge. *Journal of Medical Virology* **93**, 5729–5741 (2021).

[61] Turkahia, Y. *et al.* Pandemic-Scale Phylogenomics Reveals Elevated Recombination Rates in the SARS-CoV-2 Spike Region. *bioRxiv* 2021.08.04.455157 (2021) doi:10.1101/2021.08.04.455157.

[62] Zhang, B. *et al.* Mining of epitopes on spike protein of SARS-CoV-2 from COVID-19 patients. *Cell Research* **30**, 702–704 (2020).

[63] Aksamentov, I., Roemer, C., Hodcroft, E. & Neher, R. Nextclade: clade assignment, mutation calling and quality control for viral genomes. *Journal of Open Source Software* **6**, 3773 (2021).

[64] Hadj Hassine, I. Covid-19 vaccines and variants of concern: A review. *Reviews in Medical Virology* (2021) doi:10.1002/rmv.2313.

[65] University, S. VOC - Omicron.

[66] VanBlargan, L. A. *et al.* An infectious SARS-CoV-2 B.1.1.529 Omicron virus escapes neutralization by therapeutic monoclonal antibodies. *Nature Medicine* **28**, 490–495 (2022).

[67] Meng, B. *et al.* Recurrent emergence of SARS-CoV-2 spike deletion H69/V70 and its role in the Alpha variant B.1.1.7. *Cell Reports* **35**, 109292 (2021).

[68] Hou, Y. J. *et al.* SARS-CoV-2 D614G variant exhibits efficient replication ex vivo and transmission in vivo. *Science* **370**, 1464–1468 (2020).

[69] Weisblum, Y. *et al.* Escape from neutralizing antibodies 1 by SARS-CoV-2 spike protein variants. *eLife* **9**, 1 (2020).

[70] Lubinski, B. *et al.* Functional evaluation of the P681H mutation on the proteolytic activation of the SARS-CoV-2 variant B.1.1.7 (Alpha) spike. *iScience* **25**, 103589 (2022).

[71] Zahradník, J. *et al.* SARS-CoV-2 variant prediction and antiviral drug design are enabled by RBD in vitro evolution. *Nature Microbiology* **6**, 1188–1198 (2021).

[72] Berman, H. M. The Protein Data Bank. *Nucleic Acids Research* **28**, 235–242 (2000).

[73] Wang, Y. *et al.* Structural basis for SARS-CoV-2 Delta variant recognition of ACE2 receptor and broadly neutralizing antibodies. *Nature Communications* **13**, 871 (2022).

[74] Lan, J. *et al.* Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* **581**, 215–220 (2020).

[75] Collier, D. A. *et al.* Sensitivity of SARS-CoV-2 B.1.1.7 to mRNA vaccine-elicited antibodies. *Nature* **593**, 136–141 (2021).

[76] Masre, S. F., Jufri, N. F., Ibrahim, F. W. & Abdul Raub, S. H. Classical and alternative receptors for SARS-CoV-2 therapeutic strategy. *Reviews in Medical Virology* **31**, 1–9 (2021).

[77] Sartore, G. *et al.* In silico evaluation of the interaction between ACE2 and SARS-CoV-2 Spike protein in a hyperglycemic environment. *Scientific Reports* **11**, 22860 (2021).

[78] Jin, Z. *et al.* Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. *Nature 2020 582:7811* **582**, 289–293 (2020).

[79] Xia, B. & Kang, X. Activation and maturation of SARS-CoV main protease. *Protein & cell* **2**, 282–290 (2011).

[80] Snijder, E. J., Decroly, E. & Ziebuhr, J. The Nonstructural Proteins Directing Coronavirus RNA Synthesis and Processing. *Advances in Virus Research* **96**, 59–126 (2016).

[81] Fornasier, E. *et al.* A new inactive conformation of SARS-CoV-2 main protease. *Acta Crystallogr D Struct Biol* **78**, 363–378 (2022).

[82] Wu, F. *et al.* A new coronavirus associated with human respiratory disease in China. *Nature 2020 579:7798* **579**, 265–269 (2020).

[83] Anand, K. *et al.* Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra alpha-helical domain. *EMBO J* **21**, 3213–3224 (2002).

[84] Chen, H. *et al.* Only one protomer is active in the dimer of SARS 3C-like proteinase. *Journal of Biological Chemistry* **281**, 13894–13898 (2006).

[85] Anand, K., Ziebuhr, J., Wadhwani, P., Mesters, J. R. & Hilgenfeld, R. Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs. *Science* **300**, 1763–1767 (2003).

[86] Ullrich, S. & Nitsche, C. The SARS-CoV-2 main protease as drug target. *Bioorg Med Chem Lett* **30**, 127377 (2020).

[87] Bassani, D., Pavan, M., Bolcato, G., Sturlese, M. & Moro, S. Re-Exploring the Ability of Common Docking Programs to Correctly Reproduce the Binding Modes of Non-Covalent Inhibitors of SARS-CoV-2 Protease Mpro. *Pharmaceuticals* **15**, 180 (2022).

[88] Bissaro, M. *et al.* Inspecting the Mechanism of Fragment Hits Binding on SARS-CoV-2 M ^{pro} by Using Supervised Molecular Dynamics (SuMD) Simulations. *ChemMedChem* **16**, 2075–2081 (2021).

[89] Luttens, A. *et al.* Ultralarge Virtual Screening Identifies SARS-CoV-2 Main Protease Inhibitors with Broad-Spectrum Activity against Coronaviruses. *J Am Chem Soc* **144**, 2905–2920 (2022).

[90] Zhang, C. H. *et al.* Potent Noncovalent Inhibitors of the Main Protease of SARS-CoV-2 from Molecular Sculpting of the Drug Perampanel Guided by Free Energy Perturbation Calculations. *ACS Cent Sci* **7**, 467–475 (2021).

[91] Flynn, J. M. *et al.* Comprehensive fitness landscape of SARS-CoV-2 M^{pro} reveals insights into viral resistance mechanisms. *bioRxiv* 2022.01.26.477860 (2022) doi:10.1101/2022.01.26.477860.

[92] Greasley, S. E. *et al.* Structural basis for Nirmatrelvir in vitro efficacy against SARS-CoV-2 variants. *bioRxiv* 2022.01.17.476556 (2022) doi:10.1101/2022.01.17.476556.

[93] Xue, X. *et al.* Production of authentic SARS-CoV M(pro) with enhanced activity: application as a novel tag-cleavage endopeptidase for protein overproduction. *J Mol Biol* **366**, 965–975 (2007).

[94] Ho, B. L. *et al.* Critical Assessment of the Important Residues Involved in the Dimerization and Catalysis of MERS Coronavirus Main Protease. *PLoS One* **10**, (2015).

Implementing a scoring function based on interaction fingerprint for Autogrow4: Protein Kinase CK1δ as a case study

Matteo Pavan, Silvia Menin, **Davide Bassani**, Mattia Sturlese, Stefano Moro

Pavan, M., Menin, S., Bassani, D., Sturlese, M., & Moro, S. (2022). Implementing a Scoring Function Based on Interaction Fingerprint for Autogrow4: Protein Kinase CK1δ as a Case Study. *Frontiers in Molecular Biosciences*, *0*, 629. https://doi.org/10.3389/FMOLB.2022.909499

Abstract

In the last twenty years, Fragment-Based Drug Discovery (FBDD) has become a popular and consolidated approach within the drug-discovery pipeline, due to its ability to bring several drug candidates to clinical trials, some of them even being approved and introduced to the market.

A class of targets that have proven to be particularly suitable for this method is represented by kinases, as demonstrated by the approval of BRAF inhibitor Vemurafenib. Within this wide and diverse set of proteins, protein kinase CK1δ is a particularly interesting target for the treatment of several widespread neurodegenerative diseases such as Alzheimer's disease, Parkinson's disease, and amyotrophic lateral sclerosis.

Computational methodologies such as molecular docking are already routinely and successfully applied in fragment-based drug discovery campaigns alongside experimental techniques, both in the hit-discovery and in the hit-optimization stage. Concerning this, the open-source software Autogrow, developed by the Durrant lab, is a semi-automated computational protocol that exploits a combination between a genetic algorithm and a molecular docking software for de-novo drug design and lead optimization.

In the present work, we present and discuss a modified version of the Autogrow code that implements a custom scoring function based on the similarity between the interaction fingerprint of investigated compounds to a crystal reference. To validate its performance, we performed both a *denovo* and a lead-optimization run (as described in the original publication), evaluating the ability of

our fingerprint-based protocol to generate compounds similar to known CK1δ inhibitors based on both the predicted binding mode and electrostatic and shape similarity in comparison with the standard Autogrow protocol.

1. Introduction

Protein kinase CK1δ is a Ser/Thr protein kinase belonging to the Casein Kinase 1 family. In mammals, 7 distinct genes encoding for Casein kinase proteins are present, each producing a different isoform (α, β, γ1, γ2, γ3, δ, and ε)[1]. CK1 family proteins use exclusively ATP as a phosphate source for their kinase activity, which is carried out by the protein in its monomeric form. Each isoform is constitutionally active and does not require the presence of a cofactor to exert its activity[2].

From a biological function point of view, the members of this family have been historically related to different physiological mechanisms, such as cell replication[3], DNA repair[4], and circadian rhythm[5].

From a structural perspective, the members of the CK1 family are characterized by the typical bilobed structure of the globular Ser-Thr kinase proteins, with the N-term lobe consisting mainly of β-sheets and a larger C-term lobe, constituted primarily of α-helices. The two domains are connected by a protein region named the “hinge region”, which forms a highly conserved pocket for ATP binding[2].

As for other members of the CK1 family, CK1δ recognizes the canonical phospho-primed structural motif *pSer/pThr-X_{1-2}-Ser/Thr*, where X stands for any amino acid and pSer/pThr represents the phospho-primed residue[6]. The CK1 kinases are also able to recognize non-phosphorylated sequences, as far as they contain strongly acidic residues (Asp or Glu) that can make up for the absence of the phosphorylated residue[3]. The structural motif that can be recognized by the CK1 proteins is widespread in many cellular proteins and, because of this, over 140 substrates have been reported both *in vitro* and *in vivo*[2], underlining the pleiotropic character of this protein family. Due to the great variability of its substrates, CK1δ is involved in many cellular pathways, among which the main ones are the Wnt-pathway, the Hippo pathway, the p53 regulation pathway, and the Hedgehog pathway[3].

The endogenous regulation of CK1δ, on the other hand, can be carried out through various mechanisms, including autophosphorylation or phosphorylation by other protein kinases[7], [8], interactions with other protein and/or cellular components, and subcellular sequestration[3], [9]. In addition, homodimerization excludes ATP from the binding site, thus inhibiting kinase activity[10], [11].

In recent years, several studies have highlighted the importance of CK1δ in neurodegenerative diseases, particularly tauopathies such as Alzheimer's disease (AD), Parkinson's disease (PD), and amyotrophic lateral sclerosis (ALS)[12]. In addition to having unknown etiology, these illnesses are all characterized by loss of neuronal function, with neurotransmitter deficiency, misfolding, and protein aggregation[13]. Clinical symptoms are manifested differently, depending on the neuronal area involved[14].

AD is a progressive neurodegenerative disorder that mainly involves the neurons of the hippocampus[15]. On the extracellular side, the main marker of the disease is represented by the accumulation of β-amyloid peptides, produced by β-secretase 1 and γ-secretase enzymes, which leads to neuronal death[16]. Meanwhile, on the intracellular part, the illness presents lesions related to both cytoplasmic accumulations of vacuoles with abnormal dimensions and dense granular content and the assembly of fibrils and filaments within the neuronal body. These types of lesions are both characterized by the accumulation of hyperphosphorylated Tau protein in the filaments, but also within the vacuoles[17].

The correlation between CK1δ activity and tau protein aggregates in various neurodegenerative diseases has been confirmed by co-immunoprecipitation studies, which highlight that the presence of CK1δ is associated with hyperphosphorylated tau aggregates[18], [19]. CK1δ phosphorylates tau protein at the Ser202 / Thr205 and Ser369 / Ser404 residues *in vitro*[12], [20]. The phosphorylation sites are the same as those involved in binding with tubulin, highlighting the key role of kinase in the pathogenesis of AD[18]. It is not clear whether the hyperactivity of CK1δ is due to an over-transcription of its gene, altered protein turnover, or both causes, but it has been observed that the concentration of the protein CK1δ in an AD-affected hippocampus is 30 times higher than normal[17].

In PD, on the other hand, the pathology is characterized by the accumulation of Lewy bodies, consisting of aggregates of α-synuclein hyperphosphorylated by CK1δ at the level of Ser129

residues[21]. This process determines a massive loss of neuronal function at the substantia nigra level[22].

CK1δ also plays a key role in Amyotrophic Lateral Sclerosis (ALS), a neurodegenerative disorder in which intracellular inclusions of TDP-43 (TAR DNA-binding Protein) are found in the frontotemporal lobe. It was established that TDP-43 can be phosphorylated by CK1δ at 29 different sites[23].

These pathologies are all characterized by the absence of effective pharmacological therapy: in fact, there are no EMA-approved drugs on the market that can solve, and therefore cure, these diseases, but there are only palliative therapies for the temporary improvement of the patient's quality of life, thus resulting in a high social cost[24]. For these reasons, CK1δ appears as an interesting therapeutical target in the field of neurodegeneration, as witnessed by the increasing interest in the research for inhibitory candidates for this protein during the last 15 years.

Concerning the identification of novel kinase inhibitors, an approach that has proven to be particularly successful is the so-called Fragment-Based Drug Discovery (FBDD), as demonstrated by the approval of the BRAF inhibitor Vemurafenib[25] (employed in the treatment of metastatic melanoma) and by several other kinase inhibitors which are at various stage of clinical trials[26], [27].

This approach revolves around the exploitation of "fragments", i.e., compounds that respect the "Rule of Three" (molecular weight < 300, number of hydrogen bond donor/acceptor <= 3, logP<=3), as a starting point for the rational development of novel mature, drug-like, active molecules[28], [29].The main reason for the success of FBDD is the ability to sample a larger portion of the chemical space compared to the one occupied by drug-like molecules, thus increasing the success rate in finding novel scaffolds for targets of interest[30].

This methodology heavily relies on very sensitive biophysical methods such as X-ray crystallography (XRC), nuclear magnetic resonance (NMR), or surface plasmon resonance (SPR), to perform large screening campaigns on libraries composed of molecules with low molecular weight and high solubility, to find hit compounds[31], [32]. These hit fragments have usually a low affinity for the target, ranging from low mM to high μM (hence the need for very sensitive screening techniques), but usually have a higher binding efficiency compared to traditional drug-like molecules, being able to establish high-quality interaction with the target[33]. Fragment hits can then be easily combined

(either through a linking or a merging process) or chemically modified (growing) to increase their affinity for the target, allowing for the development of potent and selective active compounds[34].

Alongside the aforementioned experimental techniques, in the last decade, a prominent role in FBDD campaigns has been played by computer-aided drug discovery (CADD) techniques such as molecular docking or molecular dynamics[35]. These computational approaches have been routinely and successfully applied for performing large screening on virtual fragment libraries, for the characterization of the fragment interaction mode with the target and to aid the fragment-to-lead optimization in a less time-consuming, more rational, and more efficient way. Some examples of software developed specifically designed for FBDD are LUDI[36], HOOK[37], CAVEAT[38], RECORE[39]. Moreover, commercial drug discovery suites such as Schrödinger, MOE, and OpenEye have implemented several tools related to the fragment optimization process.

Among the plethora of software available for FBDD, the open-source software Autogrow, developed by the Durrant lab, is particularly interesting. As thoroughly described in the work of Spiegel et al.[40], the open-source software Autogrow is a Python written code that combines a genetic algorithm with docking calculation based on the Vina[41] docking software to perform a semi-automatized process for both de-novo drug design and lead optimization. The latest release of the Autogrow (version 4.0.3, the one used in this work), was developed with the idea of making the codebase modular thus allowing the third-party implementation of different conversion scripts, molecular docking programs, scoring functions, and reaction libraries, to better suit the need of different research groups.

A recent scientific work published by our laboratory led to the identification of seven novel fragment compounds that bind the hinge region of CK1δ with a low-micromolar IC_{50}[42]. Attracted by the idea of exploiting a semi-automatized computational protocol for the optimization of our newly discovered fragment compounds, we decided to investigate if this protocol would be suitable for our needs. Since it is notorious that molecular docking programs are usually very efficient and optimized with regards to the conformational search, but are usually lacking in the scoring phase[43], [44] (especially for molecules like fragments that deviates from the drug-like chemical space on which these scoring functions have been trained[45], [46]), we decided to investigate if the implementation of a different scoring protocol based on protein-ligand interaction fingerprint would improve the performance of

the Autogrow protocol, concerning the ability of the program to generate compounds similar to known inhibitors based on their interaction scheme and electrostatic and shape similarity.

2. Materials and Methods

2.1. Hardware Overview

Each general molecular modeling operation has been performed on a Linux Workstation equipped with an 8 core Intel Xeon® CPU E5-1620 CPU. For more intensive calculations, such as the Autogrow runs, a 64 cores AMD Opteron™ Processor 6376 CPU cluster was exploited. Both the workstation and the cluster run Ubuntu 16.04 as their operating system.

2.2. Structures Preparation

In the case of protein kinase CK1δ, 23 protein-ligand complexes between the protein and small drug-like molecules are available in the Protein Data Bank[47](PDB ID; 3UYT, 3UZP, 4HGT, 4HNF, 4KB8, 4KBA, 4KBC, 4KBK, 4TN6, 4TW9, 4TWC, 5IH5,5IH6, 5MQV, 5OKT, 5W4W, 6F1W, 6F26, 6GZM, 6HMP, 6HMR, 6RCG, 6RCH). In the context of this work, the crystals with codes 6RU6, 6RU7, and 6RU8 were not considered in the study because they contain the natural substrate adenosine-5'-diphosphate. One of the structures (PDB ID: 4KB8[48]) is composed of two different CK1δ-ligand complexes. For this reason, the system has been separated into two different entries (namely 4KB8-A and 4KB8-B). Because of this, the total number of complexes considered in our study is 24.

Each of the mentioned complexes has been downloaded and properly prepared for subsequent computational analysis with the “Structure Preparation” tool implemented in the Molecular Operating Environment (MOE)[49] 2019.01 suite. The missing hydrogen atoms were appropriately added with the MOE “Protonate 3D” program (setting the pH for the protonation at a value of 7.4) and were then energetically minimized according to the AMBER10: EHT[50] force field implemented in MOE. After the preparation phase, the protein-ligand complexes have been properly aligned and superposed with the MOE dedicated tool, to make the binding site coordinates coherent among the different crystallographic structures. These complexes were saved and used at a later stage for the generation of the pharmacophore model (see section 2.4).

Afterward, each ligand has been separated from its respective protein. All the small molecules were collected in a database and prepared for docking calculations exploiting several packages from the QUACPAC OpenEye[51] suite. For each molecule the most probable tautomeric state was selected

with the “tautomers” program, the three-dimensional coordinates were rebuilt using the “Omega” tool, the partial charges were attributed with the “MolCharge” program according to the MMFF94 force field, and finally, the dominant protonation state at pH 7.4 was determined by the “FixPka” tool.

2.3. Cross-docking

Each of the aforementioned 24 CK1δ crystallographic ligands, prepared as described in section 2.2, was docked inside each of the correspondent 24 CK1δ protein structures exploiting two different molecular docking pieces of software, namely GOLD[52] (based on a genetic algorithm, developed and distributed with a commercial license from CCDC) and PLANTS[53] (an Ant-Colony-Optimization docking algorithm, developed by the University of Tübingen and free-for-use for academics).

This approach was chosen to follow the principles of “consensus docking”[54], which is based on the fact that data obtained combining results coming from docking programs that operate in an orthogonal way are associated with higher robustness.

For both GOLD and PLANTS, 10 poses per molecule were collected. The default parameters were used for both protocols. Concerning the choice of the scoring function, Chemscore was selected for GOLD, while $PLANTS_{ChemPLP}$ was selected for PLANTS.

A total of 1152 (24 ligands x 24 proteins x 2 docking protocols) independent docking runs were performed, and the results were then analyzed by making use of an in-house Python script. The script collects the RMSD between each docking pose and the correspondent crystal reference pose, outputting two different plots. The first plot is a heatmap that illustrates the RMSD values for the best docking pose generated for each ligand onto each protein. The second plot is a histogram that re-elaborates the previous results to give a visual representation of the “success rate” of each protein: a successful docking run is obtained when the RMSD between the docking pose and the crystal reference is below the arbitrary chosen 2 Å threshold value so that the “success-rate” is defined by the percentage ratio of the successful docking runs for each protein (i.e., the percentage of docking experiments where the RMSD falls below the threshold value).

2.4. Pharmacophore Modeling

Based on previously published works on the same target, we took advantage of the structural information about known inhibitors of CK1δ in the form of crystal structures of their complex with the kinase deposited in the Protein Data Bank. The same 24 protein-ligand complexes mentioned in section 2.2 were subjected to the MOE Pharmacophore model tool: shared interaction features (with a 50% threshold value for feature retention) were then used in the generation of the pharmacophore model.

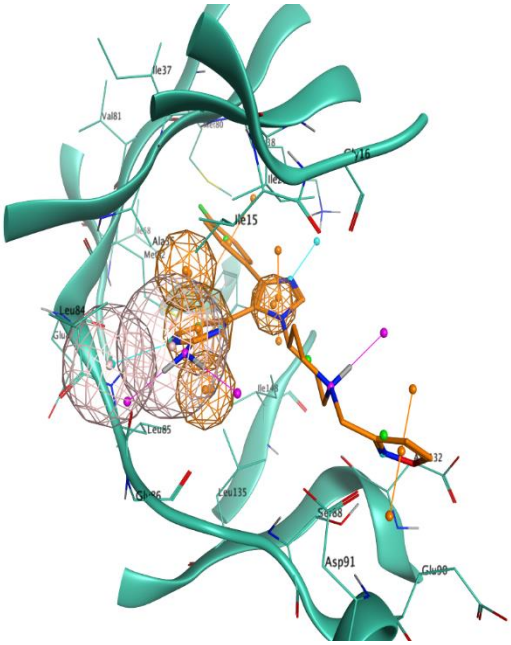


Figure 1. *Visual representation of the pharmacophore model used in this scientific work. Features are represented as spheres. Orange spheres indicate an aromatic ring, with an orientation determined by the small orange pin, while the pink spheres indicate a hydrogen bond donor/acceptor. For visual reference, the 4TN6 complex is also reported in this figure, with the protein represented in teal ribbons and the PFO ligand represented as orange sticks.*

As can be seen in Figure 1, the final model consisted of 4 features (represented as spheres in the image), namely a hydrogen bond donor and a hydrogen bond acceptor interacting with Leu85, an aromatic ring in the proximity of the hinge region, and another aromatic ring adjacent to the first one in the inner part of the binding pocket.

2.5. Autogrow

Autogrow4[55] is a fully open-source code written in Python and developed by the Durrant lab that combines a genetic algorithm with docking calculation based on the Vina[56] docking software (version 1.2.0) to perform a semi-automatized process for both de-novo drug design and lead optimization.

Molecules are submitted to the program in the form of SMILES strings. The genetic algorithm part of the code uses a series of synthetically feasible reactions to perform a defined number of mutation and crossover operations (i.e., growing and merging) on submitted chemical entities, creating a full population (called generation) of molecules to feed to the molecular docking program.

This generation is then docked using the Vina docking software. After the docking stage, the genetic algorithm retrieves the score for each docking pose, which it uses to rank molecules and pick the most fitted members of the generation to promote them to the next generation. This iterative process is repeated for a user-defined number of generations, or until an earlier termination criterion is met.

The code is released under the Apache2 license, is freely available at https://durrantlab.pitt.edu/autogrow4/, and works both in a Python 2.7 and >=3.6 environment. A detailed description of how the latest Autogrow release works is provided in the work of Spiegel et al.[40].

Two different versions of the Autogrow code were used in this scientific work. The first one was downloaded from the official repository and used as is, without any modifications to the source code. The second one was the result of an in-house modification of the source code performed to customize the scoring stage of the docking process. The traditional Autogrow protocol uses the Vina standard scoring function (from now on, defined as VINA), which encompasses some elements of knowledge-based potentials and others of a typical empiric scoring function[56]. Instead, our modified version of the Autogrow code implements an alternative scoring function (from now on, defined as IFP_{CS}) based on the similarity between protein-ligand interaction fingerprints.

The crystal complex of a known inhibitor is chosen as reference (in our case, the ligand PFO from complex 4TN6 was chosen) and its binding mode is codified into a bit vector exploiting the InteractionFingerprint function from the fingerprint module of the Open Drug Discovery Toolkit[57]

Python Library. This function converts the protein-ligand interaction into a bit array according to the residue of choice and the type of interaction. Each protein residue is represented by eight bits, one for each type of interaction considered (hydrophobic contacts, aromatic face to face, aromatic edge to face, hydrogen bond with protein acting as donor, hydrogen bond with protein acting as acceptor, salt bridge with protein acting as the positively charged member, salt bridge with protein acting as the positively negative member and ionic bond with a metal ion), so that the final vector will have a size of r*8, where r stands for the number of protein residues.

During the scoring phase of our custom Autogrow run, each docking pose is also codified into an Interaction Fingerprint vector, the same way as for the crystal reference. Then, the two vectors are transformed from sparse to dense making use of the appropriate functions from the Numpy Python library, before the comparison between the reference and the query fingerprint is executed using the cosine similarity metrics, exploiting the appropriate function of the Scikit-learn Python library. The resulting score, which ranges from 1 (indicating a complete agreement and coherence between the two binding modes) to 0 (indicating that the two binding modes are not coherent), is then multiplied by -1 to comply with the selection mechanism of Autogrow genetic algorithm, which favors the most negative scores, as is usually the case for most classic scoring functions, like the one used by Vina.

$$IFP_{CS} = \frac{A \cdot B}{\|A\|\|B\|} * (-1)$$

***Equation 1**: mathematic formulation of the IFP_{CS} scoring function. The IFP_{CS} scoring function is the inverse of the cosine similarity between two vectors, A and B, representing the Interaction Fingerprint for the reference and the query ligand respectively. Values range from -1 (indicating maximum coherence between the two binding modes) to 0 (indicating the lowest possible correspondence between the two binding modes).*

3. Results

3.1. Cross-docking

Since 24 different protein-ligand complexes were available for CK1δ (the target for our computational study), but only one at a given time can be used for docking calculations, we had to carefully evaluate which one was the most suitable for our needs. The choice of the protein structure to use for docking calculation is not trivial, for several reasons. When a ligand gets in contact with a protein, the binding event may cause a change in the structure of the protein itself[58]. These modifications are mainly depictable in the binding site and may also be extended to other regions. In a crystallographic complex, this effect is highlightable by differences in the shape of the binding site among the different crystal structures available for a single protein[59].

One of the possible approaches to accomplish this task, which is the one that we used in our workflow, is known as “cross-docking”[60]. This technique consists in taking all the protein-ligand complexes available for a target, separating all the ligands from their respective co-crystallized structure, and docking all the different ligands in the binding site of each different protein structure. By analyzing the docking results, it is possible to define which is the crystallographic protein structure that has the highest tendency to correctly reproduce ligands' crystallographic conformation.

For these reasons, we performed a cross-docking experiment on our 24 CK1δ complexes to decide which one to pick for subsequent calculation. Each ligand was docked into each protein structure using two different docking protocols, GOLD-Chemscore and PLANTS-PLANTS$_{ChemPLP}$, for a total of 1152 independent docking runs. For each ligand, the root-mean-square deviation (RMSD) between each docking pose and the crystallographic conformation was calculated. The poses with the lowest RMSD in each docking run were selected and their RMSDs were plotted, obtaining the graphs represented in Figure 2. A detailed description of the methodology used for the cross-docking experiment is provided in the Materials and Methods, section 2.3.

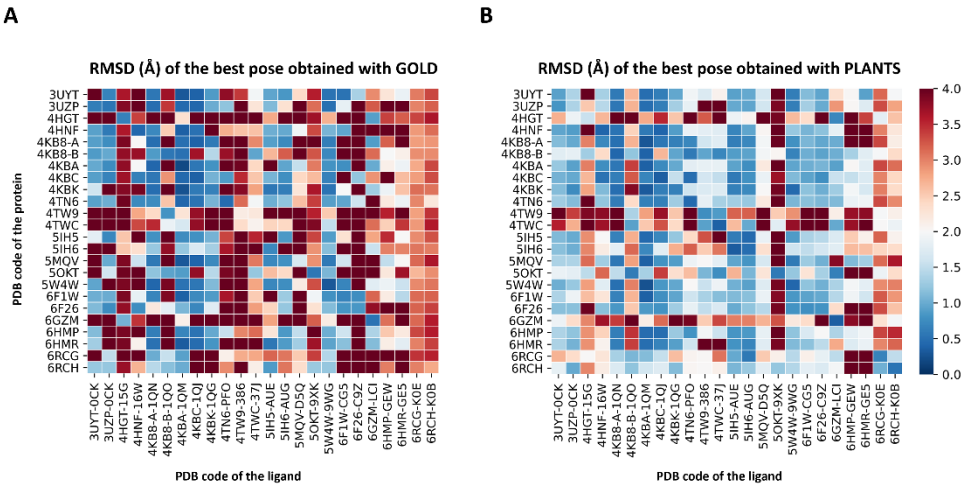


Figure 2. *This figure contains two heatmaps that summarize the results of the cross-docking experiment performed before the Autogrow runs to select the protein structure to use for subsequent calculations. Panel A reports the results for the GOLD-Chemscore protocol, while Panel B encompasses the results of the PLANTS-PLANTS$_{ChemPLP}$ one. On the vertical axis, the PDB code of the protein is reported, while on the horizontal axis the PDB code of the ligand is indicated. The colored squares report the RMSD values for the best docking pose generated by the two docking protocols according to the color bar located on the right side of the image: color ranges from Blue (indicating a low RMSD (minimum value is 0 Å, indicating a perfect superposition between the docking pose and the crystal reference) to Red (maximum value is 4 Å, indicating a high deviation between the docking pose and the crystal reference).*

To visualize the results more clearly, the data from the plots reported in Figure 2 were re-elaborated to obtain a single indicator of the performance of each protein in reproducing the correct binding mode for docked ligands. We opted for calculating the "success rate" for each protein structure: a 2 Å threshold value was chosen to discriminate between successful and unsuccessful docking runs. For each protein, the percentage of successful docking runs (the "success rate") was calculated accordingly and plotted in a histogram.

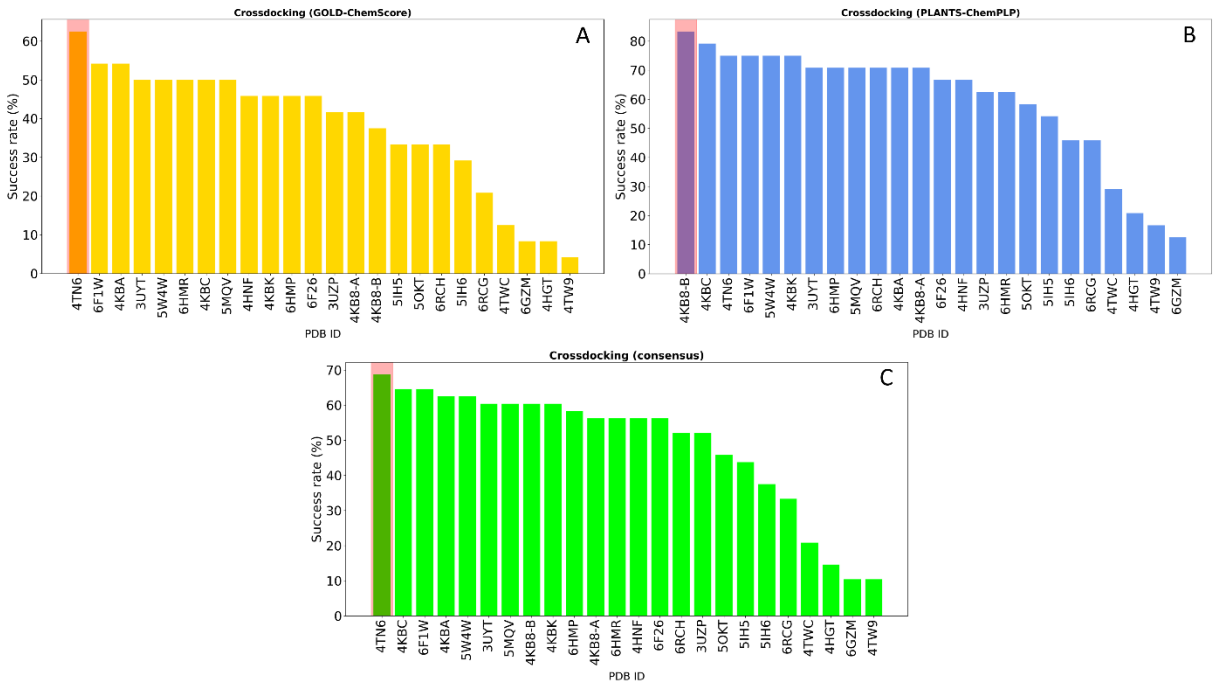


Figure 3. *This figure shows the overall "success rate" in reproducing the correct crystallographic binding mode for each of the 24 CK1δ complexes considered in the study. The "success rate" is defined as the percentage of successful docking runs for each protein in the cross-docking experiment, where a successful docking run is defined as a docking calculation where the RMSD between the best docking pose and the crystal reference falls below an arbitrarily chosen threshold value of 2 Å. Panel A reports the results for the GOLD-Chemscore protocol. Panel B reports the results for the PLANTS-PLANTS$_{ChemPLP}$ protocol. Panel C encompasses the combined "success rate" for each protein, defined as the average between the success rate for each protocol. Protein from the complex 4TN6 was chosen as the most representative CK1δ structure for successive calculations.*

Figure 3 encompasses the results of this second analysis, reporting the success rate for both the GOLD-Chemscore and PLANTS-PLANTS$_{ChemPLP}$ protocols. Moreover, since we adopted the principle of "consensus docking", as mentioned in section 2.3, we decided to calculate the average success rate between the two docking protocols. As can be seen in Figure 3, the overall "success rate" obtained by the combination of data from the two docking protocols indicates the protein from the complex 4TN6 as the protein that is, on average, more able than the other ones to correctly reproduce the crystallographic binding mode of docked ligands. Although the difference in the success rate between the first and the second protein is low, in the context of several consequential docking runs where thousands of compounds are considered at a given time even small differences in the percentage success rate could have a big impact on the quality of the run, considering that the prioritization of compounds from one generation to another is based upon their docking-predicted ability to retain

the interaction features that characterize the binding mode of known inhibitors. For this reason, we used the protein 4TN6 as a representative CK1δ structure for our subsequent calculations with Autogrow.

3.2. Benchmark DeNovo Run

To assess the performance of our alternative, fingerprint-based, Autogrow protocol (defined as IFP_{CS}, while the traditional one is defined from now on as VINA), we first performed a benchmark *denovo* run, using the same conditions as the ones described in the work of Spiegel et al.[40]

A 30-generation run was performed for each protocol, using the "Fragment_MW_100_to_150.smi" library provided in the Autogrow repository and described in the original publication. Configuration files for both *denovo* runs in the JSON format are available in the Supplementary Materials, while a detailed description of both Autogrow and our alternative scoring approach is described in Materials and Methods, section 2.5.

In order to validate the performance of both protocols, we opted for evaluating the quality of the generated compounds by filtering each generation of poses using a pharmacophore model. This filter, which has already proved to identify true binders in previous related works[42], [61], was used to retain only those poses which complied with known requirements for binding to the CK1δ pocket. This metric was used to determine if there is any advantage in incorporating a knowledge-based element in the generation of novel potential inhibitors of CK1δ, steering the compound selection process towards the ones that assume a pharmacophore-like binding mode. These pharmacophore-like compounds were then characterized by calculating their molecular weight and the similarity of their shape and electrostatic properties to crystal CK1δ inhibitors taken as reference. For this purpose, the EON[56] package from the OpenEye suite was used. Each compound passing the pharmacophore filter was compared with each crystallographic ligand, calculating the electrostatic and shape similarity (ET_{combo}). The best value for each ligand was extracted and used for the elaboration of the a posteriori analysis, whose results are reported in Figure 4 and Figure 5.

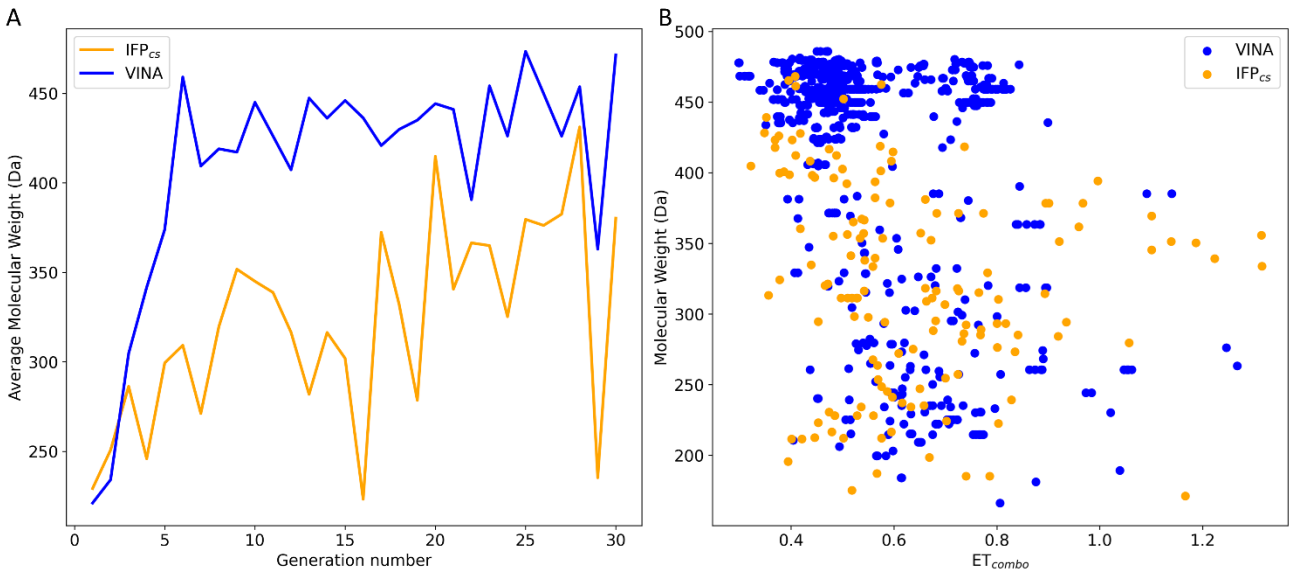


Figure 4. *This figure compares the performance of the two Autogrow protocols in the benchmark denovo runs regarding their ability to generate compounds that pass the pharmacophore filter. The VINA protocol is reported as a blue line, while the IFP$_{CS}$ one is reported as an orange line. Panel A depicts, for each protocol, the average molecular weight of compounds within the population that pass the pharmacophore filter on a per-generation basis. The vertical axis reports the molecular weight, while the horizontal axis reports the generation number. Panel B depicts, for each protocol, the distribution of generated compounds that pass the pharmacophore filter regarding their molecular weight and the similarity of shape and electrostatic properties to crystal inhibitors taken as reference. The vertical axis reports the average molecular weight in Da, while the horizontal axis reports the ET$_{combo}$ value. Blue dots represent compounds generated by the VINA protocol, while orange dots represent compounds generated by the IFP$_{CS}$ one.*

As can be seen in Figure 4 (panel A), which shows the average molecular weight of compounds that pass the pharmacophore filter for each generation, the VINA protocol rapidly reaches the peak of the average molecular weight (around generation 6), while our IFP$_{CS}$ protocol has a slower but regular growth that reaches values comparable to the VINA protocol from around generation 27 onwards. This difference is probably related to the fact that the VINA scoring function is biased towards the selection of larger compounds, which can make a good number of non-specific interactions with the target, while our IFP$_{CS}$ one is biased towards the selection of compounds that have a similar interaction pattern compared to a reference compound, regardless of their dimensions.

As depicted in Figure 4 (panel B), which illustrates the distribution of generated compounds across all generations concerning their molecular weight and their electrostatic and shape similarity with crystal

CK1δ inhibitors, this different selection process results in the production of compounds with different properties: the blue dots, which represent the compounds generated by the VINA protocol, are mostly located in the left-upper portion of the graph, indicating that most of the compounds generated by the traditional protocol have a high molecular weight but a low level of similarity with known inhibitors. On the contrary, the upper-right part of the graph (high molecular weight, high electrostatic, and shape similarity) is mostly populated with orange dots, which represent the compounds generated by our IFP_{CS} protocol.

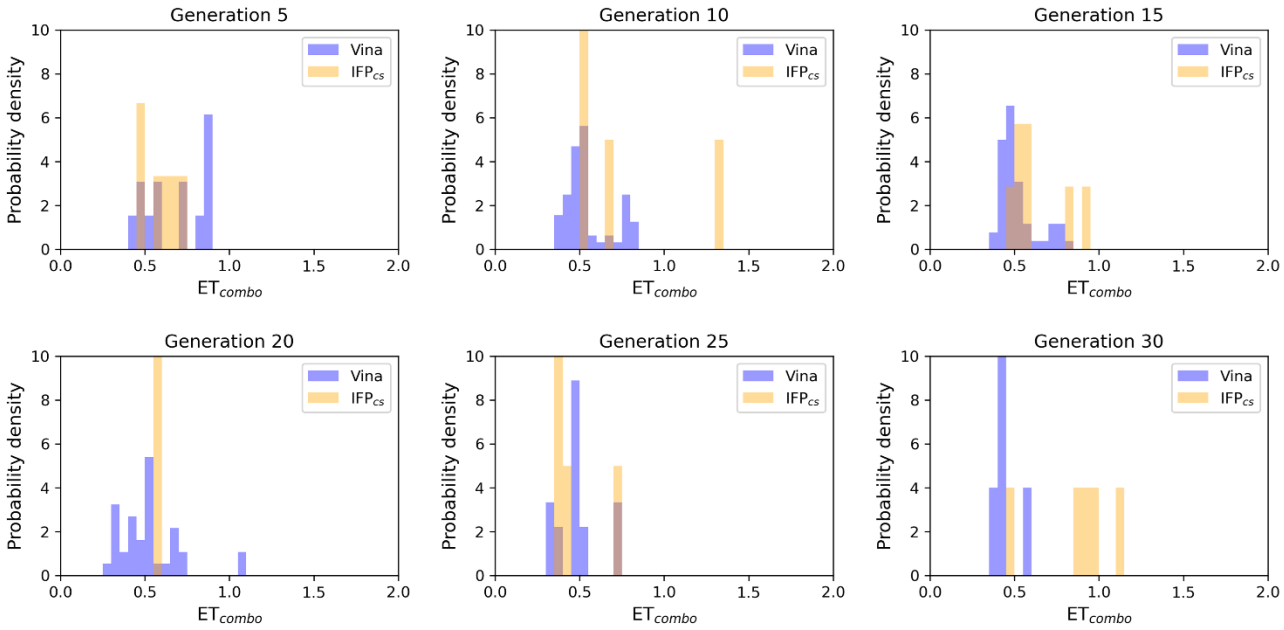


Figure 5. *This figure encompasses the ability of the two Autogrow protocols in the benchmark denovo run to produce compounds that have a high degree of similarity concerning shape and electrostatic properties to the crystallographic ligands, chosen as reference. The probability distribution of the ET_{combo} score for compounds populating each generation is reported as a histogram, where the vertical axis reports the probability density while the horizontal axis reports the ET_{combo} value. Two distributions are reported within each plot: the blue bars refer to compounds generated with the VINA protocol, while the orange bars refer to compounds generated with the IFP_{CS} one.*

The difference in the selection process is also highlighted in Figure 5, which illustrates the distribution of compounds across a representative subset of generations concerning their electrostatic and shape similarity: the graph clearly shows how the VINA protocol does not improve the similarity of generated compounds while increasing the number of generations. On the contrary, the orange

population (which represents the compounds generated by the IFP_{CS} protocol) gradually shifts towards the right part of the plot passing from earlier to later stage generations, indicating that the compounds passing the pharmacophore filter increase their electrostatic and shape similarity passing from one generation to another. Another comparison of the performances of the two protocols is given in Figure 6, which reports the progressive enrichment in compounds with a high degree of similarity to reference inhibitors within the total population. An example of a high-scoring compound generated by our IFP_{CS} protocol is reported in Figure 7, where its chemical structure and the comparison between its docking-predicted binding mode and crystal pose of the PFO ligand from reference crystal complex 4TN6 is shown.

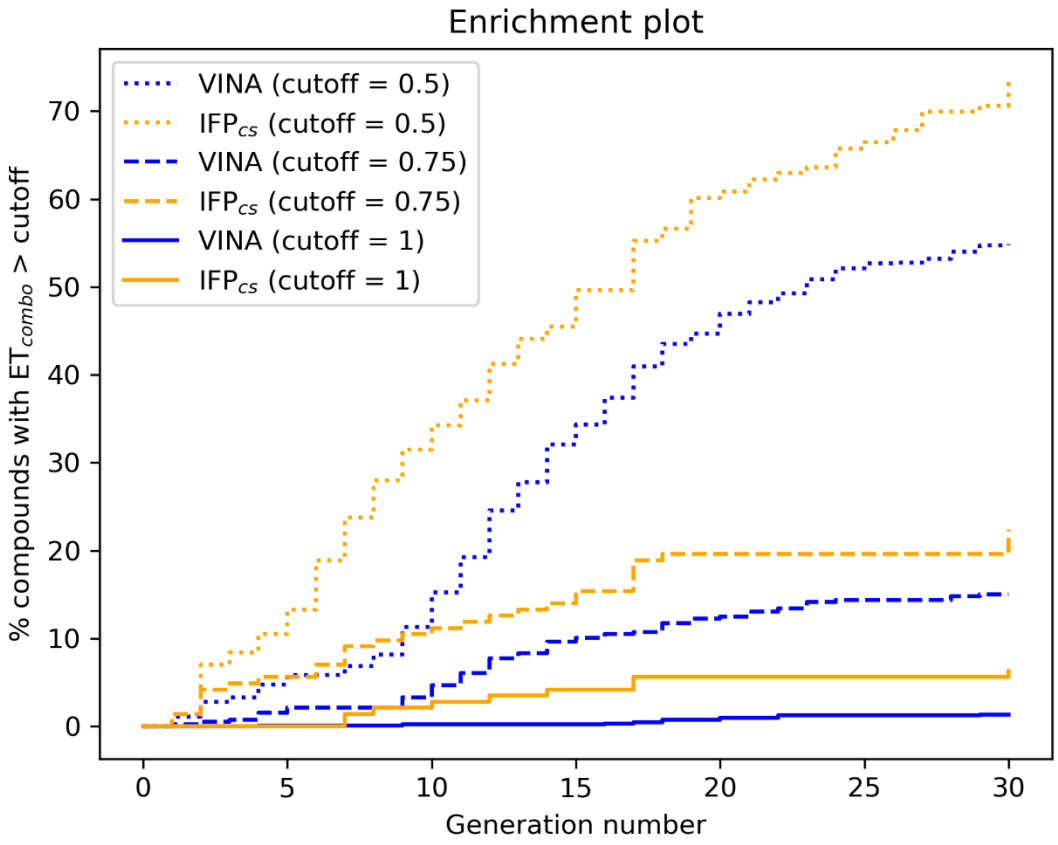


Figure 6. *This figure illustrates the capability of the two different Autogrow protocols in the benchmark denovo run to produce compounds that have a high degree of similarity concerning shape and electrostatic properties to the crystallographic ligands, chosen as reference. For each generation, the percentage of compounds within the total population whose ET_{combo} exceeds a defined threshold value is reported. Three different cutoff values are reported, 0.50, 0.75, and 1.00 respectively.*

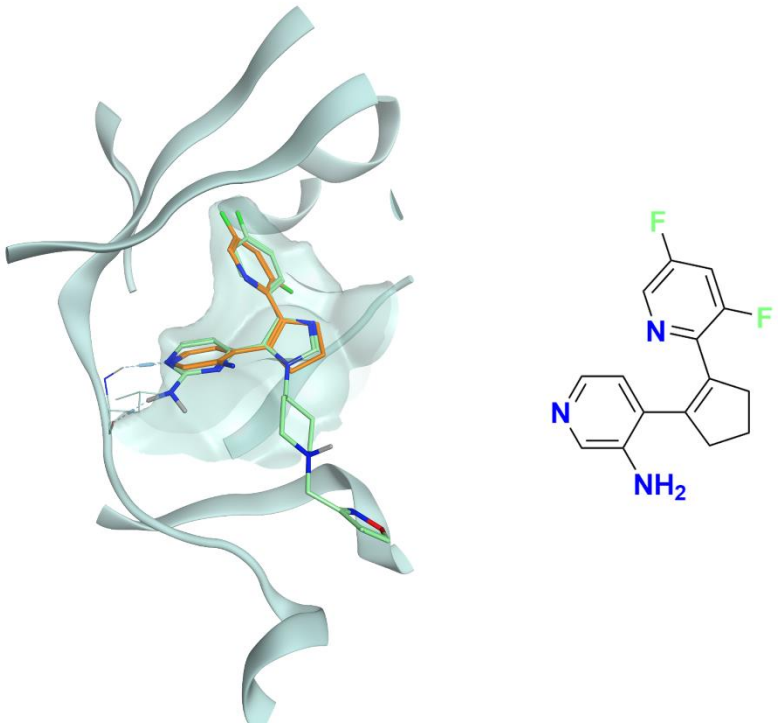


Figure 7. *This figure reports the superposition between the docking-predicted binding mode of a high scoring compound (MMS1) from the benchmark denovo run performed with the IFP$_{CS}$ scoring protocol and the reference crystal binding pose of compound PFO from the structure deposited in the Protein Data Bank with accession code 4TN6. On the left part of the image, the protein kinase CK1δ ATP binding site is reported in teal ribbon, the pose of compound MMS1 is shown in orange sticks while the pose of compound PFO is shown in green sticks. On the right part of the image, the chemical structure of compound MMS1 is reported.*

3.3. Benchmark Lead Optimization Run

To further evaluate the validity of our custom scoring protocol, we performed also a benchmark lead optimization run, using once again the same conditions as the ones reported in the work of Spiegel et al.[40] A 5-generation run was performed for each protocol, using a library composed of the 24 crystallographic ligands mentioned in the previous sections and other 316 fragments obtained from the fragmentation of crystallographic ligands exploiting the "fragmenter_of_smi_mol.py" Python script provided by the Autogrow developers, using the BRICS fragmentation rule, for a total of 340 compounds fed to the algorithm. In this case, also, configuration files for both benchmark runs in the JSON format are available in the Supplementary Materials.

To assess the performance of both protocols, we applied the same criteria described previously for the *denovo* runs, focusing once again on compounds passing the pharmacophore filter described in

section 2.4 and characterizing them about their molecular weight and electrostatic and shape similarity compared to crystal CK1δ inhibitors.

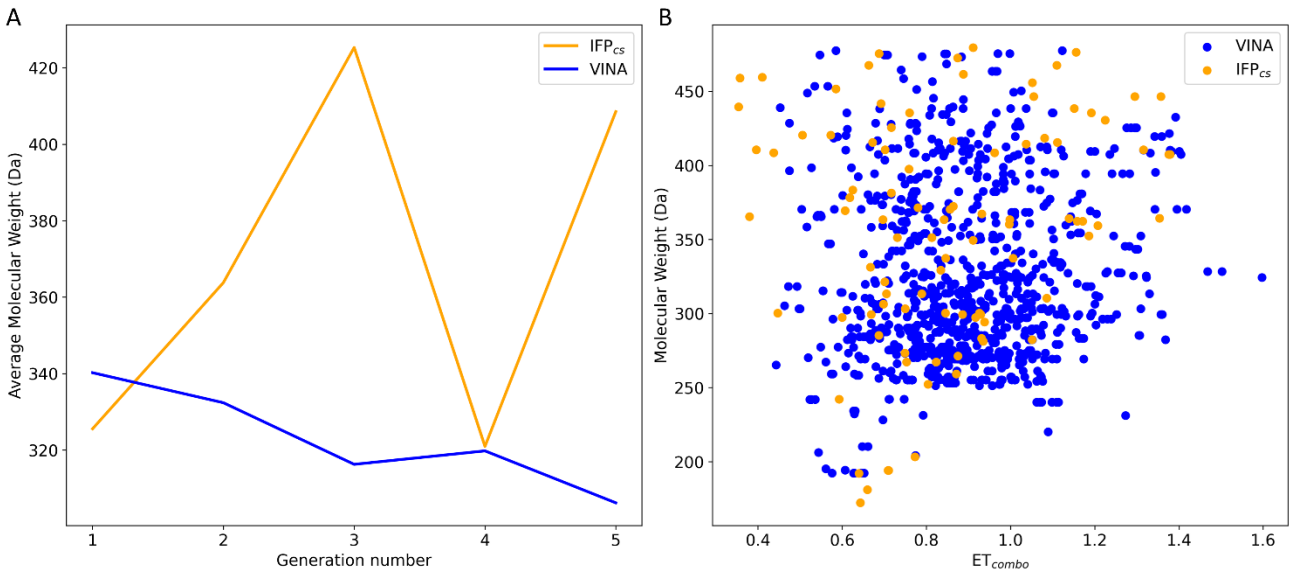


Figure 8. *This figure compares the performance of the two Autogrow protocols in the benchmark lead optimization runs regarding their ability to generate compounds that pass the pharmacophore filter. The VINA protocol is reported as a blue line, while the IFP$_{CS}$ one is reported as an orange line. Panel A depicts, for each protocol, the average molecular weight of compounds within the population that pass the pharmacophore filter on a per-generation basis. The vertical axis reports the molecular weight, while the horizontal axis reports the generation number. Panel B depicts, for each protocol, the distribution of generated compounds that pass the pharmacophore filter regarding their molecular weight and the similarity of shape and electrostatic properties to crystal inhibitors taken as reference. The vertical axis reports the average molecular weight in Da, while the horizontal axis reports the ET$_{combo}$ value. Blue dots represent compounds generated by the VINA protocol, while orange dots represent compounds generated by the IFP$_{CS}$ one.*

Figure 8 (panel B) illustrates the distribution of compounds across all five generations regarding their ET$_{combo}$ and their molecular weight: as can be seen, there is little to no difference between the two protocols, with the two populations being practically superimposable. However, contrary to what might be suggested by this plot, there is a significant difference in the performances of the two protocols, which is highlighted in both Figure 8 (panel A), Figure 9, and Figure 10.

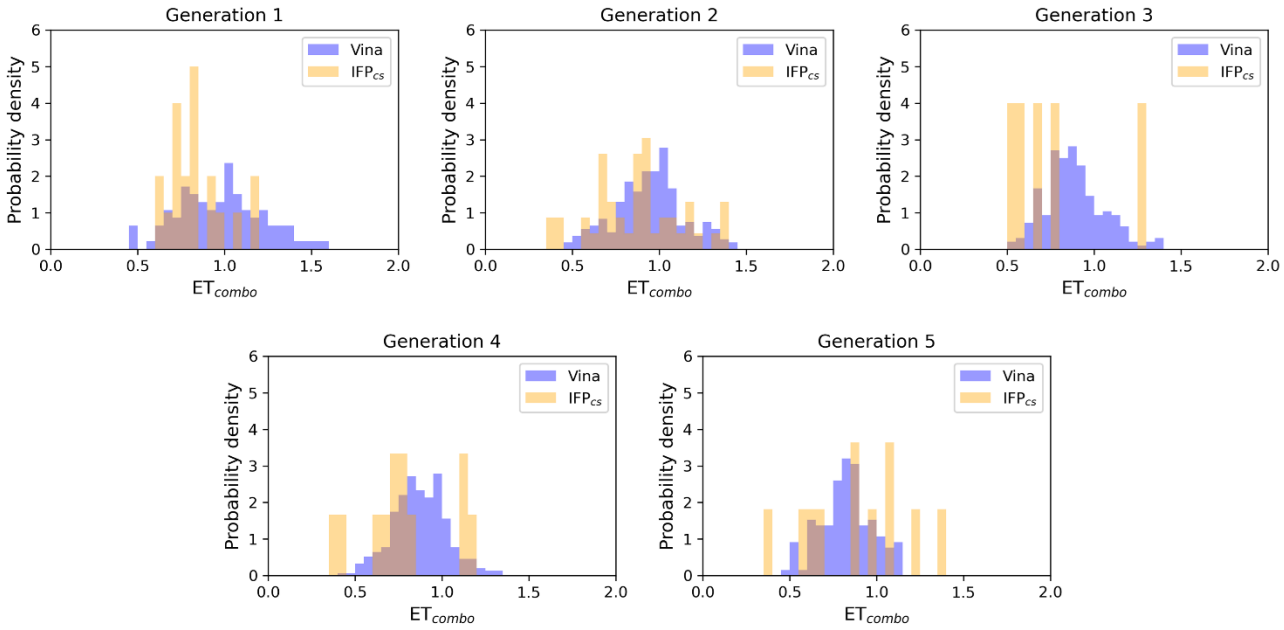


Figure 9. *This figure encompasses the ability of the two Autogrow protocols in the benchmark lead optimization run to produce compounds that have a high degree of similarity with regards to shape and electrostatic properties to the crystallographic ligands, chosen as reference. The probability distribution of the ET$_{combo}$ score for compounds populating each generation is reported as a histogram, where the vertical axis reports the probability density while the horizontal axis reports the ET$_{combo}$ value. Two distributions are reported within each plot: the blue bars refer to compounds generated with the VINA protocol, while the orange bars refer to compounds generated with the IFP$_{CS}$ one.*

As can be noticed in Figure 8 (panel A), the average molecular weight of compounds passing the pharmacophore filter grows by about 90 Da passing from the first to the last generation in the case of our IFP$_{CS}$ protocol. On the contrary, not only the average molecular weight of pharmacophore-like compounds generated by the traditional VINA protocol does not increase with the number of generations, but slightly decreases over time, falling even below the average molecular weight of the first generation derived from the IFP$_{CS}$ protocol. Furthermore, Figure 9 illustrates how, as previously seen in the benchmark *denovo* run, the similarity of compounds passing the pharmacophore filter increases over time when the IFP$_{CS}$ scoring protocol is adopted, while it slightly decreases and does not improve over time in the case of the traditional VINA scoring protocol.

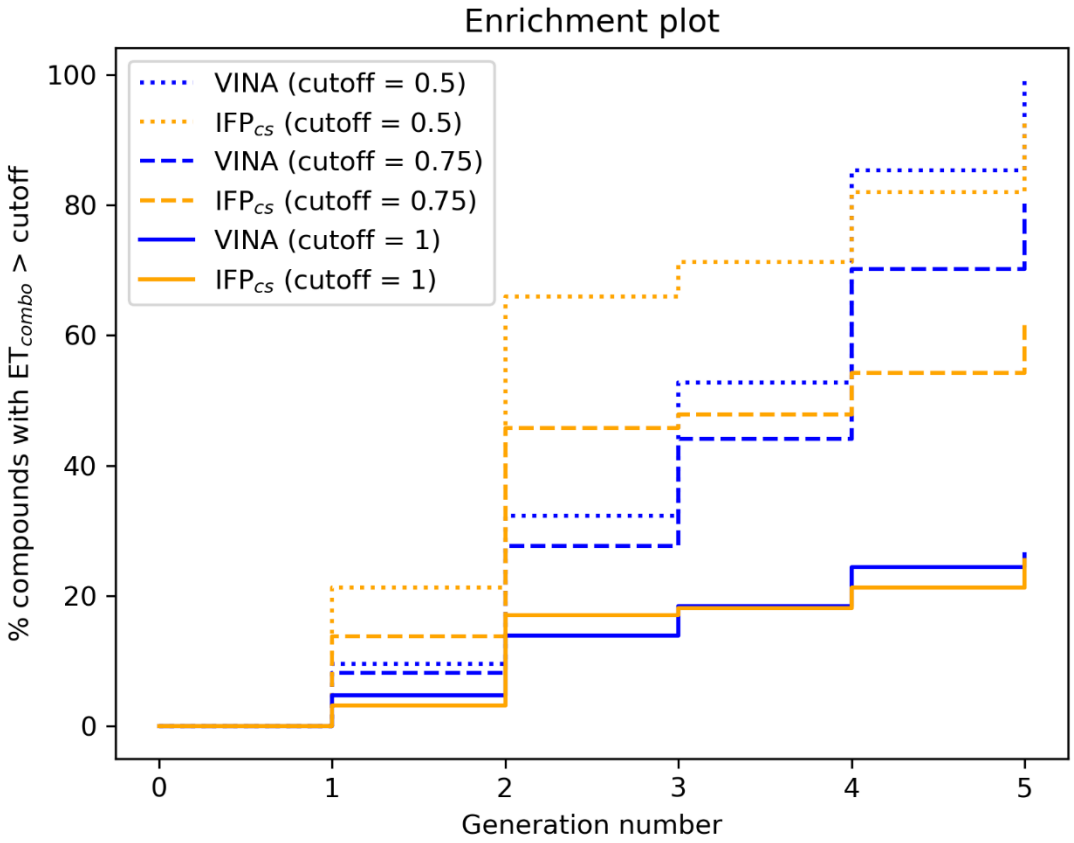


Figure 10. *This figure illustrates the capability of the two different Autogrow protocols in the benchmark lead optimization run to produce compounds that have a high degree of similarity concerning shape and electrostatic properties to the crystallographic ligands, chosen as reference. For each generation, the percentage of compounds within the total population whose ET_{combo} exceeds a defined threshold value is reported. Three different cutoff values are reported, 0.50, 0.75, and 1.00 respectively.*

Particularly, this trend is also confirmed by Figure 10, which shows how the IFP_{CS} protocol can produce a quicker enrichment of the population in high similarity compounds compared to the traditional VINA one. As for the previous case, an example of a high-scoring compound generated in the last and final generation of the IFP_{CS} run is reported in Figure 11.

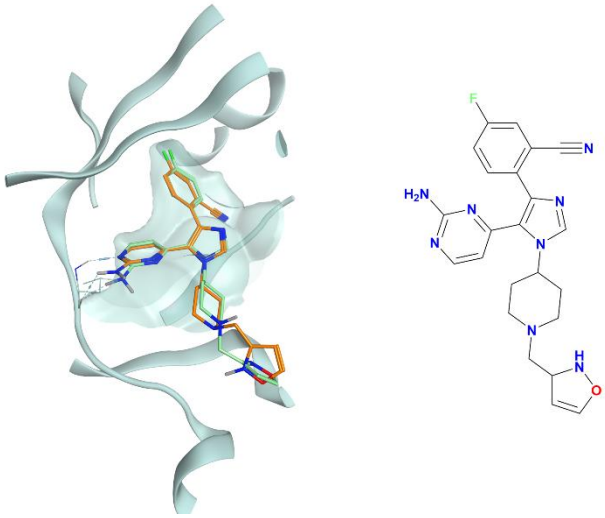


Figure 11. This figure reports the superposition between the docking-predicted binding mode of a high scoring compound (MMS2) from the benchmark lead optimization run performed with the IFP_{CS} scoring protocol and the reference crystal binding pose of compound PFO from the structure deposited in the Protein Data Bank with accession code 4TN6. On the left part of the image, the protein kinase CK1δ ATP binding site is reported in teal ribbon, the pose of compound MMS2 is shown in orange sticks while the pose of compound PFO is shown in green sticks. On the right part of the image, the chemical structure of compound MMS2 is reported.

3.4. Prospective DeNovo Run

Encouraged by the results of our benchmark runs, we decided to perform a prospective run with the IFP_{CS} protocol, applying the same operating conditions as before. This time, the starting library was modified to add to the compounds used for the benchmark runs also 7 fragment ATP-competitive CK1δ inhibitors identified during a previous virtual screening campaign from our laboratory[42]. The idea behind this run was to evaluate the ability of our IFP_{CS} scoring protocol to generate interesting novel potential CK1δ inhibitors derived from in-house, readily available compounds.

The chemical structure of the seven fragments used in this run is reported in Figure 12.

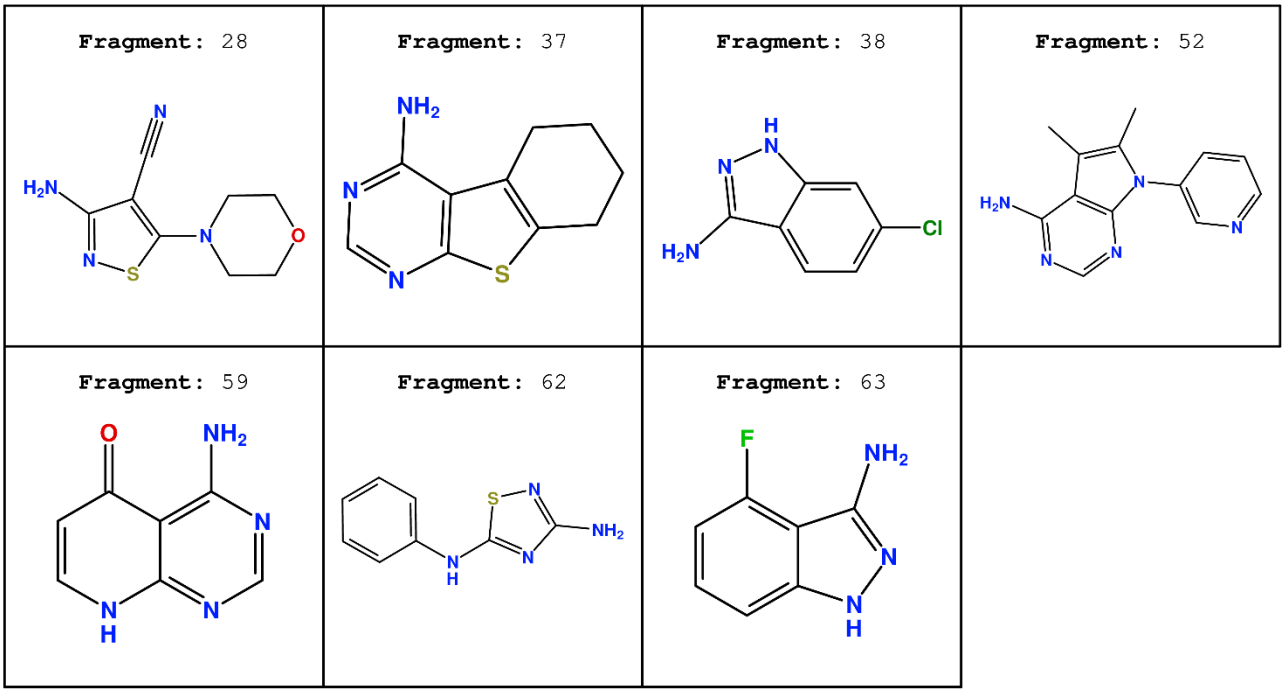


Figure 12. *Chemical structure of the seven fragment CK1δ inhibitors derived from the work of Bolcato et al.*[42]

To verify the quality of this run, we performed the same analysis as for the benchmark runs. The results of this analysis are summarized in Figures 13, 14, and 15 respectively.

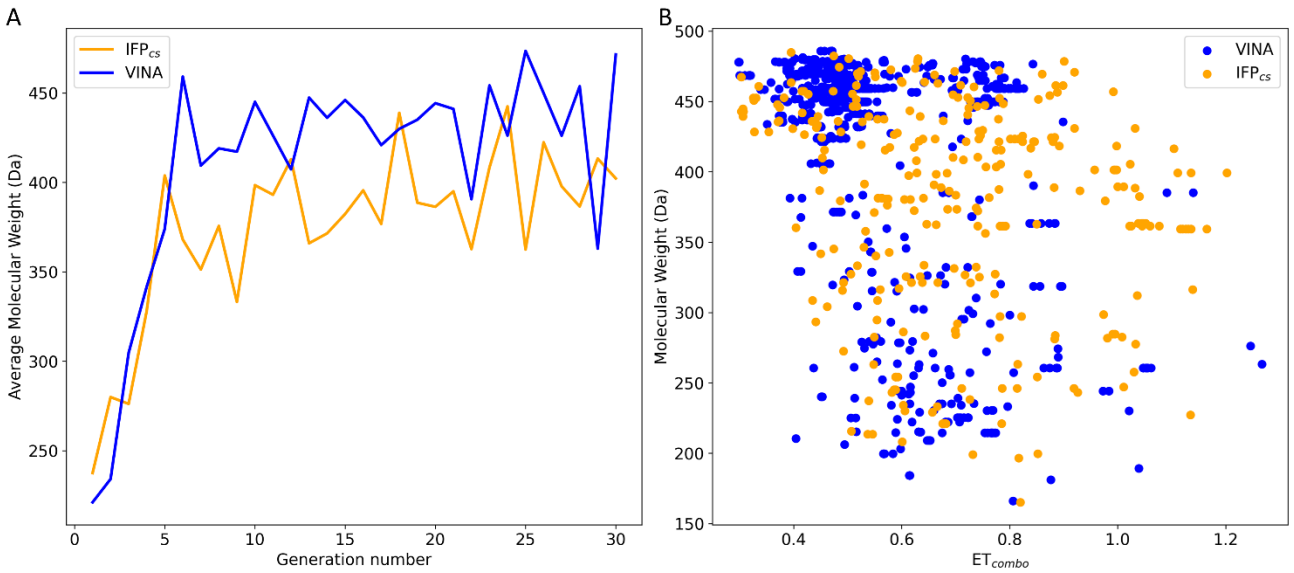


Figure 13. *This figure compares the performance of the two Autogrow protocols in the prospective denovo runs regarding their ability to generate compounds that pass the pharmacophore filter. The VINA protocol is reported as a blue line, while the IFP$_{CS}$ one is reported as an orange line. Panel A depicts, for each protocol, the average molecular weight of compounds within the population that pass the pharmacophore filter on a per-generation basis. The vertical axis reports the molecular weight, while the horizontal axis reports the generation number. Panel B depicts, for each protocol, the distribution of generated compounds that pass the pharmacophore filter regarding their molecular weight and the similarity of shape and electrostatic properties to crystal inhibitors taken as reference. The vertical axis reports the average molecular weight in Da, while the horizontal axis reports the ET$_{combo}$ value. Blue dots represent compounds generated by the VINA protocol, while orange dots represent compounds generated by the IFP$_{CS}$ one.*

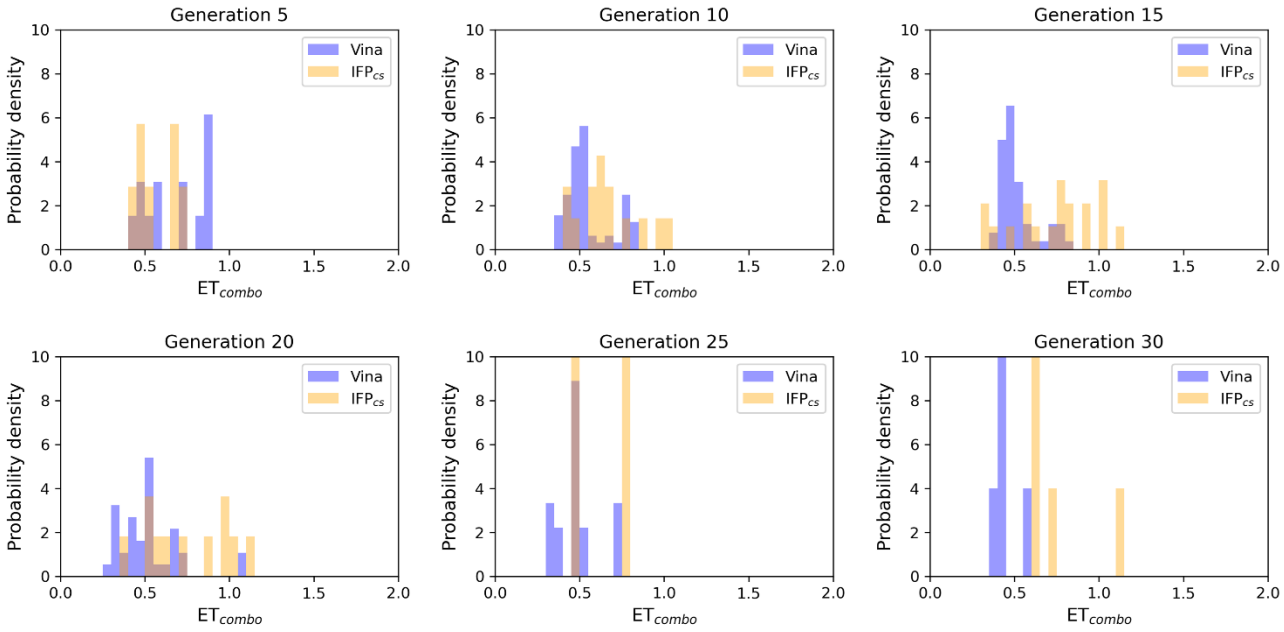


Figure 14. *This figure encompasses the ability of the two Autogrow protocols in the prospective denovo run to produce compounds that have a high degree of similarity with regards to shape and electrostatic properties to the crystallographic ligands, chosen as reference. The probability distribution of the ET$_{combo}$ score for compounds populating each generation is reported as a histogram, where the vertical axis reports the probability density while the horizontal axis reports the ET$_{combo}$ value. Two distributions are reported within each plot: the blue bars refer to compounds generated with the VINA protocol, while the orange bars refer to compounds generated with the IFP$_{CS}$ one.*

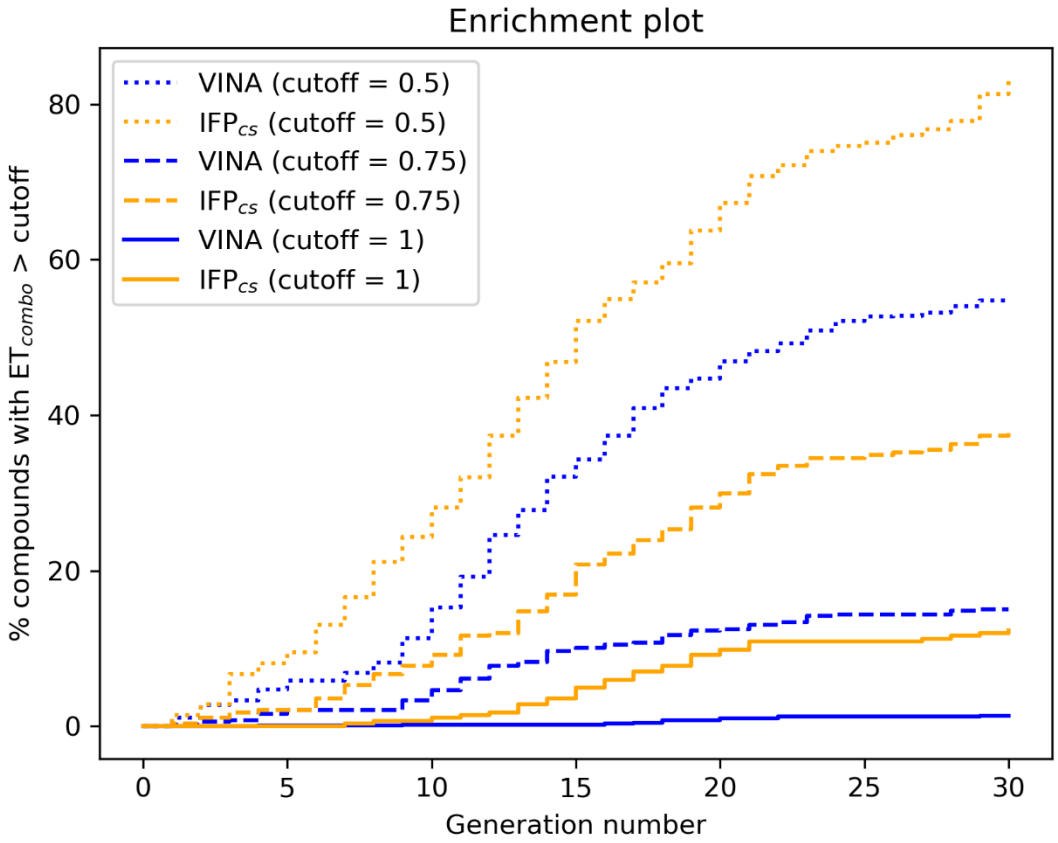


Figure 15. *This figure illustrates the capability of the two different Autogrow protocols in the prospective denovo run to produce compounds that have a high degree of similarity concerning shape and electrostatic properties to the crystallographic ligands, chosen as reference. For each generation, the percentage of compounds within the total population whose ET$_{combo}$ exceeds a defined threshold value is reported. Three different cutoff values are reported, 0.50, 0.75, and 1.00 respectively.*

As remarked in Figure 14, the same trend seen in the benchmark *denovo* run can be observed also in the case of this prospective run: while the VINA protocol is not able to increase the shape and electrostatic similarity to known inhibitors over time, the IFP$_{CS}$ protocol can produce a shift of the orange population towards higher ET$_{combo}$ values. As illustrated by Figure 13, which reports a comparison between the benchmark *denovo* run performed with the VINA protocol and the prospective *denovo* run carried out with the IFP$_{CS}$ protocol, the trend in both the distribution of compounds regarding their molecular weight and ET$_{combo}$ and the growth of molecular weight over time are similar to the benchmark *denovo* run. Figure 13 (panel B) clearly shows how the upper-right portion of the graph, which hosts compounds with both high molecular weight and ET$_{combo}$ values, is

populated exclusively by orange dots, which represent compounds generated with the IFP_{CS} scoring protocol. Interestingly, Figure 14 (panel A) highlights how there is much less difference in the growth rate of molecular weight between the IFP_{CS} run (which is contaminated by the presence of our 7 CK1δ-inhibiting fragments) and the benchmark VINA run, suggesting that performances of the IFP_{CS} could improve if some high-quality pharmacophore-like fragments are included in the starting library. However, despite the quicker growth of molecular weight, the quality of generated compounds follows the same trend seen in the benchmark *denovo* run, as reported in Figure 15. As for the previous cases, an example of a high-scoring compound is reported in Figure 16.

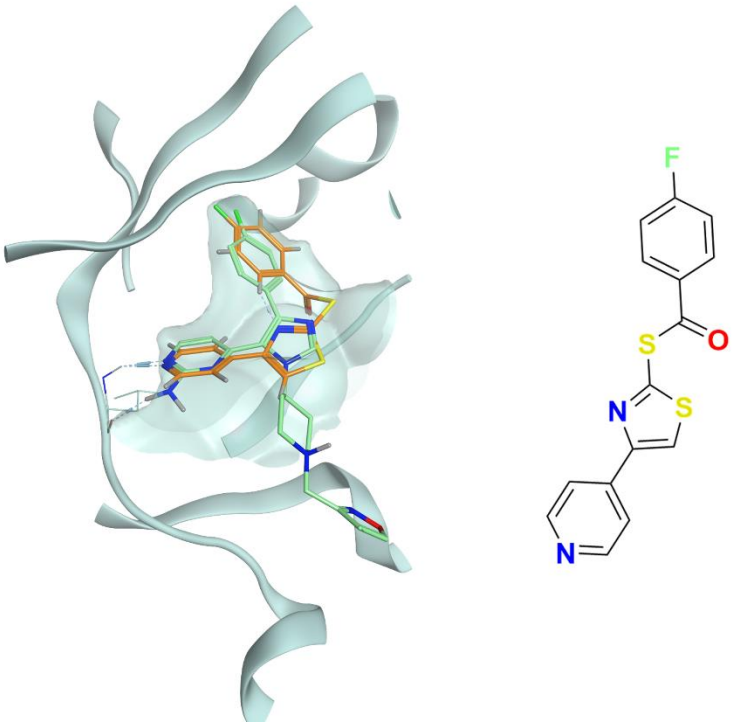


Figure 16. *This figure reports the superposition between the docking-predicted binding mode of a high scoring compound (MMS3) from the benchmark denovo run performed with the IFP_{CS} scoring protocol and the reference crystal binding pose of compound PFO from the structure deposited in the Protein Data Bank with accession code 4TN6. On the left part of the image, the protein kinase CK1δ ATP binding site is reported in teal ribbon, the pose of compound MMS3 is shown in orange sticks while the pose of compound PFO is shown in green sticks. On the right part of the image, the chemical structure of compound MMS3 is reported.*

4. Discussion

The open-source software Autogrow4[55] is an interesting piece of code that utilizes a combination between a genetic algorithm and the Vina[41] molecular docking software to semi-automatize the processes of fragment growing and lead optimization. Thanks to the modular nature of the codebase, we implemented an alternative scoring protocol (IFP_{CS}) based on the similarity of protein-ligand interaction fingerprint between a crystal reference and query compounds, exploiting the appropriate function from the open-source library Open Drug Discovery Toolkit[57], and we compared its performances with the traditional Autogrow scoring protocol (VINA), which is based on the Autodock Vina scoring function.

The VINA protocol uses a scoring function that incorporates some elements of knowledge-based potentials and some others of empirical scoring functions. As is the case for many scoring functions, the score is biased towards higher molecular weight compounds, which can establish a higher number of non-specific interactions with the target[44]. For this reason, usually, molecular docking programs are efficient in sampling the conformational space available for the ligand within the binding site but are weaker in prioritizing the right binding mode among a set of reasonable hypotheses generated by the search algorithm[43]. This is especially true in the case of fragments, which deviate from the drug-like properties of compounds upon which traditional scoring functions have been trained[46].

As thoroughly discussed in the work of Bolcato et al.[42], one possible solution to the scoring problem is to apply a pharmacophore filter to poses generated by the molecular docking program. When several structural pieces of information are available in the form of protein-ligand crystal complexes for a certain target (as is the case for protein kinase CK1δ, the case study for this work), a good solution to reduce the false positive rate of molecular docking programs is to build a pharmacophore model that encompasses the most prominent interaction features that are required to ligands to bind to the target active site[62]. In the case of a program like Autogrow, where the selection mechanism that determines which compounds to promote to the next generation is based on the docking score, we thought it would be interesting to incorporate a knowledge-based element in the pose selection mechanism in the form of a comparison between the interaction fingerprint of query compounds and

known inhibitors, to bias the selection mechanism towards molecules that respect the required features to bind to the target.

To validate our IFP_{CS} scoring protocol, we performed both a *denovo* and a lead-optimization benchmark run, using the same operative conditions described in the original work of Spiegel et al.[40] but on a different target. The protein target of choice was the protein kinase CK1δ, a pharmaceutically relevant target in the field of neurodegenerative diseases for which several crystal complexes with inhibitors are available in the Protein Data Bank. The benchmark *denovo* run was performed on a library composed of 6103 fragment compounds whose molecular weight falls between 100 and 150 Da, while the benchmark lead-optimization run was carried out on a library composed of 24 crystallographic ligands of the protein kinase CK1δ and 316 fragments derived from the fragmentation of crystallographic ligands using the BRICS rule. To compare the capabilities of the two protocols, we filtered each generation of compounds with the same pharmacophore filter already utilized in previous scientific works on the target[42], [61]. We then proceeded to evaluate the quality of compounds that pass the pharmacophore filter, considering both the size and the similarity of shape and electrostatic properties of query compounds compared to the crystallographic ligands taken as reference.

As illustrated by the results of our analysis (Section 3.3 and 3.4, respectively), there is a substantial difference in the performances of the two protocols: while both protocols can generate a certain amount of compounds that pass the pharmacophore filter (therefore possessing the right structural features that are required for the interaction with the target), in both scenarios the IFP_{CS} scoring function outperforms the traditional VINA one regarding the ability to select and prioritize pharmacophore-like compounds that have a similar shape and electrostatic properties compared to known inhibitors of the protein kinase CK1δ. This is particularly evident in the lead-optimization scenario, where within each generation passage, the average molecular weight of compounds that pass the pharmacophore filter steadily increases, passing from the typical MW of a fragment-like compound to the MW of a grown, mature, lead candidate, while the contrary happens in the case of the VINA protocol, with the average MW of the compounds that pass the pharmacophore filter steadily decreasing, falling even below the value of the first generation from the IFP_{CS} protocol. Moreover, when poses from each generation are compared with the ones of crystallographic ligands

concerning the shape and electrostatic similarity, a similar trend can be noticed. While the VINA protocol can select high-quality compounds in the first generation, compared to the IFP_{CS} one, at later stages during the run a progressive reduction in the similarity between the query and reference compounds can be noticed, contrary to what happens when the IFP_{CS} scoring protocol is utilized. This can be explained considering the different nature of the two scoring functions: the VINA protocol is biased towards bigger, therefore higher scoring, compounds, while the IFP_{CS} protocol favors compounds that respect the interaction pattern of the reference crystallographic ligand, regardless of their size. For this reason, the IFP_{CS} protocol tends to favor smaller compounds in the first generations, as long as they are complying with the constraint imposed by the reference interaction fingerprint, increasing the possibility to maintain in the population high-quality fragment to be optimized via the mutation and crossover operation of the genetic algorithm. On the contrary, the VINA protocol selects high MW compounds in the first generation that have little to no space for meaningful chemical modifications, giving low priority to those smaller compounds that will have a lower number of interactions with the target, thus resulting in lower docking score. Overall, our IFP_{CS} protocol seems preferable in those cases where structural data is available in the form of protein-ligand complex structure, as is the case for a good number of targets nowadays, while the traditional protocol seems a valid choice in those cases where such structural information is missing.

A recent virtual screening campaign performed in our laboratory led to the identification of 7 novel fragment compounds that are ATP competitive CK1δ inhibitors[42]. Curious to see if our protocol would have been able to produce novel potential CK1δ that incorporate structural features of our 7 fragments, we performed a second *denovo* run, using the same conditions as for the benchmark one, except for the introduction in the starting library of those 7 fragment compounds. The same analysis performed on the benchmark runs showed that the performance of our IFP_{CS} scoring protocol is even better when the Autogrow protocol is seeded with high-quality fragments that have the right structural feature to interact with the target. Usually, in a typical fragment-based drug discovery campaign, the identification of fragment binders either through virtual or experimental screening leads to the discovery of several potential starting points for the hit-to-lead fragment optimization phase. Our preliminary study showed that it is possible to obtain meaningful results even in those cases where the starting library is populated by fragments that are randomly selected and not

specifically tuned for the target of choice, but it certainly benefits from the contamination of the starting library with fragments that are known binders, indicating that the application of the IFP_{CS} protocol could lead to some interesting results in those cases where the known binders constitute a bigger fraction of the starting library. Concerning this, this approach could be utilized to evaluate the competitiveness of newly found scaffolds with already existing ones, based on the simplicity to derive those scaffolds with common and feasible chemical reactions, therefore, producing a good number of derivatives with increased affinity for the target.

5. Conclusions

In the present work, we presented and benchmarked a custom version of the open-source Autogrow4 which implements an alternative scoring protocol based on the similarity between protein-ligand interaction fingerprint of query compounds compared to a crystal reference. To demonstrate the applicability of our protocol in a pharmaceutically relevant scenario, we tested its capability to generate compounds that have similar binding and structural features to known inhibitors of the protein kinase CK1δ, a protein that is involved in several neurodegenerative diseases such as Alzheimer's Disease, Parkinson's Disease, and Amyotrophic Lateral Sclerosis.

A benchmark *denovo* run and a lead-optimization one were both carried out to compare the performance of our IFP_{CS} scoring protocol against the traditional one implemented in the original version of the Autogrow code, using the same conditions as the one reported in the original publication by Spiegel et al. Compared to the traditional Autogrow protocol, which uses the default scoring function of the Vina docking software, our IFP_{CS} protocol was able to generate, on average, compounds that were bigger and more similar from a shape and electrostatic properties point of view to crystallographic ligands, while retaining the key protein-ligand interaction features required for the inhibition of CK1δ.

The custom Autogrow version used in this work, which implements our alternative IFP_{CS} scoring protocol, along with the JSON configuration files used for each run and a YAML file to reconstitute the Python environment to run the custom version of the code, are available in the Supplementary Material.

References

[1] U. Knippschild, A. Gocht, S. Wolff, N. Huber, J. Löhler, and M. Stöter, "The casein kinase 1 family: participation in multiple cellular processes in eukaryotes," *Cell. Signal.*, vol. 17, no. 6, pp. 675–689, Jun. 2005, doi: 10.1016/j.cellsig.2004.12.011.

[2] U. Knippschild *et al.*, "The CK1 Family: Contribution to Cellular Stress Response and Its Role in Carcinogenesis," *Front. Oncol.*, vol. 4, May 2014, doi: 10.3389/fonc.2014.00096.

[3] P. Xu *et al.*, "Structure, regulation, and (patho-)physiological functions of the stress-induced protein kinase CK1 delta (CSNK1D)," *Gene*, vol. 715, Oct. 2019, doi: 10.1016/j.gene.2019.144005.

[4] L. Behrend *et al.*, "Interaction of casein kinase 1 delta (CK1δ) with post-Golgi structures, microtubules and the spindle apparatus," *Eur. J. Cell Biol.*, vol. 79, no. 4, pp. 240–251, Apr. 2000, doi: 10.1078/S0171-9335(04)70027-8.

[5] H. Lee, R. Chen, Y. Lee, S. Yoo, and C. Lee, "Essential roles of CKIδ and CKIε in the mammalian circadian clock," *Proc. Natl. Acad. Sci.*, vol. 106, no. 50, pp. 21359–21364, Dec. 2009, doi: 10.1073/pnas.0906651106.

[6] F. Meggio, J. W. Perich, E. C. Reynolds, and L. A. Pinna, "A synthetic β-casein phosphopeptide and analogues as model substrates for casein kinase-1, a ubiquitous, phosphate directed protein kinase," *FEBS Lett.*, 1991, doi: 10.1016/0014-5793(91)80614-9.

[7] J. Bischof, S. J. Randoll, N. Süßner, D. Henne-Bruns, L. A. Pinna, and U. Knippschild, "CK1δ Kinase Activity Is Modulated by Chk1-Mediated Phosphorylation," *PLoS One*, 2013, doi: 10.1371/journal.pone.0068803.

[8] P. R. Graves and P. J. Roach, "Role of COOH-terminal phosphorylation in the regulation of casein kinase Iδ," *J. Biol. Chem.*, 1995, doi: 10.1074/jbc.270.37.21689.

[9] D. M. Milne, P. Looby, and D. W. Meek, "Catalytic activity of protein kinase CK1 delta (casein kinase 1delta) is essential for its normal subcellular localization," *Exp. Cell Res.*, vol. 263, no. 1, pp. 43–54, Feb. 2001, doi: 10.1006/EXCR.2000.5100.

[10] K. L. Longenecker, P. J. Roach, and T. D. Hurley, "Crystallographic studies of casein kinase I delta toward a structural understanding of auto-inhibition," *Acta Crystallogr. D. Biol. Crystallogr.*, vol. 54, no. Pt 3, pp. 473–475, 1998, doi: 10.1107/S0907444997011724.

[11] H. Hirner *et al.*, "Impaired CK1 delta activity attenuates SV40-induced cellular transformation in vitro and mouse mammary carcinogenesis in Vivo," *PLoS One*, 2012, doi: 10.1371/journal.pone.0029709.

[12] D. I. Perez, C. Gil, and A. Martinez, “Protein kinases CK1 and CK2 as new targets for neurodegenerative diseases,” *Medicinal Research Reviews*. 2011, doi: 10.1002/med.20207.

[13] Z. Breijyeh and R. Karaman, “Comprehensive Review on Alzheimer’s Disease: Causes and Treatment,” *Molecules*, vol. 25, no. 24, p. 5789, Dec. 2020, doi: 10.3390/molecules25245789.

[14] V. M. Y. Lee, M. Goedert, and J. Q. Trojanowski, “Neurodegenerative tauopathies,” *Annu. Rev. Neurosci.*, vol. 24, pp. 1121–1159, 2001, doi: 10.1146/ANNUREV.NEURO.24.1.1121.

[15] D. J. Selkoe, “Alzheimer’s disease: Genes, proteins, and therapy,” *Physiol. Rev.*, vol. 81, no. 2, pp. 741–766, 2001, doi: 10.1152/PHYSREV.2001.81.2.741/ASSET/IMAGES/LARGE/9J0210134004.JPEG.

[16] H. J and S. DJ, “The amyloid hypothesis of Alzheimer’s disease: progress and problems on the road to therapeutics,” *Science*, vol. 297, no. 5580, pp. 353–356, Jul. 2002, doi: 10.1126/SCIENCE.1072994.

[17] N. Ghoshal *et al.*, “A New Molecular Link between the Fibrillar and Granulovacuolar Lesions of Alzheimer’s Disease,” *Am. J. Pathol.*, vol. 155, no. 4, pp. 1163–1172, Oct. 1999, doi: 10.1016/S0002-9440(10)65219-4.

[18] G. Li, H. Yin, and J. Kuret, “Casein Kinase 1 Delta Phosphorylates Tau and Disrupts Its Binding to Microtubules,” *J. Biol. Chem.*, vol. 279, no. 16, pp. 15938–15945, Apr. 2004, doi: 10.1074/jbc.M314116200.

[19] C. Schwab, A. J. Demaggio, N. Ghoshal, L. I. Binder, J. Kuret, and P. L. McGeer, “Casein kinase 1 delta is associated with pathological accumulation of tau in several neurodegenerative diseases,” *Neurobiol. Aging*, 2000, doi: 10.1016/S0197-4580(00)00110-X.

[20] G. Li, H. Yin, and J. Kuret, “Casein Kinase 1 Delta Phosphorylates Tau and Disrupts Its Binding to Microtubules,” *J. Biol. Chem.*, 2004, doi: 10.1074/jbc.M314116200.

[21] M. Okochi *et al.*, “Constitutive Phosphorylation of the Parkinson’s Disease Associated α-Synuclein,” *J. Biol. Chem.*, vol. 275, no. 1, pp. 390–397, Jan. 2000, doi: 10.1074/jbc.275.1.390.

[22] D. J. Surmeier, “Determinants of dopaminergic neuron loss in Parkinson’s disease,” *FEBS J.*, vol. 285, no. 19, pp. 3657–3668, Oct. 2018, doi: 10.1111/febs.14607.

[23] F. Kametani *et al.*, “Identification of casein kinase-1 phosphorylation sites on TDP-43,” *Biochem. Biophys. Res. Commun.*, vol. 382, no. 2, pp. 405–409, May 2009, doi: 10.1016/j.bbrc.2009.03.038.

[24] “‘Dementia’, WHO, https://www.who.int/news-room/fact-sheets/detail/dementia.” .

[25] G. Bollag *et al.*, “Vemurafenib: The first drug approved for BRAF-mutant cancer,” *Nature Reviews Drug Discovery*. 2012, doi: 10.1038/nrd3847.

[26] J. Schoepfer *et al.*, “Discovery of Asciminib (ABL001), an Allosteric Inhibitor of the Tyrosine Kinase Activity of BCR-ABL1,” *J. Med. Chem.*, 2018, doi: 10.1021/acs.jmedchem.8b01040.

[27] D. A. Erlanson, S. W. Fesik, R. E. Hubbard, W. Jahnke, and H. Jhoti, “Twenty years on: The impact of fragments on drug discovery,” *Nature Reviews Drug Discovery*, vol. 15, no. 9. Nature Publishing Group, pp. 605–619, Sep. 2016, doi: 10.1038/nrd.2016.109.

[28] H. Jhoti, G. Williams, D. C. Rees, and C. W. Murray, “The ‘rule of three’ for fragment-based drug discovery: Where are we now?,” *Nat. Rev. Drug Discov.*, vol. 12, no. 8, p. 644, 2013, doi: 10.1038/nrd3926-c1.

[29] P. J. Hajduk, “Fragment-based drug design: How big is too big?,” *J. Med. Chem.*, vol. 49, no. 24, pp. 6972–6976, 2006, doi: 10.1021/jm060511h.

[30] R. J. Hall, P. N. Mortenson, and C. W. Murray, “Efficient exploration of chemical space by fragment-based screening.,” *Prog Biophys Mol Biol*, vol. 116, no. 2–3, pp. 82–91, Dec. 2014, doi: 10.1016/j.pbiomolbio.2014.09.007.

[31] D. A. Erlanson, R. S. McDowell, and T. O’Brien, “Fragment-Based Drug Discovery,” *J. Med. Chem.*, vol. 47, no. 14, pp. 3463–3482, Jul. 2004, doi: 10.1021/JM040031V.

[32] C. W. Murray and D. C. Rees, “The rise of fragment-based drug discovery,” *Nat. Chem. 2009 13*, vol. 1, no. 3, pp. 187–192, Jun. 2009, doi: 10.1038/nchem.217.

[33] S. Schultes, C. De Graaf, E. E. J. Haaksma, I. J. P. De Esch, R. Leurs, and O. Krämer, “Ligand efficiency as a guide in fragment hit selection and optimization,” *Drug Discov. Today Technol.*, vol. 7, no. 3, pp. e157–e162, Sep. 2010, doi: 10.1016/J.DDTEC.2010.11.003.

[34] D. C. Rees, M. Congreve, C. W. Murray, and R. Carr, “Fragment-based lead discovery.,” *Nat. Rev. Drug Discov.*, vol. 3, no. 8, pp. 660–72, Aug. 2004, doi: 10.1038/nrd1467.

[35] M. Bissaro, M. Sturlese, and S. Moro, “The rise of molecular simulations in fragment-based drug design (FBDD): an overview,” *Drug Discov. Today*, vol. 25, no. 9, pp. 1693–1701, Sep. 2020, doi: 10.1016/j.drudis.2020.06.023.

[36] H. J. Böhm, “The computer program LUDI: A new method for the de novo design of enzyme inhibitors,” *J. Comput. Aided. Mol. Des.*, 1992, doi: 10.1007/BF00124387.

[37] M. B. Eisen, D. C. Wiley, M. Karplus, and R. E. Hubbard, “HOOK: A program for finding novel molecular architectures that satisfy the chemical and steric requirements of a macromolecule binding site,” *Proteins Struct. Funct. Bioinforma.*, 1994, doi: 10.1002/prot.340190305.

[38] G. Lauri and P. A. Bartlett, “CAVEAT: A program to facilitate the design of organic molecules,” *J. Comput. Aided. Mol. Des.*, 1994, doi: 10.1007/BF00124349.

[39] P. Maass, T. Schulz-Gasch, M. Stahl, and M. Rarey, “Recore: A fast and versatile method for scaffold hopping based on small molecule crystal structure conformations,” *J. Chem. Inf. Model.*, 2007, doi: 10.1021/ci060094h.

[40] J. O. Spiegel and J. D. Durrant, “AutoGrow4: an open-source genetic algorithm for de novo drug design and lead optimization,” *J. Cheminformatics 2020 121*, vol. 12, no. 1, pp. 1–16, Apr. 2020, doi: 10.1186/S13321-020-00429-4.

[41] O. Trott and A. J. Olson, “AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading,” *J. Comput. Chem.*, vol. 31, no. 2, p. 455, 2010, doi: 10.1002/JCC.21334.

[42] G. Bolcato *et al.*, “A Computational Workflow for the Identification of Novel Fragments Acting as Inhibitors of the Activity of Protein Kinase CK1δ,” *Int. J. Mol. Sci.*, vol. 22, no. 18, p. 9741, Sep. 2021, doi: 10.3390/ijms22189741.

[43] L. Chaput and L. Mouawad, “Efficient conformational sampling and weak scoring in docking programs? Strategy of the wisdom of crowds,” *J. Cheminform.*, 2017, doi: 10.1186/s13321-017-0227-x.

[44] Y. C. Chen, “Beware of docking!,” *Trends in Pharmacological Sciences*, vol. 36, no. 2. Elsevier Ltd, pp. 78–95, Feb. 2015, doi: 10.1016/j.tips.2014.12.001.

[45] L. R. de Souza Neto *et al.*, “In silico Strategies to Support Fragment-to-Lead Optimization in Drug Discovery,” *Frontiers in Chemistry*, vol. 8. Frontiers Media S.A., p. 93, Feb. 2020, doi: 10.3389/fchem.2020.00093.

[46] M. L. Verdonk, I. Giangreco, R. J. Hall, O. Korb, P. N. Mortenson, and C. W. Murray, “Docking Performance of Fragments and Druglike Compounds,” *J. Med. Chem.*, vol. 54, no. 15, pp. 5422–5431, Aug. 2011, doi: 10.1021/jm200558u.

[47] H. M. Berman *et al.*, “The Protein Data Bank,” *Nucleic Acids Res.*, vol. 28, no. 1, p. 235, Jan. 2000, doi: 10.1093/NAR/28.1.235.

[48] S. Mente *et al.*, “Ligand–Protein Interactions of Selective Casein Kinase 1δ Inhibitors,” *J. Med. Chem.*, vol. 56, no. 17, pp. 6819–6828, Sep. 2013, doi: 10.1021/jm4006324.

[49] “Molecular Operating Environment (MOE), 2019.01; Chemical Computing Group ULC, 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2021.” .

[50] D.A. Case et al., “Amber 10, University of California, San Francisco.” .

[51] “QUACPAC 2.1.3.0: OpenEye Scientific Software, Santa Fe, NM. http://www.eyesopen.com.” .

[52] G. Jones, P. Willett, R. C. Glen, A. R. Leach, and R. Taylor, “Development and validation of a genetic algorithm for flexible docking 1 1Edited by F. E. Cohen,” *J. Mol. Biol.*, vol. 267, no. 3, Apr. 1997, doi: 10.1006/jmbi.1996.0897.

[53] O. Korb, T. Stützle, and T. E. Exner, “PLANTS: Application of Ant Colony Optimization to Structure-Based Drug Design,” 2006.

[54] D. R. Houston and M. D. Walkinshaw, “Consensus Docking: Improving the Reliability of Docking in a Virtual Screening Context,” *J. Chem. Inf. Model.*, vol. 53, no. 2, Feb. 2013, doi: 10.1021/ci300399w.

[55] “AutoGrow4 – Durrant Lab.” .

[56] J. Eberhardt, D. Santos-Martins, A. F. Tillack, and S. Forli, “AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings,” *J. Chem. Inf. Model.*, vol. 61, no. 8, pp. 3891–3898, Aug. 2021, doi: 10.1021/ACS.JCIM.1C00203/SUPPL_FILE/CI1C00203_SI_002.ZIP.

[57] M. Wójcikowski, P. Zielenkiewicz, and P. Siedlecki, “Open Drug Discovery Toolkit (ODDT): a new open-source player in the drug discovery field,” *J. Cheminform.*, vol. 7, no. 1, p. 26, Dec. 2015, doi: 10.1186/s13321-015-0078-2.

[58] D. Li and B. Ji, “Protein conformational transitions coupling with ligand interactions: Simulations from molecules to medicine,” *Med. Nov. Technol. Devices*, vol. 3, p. 100026, Sep. 2019, doi: 10.1016/j.medntd.2019.100026.

[59] P. Csermely, R. Palotai, and R. Nussinov, “Induced fit, conformational selection and independent dynamic segments: an extended view of binding events,” *Trends Biochem. Sci.*, vol. 35, no. 10, pp. 539–546, Oct. 2010, doi: 10.1016/j.tibs.2010.04.009.

[60] S. D. Wierbowski, B. M. Wingert, J. Zheng, and C. J. Camacho, “Cross-docking benchmark for automated pose and ranking prediction of ligand binding,” *Protein Sci.*, vol. 29, no. 1, pp. 298–305, Jan. 2020, doi: 10.1002/pro.3784.

[61] E. Cescon *et al.*, “Scaffold Repurposing of in-House Chemical Library toward the Identification of New Casein Kinase 1 δ Inhibitors,” *ACS Med. Chem. Lett.*, vol. 11, no. 6, Jun. 2020, doi: 10.1021/acsmedchemlett.0c00028.

[62] M. L. Peach and M. C. Nicklaus, “Combining docking with pharmacophore filtering for improved virtual screening,” *J. Cheminform.*, vol. 1, no. 1, Dec. 2009, doi: 10.1186/1758-2946-1-6.

Computational strategies to identify new drug candidates against neuroinflammation

Matteo Pavan, **Davide Bassani**, Giovanni Bolcato, Maicol Bissaro, Mattia Sturlese and Stefano Moro

M. Pavan, **D. Bassani**, G. Bolcato, M. Bissaro, M. Sturles, and S. Moro, "Computational strategies to identify new drug candidates against neuroinflammation.," *Curr. Med. Chem.*, vol. 29, Feb. 2022, doi: 10.2174/0929867329666220208095122.

Abstract

The even more increasing application of computational approaches in these last decades has deeply modified the process of discovery and commercialization of new therapeutic entities. This is especially true in the field of neuroinflammation, in which both the peculiar anatomical localization and the presence of the blood-brain barrier make it mandatory to finely tune the candidates' physicochemical properties from the early stages of the discovery pipeline. The aim of this review is therefore to provide a general overview to the readers about the topic of neuroinflammation, together with the most common computational strategies that can be exploited to discover and design small molecules controlling neuroinflammation, especially those based on the knowledge of the three-dimensional structure of the biological targets of therapeutic interest. The techniques used to describe the molecular recognition mechanisms, such as molecular docking and molecular dynamics, will therefore be eviscerated, highlighting their advantages and their limitations. Finally, we report several case studies in which computational methods have been applied in drug discovery on neuroinflammation, focusing on the last decade's research.

1. Introduction

Inflammation is a vital host defense response to stimuli that undermine the homeostasis and integrity of tissues, such as traumatic injury, infection by pathogens, autoimmune responses, or hypoxic conditions. From a mechanistic point of view, it can be described as a multi-factor process that involves the invasion of circulating immune cells such as lymphocytes and monocytes and the induction of inflammatory mediators like cytokines and prostanoids[1].

Usually, inflammation consists of a cascade of events known as acute phase response aimed at resolving the initiating stimulus and restoring tissue's integrity and functionality. However, if the

condition persists for a prolonged period and exceeds the physiological response it becomes detrimental to the tissue's health and function[2].

Uncontrolled inflammatory processes are associated with several neurodegenerative diseases, including both proteinopathies such as Alzheimer's disease (AD), Parkinson's disease (PD), amyotrophic lateral sclerosis (ALS), or Huntington's disease (HD), and lysosomal storage diseases (LSDs) such as Niemann-Pick type C (NPC), Gaucher Disease (GD), Mucopolysaccharidoses (MPS) and Neuronal Ceroid Lipofuscinosis (NCL), resulting in a condition defined as neuroinflammation[3][4].

Although it is still debated whether this condition has a role in the early stages of the aforementioned pathologies, contributing to their etiology, several pieces of evidence point out that sustained inflammatory states in the Central Nervous System (CNS) decisively contribute to the progression of such diseases[5][6]. Moreover, some studies even suggest that inflammation could play a protective role in the first stages of the disease[7][8], further underlying the complexity of the topic and the intricacy of the mechanisms that regulate these processes.

From an immune system point of view, the brain is compartmentalized and separated from the peripheral blood circulation by the blood-brain-barrier (BBB), therefore indicating that a critical role in the inflammatory process is portrayed by locally resident cells, namely astrocytes and microglia, rather than by leucocytes infiltrating from the periphery[9][10]. This is particularly true in the initiating phases of inflammation, while in later stages the disruption of the BBB provoked by the inflammatory cascade and by proteases such Matrix Metalloproteinases (MMPs) causes the involvement of non-resident immune cells such as monocytes, neutrophils, and CD4+/CD8+ T cells[11–14].

In healthy brain tissue, astrocytes serve as key components of the BBB. Moreover, they are involved in ionic homeostasis, removal of excess neurotransmitters, and secretion of neurotrophic factors that are vital for neuron survival[15]. Astrocytes can exist in a continuous spectrum of different phenotypes, ranging from neuroprotective to pro-inflammatory[16]. Although their pro-inflammatory contribution is limited compared to that of microglia, under immunologic challenges or injuries, astrocytes assume an activated phenotype that leads to an increase in the production of glial fibrillary acidic proteins that eventually hinders axonal regeneration[17], [18]. Furthermore, reactive phenotype astrocytes can induce the production of factors such as Interleukin-1β (IL-1β), Tumor Necrosis Factor (TNF), and nitric oxide (NO) that enhance inflammation and neuronal death[19].

Microglia are the predominant immune cell species located in the brain[20]. Under physiological conditions, these macrophage-like cells exist in a deactivated ramified phenotype, and they are engaged in a process of environmental surveillance to maintain tissue homeostasis, producing neurotrophic and anti-inflammatory factors[21][22]. Under the influence of external stimuli such as pathogen invasion or tissue damage, they can switch to an activated amoeboid phenotype that promotes an inflammatory response to address the situation and promote tissue reparation[23]. If a persistent stimulus from either an environmental or an endogenous factor is perceived by the immune system as a threat, an uncontrolled inflammatory state may cause the production of several neurotoxic factors such as superoxide, NO, and TNF that contribute to the progression of underlying disease states[24]. A general depiction of the common mechanisms underlying neuroinflammation is reported in Figure 1.

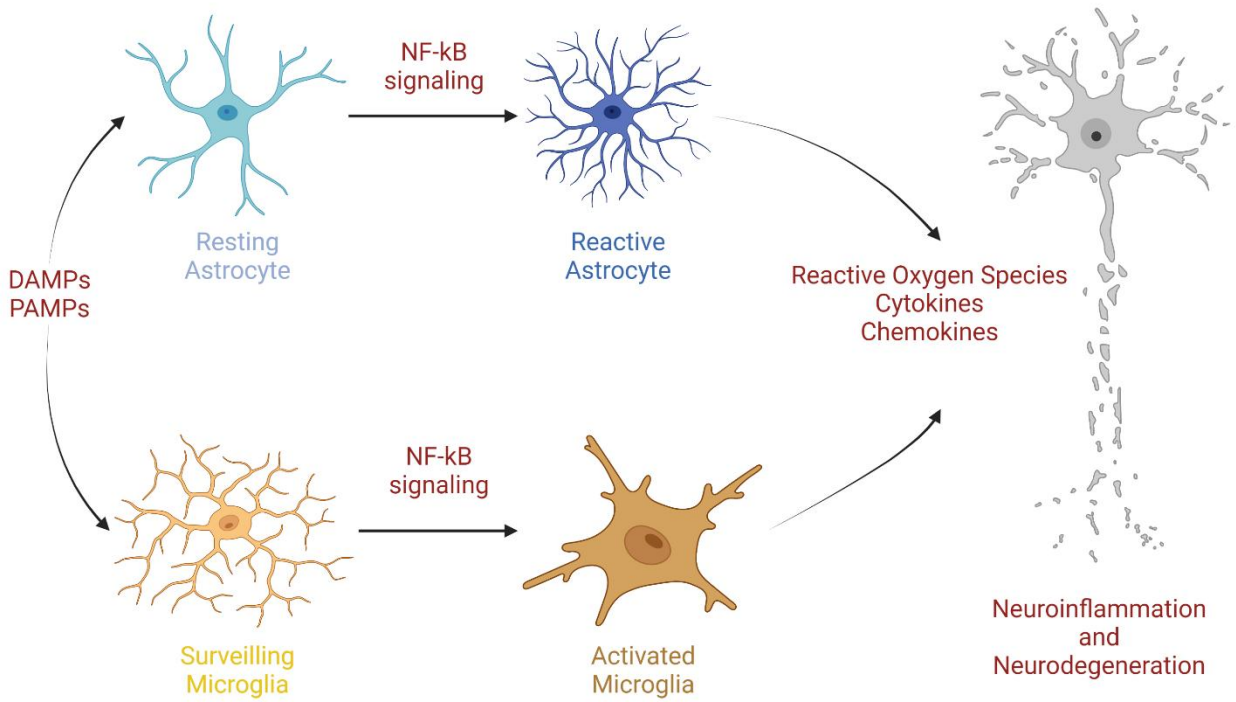


Figure 1. *General scheme of the common mechanisms involved in neuroinflammation. The figure was created with 'BioRender.com'.*

This inflammatory response is usually initiated by pattern recognition receptors such as Toll-Like Receptors (TLRs)[25], but they can also be started by purinergic[26] and scavenger receptors[27]. Particularly, TLRs such as TLR4 are overexpressed on the surface of CNS resident immune cells such

as microglia and are responsible for the recognition of both pathogen-associated molecular patterns (PAMPs) and damage/danger-associated wide-array of patterns (DAMPs) coming from the host itself[25]. The inflammatory signal is then passed on to several transduction systems, such as Myeloid differentiation primary response 88 (MyD88), TIR-domain-containing adapter-inducing interferon-β **(**TRIF), phosphoinositide 3-kinases (PI3Ks), IkB kinases (IKK), and mitogen-activated protein kinases (MAPKs), that cooperate in a combinatorial manner to regulate hundreds of different genes based upon the activated target cells employing transcription factors such as the ubiquitous nuclear factor kappa-light-chain-enhancer of activated B cells *(*NF-kB)[28]–[33]. Among these signal transduction pathways, particular importance is portrayed by the Wnt signaling, whose pro-inflammatory activity is finely regulated by several mechanisms such as the formation of the "β-catenin destruction complex" which involves neurodegenerative disease-related kinases such as casein kinase 1 delta (CK1) and glycogen synthase kinase 3 (GSK-3)[34]–[36]. Finally, gene expression then leads to the signal amplification through prostaglandins, cytokines such as TNF or interleukin 1 beta (IL-1β), and chemokines such as monocyte chemoattractant protein-1 (MCP-1), C-C motif chemokine ligand 2 (CCL2), and C-X-C Motif Chemokine Ligand 5 (CXCL5) that allow for the recruitment of other immune cells[37][38]. Furthermore, the generation of reactive oxygen species (ROS) mainly through the nicotinamide adenine dinucleotide phosphate oxidase (NADPH oxidase) system and of nitric oxide through the inducible nitric oxide synthase (iNOS) further contribute to the defense response in an unspecific manner[39][40].

By contrast, several negative feedback mechanisms exist to attenuate or arrest this inflammatory cascade. These involve proteins with different functions, such as inhibition of signal transduction (SOCS proteins), induction of transcriptional repressors (Nurr1), and production of soluble anti-inflammatory factors (IL-10, TGF-β)[41]–[43]. Furthermore, resolution of inflammation is also modulated by a family of specialized pro-resolving lipid mediators (SPMs) formed by lipoxins, resolvins, protectins, and maresins, either by inhibition of pro-inflammatory pathways or by activation of protective ones through interaction with a series of G-protein coupled receptors (GPCRs)[44][45].

Although a general convergence in the mechanism of the inflammatory processes that lead to neurotoxicity can be found, each pathology has its own set of markers that characterize the specific role of neuroinflammation in the insurgence and progression of the disease[46].

Alzheimer's disease was the first neurodegenerative disease to be linked to an inflammatory response[47]. Characterized by symptoms such as memory loss and cognitive impairment, from a

pathological point of view this illness is associated with extracellular amyloid plaques formed by cleavage products of the amyloid precursor protein (APP) such as the beta-amyloid peptide (Aβ peptide) and with intracellular neurofibrillary tangles (NFTs) that are composed of hyperphosphorylated forms of the microtubule-binding protein tau[48]. The hallmarks of the inflammatory response in AD include a change in the microglia phenotype, which switches from a resting to an activated pro-inflammatory one, astrogliosis, which consists of an increase of astrocytes number, dimensions, and motility, other than high levels of cytokines and chemokines found in senile plaques[49].

Particularly, aggregates of Aβ peptide can activate microglia through TLRs such as TLR4 and TLR2, scavenger receptors CD36, CD47, and SCARA1, receptor for advanced glycoxidation end-products (RAGE) and NOD-like receptors (NLRs) like NALP3 or NALP1, leading to the phagocytosis of Aβ plaques, degradation of Aβ fibrils through the endolysosomal pathway and the subsequent activation of an inflammatory cascade that promotes neuronal death[50]. Not only pro-inflammatory cytokines can induce neuronal apoptosis directly, but they can also activate an astrocyte response that further enhances microglia-driven inflammation[17]. On the contrary, a protective role against this inflammatory cascade is played by TREM2[51] (triggering receptor expressed on myeloid cells 2), a transmembrane glycoprotein exclusively expressed on immune cells that represses microglia-driven cytokine production and secretion[52] and regulate the phagocytic pathways involved in the removal of neuronal debris[53]. Several variants of the gene encoding for this receptor were associated with an increased risk of developing AD and other neurodegenerative diseases, such as PD and ALS, indicating that targeting TREM2 could be vital in the fight against these diseases[54][55].

Parkinson's disease is the second most common neurodegenerative disease after AD[56]. Differently from AD, it is mainly identified by motor symptoms such as bradykinesia, tremor, and rigidity, while from a neuropathological point of view it is characterized by intracellular inclusions of a misfolded protein called α-synuclein defined as Lewy bodies accompanied by the loss of dopaminergic neurons in the substantia nigra[57]. Regarding inflammation, microglial activation can be observed in PD, accompanied by astrogliosis and lymphocyte infiltration[58], [59]. The presence of α-synuclein aggregates is sensed and internalized by microglia, causing the activation of NADPH oxidase and iNOS which leads to the release of ROS and NO that indeed causes the switch of microglia to the activated phenotype[60]. Pro-inflammatory microglia then release factors such as TNF, NO, and IL-1β that modulate and drive the neuroinflammatory

process[61][62]. Moreover, high levels of neuromelanin derived from an increased catecholamine metabolism further contribute to neuron oxidative stress and microglial activation[63]. On the contrary, C-X3-C Motif Chemokine Receptor 1 (CX3CR1), CD200R, and Nurr1 can mediate anti-inflammatory responses that mitigate the negative effect of microglia activation in PD[64–66].

Amyotrophic Lateral Sclerosis is a neurodegenerative disease identified for the first time more than 150 years ago[67]. Clinically, it is characterized by a progressive motor neurons degeneration that eventually leads to the patient's death caused by respiratory muscles' paralysis[68]. The most prominent pathological marker of ALS is a strong inflammatory reaction caused by the immunoreactive inclusions of ubiquitin in the cytoplasm of degenerating neurons, which are mainly composed of TAR DBA-binding protein 43 (TDP-43)[69][70]. This inflammation process involves both astrocytes and microglia, whose activation causes the release of a high quantity of cytotoxic and pro-inflammatory molecules such as ROS, prostanoids, ILs, and TNF[71][72]. A prominent role in familiar forms of ALS is portrayed by mutant forms of superoxide dismutase 1 (SOD1), which cause an incremented NADPH oxidase-dependent oxidative stress that boosts microglia activation and subsequent inflammatory response[73]. Neuron cell death, in this case, is induced by both NO and Fas ligand (FasL), while a protective role against this process is carried out by infiltrating Th2 cells by inducing microglia to produce neurotrophic factors[74–76].

Huntington's disease has been known for centuries, but only recently its causative agent has been identified[77]. HD is caused by a CAG triplet expansion in the huntingtin (HTT) gene[78]. This expansion leads to a conformational alteration of HTT protein which assumes a toxic phenotype, provoking neuronal damage and subsequent death either through the formation of intranuclear inclusion bodies, formation of toxic fragments, or impairment of metabolic pathways[79–82]. The main clinical feature of HD is the extensive striatal neuronal cell death, but also white matter is involved at various stages in the degeneration process[83]. This disease manifests itself with a set of clearly defined symptoms, such as chorea, bradykinesia, lack of coordination, and motor impairment[84]. An inflammatory response carried out by activated microglia is correlated with disease severity and progression: the up-regulation of pro-inflammatory proteins such as complement proteins and clusterin, and an increased level of cytokines are the main markers of inflammation in HD[85], [86]. 8-OHdG, an oxidative stress marker circulating in the blood, is also up-regulated in patients with HD[87]. Finally, a microglial enzyme involved in excitotoxicity and ROS generation known as kynurenine 3-monooxygenase (KMO) has been found implicated in the regulation of HTT induced neurotoxicity[88].

Mucopolysaccharidoses are the most representative group of lysosomal storage diseases, accounting for 30% of all LSDs[89]. It is a heterogeneous group of pathologies caused by a deficiency in the lysosomal clearance of glycosaminoglycan (GAG)[90]. As for other LSDs, the alteration of lysosome function and integrity leads to the partial release of its content within the cellular lumen, with cathepsins and hydrolases acting as DAMPs and subsequently triggering the inflammatory cascade that leads to neuronal damage[91].

Although clinical manifestations of MPS can vary heavily depending on the illness type, peripheral organ dysfunction is generally a common trait across different MPS, with neurodegeneration and related cognitive and behavioral deficits being another quite common clinical marker of this set of pathologies, particularly in the most aggressive forms of MPS type I (Hurler syndrome) and MPS type III (Sanfilippo syndrome)[90], [92–94].

From an immune system point of view, MPS are associated with microglia-induced inflammatory response, characterized by high levels of chemokines such as CCL3 and CCL4 in the brain[95][96]. Moreover, treatments with anti-inflammatory/immunosuppressive drugs such as aspirin or prednisolone were able to lower the levels of both cytokines and chemokines associated with the inflammatory response[95][97].

Neuronal Ceroid Lipofuscinosis are a group of LSDs caused by mutation of one of 14 ceroid lipofuscinosis genes[98]. These mutations lead to alteration in the composition and structure of lysosomes[99], [100]. The most recurrent symptoms of these pathologies are visual failure, neurocognitive and motor decline, seizures, and premature death[101]. The most deadly forms of these diseases are the ones associated with a more prominent neuroinflammatory response, characterized by increased microglial and astrocyte activation leading to high levels of pro-inflammatory mediators such as TNF and various interleukins[102][103]. Some studies suggested that anti-inflammatory treatment may be beneficial in hampering inflammation-caused neurodegeneration[104].

Gaucher Disease, the most common lysosomal storage disease, arises from a mutation in the lysosomal enzyme glucosylceramidase, which leads to intracellular accumulation of glucosylceramide[89], [105][106]. Out of the three types of GD, only Type 2 and 3 are associated with neurological symptoms, which in the most severe cases can even lead to the premature death of the patient[92], [107][108]. Neurodegeneration associated with inflammatory response is, in this case, triggered by an alteration of calcium homeostasis, which causes increased sensitivity of

neurons to neurotoxic agents, thus inducing the release of DAMPs that induces the inflammatory response[109].

Niemann-Pick type C disease is caused by a mutation in either NPC1 or NPC2 gene[110]. These two genes codify for proteins involved in the regulation of intracellular trafficking, other than the processing of cholesterol and other lipids[111][112]. The alteration of their function leads to a lysosomal accumulation of cholesterol and sphingolipids, with lysosomal perturbation eventually resulting in the activation of the subsequent inflammatory cascade, most probably due to the release in the cytoplasm of lysosome content (e.g. hydrolases and cathepsins) that acts as DAMPs[91][113]. NPC shows an age-dependent evolution of the main symptomatologic manifestation, with peripheral symptoms such as hepatosplenomegaly being predominant in early stages, while neurological manifestation such as cerebellar ataxia, dementia, cognitive decline, and seizures are more prominent at later stages of the disease, with the most severe forms leading to the patient's early death[114][115]. The involvement of a neuroinflammatory process in the progression of the pathology is highlighted by the high levels of TNF and IL-1β in brain regions associated with major neuron loss at later stages of the disease[116][117]. Furthermore, the disease progression can be altered and slowed down by exploiting anti-inflammatory compounds such as non-steroid anti-inflammatory drugs (NSAIDs)[118][119].

Considering the huge impact that these pathologies have on patient's expectancy and quality of life[120], the social cost for the treatment of such diseases[121] and the current lack of cure for this deadly neurodegenerative illnesses[122], the possibility to rationally develop pharmacological tools that modulate inflammatory response in neurodegenerative disease is highly appealing.

2. Computational methodologies in neurodegeneration

It is nowadays widely accepted that Computational-Aided Drug Discovery (CADD) techniques have given a huge impact on both the time and money required for bringing new therapeutical entities to the market[123].

The computational approach to drug design can be divided into two main classes: structure-based drug design (SBDD) and ligand-based drug design (LBDD)[124].

The first family of techniques strongly relies on the availability of structural data about the target investigated for the pathology considered[125]. These data mainly come from X-ray crystallography (XRC) nowadays (the public database Protein Data Bank contains 180207 structures at the present day) but can also derive from NMR spectroscopy. XRC allows to obtain, with a high resolution,

objects with low limitations in terms of dimensions (roughly 150 kDa), it cannot resolve hydrogen atoms and gives just a “static” overview of the substrate. On the other hand, NMR spectroscopy can give access to data of the object which is much more similar to reality (dynamic data in solution) but it cannot be reliable with structures above 50 kDa, other than requiring lots of runs to achieve an acceptable signal-to-noise ratio[126]. Other techniques have emerged in the latest decades, such as Cryo-Electron Microscopy, which is acquiring a lot of interest for its ability to resolve complicated (such as membrane proteins) and high-molecular-weight (such as entire mitoribosomes[127]) structures[128].

The ligand-based drug discovery comprises a variety of techniques that can be both used alone -if no structural data of the target is available[129] and in consensus with SBDD protocols[130]. LBDD, as the name suggests, is based on the evaluation of known ligands of a certain target or simply on ligands that are active against a pathology on which no mechanism has been elucidated yet (so no target is known). The known actives are evaluated in several features, such as the 3D conformation, the electrostatic fields, the molecular descriptors, etc. After having obtained enough information, some algorithms can be developed to search molecules that are predicted to be similar to the known actives[131]. To pursue this task, in the latest years the experts in the field of machine learning and artificial intelligence have put some effort into developing reliable prediction models, both in industry and in academia[132].

In the case of neuroinflammation, considering the targets that we have summed up in the previous section, we can assess that SBDD can be implemented for drug discovery purposes, considering the considerable amount of structural data present for the targets of interest, as the PDB database witnesses[133][134][135][136].

Here we briefly report the main techniques which can be used for SBDD on neuroinflammation (which are also depicted in Figure 2), with a description of the tools and some cases in which they were adopted.

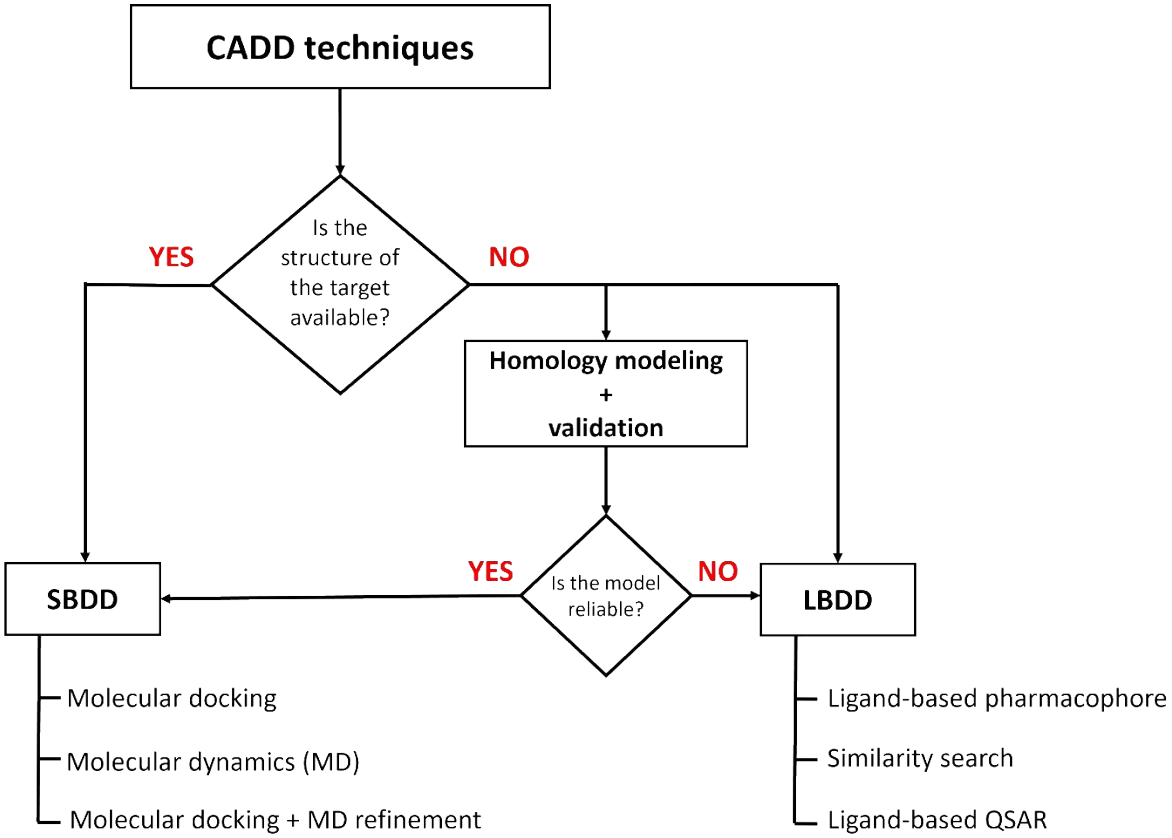


Figure 2. *Representation a general workflow scheme of the computational techniques that can be used to find new candidate molecules for neuroinflammation and related diseases.*

2.1. Structure preparation

Just as an introduction, is always useful to remember that all the computational tools require both the ligands and the targets to be properly prepared for the calculation. The ligands must be in 3D shape, with a proper protonation state and, depending on the cases, with a partial charge assigned to each atom. The biological targets must also be suitable for calculation: the structure must be complete, the hydrogen atoms added (if the structure comes from XRC) and a protonation state assigned to all amino acids. It is also advisable for added hydrogens to be energetically minimized[137]. If some portions of the targets are missing, some programs may help in building them (through homology modeling, see chapter below) or, if it is the case, they can be capped (just for a small loop). This second option has reasonably been discarded if the capping site is very near to the active site, and mostly in docking calculations.

2.2. Homology modeling

If no structural data is available for a certain target, the SBDD approach can be used if a proper model is generated. Homology modeling is a technique used to predict the structure of a certain entity (mainly proteins and nucleic acids, in the case of medicinal chemistry) using available data[138]. The approach in these cases can be dual: in the case in which some structures with reasonable sequence similarity (usually around 30% at least) are present, is possible to build the target (or even just a portion of it) using that as a template[139]. On the other hand, if no structure of the desired target is available at all, is possible to predict its 3D structure using different methodologies, such as the use of comparative models[140]. Many tools allow performing these tasks, such as MODELLER[141], Schrödinger Prime[142], MOE homology modeler[143], SWISS-MODEL[144], and others[145].

Another approach that has gained relevance recently is the prediction of protein structure based on artificial intelligence techniques. A famous example of this is represented by AlphaFold[146], a program based upon a deep learning algorithm (which has recently been updated to its second version, AlphaFold2[147]). The AlphaFold neural network has been properly trained to assign a three-dimensional shape to a sequence of amino acids based upon their similarities with existing crystallographic homologues[147]. This program has already proven to be very accurate in predicting the secondary structure of several targets[148][149][150]. Nevertheless, the application of models coming from AlphaFold to molecular docking experiments still needs some further optimization, mainly because of the higher difficulty in foretelling the conformation of a ligand-bound protein, which represents the main case of interest for drug discovery scientists[151].

It is a good practice to assess the stability and the reliability of the models produced, and this can be done through Molecular Dynamics (MD) refinement[152], evaluating various energetics and geometric parameters of the trajectory obtained.

There are some interesting situations regarding neuroinflammation in which no structural data is available for a target. When this circumstance takes place, both homology modeling and LBDD techniques can be evaluated. This is the case for the cited TSPO, on which some groups put efforts in generating valuable homology models, e.g., Bhargavi et al. used Modeller to create a virtual structure suitable for docking calculations[153]. Nguyen T. and Lai T. created and validated, through molecular dynamics simulations, a homology model for hTSPO based on the RsTSPO structure[154].

2.3. Molecular docking

Maybe the most used computational technique applied to drug discovery so far[155][156], molecular docking is a procedure consisting of generating several conformations of a molecule and evaluating the best among them for binding to a substrate (in our case, a protein or a nucleic acid). The docking programs can be roughly divided into two parts: a search algorithm and a scoring function. The first has the task to search the conformational space of the ligand, while the second has the objective to select, for each docking run, the best options among the conformations generated to fit with the target[157]. After several runs, the program returns the conformations (known as "poses") which are predicted to be the best for the scoring function applied. The docking programs can be differentiated mainly by how they search, optimize, and score the conformations of the molecules examined. A well-known family of programs use "genetic algorithms" to pursue this task (the most famous are CCDC GOLD[158] and AutoDock[159]), other use "systematic search algorithms" (such as Schrödinger Glide[160]) and other programs exploit more peculiar algorithms (such as the ACO used by PLANTS[161] and the molecular dynamics-based algorithm used by CDOCKER[162]). Several benchmarks have tried to highlight the best protocol to be used[163][164], but the most shared opinion is that each docking run must be optimized looking at the target considered[165]. A tool named "DockBench" was developed by Cuzzolin et. al to compare the efficiency of several program-scoring function pairs to reproduce the crystallographic pose of the ligand inside its substrate. This allows also us to select the best crystallographic structure to use for docking calculation (if more are present)[166].

To select among the poses obtained with molecular docking, there are several options. The first is to exclusively rely on the scoring function of the algorithm[167], another is to exclusively select poses in respect to certain parameters (such as electrostatic interaction, solvent exposure, etc.), but one of the most reliable methods remains the pharmacophore implementation[168]. A pharmacophore is an ensemble of features that are considered to be necessary for molecular recognition. If many XRC structures are present for a target, many tools offer the possibility to create a consensus "pharmacophore" (e.g., Schrödinger Phase[169] or MOE Pharmacophore editor), otherwise the features can be selected and customed with the scientist knowledge. Another approach for docking poses scoring is the molecular dynamics refinement[170], of which the details are reported in the following section.

Molecular docking offers several advantages, such as flexibility, wide applicability, and speed[171]. With the technologies available nowadays, several billion molecules can be docked in a few

days[172]. On the other hand, this technique has several limitations[173]. First of all, the scientist has to remind that the “classical” docking calculations are performed in a vacuum, which is does not take into account the presence of the water molecules solvating the biological target. Moreover, even if the solvent is considered, the water molecules should be classified for their ability to be displaced by the ligand[174], and special consideration should be given to the crystallographic waters around the ligand[175]. Another strong limitation is the number of false positives given by the algorithms. The docking programs have been developed making them very efficient in finding a good conformation for the binding, even if it is not consistent. To attenuate this feature, a particular protocol is known as “consensus docking” can be used. This approach consists of repeating the same docking calculation with different programs, which must implement different search-score algorithms[176][177]. The molecules that are selected from all the programs are more prone to be “real” positives.

In the field of neuroinflammation research, molecular docking has been applied several times. Cheng et al. implemented this technique for the development of novel MAPK inhibitors[178], while Rippin et al. used Glide docking for the design of novel GSK3β inhibitors[179].

Several efforts have been put into the discovery of CK1δ inhibitors through molecular docking-based methods. Cescon et al. were successful in repurposing two molecules on CK1δ through a docking-based protocol[180], while in a recent paper the same technique allowed Redenti et al. to design a dual GSK3β covalent/CK1δ reversible inhibitor active in the very low micromolar range[181].

Tandon and Sinha used AutoDock and Flexidock to rationalize the binding mode of three groups of known MMP-9 inhibitors[182], while Razak et al. used the same software to rationalize the activity of their compound of interest (2-(5-methoxy-2-methyl-1H-indol-3-yl)-N[(E)-(3-nitrophenyl) methylidene] acetohydrazide, called MMINA) against various targets of neuroinflammation, such as COX-2, STAT3, and TNF[183]. Docking-based virtual screening (VS) technique was chosen by Liu et al. to select (among over 90000 natural compounds) an inhibitor of STAT3 protein DNA-binding activity and dimerization[184], while Ray et al., through a large Glide-based VS, selected 15 molecules potentially able to stabilize SOD1 dimer, preventing its aggregation[185].

Looking at the TLRs, CDOCKER, and AutoDockVina were included by Durai et al. in the workflow for the discovery of two new TLR-2 antagonists active in the micromolar concentration range [186]. Mahita et al. applied molecular docking to evaluate the possibility of abrogating the TRAM-mediated signaling on Toll-Like Receptor 4[187]. Yilmazer et al. successfully implemented a virtual

screening approach based on the Glide docking program to select the most promising candidates as novel chaperones for beta-glucocerebrosidase (GBA), one of the main pharmacological targets of Gaucher's disease[188].

El-Zohairy et al. recently implemented a GOLD-based virtual screening protocol comprising a pharmacophore filtration for the design of novel low-micromolar inhibitors of CCR5[189], and a similar workflow (using MolDock instead of GOLD) was applied by Ahmad et al. for the selection of potential inhibitors of caspase 8 for treating neurodegeneration and neuroinflammation[190].

García-Aranda et al., with AutoDock 4.2 docking, evaluated and compared the candidates of a series of 1,2-diphenylbenzimidazoles (DPBI) with different aromatic substitutions for their activity on bot COX-2 and h-iNOS active sites[191].

2.4. Molecular dynamics

If Molecular docking offers a static representation of the “binding event” between ligand and receptor, Molecular Dynamics (MD) allows exploring the behavior of the complex during the time. In MD, a system (protein-ligand, but also protein alone or ligand alone) is generated, solvated, relaxed, and simulated during the time to see the evolution in time of some desired observables. In the case of medicinal chemistry, MD is implemented in various steps of the drug discovery process[192]. It can be used, as cited above, to assess the stability of a certain model created for a biological entity of interest. For model refinement, the trajectory obtained should show any huge change in RMSF of the backbone, as it could be a signal for model instability[193]. Even some ligand-protein structural data are present for a target, MD can help to understand the importance of ligand present in the complex (evaluating the geometrical and energetics parameters of the protein deprived of the ligand).

One of the uses of MD in SBDD is the docking poses refinement[170]. It is reasonable for docking poses to be compared for their ability to preserve their shape in a dynamic environment. The poses in which the ligands preserve the conformation obtained with the docking procedure are associated with a higher “virtual stability” and scored higher in respect to the poses which present a higher RMSD during the trajectory[194].

To perform MD simulations, several packages are available nowadays. Among the others, the most famous are Amber[195], GROMACS[196], CHARMM[197], and Schrödinger Desmond[198]. All these suites offer specific tools for system preparation, equilibration, and analysis (e.g., Amber relies on AmberTools for these purposes).

The classic MD protocols are based on molecular mechanics, so each parameter of the system (bonds, angles, partial charges, etc.) is set up at the beginning and does not change during the simulation. If the case considered by the scientist involves the presence of coordination bonds or the formation/break of bonds, different MD techniques must be implemented. This is the case of Quantum-Mechanics/Molecular-Mechanics (QM/MM) simulations, in which the site interested in the events cited above is treated with a QM level of theory, while the rest of the system is left to MM[199]. This allows obtaining reasonable results with a modest computational effort, which would become huge if all the systems would be considered with QM theory.

Molecular dynamics is also applied to elucidate the mechanistic aspect of some biological events of difficult experimental evaluation, as several studies have witnessed[200]. Indeed, to investigate the inhibition mechanism of CCR5, Salmaso et al. used MD, and the trajectories were processed with a principal component analysis and clustering method[201]. Banu et al., in a recent study, evaluated the interaction between NALP3 and Apoptosis-associated speck-like protein (ASC) in the inflammasome exploiting NAMD as the software for MD simulations[202]. A molecular dynamics approach was implemented by Tanwar et al. to explore the effect of different mutations on the structural behavior of N-sulphoglucosamine sulphohydrolase (SGSH), the main protein involved in mucopolysaccharidosis type IIIA[203]. Remaining in the field of LSDs, Hodošček et al. used MD simulations to investigate on the stability of the complex between the Niemann Pick type C proteins (NPC1 and NPC2) and cholesterol, considering the cases in which this molecule is bound first to one and then to the other of the pockets available in the system[204].

MD refinement of docking poses has been applied in various studies regarding neuroinflammation. Among the cited works, Ahmad et al.[190] and Tandon-Sinha[182] implemented this technique in the last part of their drug discovery workflow. Szefler and Czelen used an Amber-based MD approach for evaluating the most promising candidates in a library of oxindole derivatives as inhibitors of GSK3[205], while Kalva et al. analyzed the most promising candidates for MMP-9 inhibition derived from a cross-docking selectivity study[206]. Özkılıç and Tüzün analyzed with MD the best candidates resulting from a massive VS, which were firstly filtrated for ADME properties, highlighting two promising inhibitors of KMO[207].

In a recent methodological study, Jamal et al. implemented MD simulations for the evaluation of the structural and thermodynamic stability of the best candidates from a Glide docking VS[208].

3. The importance of ADMET in computational drug discovery for neuroinflammation

In drug discovery, and mainly in lead optimization, much importance must be given to the Blood-Brain-Barrier (BBB) passage of the candidate molecules. The BBB is a complex structure composed predominantly of endothelial cells, which are typically in contact with one another through tight junctions, which grant a very high filtration efficiency for the substances passing through this border.

Said this, it appears clear that a proper pharmacokinetic profile prediction is needed when the activity of a molecule versus a certain target located in the CNS is assessed. Moreover, looking at the topic of neuroinflammation, it is obvious that a molecule acting in the CNS is required to pass the BBB and reach a therapeutical concentration at the site of action. To exactly foretell if this is going to happen a priori is not a trivial task.

Moreover, the active transport mechanisms are active across the BBB in a very distinct way[209]. Among these systems, one of the most known and studied is the P-glycoprotein (P-gp) mediated efflux. The ability of this protein engine to transport potential bioactive molecules from inside to the outside of the BBB granted to this machinery the second name of Multidrug Resistance Protein 1 (MDR1)[210]. One of the most known examples in which the P-gp importance appears clear is the one related to the loperamide, an opioid-receptor agonist mainly acting on the μ-opioid receptors. The high lipophilicity of this molecule allows it to easily pass the BBB, but the P-gp efflux system can efficiently pump the drug against gradient[211], almost eliminating its cerebral side effects at the therapeutical dosage and limiting its action to the intestine.

From a computational point of view, a proper ADMET prediction can be achieved in various ways. An early evaluation can certainly be done on various molecular descriptors, coming both from the 2D structure of a molecule and from the 3D arrangement of its atoms. Among the 2D descriptors, the combination of the ones involved in Lipinski's rule of five (molecular weight, logarithm of partition coefficient, number of hydrogen bond acceptors, and donors) can certainly be considered[212]. Of course, the very basic knowledge coming from a 2D descriptor must be enriched with data coming from several other sources. Furthermore, the success rate of ADMET models can be improved if also high quality in vitro and in vivo data are included[213]. At the present day, machine learning and artificial intelligence-driven protocols are widely used, both in academics and in industry[214].

As for everything related to data, the efficiency of these model-based methods relies on the quality (and, to a lesser extent, on the quantity) of data used[215]. Different assays, different assay conditions, and differences in procedure protocols are all factors that strongly impact the final overall quality of the dataset. This makes industries the most advantaged environment for the creation of clean datasets, mainly due to the remarkable consistency in their assay conditions and proceeding.

Some of the most known software used to predict the pharmacokinetic profile of molecules are Schrödinger QikProp[216], OptibriumStarDrop, Metasite[217], and SwissADME[218]. An exhaustive list of such tools can be found on more specialistic articles, like the one made by Peach et al[219]. Many of these tools do not simply calculate descriptors but are also able to model them for several ADMET predictions. These programs are in some cases also able to depict important features in terms of CYP interaction and molecule metabolism. In the last decades, the machine learning approaches allowed even to obtain models for P-gp interactions, as the scientific literature witnesses[220][221][222]. On the other hand, if the ability of prediction from data is in a constant increase, is always useful to remember the complexity of the biological systems considered in drug discovery. Even if the computational tools converge in one direction, the scientist has always to remember that all the predictive outputs coming from them are based on a limited (even if large, in some cases) dataset. It is also important to notice that the majority of the tools nowadays present can perform well in predicting molecules distribution based on passive transport in the various tissues of the organism but foretelling the effects of the active transport mechanisms remain a tough task[223]. The efficiency of the prediction relies on several key factors: the diversity of the dataset, the number of elements in the dataset which are similar to the examined molecules, the quality of the data, etc. Much importance has also to be given to the right choice of the variables to build the model upon. Indeed, it is not necessary the case that more variables considered equals a better model. The most suitable models, in fact, are generally simple, easily interpretable[224] and built on few, well-chosen variables (which could be descriptors or data coming from physicochemical, biological, or clinical data). A scheme of the features that a proper prediction model for BBB permeation should present is reported in Figure 3.

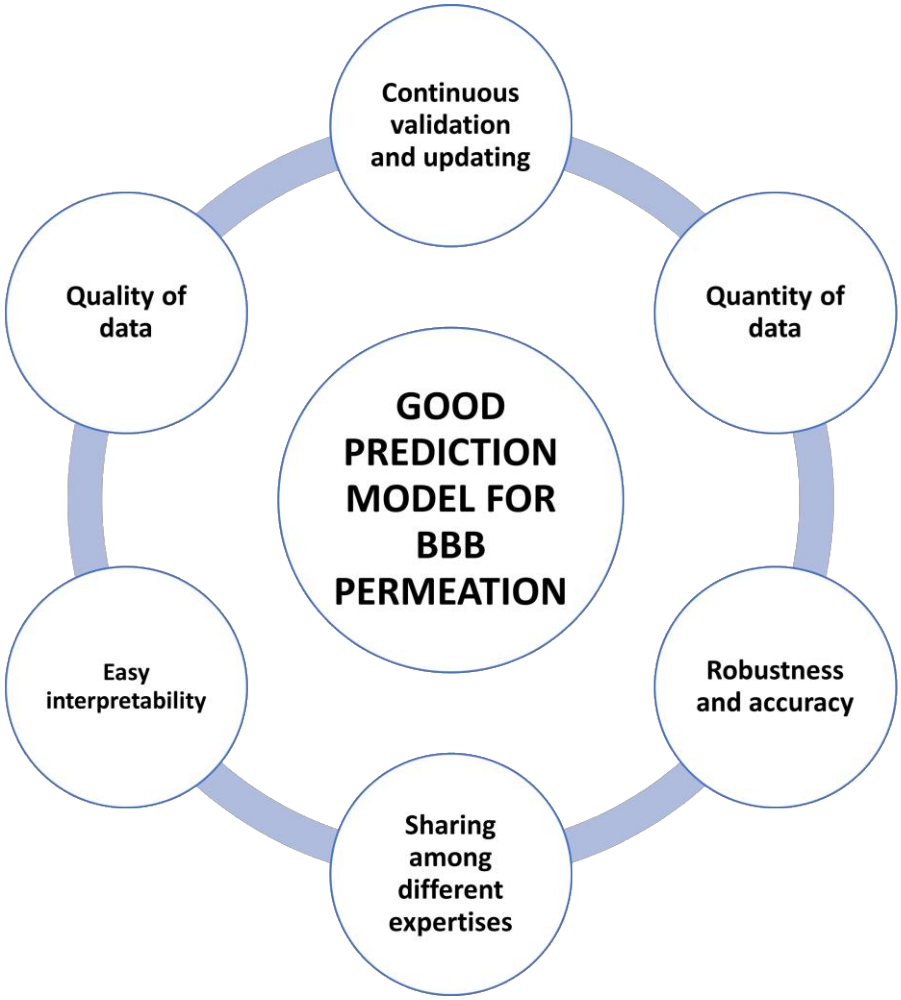


Figure 3. *Scheme reporting the main characteristics that a proper model for blood-brain barrier (BBB) passage prediction should present.*

A good prediction system for BBB passage based on molecular descriptors was developed by Hemmateenejad et al[225], exploiting a genetic algorithm-based artificial neural network, while Muehlbacher et al used a random forest-based method[226]. An ensemble method was recently applied in BBB passage prediction by Liu et al[227].

In the field of neuroinflammation, computational prediction of BBB passage has already been applied in various drug discovery protocols. Among these, Shahbazi et al. used Schrödinger QikProp for the development of potential NO-releasing anti-inflammatory molecules[228], while Dileep et al, with the same software, evaluated the BBB passage prediction of in silico designed potential multi-target ligands for neuroinflammation[229]. QikProp was also applied by Elrayess et al. for the ADME evaluation of potential triazole inhibitors of COX-2 against neurodegeneration[230].

4. Conclusion

In this review we highlighted the complexity beneath neuroinflammation, elucidating the main biological actors which play a role in its development. We then tried to give a comprehensive overview of the fundamental instruments and approaches nowadays applied from computational medicinal chemists for the discovery and the development of innovative therapeutical entities, underlying for each its advantages and its limitations. For each technique, we reported some cases in which its application was implemented for drug discovery in neuroinflammation, looking mainly at the last decade of research. In the years to come, we hope that a deeper knowledge of the neuroinflammation biological mechanism, together with a further improvement in both the efficiency and the predictivity of the computational tools will be useful in finding innovative, valid, and safe therapies for the pathological states associated to neuroinflammation and neurodegeneration.

References

[1] S.-M. Lucas, N. J. Rothwell, and R. M. Gibson, “The role of inflammation in CNS injury and disease,” *Br. J. Pharmacol.*, vol. 147, no. S1, pp. S232–S240, Jan. 2006, doi: 10.1038/sj.bjp.0706400.

[2] D. Furman *et al.*, “Chronic inflammation in the etiology of disease across the life span,” *Nat. Med. 2019 2512*, vol. 25, no. 12, pp. 1822–1832, Dec. 2019, doi: 10.1038/s41591-019-0675-0.

[3] C. K. Glass, K. Saijo, B. Winner, M. C. Marchetto, and F. H. Gage, “Mechanisms Underlying Inflammation in Neurodegeneration,” *Cell*, vol. 140, no. 6, pp. 918–934, Mar. 2010, doi: 10.1016/j.cell.2010.02.016.

[4] C. M. Gambino, B. Lo Sasso, G. Bivona, L. Agnello, and M. Ciaccio, “Aging and Neuroinflammatory Disorders: New Biomarkers and Therapeutic Targets,” *Curr. Pharm. Des.*, vol. 25, no. 39, Dec. 2019, doi: 10.2174/1381612825666191112093034.

[5] M. L. Block and J.-S. Hong, “Microglia and inflammation-mediated neurodegeneration: Multiple triggers with a common mechanism,” *Prog. Neurobiol.*, vol. 76, no. 2, pp. 77–98, Jun. 2005, doi: 10.1016/j.pneurobio.2005.06.004.

[6] D. Kempuraj *et al.*, “Neuroinflammation Induces Neurodegeneration,” *J. Neurol. Neurosurg. spine*, vol. 1, no. 1, 2016.

[7] N. S. Woodling and K. I. Andreasson, “Untangling the Web: Toxic and Protective Effects of Neuroinflammation and PGE2 Signaling in Alzheimer’s Disease,” *ACS Chem. Neurosci.*, vol. 7, no. 4, pp. 454–463, Apr. 2016, doi: 10.1021/ACSCHEMNEURO.6B00016.

[8] K. G. Hooten, D. R. Beers, W. Zhao, and S. H. Appel, “Protective and Toxic Neuroinflammation in Amyotrophic Lateral Sclerosis,” *Neurotherapeutics*, vol. 12, no. 2, p. 364, Apr. 2015, doi: 10.1007/S13311-014-0329-3.

[9] B. Obermeier, R. Daneman, and R. M. Ransohoff, “Development, maintenance and disruption of the blood-brain barrier,” *Nat. Med.*, vol. 19, no. 12, pp. 1584–1596, Dec. 2013, doi: 10.1038/nm.3407.

[10] M. J. Carson, J. C. Thrash, and B. Walter, “The cellular response in neuroinflammation: The role of leukocytes, microglia and astrocytes in neuronal death and survival,” *Clin. Neurosci. Res.*, vol. 6, no. 5, p. 237, Dec. 2006, doi: 10.1016/J.CNR.2006.09.004.

[11] J. P. Patel and B. N. Frey, “Disruption in the Blood-Brain Barrier: The Missing Link between Brain and Body Inflammation in Bipolar Disorder?,” *Neural Plast.*, vol. 2015, pp. 1–12, 2015, doi: 10.1155/2015/708306.

[12] D. C. Wraith and L. B. Nicholson, “The adaptive immune system in diseases of the central nervous system,” *J. Clin. Invest.*, vol. 122, no. 4, pp. 1172–1179, Apr. 2012, doi: 10.1172/JCI58648.

[13] T. Korn and A. Kallies, “T cell responses in the central nervous system,” *Nat. Rev. Immunol.*, vol. 17, no. 3, pp. 179–194, Mar. 2017, doi: 10.1038/nri.2016.144.

[14] G. B. Fields, “The Rebirth of Matrix Metalloproteinase Inhibitors: Moving Beyond the Dogma,” *Cells*, vol. 8, no. 9, Aug. 2019, doi: 10.3390/cells8090984.

[15] E. Colombo and C. Farina, “Astrocytes: Key Regulators of Neuroinflammation,” *Trends Immunol.*, vol. 37, no. 9, pp. 608–620, Sep. 2016, doi: 10.1016/j.it.2016.06.006.

[16] S. A. Liddelow and B. A. Barres, “Reactive Astrocytes: Production, Function, and Therapeutic Potential,” *Immunity*, vol. 46, no. 6, pp. 957–967, Jun. 2017, doi: 10.1016/J.IMMUNI.2017.06.006.

[17] H. S. Kwon and S.-H. Koh, “Neuroinflammation in neurodegenerative disorders: the roles of microglia and astrocytes,” *Transl. Neurodegener. 2020 91*, vol. 9, no. 1, pp. 1–12, Nov. 2020, doi:

10.1186/S40035-020-00221-2.

[18] J. Stephenson, E. Nutma, P. van der Valk, and S. Amor, “Inflammation in CNS neurodegenerative diseases,” *Immunology*, vol. 154, no. 2, p. 204, Jun. 2018, doi: 10.1111/IMM.12922.

[19] S. MV, “Molecular dissection of reactive astrogliosis and glial scar formation,” *Trends Neurosci.*, vol. 32, no. 12, pp. 638–647, Dec. 2009, doi: 10.1016/J.TINS.2009.08.002.

[20] M. L. Block, L. Zecca, and J.-S. Hong, “Microglia-mediated neurotoxicity: uncovering the molecular mechanisms,” *Nat. Rev. Neurosci.*, vol. 8, no. 1, pp. 57–69, Jan. 2007, doi: 10.1038/nrn2038.

[21] A. Nimmerjahn, “Resting Microglial Cells Are Highly Dynamic Surveillants of Brain Parenchyma in Vivo,” *Science (80-.).*, vol. 308, no. 5726, pp. 1314–1318, May 2005, doi: 10.1126/science.1110647.

[22] L. Fetler, “NEUROSCIENCE: Brain Under Surveillance: The Microglia Patrol,” *Science (80-.).*, vol. 309, no. 5733, pp. 392–393, Jul. 2005, doi: 10.1126/science.1114852.

[23] R. B. Rock *et al.*, “Role of Microglia in Central Nervous System Infections,” *Clin. Microbiol. Rev.*, vol. 17, no. 4, pp. 942–964, Oct. 2004, doi: 10.1128/CMR.17.4.942-964.2004.

[24] P. E and C. A, “Reciprocal interactions between microglia and neurons: from survival to neuropathology,” *Rev. Neurosci.*, vol. 13, no. 3, pp. 221–242, 2002, doi: 10.1515/REVNEURO.2002.13.3.221.

[25] V. Kumar, “Toll-like receptors in the pathogenesis of neuroinflammation,” *J. Neuroimmunol.*, vol. 332, Jul. 2019, doi: 10.1016/j.jneuroim.2019.03.012.

[26] D. V. F, C. S, B. P, and A. MP, “Purinergic signalling in inflammation of the central nervous system,” *Trends Neurosci.*, vol. 32, no. 2, pp. 79–87, Feb. 2009, doi: 10.1016/J.TINS.2008.11.003.

[27] H. J, L. JD, A. R, F. M, and S. SC, “Scavenger receptors in neurobiology and neuropathology: their role on microglia and other cells of the nervous system,” *Glia*, vol. 40, no. 2, pp. 195–205, Nov. 2002, doi: 10.1002/GLIA.10148.

[28] B. Kaminska, M. Mota, and M. Pizzi, “Signal transduction and epigenetic mechanisms in the control of microglia activation during neuroinflammation,” *Biochim. Biophys. Acta - Mol. Basis Dis.*, vol. 1862, no. 3, pp. 339–351, Mar. 2016, doi: 10.1016/J.BBADIS.2015.10.026.

[29] T. Liu, L. Zhang, D. Joo, and S.-C. Sun, “NF-κB signaling in inflammation,” *Signal Transduct. Target. Ther. 2017 21*, vol. 2, no. 1, pp. 1–9, Jul. 2017, doi: 10.1038/sigtrans.2017.23.

[30] R.-R. Ji, Z.-Z. Xu, and Y.-J. Gao, “Emerging targets in neuroinflammation-driven chronic pain,” *Nat. Rev. Drug Discov.*, vol. 13, no. 7, Jul. 2014, doi: 10.1038/nrd4334.

[31] A. Cianciulli, C. Porro, R. Calvello, T. Trotta, D. D. Lofrumento, and M. A. Panaro, “Microglia Mediated Neuroinflammation: Focus on PI3K Modulation,” *Biomolecules*, vol. 10, no. 1, p. 137, Jan. 2020, doi: 10.3390/biom10010137.

[32] M. Karin, “How NF-κB is activated: the role of the IκB kinase (IKK) complex,” *Oncogene*, vol. 18, no. 49, Nov. 1999, doi: 10.1038/sj.onc.1203219.

[33] S. Ghosh and M. S. Hayden, “New regulators of NF-κB in inflammation,” *Nat. Rev. Immunol.*, vol. 8, no. 11, Nov. 2008, doi: 10.1038/nri2423.

[34] I. Jridi, K. Canté-Barrett, K. Pike-Overzet, and F. J. T. Staal, “Inflammation and Wnt Signaling: Target for Immunomodulatory Therapy?,” *Front. Cell Dev. Biol.*, vol. 8, Feb. 2021, doi: 10.3389/fcell.2020.615131.

[35] L. Jia, J. Piña-Crespo, and Y. Li, “Restoring Wnt/β-catenin signaling is a promising therapeutic strategy for Alzheimer’s disease,” *Mol. Brain 2019 121*, vol. 12, no. 1, pp. 1–11, Dec. 2019, doi:

10.1186/S13041-019-0525-5.

[36] J. L. Stamos and W. I. Weis, “The β-Catenin Destruction Complex,” *Cold Spring Harb. Perspect. Biol.*, vol. 5, no. 1, Jan. 2013, doi: 10.1101/CSHPERSPECT.A007898.

[37] B. Becher, S. Spath, and J. Goverman, “Cytokine networks in neuroinflammation,” *Nat. Rev. Immunol. 2016 171*, vol. 17, no. 1, pp. 49–59, Dec. 2016, doi: 10.1038/nri.2016.123.

[38] G. Ramesh, A. G. Maclean, and M. T. Philipp, “Cytokines and chemokines at the crossroads of neuroinflammation, neurodegeneration, and neuropathic pain,” *Mediators Inflamm.*, vol. 2013, 2013, doi: 10.1155/2013/480739.

[39] A. JK, “Oxidative stress in neurodegeneration: cause or consequence?,” *Nat. Med.*, vol. 10 Suppl, no. 7, p. S18, 2004, doi: 10.1038/NRN1434.

[40] H. B, “Oxidative stress and neurodegeneration: where are we now?,” *J. Neurochem.*, vol. 97, no. 6, pp. 1634–1658, Jun. 2006, doi: 10.1111/J.1471-4159.2006.03907.X.

[41] H. Qin, S. A. Niyongere, S. J. Lee, B. J. Baker, and E. N. Benveniste, “Expression and Functional Significance of SOCS-1 and SOCS-3 in Astrocytes,” *J. Immunol.*, vol. 181, no. 5, pp. 3167–3176, Sep. 2008, doi: 10.4049/jimmunol.181.5.3167.

[42] S. W. Lallier, A. E. Graf, G. R. Waidyarante, and L. K. Rogers, “Nurr1 expression is modified by inflammation in microglia,” *Neuroreport*, vol. 27, no. 15, pp. 1120–1127, Oct. 2016, doi: 10.1097/WNR.0000000000000665.

[43] J.-M. Zhang and J. An, “Cytokines, Inflammation, and Pain,” *Int. Anesthesiol. Clin.*, vol. 45, no. 2, pp. 27–37, 2007, doi: 10.1097/AIA.0b013e318034194e.

[44] M. Tiberi and V. Chiurchiù, “Specialized Pro-resolving Lipid Mediators and Glial Cells: Emerging Candidates for Brain Homeostasis and Repair,” *Front. Cell. Neurosci.*, vol. 0, p. 136, Apr. 2021, doi: 10.3389/FNCEL.2021.673549.

[45] C. N. Serhan, N. Chiang, and T. E. Van Dyke, “Resolving inflammation: dual anti-inflammatory and pro-resolution lipid mediators,” *Nat. Rev. Immunol.*, vol. 8, no. 5, p. 349, May 2008, doi: 10.1038/NRI2294.

[46] L. Guzman-Martinez, R. B. Maccioni, V. Andrade, L. P. Navarrete, M. G. Pastor, and N. Ramos-Escobar, “Neuroinflammation as a Common Feature of Neurodegenerative Disorders,” *Front. Pharmacol.*, vol. 10, Sep. 2019, doi: 10.3389/fphar.2019.01008.

[47] A. H, “Inflammatory response in Alzheimer’s disease,” *Tohoku J. Exp. Med.*, vol. 174, no. 3, pp. 295–303, 1994, doi: 10.1620/TJEM.174.295.

[48] H. J and S. DJ, “The amyloid hypothesis of Alzheimer’s disease: progress and problems on the road to therapeutics,” *Science*, vol. 297, no. 5580, pp. 353–356, Jul. 2002, doi: 10.1126/SCIENCE.1072994.

[49] F. L. Heppner, R. M. Ransohoff, and B. Becher, “Immune attack: the role of inflammation in Alzheimer disease,” *Nat. Rev. Neurosci.*, vol. 16, no. 6, pp. 358–372, Jun. 2015, doi: 10.1038/nrn3880.

[50] M. T. Heneka *et al.*, “Neuroinflammation in Alzheimer’s disease,” *Lancet Neurol.*, vol. 14, no. 4, pp. 388–405, Apr. 2015, doi: 10.1016/S1474-4422(15)70016-5.

[51] T. Jiang, J. T. Yu, X. C. Zhu, and L. Tan, “TREM2 in Alzheimer’s disease,” *Mol. Neurobiol.*, vol. 48, no. 1, pp. 180–185, Feb. 2013, doi: 10.1007/S12035-013-8424-8/FIGURES/2.

[52] J. J. Bajramovic, “Regulation of Innate Immune Responses in the Central Nervous System,” *CNS Neurol. Disord. - Drug Targets*, vol. 10, no. 1, pp. 4–24, Feb. 2011, doi: 10.2174/187152711794488610.

[53] K. Takahashi, C. D. P. Rochford, and H. Neumann, “Clearance of apoptotic neurons without inflammation by microglial triggering receptor expressed on myeloid cells-2,” *J. Exp. Med.*, vol. 201, no. 4, pp. 647–657, Feb. 2005, doi: 10.1084/JEM.20041611.

[54] T. R. Jay, V. E. Von Saucken, and G. E. Landreth, “TREM2 in Neurodegenerative Diseases,” *Mol. Neurodegener. 2017 121*, vol. 12, no. 1, pp. 1–33, Aug. 2017, doi: 10.1186/S13024-017-0197-5.

[55] S. L. Zhou, C. C. Tan, X. H. Hou, X. P. Cao, L. Tan, and J. T. Yu, “TREM2 Variants and Neurodegenerative Diseases: A Systematic Review and Meta-Analysis,” *J. Alzheimers. Dis.*, vol. 68, no. 3, pp. 1171–1184, 2019, doi: 10.3233/JAD-181038.

[56] Q. Wang, Y. Liu, and J. Zhou, “Neuroinflammation in Parkinson’s disease and its potential as therapeutic target,” *Transl. Neurodegener.*, vol. 4, no. 1, p. 19, Dec. 2015, doi: 10.1186/s40035-015-0042-0.

[57] H. Braak, K. Del Tredici, U. Rüb, R. A. I. De Vos, E. N. H. Jansen Steur, and E. Braak, “Staging of brain pathology related to sporadic Parkinson’s disease,” *Neurobiol. Aging*, vol. 24, no. 2, pp. 197–211, Mar. 2003, doi: 10.1016/S0197-4580(02)00065-9.

[58] M. Oksanen, S. Lehtonen, M. Jaronen, G. Goldsteins, R. H. Hämäläinen, and J. Koistinaho, “Astrocyte alterations in neurodegenerative pathologies and their modeling in human induced pluripotent stem cell platforms,” *Cell. Mol. Life Sci.*, vol. 76, no. 14, p. 2739, Jul. 2019, doi: 10.1007/S00018-019-03111-7.

[59] N. P. Rocha, A. S. de Miranda, and A. L. Teixeira, “Insights into Neuroinflammation in Parkinson’s Disease: From Biomarkers to Anti-Inflammatory Based Therapies,” *Biomed Res. Int.*, vol. 2015, 2015, doi: 10.1155/2015/628192.

[60] W. Zhang *et al.*, “Aggregated α-synuclein activates microglia: a process leading to disease progression in Parkinson’s disease,” *FASEB J.*, vol. 19, no. 6, pp. 533–542, Apr. 2005, doi: 10.1096/FJ.04-2751COM.

[61] L. Rojanathammanee, E. J. Murphy, and C. K. Combs, “Expression of mutant alpha-synuclein modulates microglial phenotype in vitro,” *J. Neuroinflammation*, vol. 8, p. 44, May 2011, doi: 10.1186/1742-2094-8-44.

[62] E. C. Hirsch and S. Hunot, “Neuroinflammation in Parkinson’s disease: a target for neuroprotection?,” *Lancet Neurol.*, vol. 8, no. 4, pp. 382–397, Apr. 2009, doi: 10.1016/S1474-4422(09)70062-6.

[63] W. H, R. P, S. J, D. G, Z. L, and L. R, “Activation of microglia by human neuromelanin is NF-kappaB dependent and involves p38 mitogen-activated protein kinase: implications for Parkinson’s disease,” *FASEB J.*, vol. 17, no. 3, pp. 500–502, 2003, doi: 10.1096/FJ.02-0314FJE.

[64] X.-J. Wang *et al.*, “Impaired CD200–CD200R-mediated microglia silencing enhances midbrain dopaminergic neurodegeneration: Roles of aging, superoxide, NADPH oxidase, and p38 MAPK,” *Free Radic. Biol. Med.*, vol. 50, no. 9, pp. 1094–1106, May 2011, doi: 10.1016/j.freeradbiomed.2011.01.032.

[65] G. K. Sheridan and K. J. Murphy, “Neuron–glia crosstalk in health and disease: fractalkine and CX 3 CR1 take centre stage,” *Open Biol.*, vol. 3, no. 12, p. 130181, Dec. 2013, doi: 10.1098/rsob.130181.

[66] D. M, V. N, B. A, and P. T, “NURR1 in Parkinson disease--from pathogenesis to therapeutic potential,” *Nat. Rev. Neurol.*, vol. 9, no. 11, pp. 629–636, Nov. 2013, doi: 10.1038/NRNEUROL.2013.209.

[67] L. P. Rowland and N. A. Shneider, “Amyotrophic Lateral Sclerosis,” *N. Engl. J. Med.*, vol. 344, no. 22, pp. 1688–1700, May 2001, doi: 10.1056/NEJM200105313442207.

[68] M. C. Kiernan *et al.*, “Amyotrophic lateral sclerosis,” *Lancet*, vol. 377, no. 9769, pp. 942–955, Mar. 2011, doi: 10.1016/S0140-6736(10)61156-7.

[69] N. M *et al.*, “Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis,” *Science*, vol. 314, no. 5796, pp. 130–133, Oct. 2006, doi: 10.1126/SCIENCE.1134108.

[70] V. D. VM *et al.*, “TARDBP mutations in amyotrophic lateral sclerosis with TDP-43 neuropathology: a genetic and histopathological analysis,” *Lancet. Neurol.*, vol. 7, no. 5, pp. 409–416, May 2008, doi: 10.1016/S1474-4422(08)70071-1.

[71] L. J and W. F, “Role of Neuroinflammation in Amyotrophic Lateral Sclerosis: Cellular Mechanisms and Therapeutic Implications,” *Front. Immunol.*, vol. 8, no. AUG, Aug. 2017, doi: 10.3389/FIMMU.2017.01005.

[72] T. Philips and W. Robberecht, “Neuroinflammation in amyotrophic lateral sclerosis: role of glial activation in motor neuron disease,” *Lancet Neurol.*, vol. 10, no. 3, pp. 253–263, Mar. 2011, doi: 10.1016/S1474-4422(11)70015-1.

[73] Y. Liu, W. Hao, A. Dawson, S. Liu, and K. Fassbender, “Expression of Amyotrophic Lateral Sclerosis-linked SOD1 Mutant Increases the Neurotoxic Potential of Microglia via TLR2,” *J. Biol. Chem.*, vol. 284, no. 6, pp. 3691–3699, Feb. 2009, doi: 10.1074/JBC.M804446200.

[74] A. Neymotin *et al.*, “Lenalidomide (Revlimid®) administration at symptom onset is neuroprotective in a mouse model of amyotrophic lateral sclerosis,” *Exp. Neurol.*, vol. 220, no. 1, p. 191, Nov. 2009, doi: 10.1016/J.EXPNEUROL.2009.08.028.

[75] B. LH *et al.*, “A role for astrocytes in motor neuron loss in amyotrophic lateral sclerosis,” *Brain Res. Brain Res. Rev.*, vol. 47, no. 1–3, pp. 263–274, Dec. 2004, doi: 10.1016/J.BRAINRESREV.2004.05.003.

[76] I. M. Chiu *et al.*, “T lymphocytes potentiate endogenous neuroprotective inflammation in a mouse model of ALS,” *Proc. Natl. Acad. Sci.*, vol. 105, no. 46, pp. 17913–17918, Nov. 2008, doi: 10.1073/pnas.0804610105.

[77] “A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington’s disease chromosomes. The Huntington’s Disease Collaborative Research Group,” *Cell*, vol. 72, no. 6, pp. 971–983, Mar. 1993, doi: 10.1016/0092-8674(93)90585-E.

[78] R. CA and T. SJ, “Huntington’s disease: from molecular pathogenesis to clinical treatment,” *Lancet. Neurol.*, vol. 10, no. 1, pp. 83–98, Jan. 2011, doi: 10.1016/S1474-4422(10)70245-3.

[79] S. E. Browne, “Mitochondria and Huntington’s Disease Pathogenesis,” *Ann. N. Y. Acad. Sci.*, vol. 1147, no. 1, pp. 358–382, Dec. 2008, doi: 10.1196/annals.1427.018.

[80] G. Schilling *et al.*, “Characterization of Huntingtin Pathologic Fragments in Human Huntington Disease, Transgenic Mice, and Cell Models,” *J. Neuropathol. Exp. Neurol.*, vol. 66, no. 4, pp. 313–320, Apr. 2007, doi: 10.1097/nen.0b013e318040b2c8.

[81] C. A. Ross and M. A. Poirier, “What is the role of protein aggregation in neurodegeneration?,” *Nat. Rev. Mol. Cell Biol.*, vol. 6, no. 11, pp. 891–898, Nov. 2005, doi: 10.1038/nrm1742.

[82] R. Truant, R. S. Atwal, C. Desmond, L. Munsie, and T. Tran, “Huntington’s disease: revisiting the aggregation hypothesis in polyglutamine neurodegenerative diseases,” *FEBS J.*, vol. 275, no. 17, pp. 4252–4262, Sep. 2008, doi: 10.1111/j.1742-4658.2008.06561.x.

[83] S. J. Tabrizi *et al.*, “Biological and clinical manifestations of Huntington’s disease in the longitudinal TRACK-HD study: cross-sectional analysis of baseline data,” *Lancet Neurol.*, vol. 8, no. 9, pp. 791–801, Sep. 2009, doi: 10.1016/S1474-4422(09)70170-X.

[84] F. O. Walker, “Huntington’s disease,” *Lancet*, vol. 369, no. 9557, pp. 218–228, Jan. 2007, doi:

10.1016/S0140-6736(07)60111-1.

[85] A. Hodges *et al.*, “Regional and cellular gene expression changes in human Huntington’s disease brain,” *Hum. Mol. Genet.*, vol. 15, no. 6, pp. 965–977, Mar. 2006, doi: 10.1093/hmg/ddl013.

[86] Y. F. Tai *et al.*, “Microglial activation in presymptomatic Huntington’s disease gene carriers,” *Brain*, vol. 130, no. 7, pp. 1759–1766, May 2007, doi: 10.1093/brain/awm044.

[87] S. M. Hersch *et al.*, “Creatine in Huntington disease is safe, tolerable, bioavailable in brain and reduces serum 8OH2’dG,” *Neurology*, vol. 66, no. 2, pp. 250–252, Jan. 2006, doi: 10.1212/01.wnl.0000194318.74946.b6.

[88] F. Giorgini, P. Guidetti, Q. Nguyen, S. C. Bennett, and P. J. Muchowski, “A genomic screen in yeast implicates kynurenine 3-monooxygenase as a therapeutic target for Huntington disease,” *Nat. Genet.*, vol. 37, no. 5, pp. 526–531, May 2005, doi: 10.1038/ng1542.

[89] M. E. Bosch and T. Kielian, “Neuroinflammatory paradigms in lysosomal storage diseases,” *Front. Neurosci.*, vol. 9, no. OCT, p. 417, 2015, doi: 10.3389/FNINS.2015.00417/BIBTEX.

[90] L. D. Archer, K. J. Langford-Smith, B. W. Bigger, and J. E. Fildes, “Mucopolysaccharide diseases: A complex interplay between neuroinflammation, microglial activation and adaptive immunity,” *J. Inherit. Metab. Dis.*, vol. 37, no. 1, pp. 1–12, Jan. 2014, doi: 10.1007/s10545-013-9613-3.

[91] A. H. Futerman and G. van Meer, “The cell biology of lysosomal storage disorders,” *Nat. Rev. Mol. Cell Biol.*, vol. 5, no. 7, pp. 554–565, Jul. 2004, doi: 10.1038/nrm1423.

[92] R. Giugliani *et al.*, “Mucopolysaccharidosis I, II, and VI: Brief review and guidelines for treatment,” *Genet. Mol. Biol.*, vol. 33, no. 4, p. 589, 2010, doi: 10.1590/S1415-47572010005000093.

[93] J. L. Suarez-Guerrero, P. J. I. Gómez Higuera, J. S. Arias Flórez, and G. A. Contreras-García, “Mucopolysaccharidosis: clinical features, diagnosis and management,” *Rev. Chil. pediatría*, vol. 87, no. 4, pp. 295–304, Jul. 2016, doi: 10.1016/J.RCHIPE.2015.10.004.

[94] M. J. Valstar, G. J. G. Ruijter, O. P. van Diggelen, B. J. Poorthuis, and F. A. Wijburg, “Sanfilippo syndrome: A mini-review,” *J. Inherit. Metab. Dis.*, vol. 31, no. 2, pp. 240–252, Apr. 2008, doi: 10.1007/S10545-008-0838-5.

[95] J. DiRosario *et al.*, “Innate and adaptive immune activation in the brain of MPS IIIB mouse model,” *J. Neurosci. Res.*, vol. 87, no. 4, pp. 978–990, Mar. 2009, doi: 10.1002/jnr.21912.

[96] G. R. D. Villani, N. Gargiulo, R. Faraonio, S. Castaldo, E. Gonzalez y Reyero, and P. Di Natale, “Cytokines, neurotrophins, and oxidative stress in brain disease from mucopolysaccharidosis IIIB,” *J. Neurosci. Res.*, vol. 85, no. 3, pp. 612–622, Feb. 2007, doi: 10.1002/jnr.21134.

[97] A. Arfi, M. Richard, C. Gandolphe, D. Bonnefont-Rousselot, P. Thérond, and D. Scherman, “Neuroinflammatory and oxidative stress phenomena in MPS IIIA mouse model: The positive effect of long-term aspirin treatment,” *Mol. Genet. Metab.*, vol. 103, no. 1, pp. 18–25, May 2011, doi: 10.1016/j.ymgme.2011.01.015.

[98] K. Kollmann, K. Uusi-Rauva, E. Scifo, J. Tyynelä, A. Jalanko, and T. Braulke, “Cell biology and function of neuronal ceroid lipofuscinosis-related proteins,” *Biochim. Biophys. Acta - Mol. Basis Dis.*, vol. 1832, no. 11, pp. 1866–1881, Nov. 2013, doi: 10.1016/j.bbadis.2013.01.019.

[99] M. Haltia, “The neuronal ceroid-lipofuscinoses,” *J. Neuropathol. Exp. Neurol.*, vol. 62, no. 1, pp. 1–13, Jan. 2003, doi: 10.1093/JNEN/62.1.1.

[100] S. E. Mole, R. E. Williams, and H. H. Goebel, “Correlations between genotype, ultrastructural morphology and clinical phenotype in the neuronal ceroid lipofuscinoses,” *Neurogenetics*, vol. 6, no. 3, pp. 107–126, Sep. 2005, doi: 10.1007/s10048-005-0218-3.

[101] G. W. Anderson, H. H. Goebel, and A. Simonati, “Human pathology in NCL,” *Biochim. Biophys. Acta - Mol. Basis Dis.*, vol. 1832, no. 11, pp. 1807–1826, Nov. 2013, doi: 10.1016/j.bbadis.2012.11.014.

[102] S.-B. Dolisca, M. Mehta, D. A. Pearce, J. W. Mink, and B. L. Maria, “Batten Disease,” *J. Child Neurol.*, vol. 28, no. 9, pp. 1074–1100, Sep. 2013, doi: 10.1177/0883073813493665.

[103] C. Mencarelli and P. Martinez–Martinez, “Ceramide function in the brain: when a slight tilt is enough,” *Cell. Mol. Life Sci.*, vol. 70, no. 2, pp. 181–203, Jan. 2013, doi: 10.1007/s00018-012-1038-x.

[104] S. L. Macauley *et al.*, “Synergistic effects of central nervous system-directed gene therapy and bone marrow transplantation in the murine model of infantile neuronal ceroid lipofuscinosis,” *Ann. Neurol.*, vol. 71, no. 6, pp. 797–804, Jun. 2012, doi: 10.1002/ana.23545.

[105] T. D. Butters, “Gaucher disease,” *Curr. Opin. Chem. Biol.*, vol. 11, no. 4, pp. 412–418, Aug. 2007, doi: 10.1016/J.CBPA.2007.05.035.

[106] J. Ô. Stirnemann *et al.*, “A Review of Gaucher Disease Pathophysiology, Clinical Presentation and Treatments,” *Int. J. Mol. Sci.*, vol. 18, no. 2, p. 441, Feb. 2017, doi: 10.3390/IJMS18020441.

[107] A. Nagral, “Gaucher Disease,” *J. Clin. Exp. Hepatol.*, vol. 4, no. 1, p. 37, 2014, doi: 10.1016/J.JCEH.2014.02.005.

[108] L. Nalysnyk, P. Rotella, J. C. Simeone, A. Hamed, and N. Weinreb, “Gaucher disease epidemiology and natural history: a comprehensive review of the literature,” *Hematology*, vol. 22, no. 2, pp. 65–73, Feb. 2017, doi: 10.1080/10245332.2016.1240391/SUPPL_FILE/YHEM_A_1240391_SM3455.DOCX.

[109] D. M. Sama and C. M. Norris, “Calcium dysregulation and neuroinflammation: Discrete and integrated mechanisms for age-related synaptic dysfunction,” *Ageing Res. Rev.*, vol. 12, no. 4, pp. 982–995, Sep. 2013, doi: 10.1016/j.arr.2013.05.008.

[110] H. Alobaidy, “Recent Advances in the Diagnosis and Treatment of Niemann-Pick Disease Type C in Children: A Guide to Early Diagnosis for the General Pediatrician,” *Int. J. Pediatr.*, vol. 2015, pp. 1–10, 2015, doi: 10.1155/2015/816593.

[111] M. Baudry, Y. Yao, D. Simmons, J. Liu, and X. Bi, “Postnatal development of inflammation in a murine model of Niemann–Pick type C disease: immunohistochemical observations of microglia and astroglia,” *Exp. Neurol.*, vol. 184, no. 2, pp. 887–903, Dec. 2003, doi: 10.1016/S0014-4886(03)00345-5.

[112] H. D. Gallala, B. Breiden, and K. Sandhoff, “Regulation of the NPC2 protein-mediated cholesterol trafficking by membrane lipids,” *J. Neurochem.*, vol. 116, no. 5, pp. 702–707, Mar. 2011, doi: 10.1111/J.1471-4159.2010.07014.X.

[113] A. I. Rosenbaum and F. R. Maxfield, “Niemann-Pick type C disease: molecular mechanisms and potential therapeutic approaches,” *J. Neurochem.*, vol. 116, no. 5, pp. 789–795, Mar. 2011, doi: 10.1111/j.1471-4159.2010.06976.x.

[114] M. Patterson, “Niemann-Pick Disease Type C,” *GeneReviews®*, Dec. 2020.

[115] M. T. Vanier, “Niemann-Pick disease type C,” *Orphanet J. Rare Dis.*, vol. 5, no. 1, p. 16, Dec. 2010, doi: 10.1186/1750-1172-5-16.

[116] M. L. Wang *et al.*, “Identification of Surface Residues on Niemann-Pick C2 Essential for Hydrophobic Handoff of Cholesterol to NPC1 in Lysosomes,” *Cell Metab.*, vol. 12, no. 2, pp. 166–173, Aug. 2010, doi: 10.1016/j.cmet.2010.05.016.

[117] K. B. Peake, R. B. Campenot, D. E. Vance, and J. E. Vance, “Niemann-Pick Type C1 deficiency in microglia does not cause neuron death in vitro,” *Biochim. Biophys. Acta - Mol. Basis Dis.*, vol. 1812,

no. 9, pp. 1121–1129, Sep. 2011, doi: 10.1016/j.bbadis.2011.06.003.

[118] D. Smith, K.-L. Wallom, I. M. Williams, M. Jeyakumar, and F. M. Platt, “Beneficial effects of anti-inflammatory therapy in a mouse model of Niemann-Pick disease type C1,” *Neurobiol. Dis.*, vol. 36, no. 2, pp. 242–251, Nov. 2009, doi: 10.1016/j.nbd.2009.07.010.

[119] I. M. Williams, K.-L. Wallom, D. A. Smith, N. Al Eisa, C. Smith, and F. M. Platt, “Improved neuroprotection using miglustat, curcumin and ibuprofen as a triple combination therapy in Niemann–Pick disease type C1 mice,” *Neurobiol. Dis.*, vol. 67, pp. 9–17, Jul. 2014, doi: 10.1016/j.nbd.2014.03.001.

[120] A. D. Gitler, P. Dhillon, and J. Shorter, “Neurodegenerative disease: models, mechanisms, and a new hope,” *Dis. Model. Mech.*, vol. 10, no. 5, pp. 499–502, May 2017, doi: 10.1242/dmm.030205.

[121] H. Leicht *et al.*, “Predictors of Costs in Dementia in a Longitudinal Perspective,” *PLoS One*, vol. 8, no. 7, p. e70018, Jul. 2013, doi: 10.1371/JOURNAL.PONE.0070018.

[122] F. Durães, M. Pinto, and E. Sousa, “Old Drugs as New Treatments for Neurodegenerative Diseases,” *Pharmaceuticals*, vol. 11, no. 2, Jun. 2018, doi: 10.3390/PH11020044.

[123] A. B. Gurung, M. A. Ali, J. Lee, M. A. Farah, and K. M. Al-Anazi, “An Updated Review of Computer-Aided Drug Design and Its Application to COVID-19,” *Biomed Res. Int.*, vol. 2021, Jun. 2021, doi: 10.1155/2021/8853056.

[124] W. Yu and A. D. MacKerell, “Computer-Aided Drug Design Methods,” 2017.

[125] E. Lionta, G. Spyrou, D. Vassilatis, and Z. Cournia, “Structure-Based Virtual Screening for Drug Discovery: Principles, Applications and Recent Advances,” *Curr. Top. Med. Chem.*, vol. 14, no. 16, Oct. 2014, doi: 10.2174/1568026614666140929124445.

[126] V. Krishnan and B. Rupp, “Macromolecular Structure Determination: Comparison of X-ray Crystallography and NMR Spectroscopy,” in *eLS*, Chichester, UK: John Wiley & Sons, Ltd, 2012.

[127] S. K. Natchiar, A. G. Myasnikov, H. Kratzat, I. Hazemann, and B. P. Klaholz, “Visualization of chemical modifications in the human 80S ribosome structure,” *Nature*, vol. 551, no. 7681, Nov. 2017, doi: 10.1038/nature24482.

[128] X. Benjin and L. Ling, “Developments, applications, and prospects of cryo-electron microscopy,” *Protein Sci.*, vol. 29, no. 4, Apr. 2020, doi: 10.1002/pro.3805.

[129] P. Aparoy, K. Kumar Reddy, and P. Reddanna, “Structure and Ligand Based Drug Design Strategies in the Development of Novel 5- LOX Inhibitors,” *Curr. Med. Chem.*, vol. 19, no. 22, Jun. 2012, doi: 10.2174/092986712801661112.

[130] J. Vázquez, M. López, E. Gibert, E. Herrero, and F. J. Luque, “Merging Ligand-Based and Structure-Based Methods in Drug Discovery: An Overview of Combined Virtual Screening Approaches,” *Molecules*, vol. 25, no. 20, Oct. 2020, doi: 10.3390/molecules25204723.

[131] J. Shim and J. . A. D. MacKerell, “Computational ligand-based rational design: role of conformational sampling and force fields in model development,” *Medchemcomm*, vol. 2, no. 5, 2011, doi: 10.1039/c1md00044f.

[132] A. Lavecchia, “Machine-learning approaches in drug discovery: methods and applications,” *Drug Discov. Today*, vol. 20, no. 3, Mar. 2015, doi: 10.1016/j.drudis.2014.10.012.

[133] J. M. Alexander-Brett and D. H. Fremont, “Dual GPCR and GAG mimicry by the M3 chemokine decoy receptor,” *J. Exp. Med.*, vol. 204, no. 13, Dec. 2007, doi: 10.1084/jem.20071677.

[134] L. Toledo-Sherman *et al.*, “Optimization of Potent and Selective Ataxia Telangiectasia-Mutated Inhibitors Suitable for a Proof-of-Concept Study in Huntington’s Disease Models,” *J. Med. Chem.*, vol.

62, no. 6, Mar. 2019, doi: 10.1021/acs.jmedchem.8b01819.

[135] A. Long, H. Zhao, and X. Huang, “Structural Basis for the Interaction between Casein Kinase 1 Delta and a Potent and Selective Inhibitor,” *J. Med. Chem.*, vol. 55, no. 2, Jan. 2012, doi: 10.1021/jm201387s.

[136] P. A. Elkins *et al.*, “Structure of the C-terminally truncated human ProMMP9, a gelatin-binding matrix metalloproteinase,” *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 58, no. 7, Jul. 2002, doi: 10.1107/S0907444902007849.

[137] G. Madhavi Sastry, M. Adzhigirey, T. Day, R. Annabhimoju, and W. Sherman, “Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments,” *J. Comput. Aided. Mol. Des.*, vol. 27, no. 3, Mar. 2013, doi: 10.1007/s10822-013-9644-8.

[138] C. N. Cavasotto and S. S. Phatak, “Homology modeling in drug discovery: current trends and applications,” *Drug Discov. Today*, vol. 14, no. 13–14, Jul. 2009, doi: 10.1016/j.drudis.2009.04.006.

[139] Z. Xiang, “Advances in Homology Protein Structure Modeling,” *Curr. Protein Pept. Sci.*, vol. 7, no. 3, Jun. 2006, doi: 10.2174/138920306777452312.

[140] B. J. Bender, B. Marlow, and J. Meiler, “Improving homology modeling from low-sequence identity templates in Rosetta: A case study in GPCRs,” *PLOS Comput. Biol.*, vol. 16, no. 10, Oct. 2020, doi: 10.1371/journal.pcbi.1007597.

[141] A. Šali and T. L. Blundell, “Comparative Protein Modelling by Satisfaction of Spatial Restraints,” *J. Mol. Biol.*, vol. 234, no. 3, Dec. 1993, doi: 10.1006/jmbi.1993.1626.

[142] M. P. Jacobson *et al.*, “A hierarchical approach to all-atom protein loop prediction,” *Proteins Struct. Funct. Bioinforma.*, vol. 55, no. 2, Mar. 2004, doi: 10.1002/prot.10613.

[143] “Molecular Operating Environment (MOE), 2019.01; Chemical Computing Group ULC, 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2021.” .

[144] A. Waterhouse *et al.*, “SWISS-MODEL: homology modelling of protein structures and complexes,” *Nucleic Acids Res.*, vol. 46, no. W1, Jul. 2018, doi: 10.1093/nar/gky427.

[145] M. T. Muhammed and E. Aki-Yalcin, “Homology modeling in drug discovery: Overview, current applications, and future perspectives,” *Chem. Biol. Drug Des.*, vol. 93, no. 1, Jan. 2019, doi: 10.1111/cbdd.13388.

[146] R. Service, “‘The game has changed.’ AI triumphs at solving protein structures,” *Science (80-.).*, Nov. 2020, doi: 10.1126/science.abf9367.

[147] J. Jumper *et al.*, “Highly accurate protein structure prediction with AlphaFold,” *Nature*, vol. 596, no. 7873, pp. 583–589, Aug. 2021, doi: 10.1038/s41586-021-03819-2.

[148] E. Callaway, “‘It will change everything’: DeepMind’s AI makes gigantic leap in solving protein structures,” *Nature*, vol. 588, no. 7837, pp. 203–204, Dec. 2020, doi: 10.1038/d41586-020-03348-4.

[149] A. W. Senior *et al.*, “Protein structure prediction using multiple deep neural networks in the 13th Critical Assessment of Protein Structure Prediction (CASP13),” *Proteins Struct. Funct. Bioinforma.*, vol. 87, no. 12, pp. 1141–1148, Dec. 2019, doi: 10.1002/prot.25834.

[150] A. Kryshtafovych *et al.*, “Modeling SARS-CoV-2 proteins in the CASP-commons experiment,” *Proteins Struct. Funct. Bioinforma.*, vol. 89, no. 12, pp. 1987–1996, Dec. 2021, doi: 10.1002/prot.26231.

[151] A. Mullard, “What does AlphaFold mean for drug discovery?,” *Nat. Rev. Drug Discov.*, vol. 20, no. 10, pp. 725–727, Oct. 2021, doi: 10.1038/d41573-021-00161-0.

[152] H. Fan, “Refinement of homology-based protein structures by molecular dynamics simulation

techniques," *Protein Sci.*, vol. 13, no. 1, Jan. 2004, doi: 10.1110/ps.03381404.

[153] M. Bhargavi, S. K. Sivan, and S. R. Potlapally, "Identification of novel anti cancer agents by applying insilico methods for inhibition of TSPO protein," *Comput. Biol. Chem.*, vol. 68, Jun. 2017, doi: 10.1016/j.compbiolchem.2016.12.016.

[154] H. T. T. Lai and T. T. Nguyen, "Construction of dimeric hTSPO protein model using homology modeling and molecular dynamics," *J. Phys. Conf. Ser.*, vol. 1932, no. 1, May 2021, doi: 10.1088/1742-6596/1932/1/012016.

[155] S. P. Leelananda and S. Lindert, "Computational methods in drug discovery," *Beilstein J. Org. Chem.*, vol. 12, Dec. 2016, doi: 10.3762/bjoc.12.267.

[156] G. Wang and W. Zhu, "Molecular docking for drug discovery and development: a widely used approach but far from perfect," *Future Med. Chem.*, vol. 8, no. 14, Sep. 2016, doi: 10.4155/fmc-2016-0143.

[157] I. Halperin, B. Ma, H. Wolfson, and R. Nussinov, "Principles of docking: An overview of search algorithms and a guide to scoring functions," *Proteins Struct. Funct. Genet.*, vol. 47, no. 4, Jun. 2002, doi: 10.1002/prot.10115.

[158] G. Jones, P. Willett, R. C. Glen, A. R. Leach, and R. Taylor, "Development and validation of a genetic algorithm for flexible docking 1 1Edited by F. E. Cohen," *J. Mol. Biol.*, vol. 267, no. 3, Apr. 1997, doi: 10.1006/jmbi.1996.0897.

[159] G. M. Morris *et al.*, "AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility," *J. Comput. Chem.*, vol. 30, no. 16, Dec. 2009, doi: 10.1002/jcc.21256.

[160] R. A. Friesner *et al.*, "Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy," *J. Med. Chem.*, vol. 47, no. 7, Mar. 2004, doi: 10.1021/jm0306430.

[161] O. Korb, T. Stützle, and T. E. Exner, "PLANTS: Application of Ant Colony Optimization to Structure-Based Drug Design," 2006.

[162] J. K. Gagnon, S. M. Law, and C. L. Brooks, "Flexible CDOCKER: Development and application of a pseudo-explicit structure-based docking method within CHARMM," *J. Comput. Chem.*, vol. 37, no. 8, Mar. 2016, doi: 10.1002/jcc.24259.

[163] J. B. Cross *et al.*, "Comparison of Several Molecular Docking Programs: Pose Prediction and Virtual Screening Accuracy," *J. Chem. Inf. Model.*, vol. 49, no. 6, Jun. 2009, doi: 10.1021/ci900056c.

[164] Z. Wang *et al.*, "Comprehensive evaluation of ten docking programs on a diverse set of protein–ligand complexes: the prediction accuracy of sampling power and scoring power," *Phys. Chem. Chem. Phys.*, vol. 18, no. 18, 2016, doi: 10.1039/C6CP01555G.

[165] M. Lapillo, T. Tuccinardi, A. Martinelli, M. Macchia, A. Giordano, and G. Poli, "Extensive Reliability Evaluation of Docking-Based Target-Fishing Strategies," *Int. J. Mol. Sci.*, vol. 20, no. 5, Feb. 2019, doi: 10.3390/ijms20051023.

[166] A. Cuzzolin, M. Sturlese, I. Malvacio, A. Ciancetta, and S. Moro, "DockBench: An Integrated Informatic Platform Bridging the Gap between the Robust Validation of Docking Protocols and Virtual Screening Simulations," *Molecules*, vol. 20, no. 6, pp. 9977–9993, May 2015, doi: 10.3390/molecules20069977.

[167] G. L. Warren *et al.*, "A Critical Assessment of Docking Programs and Scoring Functions," *J. Med. Chem.*, vol. 49, no. 20, Oct. 2006, doi: 10.1021/jm050362n.

[168] M. L. Peach and M. C. Nicklaus, "Combining docking with pharmacophore filtering for improved

virtual screening," *J. Cheminform.*, vol. 1, no. 1, Dec. 2009, doi: 10.1186/1758-2946-1-6.

[169] S. L. Dixon, A. M. Smondyrev, and S. N. Rao, "PHASE: A Novel Approach to Pharmacophore Modeling and 3D Database Searching," *Chem. Biol. <html_ent glyph="@amp;" ascii="&"/> Drug Des.*, vol. 67, no. 5, May 2006, doi: 10.1111/j.1747-0285.2006.00384.x.

[170] G. Rastelli and L. Pinzi, "Refinement and Rescoring of Virtual Screening Results," *Front. Chem.*, vol. 7, Jul. 2019, doi: 10.3389/fchem.2019.00498.

[171] X.-Y. Meng, H.-X. Zhang, M. Mezei, and M. Cui, "Molecular Docking: A Powerful Approach for Structure-Based Drug Discovery," *Curr. Comput. Aided-Drug Des.*, vol. 7, no. 2, pp. 146–157, Jun. 2011, doi: 10.2174/157340911795677602.

[172] C. Gorgulla *et al.*, "An open-source drug discovery platform enables ultra-large virtual screens," *Nature*, vol. 580, no. 7805, Apr. 2020, doi: 10.1038/s41586-020-2117-z.

[173] G. Bolcato, A. Cuzzolin, M. Bissaro, S. Moro, and M. Sturlese, "Can We Still Trust Docking Results? An Extension of the Applicability of DockBench on PDBbind Database," *Int. J. Mol. Sci.*, vol. 20, no. 14, Jul. 2019, doi: 10.3390/ijms20143558.

[174] A. Cuzzolin, G. Deganutti, V. Salmaso, M. Sturlese, and S. Moro, "AquaMMapS: An Alternative Tool to Monitor the Role of Water Molecules During Protein-Ligand Association," *ChemMedChem*, vol. 13, no. 6, Mar. 2018, doi: 10.1002/cmdc.201700564.

[175] B. C. Roberts and R. L. Mancera, "Ligand−Protein Docking with Water Molecules," *J. Chem. Inf. Model.*, vol. 48, no. 2, Feb. 2008, doi: 10.1021/ci700285e.

[176] D. R. Houston and M. D. Walkinshaw, "Consensus Docking: Improving the Reliability of Docking in a Virtual Screening Context," *J. Chem. Inf. Model.*, vol. 53, no. 2, Feb. 2013, doi: 10.1021/ci300399w.

[177] T. Tuccinardi, G. Poli, V. Romboli, A. Giordano, and A. Martinelli, "Extensive Consensus Docking Evaluation for Ligand Pose Prediction and Virtual Screening Studies," *J. Chem. Inf. Model.*, vol. 54, no. 10, Oct. 2014, doi: 10.1021/ci500424n.

[178] B. Cheng *et al.*, "Synthesis and Anti-neuroinflammatory Activity of Lactone Benzoyl Hydrazine and 2-nitro-1-phenyl-1 *H* -Indole Derivatives as p38 *α* MAPK Inhibitors," *Chem. Biol. Drug Des.*, vol. 86, no. 5, Nov. 2015, doi: 10.1111/cbdd.12581.

[179] I. Rippin *et al.*, "Discovery and Design of Novel Small Molecule GSK-3 Inhibitors Targeting the Substrate Binding Site," *Int. J. Mol. Sci.*, vol. 21, no. 22, Nov. 2020, doi: 10.3390/ijms21228709.

[180] E. Cescon *et al.*, "Scaffold Repurposing of in-House Chemical Library toward the Identification of New Casein Kinase 1 δ Inhibitors," *ACS Med. Chem. Lett.*, vol. 11, no. 6, Jun. 2020, doi: 10.1021/acsmedchemlett.0c00028.

[181] S. Redenti *et al.*, "A Triazolotriazine-Based Dual GSK-3β/CK-1δ Ligand as a Potential Neuroprotective Agent Presenting Two Different Mechanisms of Enzymatic Inhibition," *ChemMedChem*, vol. 14, no. 3, Feb. 2019, doi: 10.1002/cmdc.201800778.

[182] A. Tandon and S. Sinha, "Structural insights into the binding of MMP9 inhibitors," *Bioinformation*, vol. 5, no. 8, Jan. 2011, doi: 10.6026/97320630005310.

[183] S. Razak *et al.*, "Molecular docking, pharmacokinetic studies, and in vivo pharmacological study of indole derivative 2-(5-methoxy-2-methyl-1H-indole-3-yl)-N'-[(E)-(3-nitrophenyl) methylidene] acetohydrazide as a promising chemoprotective agent against cisplatin induced organ damage," *Sci. Rep.*, vol. 11, no. 1, Dec. 2021, doi: 10.1038/s41598-021-84748-y.

[184] L.-J. Liu, K.-H. Leung, D. S.-H. Chan, Y.-T. Wang, D.-L. Ma, and C.-H. Leung, "Identification of a natural product-like STAT3 dimerization inhibitor by structure-based virtual screening," *Cell Death Dis.*, vol.

5, no. 6, Jun. 2014, doi: 10.1038/cddis.2014.250.

[185] S. S. Ray, R. J. Nowak, R. H. Brown, and P. T. Lansbury, “Small-molecule-mediated stabilization of familial amyotrophic lateral sclerosis-linked superoxide dismutase mutants against unfolding and aggregation,” *Proc. Natl. Acad. Sci.*, vol. 102, no. 10, Mar. 2005, doi: 10.1073/pnas.0408277102.

[186] P. Durai *et al.*, “Toll-like receptor 2 antagonists identified through virtual screening and experimental validation,” *FEBS J.*, vol. 284, no. 14, Jul. 2017, doi: 10.1111/febs.14124.

[187] J. Mahita, K. Harini, M. Rao Pichika, and R. Sowdhamini, “An *in silico* approach towards the identification of novel inhibitors of the TLR-4 signaling pathway,” *J. Biomol. Struct. Dyn.*, vol. 34, no. 6, Jun. 2016, doi: 10.1080/07391102.2015.1079243.

[188] B. Yilmazer, Z. B. Yagci, E. Bakar, B. Ozden, K. Ulgen, and E. Ozkirimli, “Investigation of novel pharmacological chaperones for Gaucher Disease,” *J. Mol. Graph. Model.*, vol. 76, pp. 364–378, Sep. 2017, doi: 10.1016/j.jmgm.2017.07.014.

[189] M. A. El-Zohairy, D. P. Zlotos, M. R. Berger, H. H. Adwan, and Y. M. Mandour, “Discovery of Novel CCR5 Ligands as Anticolorectal Cancer Agents by Sequential Virtual Screening,” *ACS Omega*, vol. 6, no. 16, Apr. 2021, doi: 10.1021/acsomega.1c00681.

[190] K. Ahmad *et al.*, “Targeting Caspase 8: Using Structural and Ligand-Based Approaches to Identify Potential Leads for the Treatment of Multi-Neurodegenerative Diseases,” *Molecules*, vol. 24, no. 9, May 2019, doi: 10.3390/molecules24091827.

[191] M. I. García-Aranda *et al.*, “Anti-inflammatory effect and inhibition of nitric oxide production by targeting COXs and iNOS enzymes with the 1,2-diphenylbenzimidazole pharmacophore,” *Bioorg. Med. Chem.*, vol. 28, no. 9, May 2020, doi: 10.1016/j.bmc.2020.115427.

[192] G. Martinez-Rosell, M. J. Harvey, and G. De Fabritiis, “Molecular-Simulation-Driven Fragment Screening for the Discovery of New CXCL12 Inhibitors,” *J. Chem. Inf. Model.*, vol. 58, no. 3, Mar. 2018, doi: 10.1021/acs.jcim.7b00625.

[193] M. De Vivo, M. Masetti, G. Bottegoni, and A. Cavalli, “Role of Molecular Dynamics and Related Methods in Drug Discovery,” *J. Med. Chem.*, vol. 59, no. 9, May 2016, doi: 10.1021/acs.jmedchem.5b01684.

[194] S. A. Hollingsworth and R. O. Dror, “Molecular Dynamics Simulation for All,” *Neuron*, vol. 99, no. 6, Sep. 2018, doi: 10.1016/j.neuron.2018.08.011.

[195] D.A. Case et al., “Amber 2021, University of California, San Francisco.” 2021.

[196] H. J. C. Berendsen, D. van der Spoel, and R. van Drunen, “GROMACS: A message-passing parallel molecular dynamics implementation,” *Comput. Phys. Commun.*, vol. 91, no. 1–3, Sep. 1995, doi: 10.1016/0010-4655(95)00042-E.

[197] B. R. Brooks *et al.*, “CHARMM: The biomolecular simulation program,” *J. Comput. Chem.*, vol. 30, no. 10, Jul. 2009, doi: 10.1002/jcc.21287.

[198] K. J. Bowers *et al.*, “Molecular dynamics---Scalable algorithms for molecular dynamics simulations on commodity clusters,” 2006, doi: 10.1145/1188455.1188544.

[199] A. Lodola and M. De Vivo, “The Increasing Role of QM/MM in Drug Discovery,” 2012.

[200] A. Bunker and T. Róg, “Mechanistic Understanding From Molecular Dynamics Simulation in Pharmaceutical Research 1: Drug Delivery,” *Front. Mol. Biosci.*, vol. 7, Nov. 2020, doi: 10.3389/fmolb.2020.604770.

[201] R. E. Salmas, M. Yurtsever, and S. Durdagi, “Investigation of Inhibition Mechanism of Chemokine Receptor CCR5 by Micro-second Molecular Dynamics Simulations,” *Sci. Rep.*, vol. 5, no. 1, Oct. 2015,

doi: 10.1038/srep13180.

[202] H. Banu, M. C. Joseph, and M. N. Nisar, “In-silico approach to investigate death domains associated with nano-particle-mediated cellular responses,” *Comput. Biol. Chem.*, vol. 75, Aug. 2018, doi: 10.1016/j.compbiolchem.2018.04.013.

[203] H. Tanwar, D. T. Kumar, C. G. P. Doss, and H. Zayed, “Bioinformatics classification of mutations in patients with Mucopolysaccharidosis IIIA,” *Metab. Brain Dis.*, vol. 34, no. 6, pp. 1577–1594, Dec. 2019, doi: 10.1007/s11011-019-00465-6.

[204] M. Hodošček and N. Elghobashi-Meinhardt, “Simulations of NPC1(NTD):NPC2 Protein Complex Reveal Cholesterol Transfer Pathways,” *Int. J. Mol. Sci.*, vol. 19, no. 9, p. 2623, Sep. 2018, doi: 10.3390/ijms19092623.

[205] P. Czeleń and B. Szefler, “The Oxindole Derivatives, New Promising GSK-3β Inhibitors as One of the Potential Treatments for Alzheimer’s Disease—A Molecular Dynamics Approach,” *Biology (Basel).*, vol. 10, no. 4, Apr. 2021, doi: 10.3390/biology10040332.

[206] S. Kalva, N. Agrawal, A. A. Skelton, and L. M. Saleena, “Identification of novel selective MMP-9 inhibitors as potential anti-metastatic lead using structure-based hierarchical virtual screening and molecular dynamics simulation,” *Mol. Biosyst.*, vol. 12, no. 8, 2016, doi: 10.1039/C6MB00066E.

[207] Y. Özkılıç and N. Ş. Tüzün, “In silico methods predict new blood-brain barrier permeable structure for the inhibition of kynurenine 3-monooxygenase,” *J. Mol. Graph. Model.*, vol. 100, Nov. 2020, doi: 10.1016/j.jmgm.2020.107701.

[208] S. Jamal, A. Grover, and S. Grover, “Machine Learning From Molecular Dynamics Trajectories to Predict Caspase-8 Inhibitors Against Alzheimer’s Disease,” *Front. Pharmacol.*, vol. 10, Jul. 2019, doi: 10.3389/fphar.2019.00780.

[209] W. Löscher and H. Potschka, “Blood-brain barrier active efflux transporters: ATP-binding cassette gene family,” *NeuroRX*, vol. 2, no. 1, Jan. 2005, doi: 10.1602/neurorx.2.1.86.

[210] R. Daneman and A. Prat, “The Blood–Brain Barrier,” *Cold Spring Harb. Perspect. Biol.*, vol. 7, no. 1, Jan. 2015, doi: 10.1101/cshperspect.a020412.

[211] R. N. Upton, “CEREBRAL UPTAKE OF DRUGS IN HUMANS,” *Clin. Exp. Pharmacol. Physiol.*, vol. 34, no. 8, Aug. 2007, doi: 10.1111/j.1440-1681.2007.04649.x.

[212] L. Z. Benet, C. M. Hosey, O. Ursu, and T. I. Oprea, “BDDCS, the Rule of 5 and drugability,” *Adv. Drug Deliv. Rev.*, vol. 101, Jun. 2016, doi: 10.1016/j.addr.2016.05.007.

[213] Y. Wang *et al.*, “*In silico* ADME/T modelling for rational drug design,” *Q. Rev. Biophys.*, vol. 48, no. 4, Nov. 2015, doi: 10.1017/S0033583515000190.

[214] B. Bhhatarai, W. P. Walters, C. E. C. A. Hop, G. Lanza, and S. Ekins, “Opportunities and challenges using artificial intelligence in ADME/Tox,” *Nat. Mater.*, vol. 18, no. 5, May 2019, doi: 10.1038/s41563-019-0332-5.

[215] A. H. Göller *et al.*, “Bayer’s in silico ADMET platform: a journey of machine learning over the past two decades,” *Drug Discov. Today*, vol. 25, no. 9, Sep. 2020, doi: 10.1016/j.drudis.2020.07.001.

[216] W. L. Jorgensen and E. M. Duffy, “Prediction of drug solubility from structure,” *Adv. Drug Deliv. Rev.*, vol. 54, no. 3, Mar. 2002, doi: 10.1016/S0169-409X(02)00008-X.

[217] G. Cruciani *et al.*, “MetaSite: Understanding Metabolism in Human Cytochromes from the Perspective of the Chemist,” *J. Med. Chem.*, vol. 48, no. 22, Nov. 2005, doi: 10.1021/jm050529c.

[218] A. Daina, O. Michielin, and V. Zoete, “SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules,” *Sci. Rep.*, vol. 7, no. 1, May

2017, doi: 10.1038/srep42717.

[219] M. L. Peach *et al.*, “Computational tools and resources for metabolism-related property predictions. 1. Overview of publicly available (free and commercial) databases and software,” *Future Med. Chem.*, vol. 4, no. 15, Oct. 2012, doi: 10.4155/fmc.12.150.

[220] O. Kadioglu and T. Efferth, “A Machine Learning-Based Prediction Platform for P-Glycoprotein Modulators and Its Validation by Molecular Docking,” *Cells*, vol. 8, no. 10, Oct. 2019, doi: 10.3390/cells8101286.

[221] R. Watanabe *et al.*, “Development of an *In Silico* Prediction Model for P-glycoprotein Efflux Potential in Brain Capillary Endothelial Cells toward the Prediction of Brain Penetration,” *J. Med. Chem.*, vol. 64, no. 5, Mar. 2021, doi: 10.1021/acs.jmedchem.0c02011.

[222] F. Montanari and G. F. Ecker, “Prediction of drug–ABC-transporter interaction — Recent advances and future challenges,” *Adv. Drug Deliv. Rev.*, vol. 86, Jun. 2015, doi: 10.1016/j.addr.2015.03.001.

[223] N. J. Yang and M. J. Hinner, “Getting Across the Cell Membrane: An Overview for Small Molecules, Peptides, and Proteins,” 2015.

[224] J. C. Madden, S. J. Enoch, A. Paini, and M. T. D. Cronin, “A Review of *In Silico* Tools as Alternatives to Animal Testing: Principles, Resources and Applications,” *Altern. to Lab. Anim.*, vol. 48, no. 4, Jul. 2020, doi: 10.1177/0261192920965977.

[225] B. Hemmateenejad, R. Miri, M. A. Safarpour, and A. R. Mehdipour, “Accurate prediction of the blood-brain partitioning of a large set of solutes using *ab initio* calculations and genetic neural network modeling,” *J. Comput. Chem.*, vol. 27, no. 11, Aug. 2006, doi: 10.1002/jcc.20437.

[226] M. Muehlbacher, G. M. Spitzer, K. R. Liedl, and J. Kornhuber, “Qualitative prediction of blood–brain barrier permeability on a large and refined dataset,” *J. Comput. Aided. Mol. Des.*, vol. 25, no. 12, Dec. 2011, doi: 10.1007/s10822-011-9478-1.

[227] L. Liu *et al.*, “Prediction of the Blood–Brain Barrier (BBB) Permeability of Chemicals Based on Machine-Learning and Ensemble Methods,” *Chem. Res. Toxicol.*, vol. 34, no. 6, Jun. 2021, doi: 10.1021/acs.chemrestox.0c00343.

[228] S. Shahbazi *et al.*, “Impact of novel *N* -aryl piperamide NO donors on NF-κB translocation in neuroinflammation: rational drug-designing synthesis and biological evaluation,” *Innate Immun.*, vol. 24, no. 1, Jan. 2018, doi: 10.1177/1753425917740727.

[229] K. V. Dileep, C. Remya, I. Tintu, and C. Sadasivan, “Designing of multi-target-directed ligands against the enzymes associated with neuroinflammation: an *in silico* approach,” *Front. Life Sci.*, vol. 7, no. 3–4, Dec. 2013, doi: 10.1080/21553769.2014.901924.

[230] R. Elrayess, M. S. Elgawish, M. Elewa, M. S. Nafie, S. S. Elhady, and A. S. A. Yassen, “Synthesis, 3D-QSAR, and Molecular Modeling Studies of Triazole Bearing Compounds as a Promising Scaffold for Cyclooxygenase-2 Inhibition,” *Pharmaceuticals*, vol. 13, no. 11, Nov. 2020, doi: 10.3390/ph13110370.

SARS-CoV-2 3CL^{pro} mutations selected in a VSV-based system confer resistance to nirmatrelvir, ensitrelvir, and GC376

Emmanuel Heilmann, Francesco Costacurta, Seyed Arad Moghadasi, Chengjin Ye, Matteo Pavan, **Davide Bassani**, Andre Volland, Claudia Ascher, Alexander Kurt Hermann Weiss, David Bante, Reuben S. Harris, Stefano Moro, Bernhard Rup, Luis Martinez-Sobrido, Dorothee von Laer

E. Heilmann et al., "SARS-CoV-2 3CL pro mutations selected in a VSV-based system confer resistance to nirmatrelvir, ensitrelvir, and GC376," Sci. Transl. Med., Oct. 2022, doi: 10.1126/scitranslmed.abq7360.

Abstract

Protease inhibitors are among the most powerful antiviral drugs. Nirmatrelvir is the first protease inhibitor against the SARS-CoV-2 protease 3CL^{pro} that has been licensed for clinical use. To identify mutations that confer resistance to this protease inhibitor, we engineered a chimeric vesicular stomatitis virus (VSV) that expressed a polyprotein composed of the VSV glycoprotein G, the SARS-CoV-2 3CL^{pro}, and the VSV polymerase L. Viral replication was thus dependent on the autocatalytic processing of this precursor protein by 3CL^{pro} and release of the functional viral polymerase L, and replication of this chimeric VSV was effectively inhibited by nirmatrelvir. Using this system, we applied nirmatrelvir to select for resistance mutations. Resistance was confirmed by retesting nirmatrelvir against the selected mutations in an additional VSV-based systems, in an independently developed cellular system, in a biochemical assay, and in a recombinant SARS-CoV-2 system. We demonstrate that some mutants are cross-resistant to ensitrelvir and GC376, whereas others are less resistant to these compounds. Furthermore, we found that most of these resistance mutations already existed in SARS-CoV-2 sequences that have been deposited in the NCBI and GISAID databases, indicating that these mutations were present in circulating SARS-CoV-2 strains.

1. Introduction

In late 2019, the zoonotic transmission of a new coronavirus, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), into the human population [1], has led to worldwide efforts to find effective treatments against the various pathologies caused by the virus. Inhibitors of viral enzymes, such as proteases, have proven to be highly potent drugs in the treatment of HIV and Hepatitis C virus infections. However, resistant viruses rapidly emerge unless the protease inhibitors are given in combination with other directly acting antivirals [2][3]. SARS-CoV-2 encodes two proteases. The

3-Chymotrypsin-like protease (3CL^{pro}) cleaves 11 sites in the viral polyproteins pp1a and pp1ab and is also referred to as the main protease (M^{pro}) or non-structural protein 5 (nsp5), indicating that it cleaves more sites than the second protease and its location within the polyproteins, respectively [4]. The second viral protease, Papain-like protease (PL^{pro}) cleaves three additional sites in pp1a and pp1ab [5]. Thus, both proteases are essential for viral replication and therefore interesting drug targets. Recently, the 3CL^{pro} inhibitor nirmatrelvir was approved for clinical use. Nirmatrelvir acts as a peptidomimetic, covalent inhibitor binding to the catalytic site cysteine (C145), thereby blocking its function [6][7][8]. Nirmatrelvir has been authorized in combination with ritonavir by the U.S. food and drug administration (FDA) for emergency use in high-risk SARS-CoV-2-infected individuals under the trade name Paxlovid (EUA 105 Pfizer Paxlovid, 22.12.2021). In the studies leading to the Paxlovid (nirmatrelvir/ritonavir) emergency use authorization (EUA), mouse hepatitis virus (MHV) 3CL^{pro} was used as a surrogate for SARS-CoV-2 3CL^{pro} to generate resistance data, which may not be the ideal system. In addition, very recently several preprints have described nirmatrelvir resistance mutations in authentic SARS-CoV-2, either generated de novo [9][10], found in isolates [11][12], or modelled in silico [13]. Working with SARS-CoV-2 requires biosafety level 3 (BSL3) installations due to its virulence [14]. Even more so, performing SARS-CoV-2 antibody or antiviral resistance studies demand utmost caution to avoid biosafety breaches and subsequent spread of mutant variants. To address these caveats, we describe in this study a BSL-2 system based on VSV that allows the selection of resistance mutations in the SARS-CoV-2 3CL^{pro} . Several mutations identified were validated in cell-based, biochemical, and recombinant SARS-CoV-2 assays and two mutations were found to be identical to (L167F) or at the same residue (Q192) as those described in the other manuscripts characterizing resistance in authentic SARS-CoV-2. We furthermore showed that some mutations selected by one 3CL^{pro} inhibitor can confer cross-resistance to other inhibitors. In contrast, other mutations appeared more inhibitor-specific. We attributed these effects to the distinct chemical structures of the inhibitors and occupation within the active site of the 3CL^{pro} . Lastly, we modelled catalytic site mutations with Robetta [15] and Molecular Operating Environment [16] to elucidate their mechanism of resistance.

2. Results

2.1. MHV 3CL^{pro} is less sensitive to nirmatrelvir than SARS-CoV2 3CL^{pro}

We compared the sensitivity of SARS-CoV-2 and MHV 3CL^{pro} to the active component of Paxlovid, nirmatrelvir, using the gain-of-signal variant of a VSV-based 3CL^{pro} measurement assay shown in fig. S1A and B and described in detail recently [17]. In brief, the coronavirus 3CL^{pro} proteases flanked by autocatalytic sites were cloned into the P protein of a red fluorescent protein (dsRed) expressing VSV. The P protein with the internal 3CL^{pro} is functional and essential for viral genome replication and dsRed expression. In the absence of protease inhibitor, the P:3CL^{pro} protein is autocatalytically cleaved and dsRed is not expressed. In the presence of a protease inhibitor, the P protein is functional, the VSV genome replicates and dsRed is expressed. Using this system, we found that MHV 3CL^{pro} showed a weaker response to nirmatrelvir than to SARS-CoV-2 3CL^{pro} (Fig. 1A). The sequence identity of the two proteins is only 50% (fig. S2), whereas the structures of the SARS-CoV-2 and MHV 3CL^{pro} enzymes are strongly conserved. However, the interaction site of nirmatrelvir (a distance of 5 Å or less from the compound) shows seven amino acid differences between the two enzymes, namely H164 - Q, M165 - L, P168 - S, V186 - R, R188 - A, T190 - V and A191 - V (counting from the first residue (serine) after the glutamine of the N-terminal cleavage site). We therefore suggest that MHV 3CL^{pro} is not an optimal proxy to study SARS resistance mutations.

2.2. A VSV-based non-gain-of-function system was generated to predict SARS-CoV-2 3CL^{pro} mutations

To generate a safer alternative to selection of drug resistant SARS-CoV-2 for studying mutants, we engineered a chimeric VSV variant, where the intergenic region between the glycoprotein (G) and the polymerase (L) was replaced by the 3CL^{pro} of SARS-CoV-2 (fig. S3A). Upon translation, G, 3CL^{pro}, and L form a surrogate polyprotein, which must be processed by 3CL^{pro} to generate the functional viral proteins G and L. This surrogate polyprotein mimics the polyprotein that is produced by SARS-CoV-2 as dimerization of 3CL^{pro} is obligate for its function [18][19] and cleavage of the cognate 3CL^{pro} N- and C-terminal motifs must occur for successful VSV replication. By applying an appropriate protease inhibitor (+PI), this processing is disturbed and therefore viral replication cannot occur (fig. S3B). Through passaging the chimeric VSV variant in presence of suboptimal concentrations of a protease inhibitor, 3CL^{pro} mutations that are generated by the error-prone viral polymerase [20][21] are selected for resistance to the inhibitor (fig. S3C). In a first proof-of-

concept study, we selected a mutant against the inhibitor GC376, which acquired the amino acid change in the 3CL^{pro} from phenylalanine to leucine at position 305 (F305L) in the autocatalytic cleavage motif at the C terminus of the protease. This virus gained a mildly faster replication kinetic and produced higher titers in the presence of GC376 and nirmatrelvir compared to the parental virus (Fig. 1B and C). Related coronaviruses have leucine at position 305 as a preferred cleavage motif (Fig. 1D); therefore the likely mechanism of the selection of F305L is autocleavage site optimization. We used the wild-type VSV-3CL^{pro} for subsequent mutation selection studies with nirmatrelvir.

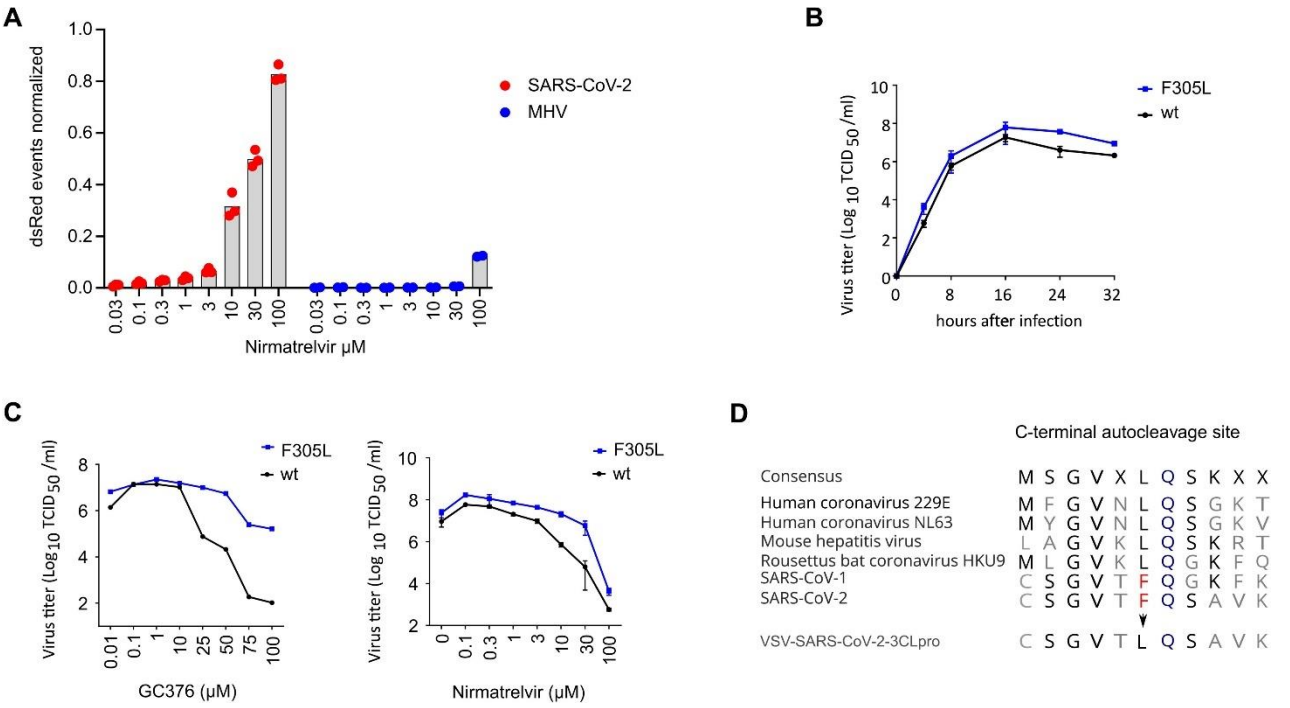


Fig. 1. *A VSV-based non-gain-of-function system was developed to predict SARS-CoV-2 3CL^{pro} mutations.* ***(A)*** *3CL^{pro} from SARS-CoV-2 and mouse hepatitis virus (MHV) were tested in a gain-of-signal assay. Data are presented as individual points of n = 3 biologically independent replicates per condition for SARS-CoV-2 3CL^{pro} and n = 2 for MHV 3CL^{pro}, average values are represented by histogram bars.* ***(B)*** *Replication kinetics are shown for wild-type (wt) VSV-G-3CL^{pro}-L and GC376-selected F305L variant. Data are presented as SD of n = 2 biologically independent replicates per condition.* ***(C)*** *GC376 and nirmatrelvir dose responses are shown for wild-type (wt) VSVG-3CL^{pro}-L and GC376-selected F305L variant. Data are presented as means of n = 2 (GC376) and n = 3 (Nirmatrelvir) biologically independent replicates per condition.* ***(D)*** *Sequence alignment of C-terminal autocleavage sites is shown for SARS-CoV-2 3CL^{pro} and related coronaviruses.*

We also included the F305L as parental virus for further selection experiments, because the F305L mutation has been found in regional outbreaks (mostly in England) and has been deposited in the global initiative on sharing avian flu data (GISAID) database [22][23][24] with three different codon usages to obtain leucine instead of phenylalanine (fig. S4A). The mutants were variants of Delta, mainly the sublineage AY.4 (fig. S4B). We therefore assumed F305L could be an advantageous

mutation that, in combination with further mutations, may give rise to protease inhibitor resistant lineages.

2.3. Nirmatrelvir resistant 3CL^{pro} mutants were selected for in the VSV-3CL^{pro} system

We next used the wild-type and F305L mutant viruses to select for nirmatrelvir resistant 3CLpro. BHK-21 cells in a 96-well plate were infected at a low multiplicity of infection (MOI; 0.01). Where cytopathic effects were visible in the first passage (25 out of 48 wells from parental wild-type and 17 out of 48 wells from parental F305L), supernatants were used for passaging individual wells with increasing concentrations (wild-type initial infection: 30 µM, second round: 40 µM, and third round: 50 µM; F305L initial infection: 50 µM, second round: 75 µM, and third round: 100 µM) of nirmatrelvir. At every passage, where cytopathic effects were observed again, supernatants were collected from the cell culture of individual 96-wells and transferred to individual new wells of a 96-well plate. At every passage, each well was sequenced individually, the target region being 3CL^{pro} and adjacent parts of G and L. We only counted mutants from unambiguous chromatogram peaks (as exemplified in fig. S3C). If, in the first or second passage there were still overlapping peaks, we sequenced the well again after the next passage. By this continuous selection pressure, the fittest mutant virus variant became dominant over the wild-type (and potential other mutants) in each well and made up the entirety of the genomic RNA, cDNA, and subsequent polymerase chain reaction (PCR) fragment. To finally exclude minority mutant populations that were not visible in a Sanger sequencing chromatogram, but could contribute to the resistance phenotype, mutations were later re-introduced into 3CL^{pro} measurement systems individually. We found 39 distinct mutations within 3CL^{pro} by Sanger sequencing. Viruses carried from one dominant mutation up to four mutations. The mutations were distributed over the entire sequence of 3CL^{pro} (Fig. 2A, table S1). We categorized them into catalytic site, near-catalytic site, dimerization interface and autocleavage site mutants. A fourth category for all mutations not fitting the first three was chosen as “allosteric” mutants. The mutants Y54C, L141F, L167F and Q192R occurred in residues in very close proximity (within 5 Å of PDB 3CLpro structure 7vh8, Fig. 2B, table S2) to nirmatrelvir. We searched for the mutants in the National Center for Biotechnology Information (NCBI) Virus data base [25] and GISAID EpiCoV [22][23][24], and found most of the mutations, or at least the same residue with a different mutation, in deposited sequences with varying coverage (Fig. 2A, table S1). We further subdivided GISAID entries into depositions made before and after the emergency use authorization of Paxlovid (nirmatrelvir/ritonavir) on 22th December, 2021 (table S3). An update of the Paxlovid EUA (18th March, 2022) included 3CL^{pro} mutants that were

retrieved from patients treated with Paxlovid (nirmatrelvir/ritonavir) (Fig. 2A). The update stated that it was unclear whether these mutations had clinical relevance [26].

2.4. Replications kinetics and dose response were analyzed for selected 3CL^{pro} mutants

To confirm these potential resistance mutations, we chose six virus samples to perform replication kinetics and dose-response experiments. Mutants were selected for further testing based on two criteria. First, we chose virus variants with catalytic site mutations because alterations in drug binding residues (direct or indirect) are more likely to alter efficacy. Second, we chose the most frequently recovered mutant outside of the catalytic site. Four samples were derived from wild-type VSV-G-3CL^{pro} -L and two were derived from the F305L variant. Supernatants for the replication kinetic experiments were collected at indicated time points after infection, and supernatants for the virus nirmatrelvir dose-response experiments were collected 24 hours after infection. The replication kinetics revealed that all variants were still capable of replicating to high titers, suggesting that resistance mutations did not result in a strong negative effect on 3CL^{pro} activity (Fig. 3). The dose responses showed that wild-type VSV-G-3CL^{pro} -L replication was inhibited by 106 -fold at 100 µM of nirmatrelvir, with a half maximal inhibitor concentration (IC50) of about 185 nM (Fig. 3A). We tested two L167F variants, because this mutant arose twice independently. The similarity of their dose responses (Fig. 3B and C), as well as the low variation of the biological replicates, suggests that the differences in the degree of resistance we observed between the mutants were not artifacts. We tested additional single mutants, namely the catalytic site mutant Y54C (Fig. 3D), a mutant from the mutation cluster shown in Fig. 2B, N203D (Fig. 3E) and the autocleavage site mutant F305L (Fig. 3F). To test if mutants that were selected from the F305L background had increased resistance, we also tested double mutants G138S/F305L (Fig. 3G) and Q192R/F305L (Fig. 3H). We observed the strongest resistance phenotype in the double mutant Q192R/F305L, replicating to high viral titers with a pronounced cytopathic effect (fig. S5) even in the presence of 100 µM nirmatrelvir.

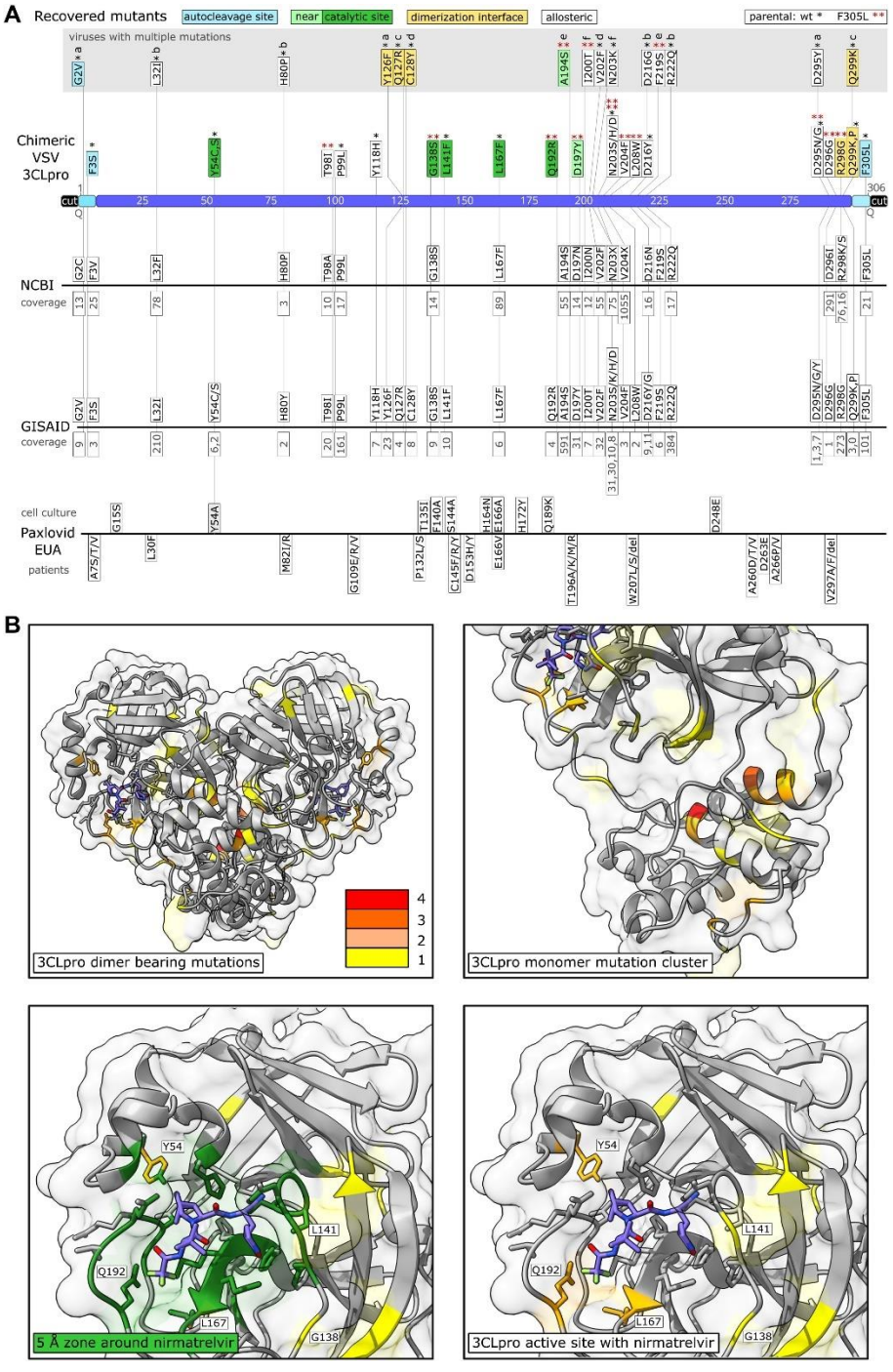


Fig. 2. *Sequencing of 3CL^{pro} escape mutants and comparison to data bases and Paxlovid EUA information.* ***(A)*** *Mutants were recovered from VSV-G-3CL^{pro}-L wildtype (*) and the F305L variant (red **). Autocleavage site mutants are colored in turquoise, catalytic site mutants in green, near catalytic site mutants in light green, dimerization interface mutants in yellow and "allosteric" mutants in white. Viruses with more than one mutation are displayed above in a gray box and named a to f. The number of mutated sequences in the databases from NCBI and GISAID are displayed below the mutations in gray. If specific mutations were not present in the database, the residue is displayed with any mutation that occurred at this position. Multiple such different amino acid changes that were not selected in our virus are displayed with X (N203X, V204X). Mutations from the Paxlovid EUA are divided into mutations found in cell culture and mutations sequenced from treated patients. The coverage of mutation entries was obtained on June second, 2022.* ***(B)*** *Visualizations of mutation-affected residues are shown. Residues that were mutated one time are highlighted in yellow, two times in light orange, three times dark orange, and four times in red. The 3CL^{pro} protease dimer with bound nirmatrelvir (blue) was visualized in ChimeraX from the Protein Data Bank structure 7VH8 (32). Catalytic center mutations are within a range of 5 Å as visualized in dark green.*

2.5. Re-introduction of 3CL^{pro} mutations confirms their resistance phenotype

As shown in table S1, VSV-induced 3CL^{pro} mutations were observed after the first passage when nirmatrelvir was applied. To validate the resistance data of replication-competent VSV-3CL^{pro} and at the same time exclude the effects of potential additional mutations arising within the dose-response experiment, we re-introduced some of the catalytic center mutations (Y54C, L167F, Q192R) into a recently developed protease activity measurement tool based on replication-incompetent VSV [17] (fig. S1A and B). In brief, the protease activity measurement tools comprise replication-incompetent VSV-dsRed variants missing either the viral phosphoprotein (ΔP) or polymerase (ΔL). These viruses are complemented with either an INTRAmolecular-3CL^{pro} tagged phosphoprotein or and INTERmolecular GFP-3CL^{pro} -L fusion protein, respectively. The P:3CLpro or GFP-3CLpro -L proteins are expressed in cells from transfected plasmids. The cells are then infected with the replication incompetent VSV-dsRed variant and treated with inhibitors. An intramolecular 3CL^{pro} tag in combination with VSV-ΔP-dsRed constitutes a gain-of-signal or “on”-assay. An intermolecular 3CL^{pro} tag in combination with VSV-ΔL-dsRed constitutes a loss-of-signal or “off”-assay.

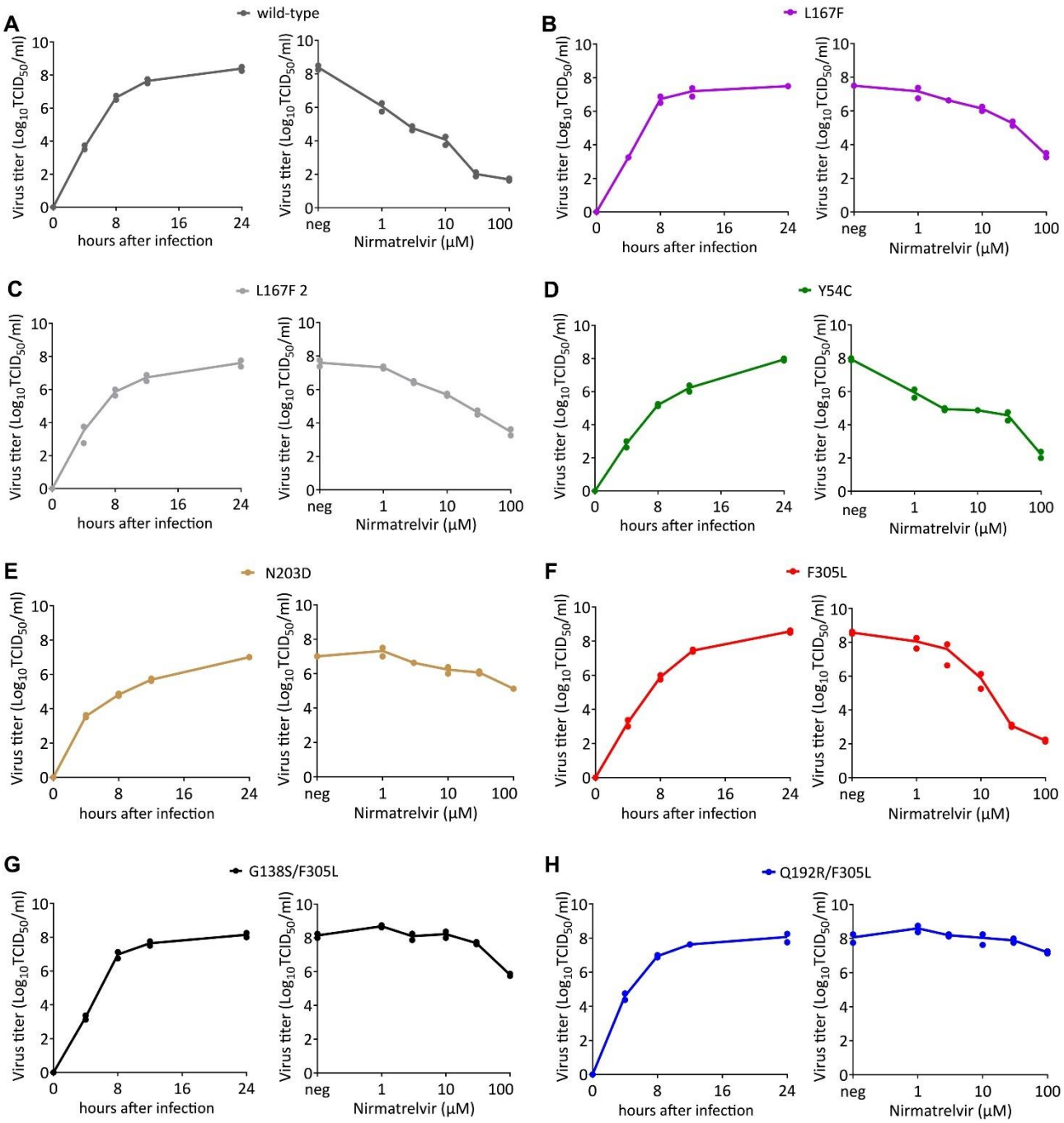


Fig. 3. *Replication kinetics and nirmatrelvir dose responses of parental VSV-G-3CL^{pro}-L and mutant variants.* ***(A to H)*** *Replication kinetics and dose responses are shown for wild-type* ***(A)****, L167F* ***(B)****, L167F-2* ***(C)****, Y54C* ***(D)****, N203D* ***(E)****, F305L* ***(F)****, G138S/F305L* ***(G)*** *and Q192R/F305L* ***(H)*** *VSV-G-3CL^{pro}-L. Supernatants for replication kinetics were collected at indicated time points. Supernatants for virus nirmatrelvir dose response were collected 24 hours after infection. (n = 2 biologically independent replicates per condition with individual data points shown and connecting lines of mean values). neg, without nirmatrelvir; TCID$_{50}$, 50% tissue culture infective dose.*

We found that the identified single catalytic center mutations indeed conferred partial resistance against nirmatrelvir of the 3CL^{pro} from the Wuhan-1 as well as the Omicron SARS-CoV-2 variant (Omicron signature mutation in 3CL^{pro} P132H) (Fig. 4A to D), which could be further enhanced by introduction of a second mutation in the autocleavage site (F305L, Fig. 4E to H). The mutation Q192R arose in the F305L parental virus. Introducing Q192R alone reduced 3CL^{pro} activity mildly,

as we observed by increased values in 3CL^{pro} -On-Q192R at low nirmatrelvir concentrations. Adding F305L as second mutation, thereby restoring the original combination from the double mutant virus, rescued this phenotype (Fig. 4G). A randomly selected combination of catalytic center mutations led to a strong loss in enzymatic activity (fig. S6A and B). We further introduced two mutants (A194S, G138S) into the 3CL^{pro} measurement assays, which also conferred resistance to nirmatrelvir (Fig. 4I and J).

2.6. Nirmatrelvir and GC376 react differently to 3CL^{pro} mutants

Comparing GC376 to nirmatrelvir in the 3CL^{pro} Y54C and L167F mutants directly revealed that these mutants react differently to the compounds. Y54C confers a similar resistance to GC376 as to nirmatrelvir (Fig. 5A). GC376 and nirmatrelvir interact similarly with the residue Y54, whereas L167 and Q192 are distant to GC376 and close to nirmatrelvir (within 5 Å) (Fig. 5B). L167F and Q192R appeared to affect the activity of GC376 less than nirmatrelvir (Fig. 5C and D, tables S4 and S5). Nirmatrelvir IC_{50} values were especially high in the 3CL^{pro} -On construct. We sought to improve the assays sensitivity by changing the read-out method from a FluoroSpot to a flow cytometry-based readout. With this approach, we could decrease the IC_{50} of the wild-type 3CL^{pro} -On to 0.91 µM of nirmatrelvir (fig. S6C, table S6).

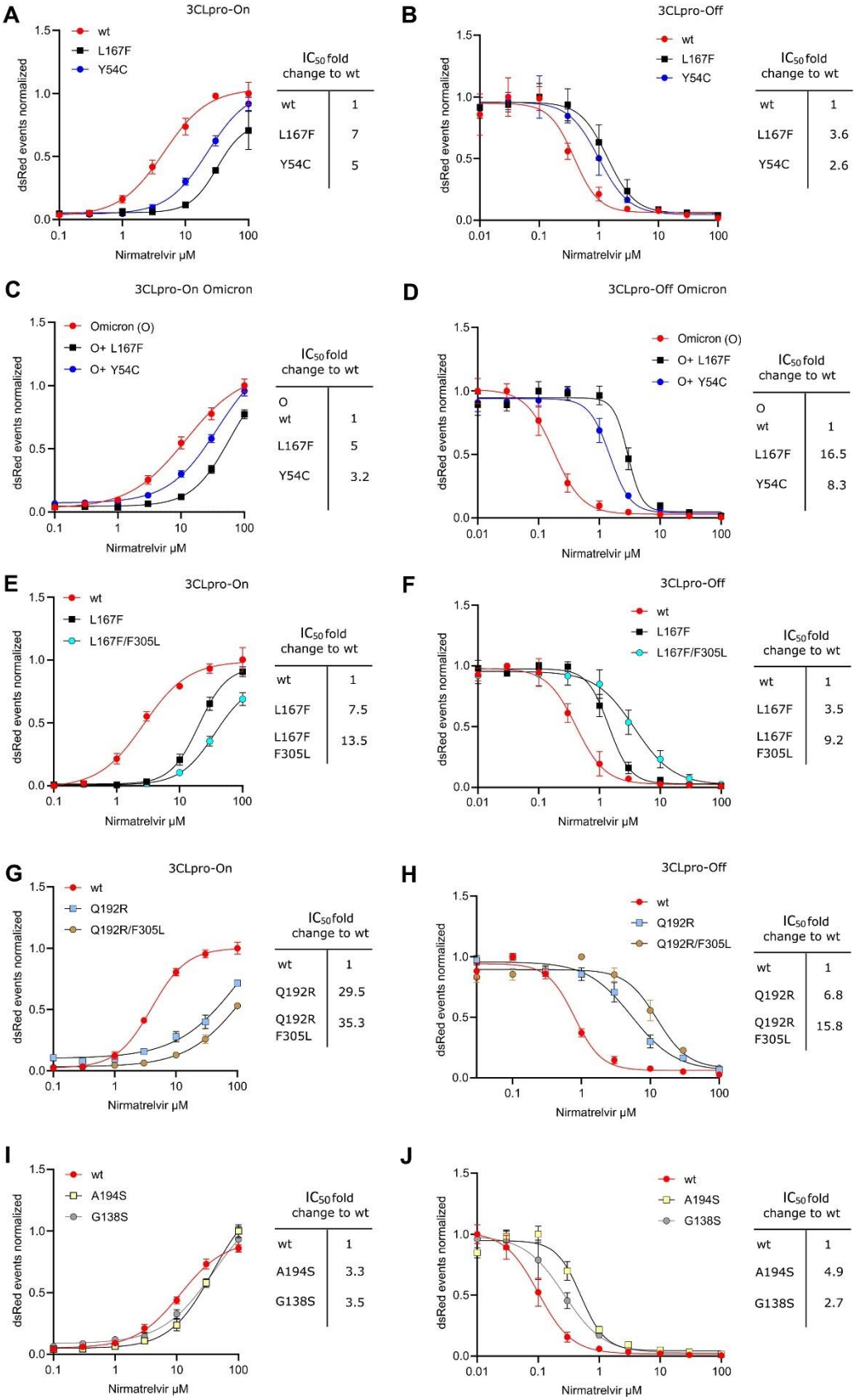


Fig. 4. *Re-introduction of individual or dual 3CL^{pro} mutations confirms their resistance phenotype. A graphic representation of the 3CL^{pro} -on and 3CL^{pro}-off system used to measure the inhibitory activity of the protease inhibitor against the different 3CL^{pro} mutants can be found in fig S3.* ***(A)*** *Gain-of-signal assay results are shown for single catalytic site mutations Y54C and L167F with nirmatrelvir. Data are presented as the standard deviation (SD) of n = 3 biologically independent replicates per condition.* ***(B)*** *Loss-of-signal assay results are shown for single catalytic site mutations Y54C and L167F with nirmatrelvir. Data are presented as the SD of n = 4 biologically independent replicates*

per condition. ***(C)*** *Gain of-signal assay results are shown for catalytic site mutations Y54C and L167F in combination with the Omicron 3CL^{pro} signature mutation P132H. Data are presented as the SD of n = 4 biologically independent replicates per condition.* ***(D)*** *Loss-of-signal assay results are shown for single catalytic site mutations Y54C and L167F in combination with the Omicron 3CL^{pro} signature mutation P132H. Data are presented as the SD of n = 4 biologically independent replicates per condition.* ***(E)*** *Gain-of-signal assay results are shown for double mutant L167F/F305L versus wild-type and single mutant L167F. Data are presented as the SD of n = 4 biologically independent replicates per condition.* ***(F)*** *Loss-of-signal assay results are shown for double mutant L167F/F305L versus wild-type and single mutant L167F. Data are presented as the SD of n = 4 biologically independent replicates per condition.* ***(G)*** *Gain-of-signal assay results are shown for double mutant Q192R-F305L versus wild-type and single mutant Q192R. Data are presented as the SD of n = 4 biologically independent replicates per condition.* ***(H)*** *Loss-of-signal assay results are shown for double mutant Q192R/F305L versus wild-type and single mutant Q192R. Data are presented as the SD of n = 4 biologically independent replicates per condition.* ***(I)*** *Gain-of-signal assay results are shown for mutants A194S and G138S versus wild-type. Data are presented as the SD of n = 4 biologically independent replicates per condition.* ***(J)*** *Loss-of-signal assay results are shown for of mutants A194S and G138S versus wild-type. Data are presented as the SD of n = 4 biologically independent replicates per condition.*

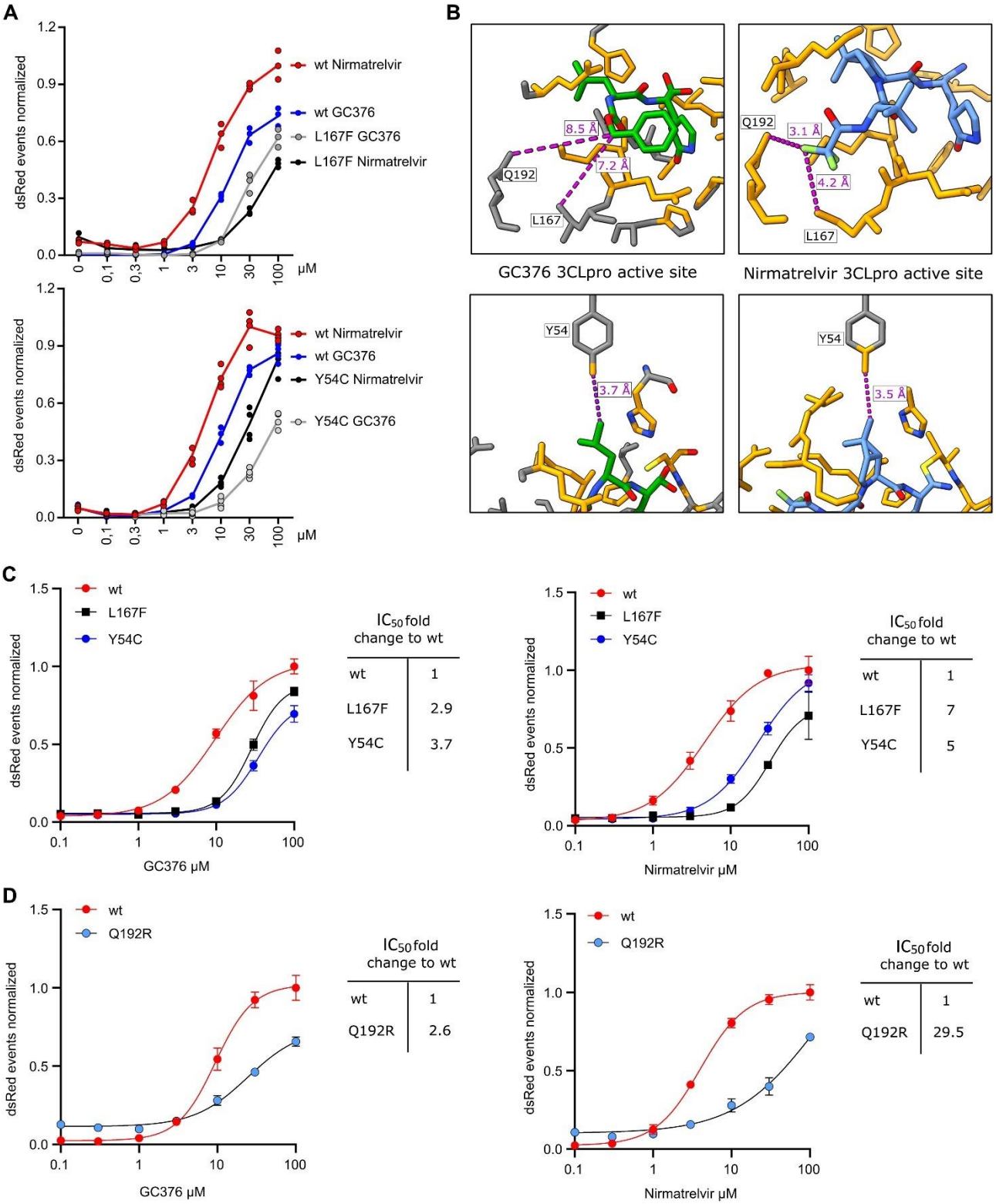


Fig. 5. *Nirmatrelvir and GC376 react differently to mutants.* ***(A)*** *Gain-of-signal assay results are shown for single mutants Y54C and L167F versus wild-type tested with GC376 and nirmatrelvir (Y54C: n = 4, L167F: n = 3 biologically independent replicates per condition).* ***(B)*** *GC376 (PDB: 7K0G) and nirmatrelvir (PDB: 7VH8) 3CL^{pro} crystal structures are shown with GC376 in green (and colored by heteroatom) and nirmatrelvir in light blue (and colored by heteroatom) and proximal residues in orange (within zone of 5 Å). Compound to residue distances are shown with dotted purple lines.* ***(C)*** *Fitting of gain-of-signal assay results are shown for single mutants Y54C and L167F versus wild-type tested with GC376 and nirmatrelvir.* ***(D)*** *Fitting of gain-of-signal assay results are shown for single mutant Q192R versus wild-type tested with GC376 and nirmatrelvir. Data in (C and D) are presented as the SD of n = 4 biologically independent replicates per condition.*

2.7. Confirmation of resistance mutations in a second cell-based assay system, biochemical assay, and with recombinant SARS-CoV-2

The resistance phenotype of L167F observed with gain- and loss-of-signal assays was confirmed using another recently published cellular system [27]. In this complementary assay, a polyprotein of Src, $3CL^{pro}$ with N-and C-terminal autocleavage sites, HIV Tat, and luciferase was used to repress transcription when $3CL^{pro}$ was active (fig. S7). Bona fide chemical inhibitors blocked $3CL^{pro}$ activity and restored luciferase signal in a dose-dependent manner (Fig. 6A to C). Similar to the results described above, L167F was more resistant to nirmatrelvir than GC376 (Fig. 6A and C, table S7). Furthermore, this mutant was most resistant to ensitrelvir, a recently developed compound in clinical trials in both the Src-$3CL^{pro}$ -Tat-Luc (Fig. 6B, table S7) [28] and our assay (fig. S8, table S8). For further confirmation of resistance phenotypes, we purified recombinant enzymes (fig. S9A to C). We tested catalytic activity with a substrate dose-response kinetic experiment with purified wild-type $3CL^{pro}$ versus mutants Y54C, L167F, and Q192R with the substrate Ac-Abu-Tle-Leu-Gln↓MCA releasing the fluorogenic molecule 7-amino-4-methylcoumarin (AMC) (fig. S10A). The ratio between the catalytic constant or turnover rate (kcat) and the Michaelis-Menten constant (Km), displayed as kcat/Km, showed that some of the mutants partially lose catalytic activity, most notable Q192R, which was in line with the cellular assays (Fig. 6D, table S9). For further resistance confirmation we applied a biochemical fluorescence resonance energy transfer (FRET) assay, which uses a quencher (DABCYL) and a fluorogenic substance (EDANS), which are connected by a peptide (KTSAVLQSGFRKME) that is recognized and cleaved by $3CL^{pro}$ (fig. S10B). Upon cleavage, fluorescence of EDANS increases. All three mutant $3CL^{pro}$ enzymes were more resistant to nirmatrelvir than the wild-type $3CL^{pro}$ (Fig. 6E, table S10). Lastly, we confirmed our findings in recombinant SARS-CoV-2 viruses expressing a reporter gene (fig. S10C) [29][30]. The recombinant SARS-CoV-2 variant expressing mCherry used for mutagenesis, aside from its transgene, was sequence identic to the Wuhan-1 variant [29][30] (Data file S1). Viruses carrying L167F alone and in combination with F305L were able to replicate, but replicated slower than wt virus and produced smaller plaques (Fig. 6F and G). The mutations introduced into recombinant SARS-CoV-2-mCherry had been found in clinical samples prior to our study (Data file S2). As expected, both L167F single and L167F/F305L double mutants were more resistant to nirmatrelvir than the wild-type (Fig. 6H, table S11).

2.8. Structural modelling of mutant 3CL^{pro} variants

To explore potential mechanisms of resistance, we performed molecular modelling with an in silico alanine mutation scanning as well as resistance mutation scanning with Molecular Operating Environment (MOE) suite [16] and the Robetta service [15]. MOE modelling was based on the PDB structure 7RFW [31], and Robetta modelling was based on 7VH8 [32], both of which are 3CL^{pro} structures with high nirmatrelvir occupancy. Experimental alanine scanning [33][34] as well as in silico alanine screening [35][36] is routinely utilized to evaluate the impact of single amino acid mutations on protein structure, and the models can provide plausible explanations for the structural basis of nirmatrelvir resistance. Fig. 7A shows that the most important losses of binding affinity are primarily related to mutation of residues whose sidechains directly contact the ligand, such as H41, M49, N142, H163, M165 and Q189, and secondarily to other residues lining the binding site like Y54, H164, E166, P168, D187, and Q192. Residues with hydrophobic sidechains, such as L27, Y54, F140, L167 and F181 seem to have a pivotal role in the structural integrity of the binding site (Fig. 7B), despite having a negligible impact on the variation of binding affinity. The primary effect of Y54C in these models is the disruption of a stabilizing inter-protein hydrogen bond to the backbone oxygen of D187 and disruption of additional weak but stabilizing interactions with surrounding hydrophobic residues such as a π-charge interaction between the phenyl ring of Y54 and the guanidinium group of R40. No major direct interaction to the 3.5 Å distant C20 methyl group of nirmatrelvir exists (Fig. 7C, fig. S11A). However, loss of the critical hydrogen bond between loop region 43 to 55 and the adjacent loop around D187 allows for a structural rearrangement destabilizing the distal part of the binding site, likely increasing the inherent plasticity of this protein region. Residue G138 lies in a solvent accessible loop, with backbone torsion angles in the β sheet region. Replacing it with a polar serine (G138S) while maintaining the same backbone conformation led to the Cβ of S138 pointing into the protein interior, and all of the preferred rotamers led to unfavorable interactions or required a rearrangement of the affected region. Formation of new hydrogen bonds, for example with a backbone hydrogen of F140 and the sulfur of C128 (Fig. 7D) likely led to a rearrangement of the S1 subpocket, which is responsible for hosting the terminal carboxamide moiety that mimics the P1 glutamine in natural peptide substrates. Supervised Molecular Dynamics simulations of the nirmatrelvir 3CL^{pro} recognition process revealed how L141 is one of the first residues that is contacted during the approach of nirmatrelvir into the binding site [37]. In the L167F mutant, the larger sidechain of phenylalanine cannot be accommodated without a structural rearrangement, which likely leads to repulsive interaction between the trifluoromethyl (CF_3) moiety of nirmatrelvir

and weakening its interactions with other proximal residues such as N142, which are thought to play a pivotal role in maneuvering the ligand entrance in the catalytic pocket [37]. As anticipated in the alanine scan, the L167F mutation seems to have an indirect effect on the binding affinity by alteration of the β sheet that constitutes the lower portion of the binding site, where a set of hydrogen bonds are established between nirmatrelvir and the backbone of both H164 and E166 (fig. S11B). A similar distortion of the binding pocket by the bulkier phenylalanine has also been described recently (10). Finally, polar Q192 stabilizes a solvent exposed loop participating in hydrogen bonds to backbone oxygen and nitrogen of V186, backbone oxygen of R188, and a stabilizing contact to the CF_3 group of nirmatrelvir. Replacement with positively charged R192 disrupted this network, which likely results in a structural rearrangement and altered binding to nirmatrelvir (Fig. 7E). The slight increase in binding affinity of Q192R concurrent with protease destabilization (Fig. 7A and B) could be explained by recontouring of the subpocket hosting the negatively polarized CF_3 moiety of nirmatrelvir and interacting with positively charged R192 (Fig. 7E, fig. S11C). Despite the predicted, marginally more favorable interaction with nirmatrelvir, an overall unfavorable effect of the mutation could still be possible due to altered sequestration of nirmatrelvir and the destabilization of the loop region lining the S2 and S4 subpocket of the catalytic site, where important residues such as Q189 are located.

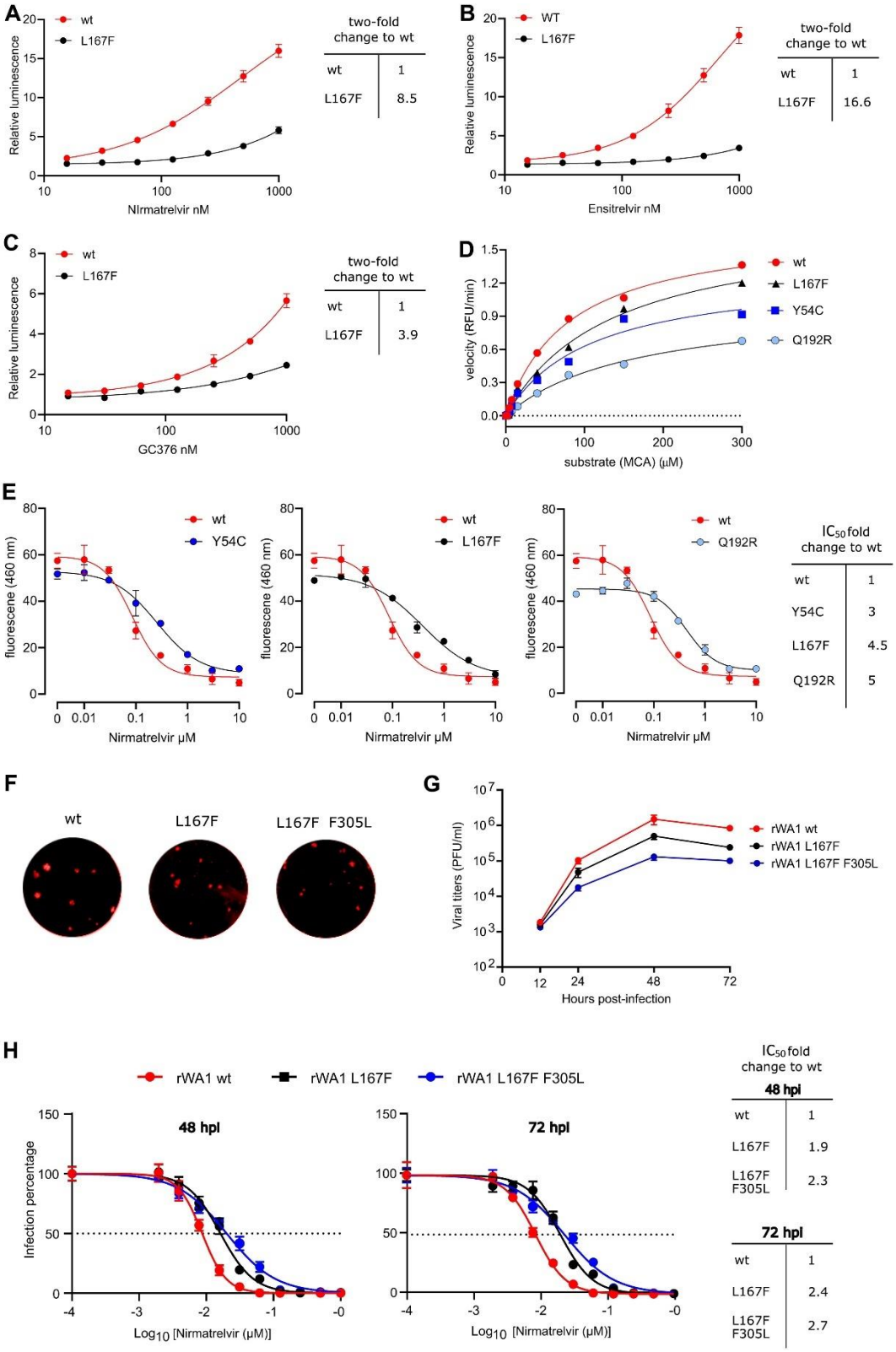
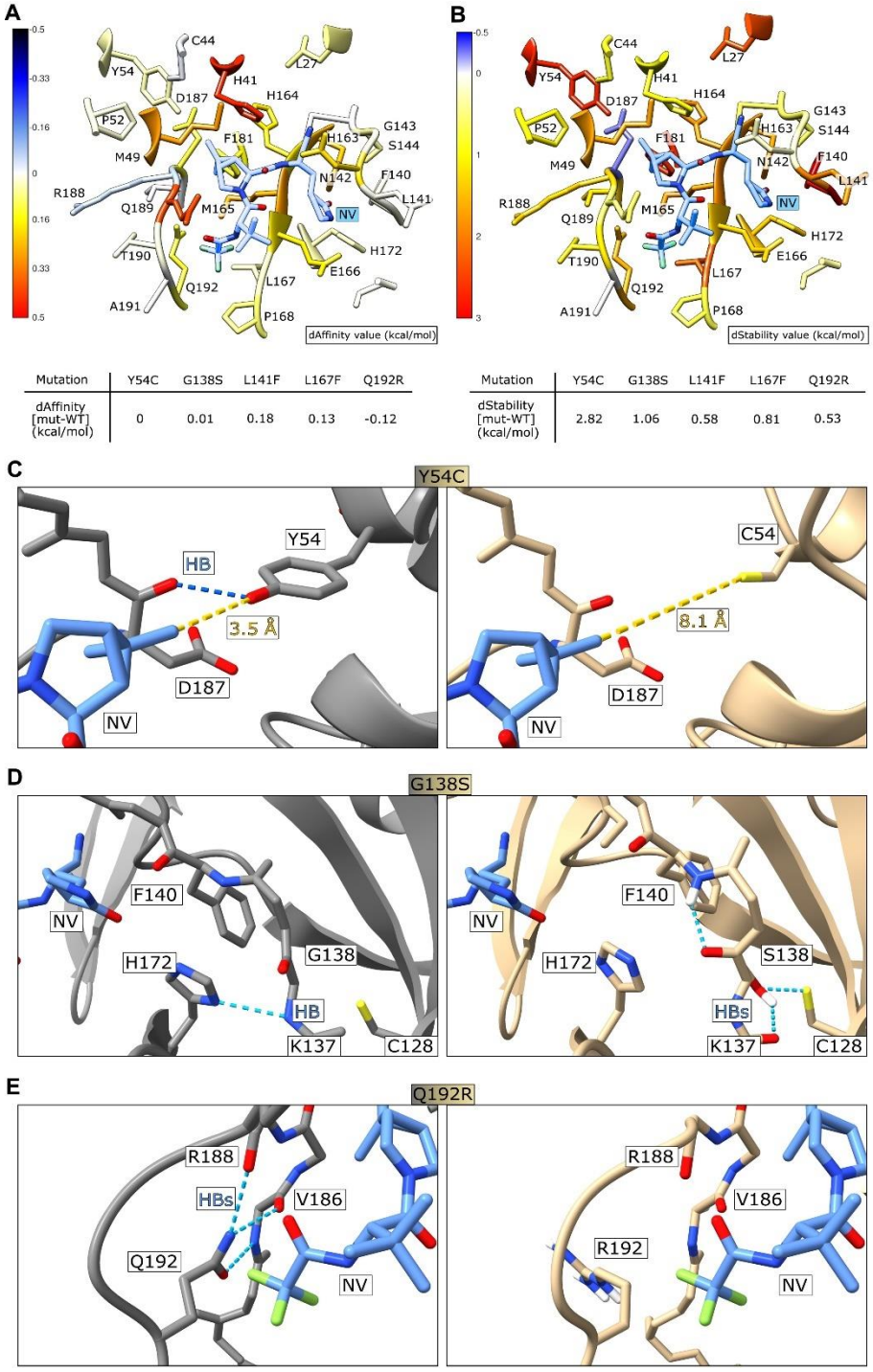


Fig. 6. *Cross-testing mutants and validation of enzyme kinetics. (A to C) Cross validation with cellular gain-of-signal assay based on Src-3CL^{pro}-Tat-Luc polyprotein. Wt Src-3CL^{pro}-Tat-Luc and L167F mutant were tested with nirmatrelvir* ***(A)****, ensitrelvir* ***(B)****, and GC376* ***(C)****. Data are presented as mean ± the standard deviation (SD) of n = 3 biologically independent replicates per condition.* ***(D)*** *Enzyme kinetics were measured for wt 3CL^{pro} and mutants with the substrate Ac-Abu-Tle-LeuGln↓MCA releasing the fluorogenic molecule MCA. Two biologically independent replicates per condition were used to calculate means for slopes. Slopes were used to compose a Michaelis Menten graph. Relative fluorescent units / minute (RFU/min) were used to plot the velocity of enzymes in the Michaelis Menten graph.* ***(E)*** *Results of a biochemical assay used for cross-validation are shown. In presence of an appropriate protease inhibitor, 3CL^{pro} cannot cleave the substrate and fluorescence is low. Without inhibitor, 3CL^{pro} cleaves the substrate peptide*

(KTSAVLQSGFRKME), quencher (DABCYL) and fluorogen (EDANS) are separated, and fluorescence increases. Nirmatrelvir dose responses are shown for wt versus mutant 3CL^{pro} variants; IC_{50} fold changes show varying resistance of the mutant enzymes. Data are presented as mean ± the standard deviation (SD) of n = 2 biologically independent replicates per condition. ***(F)*** *Recombinant wt, L167F and L167F/F305L SARS-CoV-2-mCherry (rWA1) exemplary plaques were imaged without magnification.* ***(G)*** *Replication kinetics are shown for recombinant wt SARS-CoV-2 (rWA1) versus L167F single and L167F F305L double mutant viruses. Data are presented as mean ± the standard deviation (SD) of n = 3 biologically independent replicates per condition.* ***(H)*** *Results are shown comparing the resistance of L167F single and L167F F305L double mutants to nirmatrelvir versus wt rWA1 expressing mCherry (top). The fold change in IC_{50} values are shown on the right. Data are shown for 48 and 72 hours post infection (hpi). The dotted line indicates 50% inhibition. Data are presented as mean ± SEM from quadruplicate wells of 2 independent experiments.*

3. Discussion

In our study, we selected mutations in the main protease 3CL^{pro} of SARS-CoV-2 against the protease inhibitor nirmatrelvir with a non-gain-of-function system based on VSV. The selected mutations were confirmed in two cellular assays and in one biochemical assay, along with confirmation using recombinant SARS-CoV-2. For the catalytic site mutations, a resistance mechanism was postulated based on mapping the mutations onto the co-crystal structure of 3CL^{pro} -nirmatrelvir and generating mutant models with Robetta [38]. In previous initial resistance studies leading to emergency use authorization of Paxlovid (nirmatrelvir/ritonavir), the 3CL^{pro} of a related coronavirus, MHV, was used to select for resistance mutations. The 3CL^{pro} of SARS-CoV-2 and MHV share 50% sequence identity. In this study, we compared the activity of 3CL^{pro} of SARS-CoV-2 and MHV and found that MHV 3CL^{pro} responded only mildly to nirmatrelvir in our gain-of-signal assay [17]. Although the structures of SARS-CoV-2 and MHV 3CL^{pro} are conserved, we propose that the low amino acid sequence identity alters the binding pocket affinity to nirmatrelvir sufficiently to reduce the sensitivity against the inhibitor. Key corresponding residues of the binding pocket (within 5 Å or less) are different, namely H164 - Q, M165 - L, P168 - S, V186 - R, R188 - A, T190 - V and A191 - V. Furthermore, amino acid changes that occurred in our selection experiments, Y126F and F305L are already present in the MHV 3CL^{pro} sequence. Taken together, we argue that MHV 3CL^{pro} was not an optimal proxy for resistance studies. Recently, chimeric VSV variants with SARS-CoV-2 spike were used to predict spike protein immune escape mutations by selecting against neutralizing serum [39][40][41]. The fast occurrence of mutations was facilitated in those studies by the high error rate of the VSV polymerase [20][21]. In a similar approach, we exploited this high error rate in a recombinant VSV expressing 3CL^{pro} to select for 3CL^{pro} mutations that confer resistance against protease inhibitors.



Mutation	Y54C	G138S	L141F	L167F	Q192R
dAffinity [mut-WT] (kcal/mol)	0	0.01	0.18	0.13	-0.12

Mutation	Y54C	G138S	L141F	L167F	Q192R
dStability [mut-WT] (kcal/mol)	2.82	1.06	0.58	0.81	0.53

Fig. 7. *Structural modelling of mutant 3CL^{pro} variants.* ***(A)*** *Colorimetric mapping of the dAffinity value (kcal/mol) by virtual alanine scanning with MOE suite. Residues within 5 Å of the nirmatrelvir position are displayed. Colors range from blue (negative values, indicating increased protein-ligand affinity) to red (positive values, indicating decreased protein-ligand affinity). The nirmatrelvir (NV) structure is shown in light blue.* ***(B)*** *Colorimetric mapping of the dStability value (kcal/mol), computed as above for (A). Colors range from blue (negative values, indicating increased in the protein stability) to red (positive values, indicating decreased protein stability.* ***(C)*** *The catalytic center of 3CL^{pro} from PDB structure 7VH8 is shown with nirmatrelvir bound. Y54 (left) forms a strong hydrogen bond (HB, highlighted with a blue dashed line) with D187, whereas nirmatrelvir is at a distance of 3.5 Å (yellow dashed line). The exchange of Y54 with C (right) leads to a loss of the hydrogen bond to D187 and makes room in the nirmatrelvir binding pocket due to the smaller sidechain of cysteine versus tyrosine.* ***(D)*** *G138 (left) contacts H172 with a hydrogen bond. S138 (right) forms several new hydrogen bonds with the backbone hydrogen of F140, backbone oxygen of K137 and the sulfur of C128.* ***(E)*** *Q192 (left) forms hydrogen bonds with the oxygen and nitrogen of V186, the oxygen of R188, and stabilizes the polar contact to the CF_3 group of nirmatrelvir. R192 (right) disrupts this hydrogen bond network; subsequent rearrangement could form additional interactions with the CF_3 group.*

The 3CL^{pro} was used to replace the function of an intergenic region between the viral glycoprotein (G) and the polymerase (L). The intergenic regions of VSV are responsible for separate gene expression, which in other viruses is accomplished by a polyprotein and proteases. Although this polyprotein of VSV-G-3CL^{pro} -L is only a surrogate to the one in SARS-CoV-2, the cognate cleavage sites, the requirement for dimerization of the protease for proper function (18, 19) and the context of a replicating virus in the cell make this approach an attractive proxy. Initially, we selected the 3CL^{pro} mutant F305L using GC376, which showed reduced sensitivity to GC376, as well as to nirmatrelvir. This mutation lies in the 3CL^{pro} cleavage site that flanks 3CL^{pro} at its C terminus. Interestingly, the LQ motif found in the mutated site is indeed known to be preferred over FQ as a target motif for 3CL^{pro} [42][43][44], which may explain the reduced sensitivity of the F305L mutant to the protease inhibitors. We then selected both wild-type and the F305L mutant against nirmatrelvir. We also used F305L as parental virus because we found that this variant existed already in regional outbreaks (mainly in England), underlining the viability of this mutation and its potential replicative advantage. These clusters were mainly of the Delta subvariant AY.4. Delta was replaced gradually by Omicron, which may have ended the spread of the Delta F305L. Nevertheless, we also found combinations of the Omicron signature mutation P132H with F305L. We were therefore interested in finding potential combinations of F305L with further protease inhibitor resistance mutations, assuming that the combination would show a higher degree of resistance than single mutations, which we did observe. Mutations from both wild-type and F305L were selected that ultimately allow the mutants to escape the inhibitor. Resistance phenotypes were confirmed by dose-response experiments and re-introduction of mutations into recently developed protease activity measurement systems [17] as well as alternative methods such as biochemical [45] and cellular assays [27]. We collected a total of 39 unique mutations, of which Y54C, L167F, N203D and D216Y occurred twice independently. F305L was selected for with both GC376 and nirmatrelvir. Six out of 39 occurred in the catalytic site, two near the catalytic site, seven at the dimer interface, three in the autocleavage sites and 21 in the rest of the 3CL^{pro} sequence (which we called “allosteric” mutations). First, we confirmed catalytic site mutants (Y54C, G138S, L167F and Q192R), where the resistance mechanism is likely straightforward: the steric disturbance of nirmatrelvir binding. Then, we tested the near-catalytic site substitution A194S, which is more prevalent than the previous catalytic site mutants in virus isolates. In GISAID, this particular mutation can be found in over 800 sequence depositions in the variants of concern Alpha, Gamma, Delta, Lambda, and Omicron. Changes of the residue A194 in general are frequent with over 3000 entries. Although it is not known

if this mutant was selected for by the use of nirmatrelvir in patients, the fact that it is a resistant mutant and prevalent in virus sequences makes is a variant worth tracking. We further combined L167F and Q192R with the autocleavage site mutation F305L, which further increased the resistance. The combination of Y54C and L167F with the Omicron signature mutation P132H also conferred increased resistance, highlighting the potential relevance of these mutations for the Omicron variant. The substitution F305L was described as resistance mutation in this study. An adjacent mutant, T304I, was found in nirmatrelvir selection experiments with authentic SARS-CoV-2 (9) and the suggested mechanism was autocleavage site optimization. Given F305L is also likely an autocleavage site optimizing mutant, we did not test it further in a biochemical assay, since such assays use mature protease in which autoprocessing does not play a role. Lacking an appropriate method, we therefore did not investigate the mechanism of F305L. Nevertheless, such mutants merit further study in assays that can elucidate the mechanism of action. One technical particularity in the $3CL^{pro}$ -On construct is that the nirmatrelvir IC_{50} values are higher than generally reported in the literature. However, IC_{50} values are generally higher in cell-based assays than in biochemical assays as we described previously [17]. In brief, in the excess of an inhibitor and constant renewal of protease fusion protein, signals are expected to plateau later than in a biochemical assay with a fixed amount of enzyme. Furthermore, the screening method used in this study to assess mutants was originally developed as high-throughput screening tool for $3CL^{pro}$ inhibitors using a FluoroSpot reader that allows fast sampling [17]. We improved the assays sensitivity by changing the read-out method from FluoroSpot to flow cytometry-based sampling. Flow cytometry sampling is more sensitive, but also more time-consuming. Flow cytometry read-outs captured milder degrees of inhibition and resulted in a more gradual signal increase; therefore, this resulted in lower IC_{50} values in $3CL^{pro}$ -On assay (0.91 µM of nirmatrelvir, which is closer to the published range of 74.5 (66.5 to 83.4) nM) [46]. We cross-validated several of our mutants in different assays. We confirmed the resistance data of L167F with a previously published cellular assay [27] and the mutants Y54C, L167F and Q192R with a biochemical assay [47]. We showed also in a biochemical assay that the kinetic metrics of the mutants Y54C, L167F, and Q192R are attenuated to varying degrees. However, the VSV-chimeric viruses containing resistant 3CLpro showed little fitness loss. In three recent preprints, L167F as well as various mutants at Q192 were identified to be resistance mutations in authentic SARS-CoV-2 (9, 10, 12). Lastly, we confirmed the viability and resistance of the single mutant L167F and in combination with F305L in a previously published recombinant SARS-CoV-2 expressing a reporter gene [29][30], finally confirming the validity of our mutation prediction tool based on VSV as well as the resistance mutations identified. Genetically modifying highly pathogenic viruses such

as SARS-CoV-2 can be considered as so-called gain-of-function experiments if the recombinant virus is more apt to cause disease, or if treatment is made more difficult than the wt variant. We therefore applied several safety measures such as using a virus for mutagenesis that was sequence identic to the Wuhan-1 variant. Therefore, compared to currently circulating viruses, it has not undergone extensive evolution and if set free, would unlikely be able to compete with current Omicron variants. Importantly, other previously described antivirals approved for use in humans have been shown to inhibit viral replication of this Wuhan-1 strain. Moreover, current vaccines used in humans to protect against SARS-CoV-2 have been developed based on the sequence of the Spike glycoprotein of this, or a similar, Wuhan-1 strain. Thus, neutralizing antibody responses induced by these vaccines will be able to protect against these recombinant viruses. Secondly, the plaques the mutant recombinant viruses formed in Vero E6 cells were found to be smaller than that of wild-type recombinant virus and replication kinetics also indicate mild attenuation. Third, the mutations introduced into recombinant SARS-CoV-2-mCherry had been found in clinical samples already. In this study, we identified several mutations such as Y54C, G138S, L167F, Q192R, A194S and F305L in the SARS-CoV-2 $3CL^{pro}$ that confer resistance to the $3CL^{pro}$ inhibitors nirmatrelvir and GC376. To understand these mutations in light of the Omicron variant, we combined two of our most intensively studied mutations, Y54C and L167F with the Omicron $3CL^{pro}$ signature mutation P132H. These results showed that the mutations are functional, thereby confirming their potential relevance in this context of the Omicron variant. Complementary structure modelling approaches based on Robetta and the Molecular Operating Environment (MOE) reveal potential effects of the catalytic site mutants Y54C, G138S, L141F, L167F and Q192R. Y54C and G138S seem not to directly affect nirmatrelvir binding, but may lead to a restructuring of the catalytic site, thereby indirectly affecting the binding site. L141F is likely impacting the early sampling of the catalytic pocket by nirmatrelvir. L167F may distort the distal region of the binding pocket. Lastly, Q192R could strengthen a polar interaction with nirmatrelvir, which may alter sequestration of the compound in an unfavorable position concurrent with destabilization of a loop containing important nirmatrelvir-catalytic site interaction partners such as Q189. Our study has limitations. The mutations generated in VSV occurred in an artificial polyprotein, which, like pp1a or pp1ab, comprises precursors for large protein subunits and requires $3CL^{pro}$ dimerization for autocleavage. Nevertheless, the polyprotein structure is different, which could result in mutations not relevant in authentic SARS-CoV-2. Along the same line, in this artificial polyprotein, only autocleavage or cis-cleavage occurs, whereas in SARS-CoV-2, the mature $3CL^{pro}$ additionally cleaves distant or trans-cleavage sites. It has been shown that autocleavage of $3CL^{pro}$ in coronaviruses is a stepwise process with distinct N- and C-terminal

autocleavage binding pocket confirmations [48][49], where the C-terminal autocleavage occurs after N-terminal autocleavage and might resemble a matured structure as in the transcleavage confirmation. Even so, this system could, in theory, disregard trans-cleavage specific mutants, if such exist. Finally, we did not elucidate the exact mode of resistance of the different mutants described in this study. Although we modelled catalytic site mutations and describe a plausible mechanism for autocleavage site mutants, solving crystal structures was beyond the scope of this work and remains for future studies. In conclusion, our findings argue for a highly selective application of protease inhibitors to patients at increased risk of severe disease, as extensive, unselective use is expected to rapidly lead to emergence of drug resistance. Furthermore, the combination of different drugs is a proven strategy to avoid resistance mutations, as has been shown for HIV [3] and HCV [2] therapy. As more compounds became available, combinations including classes of inhibitors targeting distinct viral functions, such as protease and polymerase inhibitors, may be an effective strategy. However, as we observed in this study, $3CL^{pro}$ mutants can react differently to specific compounds. Therefore, even the combination of different protease inhibitors could lower the risk of viral escape.

4. Materials and Methods

4.1. Study Design

The overall rationale of the study was to develop a mutation selection tool based on VSV and to describe mutants as proof-of-concept for that tool. The study was performed on cell lines and in-silico, and no animal husbandry or human participants were involved. Human and monkey cell lines with replicating BSL-1, 2 and − 3 viruses were treated with protease inhibitors to observe resistance phenotypes in appropriate facilities. Viral titers were determined using $TCID_{50}$ and plaque assays. Measurement readouts were fluorescence-based, detected by flow cytometry, ELISpot and multi-well readers. Autofluorescent fibers were excluded automatically and manually from spot counting in the ELISpot readout. Experiments were neither blinded nor randomly distributed to experimenters. We chose sample sizes empirically based on experience from former studies. At least two and up to four biologically independent replicates were performed per condition. Biologically independent meant distinct wells with the same condition, not multiple measurement of the same wells (technical replicates). Resistance phenotypes were reproduced at least twice, usually more often and in different combinations (comparing single mutants to each

other and the wild-type or wild-type to single and double mutants). Representative measurements were chosen to compile graphs and figures.

4.2. Cloning Strategies

The chimeric VSV variant with 3CL^{pro} instead of the intergenic region between G and L was cloned by Gibson assembly (New England Biolabs, NEB) [50]. A VSV-G plasmid [51] was digested with KpnI and HpaI (NEB), removing a C-terminal part of G, the intergenic region and a small N-terminal part of L. Insert fragments were generated as follows. Missing C-terminal G with an additional overhang to the N-terminal cleavage site of 3CL^{pro} was amplified with primers 33n-before-KpnI-for and G-cut1-rev. 3CL^{pro} with its N- and C-terminal cleavage sites and a C-terminal overhang to L was amplified from Wuhan-1 (NCBI Reference Sequence: NC_045512.2) cDNA with primers cut1-for and cut2-L-rev. The N-terminal missing L sequence was amplified with primers cut2- L-for and 33n-after-HpaI-rev. For subsequent Gibson assembly, the fragments were ligated in a fusion PCR using the outer primers 33n-before-KpnI-for and 33n-after-HpaI-rev with all three fragments as templates. The cloning primers for VSV vectors are shown in table S12 and the annotated sequence is shown in data file S3. 3CL^{pro} -Off and -On point mutants were generated by mutagenic Gibson assembly on parental plasmids (GenBank accession codes: 3CL^{pro} -Off: 25684003; 3CL^{pro} -On: 2568399). For 3CL^{pro} -Off mutants, a lentiviral expression plasmid expressing VSV L (identical sequence as blasticidin 3CL^{pro} -Off plasmid without GFP and 3CL^{pro}) was digested with HpaI, which removed the cPPT/CTS and CMV promoter sequences and a small N-terminal part of L. This missing sequence was replaced with the identical sequence from 3CL^{pro} -Off with the addition of the N-terminal 3CL^{pro} sequence up to the respective mutation site with primers blasticidin-for and 3CL^{pro} -*mut-x*-rev, where *mut-x* is the mutation of interest. The C-terminal part of 3CL^{pro} and the small missing fragment of L were generated by PCRs on parental vectors with primers 3CL^{pro} -*mut-x*-for and 33n-after-HpaI-rev. For 3CL^{pro} -On mutants, a lentiviral hygromycin vector (modified from Addgene pLenti CMVie-IRES-BlastR accession: #119863) was digested with NheI and PacI. N-terminal 3CL^{pro} insert fragments with vector overhangs were generated with hygro-P-for and 3CL^{pro} -*mut-x*-rev. C-terminal 3CLpro insert fragments with vector overhangs were generated with 3CL^{pro} -*mut-x*-for and Phygro-rev. Double mutants were cloned by repeating the site directed mutagenesis with a second primer pair in combination with Gibson assembly on an already mutant-bearing plasmid. Cloning primers for 3CL^{pro} -Off and -On mutant variants are shown in table S13.

4.3. Cell Lines

BHK-21 cells (American Type Culture Collection, ATCC) were cultured in Glasgow Minimum Essential Medium (GMEM) (Lonza) supplemented with 10% fetal calf serum (FCS), 5% tryptose phosphate broth, and 100 units/ml penicillin and 0.1 mg/ml streptomycin (P/S) (Gibco). 293 T cells (293tsA1609neo, ATCC), and 293- VSV (293 expressing N, P-GFP and L of VSV) [52] were cultured in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% FCS, P/S, 2% glutamine, 1x sodium pyruvate and 1x non-essential amino acids (Gibco). Vero E6 (ATCC CRL-1586) were cultured in DMEM supplemented with 5% FCS (VWR) and 1% penicillin−streptomycin−glutamine (PSG) solution (Corning). A549-hACE2 (Biomedical Resource Ontology NR53821) were grown in DMEM supplemented with 4 mM L-glutamine, 4500 mg/l glucose, 1 mM sodium pyruvate, 1500 mg/l sodium bicarbonate, 10% FCS, 1x non-essential amino acid solution (Gibco) and 100 µg/ml blasticidin (Gibco).

4.4. Virus Recovery

VSV-G-3CL^{pro} -L was rescued in 293 T cells by CaPO4 transfection of whole-genome VSV plasmids together with T7-polymerase, N-, P-, M-, G- and L expression plasmids as helper plasmids [53]. Briefly, genome and helper plasmids were transfected into 293 T in the presence of 10 µM chloroquine to avoid lysosomal DNA degradation. After 6 to 16 hours, chloroquine was removed and cells were cultured until cytopathic effects occurred. M and G proteins were used as helper plasmids; although these proteins are optional in the recovery of VSV, they were chosen here as a precaution to support the rescue of a potentially attenuated virus variant. After the rescue, viruses were passaged on 293-VSV cells and plaque purified twice on BHK-21 cells. ΔP and ΔL VSV variants expressing dsRed were produced on replication supporting 293-VSV cells. VSV-G-3CL^{pro} -L was fully replication competent and produced on BHK-21 cells.

4.5. Replication Kinetics, $TCID_{50}$ assays, and dose responses

Initial replication kinetics (wild-type versus F305L) were performed as single-step kinetics. 105 BHK-21 cells per well were seeded in 24- well plates one day before infection. Cells were infected in duplicate with an MOI of 5 of VSV 3CL^{pro} wild-type or the F305L variant. One hour after infection, the medium was removed, cells were washed with phosphate-buffered saline (PBS), and fresh medium was added. Supernatant was collected at the indicated time points and stored at −80°C until further analysis. For quantification, $TCID_{50}$ assays were performed as described previously [54].

In short, 100 µl of serial dilutions of virus were added in octuplicates to 103 BHK-21 cells seeded in a 96-well plate. Six days after infection, the $TCID_{50}$ values were read out and titers were calculated according to the Kaerber method [55]. For wild-type versus different mutants replication kinetics, multi-step growth kinetics were performed. 105 BHK-21 cells per well were seeded in 24-well plates one day before infection. Cells were infected in duplicates with an MOI of 0.5 of VSV $3CL^{pro}$ wild-type or mutant variants. For initial dose response experiments, 5 x 104 BHK-21 cells per well were seeded in 48-well plates one day before infection. Cells were infected in duplicates with a MOI of 0.05 of VSV $3CL^{pro}$ wild-type or mutant variants and indicated concentrations of nirmatrelvir were added to the wells. After 48 hours, supernatants were collected and titrated to determine the $TCID_{50}$. For mutant comparing dose response experiments, 5 x 104 BHK21 cells per well were seeded in 48-well plates one day before infection. Cells were infected in duplicates with an MOI of 0.05 of VSV $3CL^{pro}$ wild-type or mutant variants and indicated concentrations of nirmatrelvir added to the wells. To prevent initial escape or further mutation in wild-type or already mutation-bearing viruses ("intra-assay mutants"), respectively, supernatants of all viruses were collected after the first mutant (Q192R, F305L) showed a massive cytopathic effect at 100 µM nirmatrelvir (at about 24 hours after infection). Initial dose responses (wild-type versus F305L) were performed as described above, but the supernatant was collected after 48 hours.

4.6. Viral RNA isolation and $3CL^{pro}$ sequencing

VSV-G-$3CL^{pro}$ -L RNA was isolated with E.Z.N.A. Viral RNA Kit (Omega Bio-Tek Inc.) or NucleoSpin RNA Virus (MachereyNagel GmbH). BHK-21 cells were infected with VSV-G-$3CL^{pro}$ -L wild-type and F305L ($3CL^{pro}$) mutant in 96-well plates. Virus-containing supernatants were collected from individual 96-wells and the RNA was purified from the supernatants according to manufacturers' instructions. Then, cDNA was synthesized from isolated viral RNA by RevertAid RT Reverse Transcription Kit (Thermo Fisher Scientific). $3CL^{pro}$ sequence was amplified by PCR with primers (for: CTCAGGTGTTCGAACATCCTCAC and rev: GAT GTTGGGATGGGATTGGC) and sent for sequencing (MicroSynth AG). Obtained sequences were mapped to the $3CL^{pro}$ -wt (Wuhan1) reference sequence in Geneious Prime 2022.0.2 and examined for mutations.

4.7. Mutation selection assay

104 BHK-21 cells per well in a 96-well plate were seeded one day before infection with wild-type VSV-G-$3CL^{pro}$-L or VSV-G$3CL^{pro}$-L-F305L at an MOI of 0.01 and indicated nirmatrelvir doses. Each virus variant occupied 48 wells of the 96-well plate. Wells that displayed cytopathic effect after two

days (25 out of 48 from parental wild-type and 17 out of 48 from parental F305L) were further passaged with increasing concentrations of nirmatrelvir with each passage (wild-type: 30, 40 and 50 µM; F305L: 50, 75 and 100 µM). Table S1 indicates at which passage a pure mutant virus could be distinguished by Sanger sequencing, such that only one base-pair peak appeared in the chromatogram instead of a mixture with the parental virus. Only pure mutants are displayed in Fig. 2 and table S1.

4.8. Expression and purification of his-tagged $3CL^{pro}$ and point mutations

Plasmids containing cDNA of SARS-CoV-2 main protease $3CL^{pro}$ (pMCSG92 [56]) and mutants thereof were prepared as described in the following. Plasmids were cloned by site-directed mutagenesis with primers of table S13 on pMCSG92. 100 ng of each plasmid was applied to 50 µl of thawed competent BL21(DE3) TUNER E. coli (Merck) on ice in 1.5 ml tubes. Bacterial suspensions containing the plasmids were flicked and incubated for 30 minutes on ice. Subsequently, bacteria were heated to 42°C for 90 seconds in a thermomixer without shaking, and put back on ice for 5 minutes. 400 µl of NZCYM Medium produced in-house (NZ amine (Art.-Nr. CP76.1, Roth) 10.0 g, NaCl 5.0 g, casamino acids (Gibco) 1.0 g, yeast extract (Art. Nr. 2363.2, Roth) 5.0 g, $MgSO_4$ x 7 H_2O 2.0 g, ddH_2O to 1 L, adjusted to pH 7.4) was added to each bacterial suspension, and the bacteria were amplified for 1 hour at 37°C in a bacterial shaker in 1.5 ml tubes. Meanwhile, LB agar plates containing selection antibiotics (ampicillin) were prepared and incubated at room temperature. 200 µl of bacterial culture was crossed out on individual plates and incubated overnight at 37°C. Single colony formation was observed the following day. Individual colonies were picked and placed in 5 ml of NZCYM medium supplemented with selection antibiotics, and amplified overnight at 37°C. The next day, overnight cultures were amplified in 1 L of NZCYM medium supplemented with selection antibiotics to an optical density (OD) of 0.2, after which protein expression was induced by applying 1 mM isopropyl-β-D-thiogalactopyranosid (IPTG). After 5 hours bacteria were harvested by centrifugation and the supernatant discarded. Bacterial pellets were frozen at −20°C for further use. Pellets were suspended in 10 ml of Ni-NTA running buffer (20 mM Tris-HCl, 300 mM NaCl, 10 mM Imidazol, adjusted to pH 7.4) and transferred into 50 ml tubes. Bacteria were lysed using an ultrasonic probe on ice. The homogenates were centrifuged at 10,000 x g for 10 minutes and the supernatant was filtered using 0.45 µm and 0.22 µm syringe filter units. After preparing a Ni-NTA agarose column (Invitrogen, Ni-NTA Agarose R90115) and washing with 30 ml Ni-NTA running buffer, the filtered homogenate was applied to the column, and the flow-through was collected. The column was again washed with 3 x 10 ml of Ni-NTA running buffer. The His-tagged protein

bound to the Ni-NTA resin, and was then eluted with 3 x 1 ml of Ni-NTA elution buffer (20 mM Tris-HCl, 300 mM NaCl, 200 mM imidazol, adjusted to pH 7.4). The obtained protein solutions were dialyzed at 4°C overnight against a storage buffer (Tris-HCl 1 mM, NaCl 4 mM, KCl 2.2 mM, TWEEN-20 0.04 vol-%, DTT 3 mM, glycerol 20.2 vol-%, adjusted to pH 8). Eluted protein samples were further purified using size exclusion chromatography (SEC) with fast protein liquid chromatography (FPLC) (ÄKTA Pure FPLC System, Superdex 200 10/300 GL). At each step of the protocol, samples for SDS-PAGE analysis were obtained, and the successful expression of the $3CL^{pro}$ proteins was monitored by SDS-PAGE and Coomassie R staining. The final degree of protein purity was estimated to be greater than 90% based on Coomassie R staining, similar among the different preparations of $3CL^{pro}$ wild-type and mutant forms.

4.9. Screening assay with Fluorospot read-out

3 x 105 cells were seeded per well in 6-well plates and transfected one day after seeding with $3CL^{pro}$ plasmids using TransIT-PRO (Mirus Bio LLC) and incubated overnight. Then, cells were seeded into a 96-well plate with 2 x 104 cells per well in 50 µl complete growth medium. Directly after seeding, compounds and virus (MOI 0.1) were added in 50 µl complete growth medium to wells. After 48 to 72 hours, supernatants were removed, and fluorescent spots counted in a Fluoro/ImmunoSpot counter (CTL Europe GmbH). Longer incubation times of 72 hours increased the overall signal and were chosen in order to achieve a clear signal of the more resistant double mutants, which as expected have a lower signal output in $3CL^{pro}$ -On assays. The manufacturer-provided software CTL switchboard 2.7.2. was used to scan 90% of each well area concentrically to exclude reflection from the well edges, and counts were normalized to the full area. Automatic fiber exclusion was applied while scanning. The excitation wave length for dsRed was 570 nm, the D_F_R triple band filter was used to collect fluorescence. In addition, manual quality control for residual fibers was performed. To increase comparability between $3CL^{pro}$ -On and -Off signals, we normalized dsRed events with the following strategies. In $3CL^{pro}$ - On, the highest compound concentrations would not reach the same value due to the different response of each mutant. Therefore, we normalized to the highest mean of the experiment, which was the wild-type signal. In $3CL^{pro}$ -Off, untreated wells reached the same signal yield in wild-type and mutants. Therefore, we normalized the signal to each individual highest mean of the construct.

4.10. Screening assay with flow cytometry read-out

3 x 105 cells were seeded per well in 6-well plates and transfected with $3CL^{pro}$ plasmids using TransIT-PRO (Mirus Bio LLC) and incubated overnight. Then, cells were seeded into a 96-well plate with twenty thousand cells per well in 50 µl. Compound and virus (MOI 0.1) were added in 50 µl to reach desired concentrations. After two days, cells were detached with 0.05% Trypsin-EDTA (Gibco) and transferred to a 96-well round-bottom plate (TPP Techno Plastic Products AG) for automatic sampling by flow cytometry using a BD FACSCanto II. Gates were set to distinguish live and dead cells and to exclude doublets. Singlet cells were divided into dsRed positive and negative based on reference to samples, which were infected, but not treated with inhibitor [17]. Samples were analyzed using BD FACSDiva 8.0.1 (BD Biosciences).

4.11. Cross validation with orthologous cellular Src-$3CL^{pro}$-tatLuc assay

3x106 293 T were seeded per well in a 6-well dish. 24 hours later, they were transfected with 2 µg of the wild-type Src-$3CL^{pro}$-TatLuc or mutants thereof with TransIT-LT1 (Mirus, catalog number MIR 2304). Four hours after transfection, cells were washed with PBS, trypsinized, resuspended in medium and counted. 2x105 cell per well were seeded in 50 µl medium in a flat-bottom 96-well plate (Greiner). Inhibitor dilution series were added in 2-fold excess to required concentrations in 50 µl medium. After 44 hours, medium was removed and 50 µl of Bright-Glo reagent (Promega) added to each well. Cells were incubated for five minutes in the dark and then transferred to a white flat 96-well plate (CLS3600, Heilmann et al., Sci. Transl. Med. published First Release 4 October 2022 Page numbers not final at time of First Release 14 of 18 | RESEARCH ARTICLE | FIRST RELEASE Downloaded from https://www.science.org on November 12, 2022 Corning) for measuring luminescence on a Synergy H1 plate reader (Agilent). The percent inhibition was calculated with the following formula.

$$\% \text{ inhibition} = 100 - (100 / \text{relative luminescence})$$

4.12. $3CL^{pro}$ enzymatic activity

Wild type and variant proteases were produced in-house as described in $3CL^{pro}$ purification method section. Solution of wild type 3CLpro and variants at 85 ng / 30 µl were prepared in appropriate buffer (20 mM Tris/HCl pH 8, 150 mM NaCl, 0.1 mg/ml BSA, 1 mM DTT) and these 30 µL were added to each well in a black 96- well plate (BPS Biosciences) to get a final concentration of 50 nM in 50 µl/well. The substrate Ac-Abu-Tle-Leu-Gln↓MCA (Acetyl-L-αaminobutyroyl-L-tert-leucyl-L-leucyl-L-

glutamine α-(4-methylcoumaryl-7-amide)) was purchased from Peptide Inc., resuspended in dimethyl sulfoxide at 5 mM concentration. 20 µl buffer with diluted substrate was then added to the protein solutions at different concentrations. The plate was immediately placed inside in a GloMax Explorer reader (Promega) and fluorescence emission measured with a substrate-appropriate filter. To determine enzymes initial velocities, we plotted RFU (relative fluorescent units) on the y-axis and time (min) on the x-axis. We performed a simple linear regression analysis. Fitting values from zero up to 60 minutes were used in a range that had a linear increase. The resulting slopes represented the initial velocity expressed as RFU/min for each protein variant at each substrate concentration. Slope values were plotted (y-axis) against the substrate concentration (x-axis). Finally, the obtained values were fitted using the "Michaelis-Menten" equation built-into GraphPad Prism 9 to extrapolate the kinetic parameters K_m and V_{max}.

$$Y = \frac{Vmax * X}{Km + X}$$

k_{cat} was calculated dividing V_{max} by [ET], where [ET] is the give enzyme concentration. Wild type and variant $3CL^{pro}$ catalytic efficiencies were determined as k_{cat}/KM.

4.13. Cross validation with biochemical $3CL^{pro}$ inhibition assay

The biochemical assay used to confirm mutations was based on the $3CL^{pro}$ activity assay from BPS Biosciences, catalogue number #78042–2. The $3CL^{pro}$ in the kit was replaced by an in-house produced $3CL^{pro}$ and mutants thereof, as described in $3CL^{pro}$ purification. Solutions of wild-type $3CL^{pro}$ and mutants at 5 ng/µl in 30 µl buffer (composition described above) were prepared according to the kit's manual. Ten µl of five-fold excess to tested nirmatrelvir concentrations were added to the 30 µl of $3CL^{pro}$ solution and incubated for 30 minutes. Then, 10 µl of fluorogenic substrate (DABCYL-KTSAVLQSGFRKME-EDANS) was added (generating a in total a 1:5 dilution of the excess nirmatrelvir and therefore final concentrations) and incubated for 4 hours. Fluorescence was induced with 365 nm UV-light and read at 460 nm in a Glomax Explorer (Promega). Blanks (assay buffer plus substrate) values were subtracted from sample values.

4.14. Replication kinetics with recombinant SARS-CoV-2 expressing mCherry

Monolayers of Vero E6 cells (6-well plate, 106 cells per well, triplicates) were infected with the indicated viruses at MOI 0.01. After viral absorption for 3 hours at 37°C, the supernatant was discarded, the cells were washed three times with PBS, and post-infection media (3 ml/well) was

added. At the indicated time points, the supernatant (300 µl/well) was collected and titrated by plaque assay (29).

4.15. Cross-validation with recombinant SARS-CoV-2 (rWA1) expressing mCherry

A monolayer of A549-hACE2 cells was infected with 300 plaque-forming units (PFU) of indicated viruses in quadruplicates at 37°C. After viral adsorption for 1 hour, the supernatant was discarded and the cells were washed twice with PBS. Then, phenol red-free post-infection medium (DMEM +2% fetal bovine serum +1% penicillin-streptomycin-glutamine (PSG)) containing the indicated concentrations of nirmatrelvir was added to each well. The mCherry intensity was determined at 48 and 72 hours post-infection under a Synergy LX Multimode Reader (Agilent). Wells without drug or virus were used as negative controls or baseline signal. Positive controls were wells with virus, but no drug. Infection percentages of wells with different amounts of inhibitor were calculated by subtracting the negative control (mean of wells without virus or drug) and then dividing by the positive control (mean of wells with virus but without drug). Data were analyzed in GraphPad Prism 9 and IC_{50} values were calculated as the highest dilution of the nirmatrelvir-containing sample that prevents 50% plaque formation in infected cells, determined by a sigmoidal dose-response curve (see statistical analysis section).

4.16. IC_{50} and EC_{50} calculations

In this study, different assay systems were used to generate resistance data, namely VSV-based cellular assays with FluoroSpot and flow cytometry read-outs, an orthologous cell-based assay with a luciferase read-out, as well as a biochemical assay and SARS-CoV-2-mCherry assay with fluorescence read-outs. Although the magnitudes of resistance are different in these assays, the tendencies agree. We expected the dynamic range of the $3CL^{pro}$ cellular assays to be greater than in a biochemical assay, where there is a fixed amount of enzyme. In cells, the continuous renewal of protease-viral fusion proteins in an excess of inhibitor likely led to a later plateauing of the signal. At lower concentrations, compound molecules are depleted and the signal plateaus. In FluoroSpot read-outs, the $3CL^{pro}$ -On assay data were normalized to the highest mean value in an experiment. $3CL^{pro}$ -Off data were normalized to the highest value of each construct in an experiment. In the flow cytometry experiments, $3CL^{pro}$ -On assay data were also normalized to the highest value of each construct in an experiment. For purified wild-type and mutant enzymes, IC_{50} values were determined using the biochemical assay “3CL Protease, Untagged (SARS-CoV-2)” from BPS Biosciences with the assays substrate DABCYL-KTSAVLQSGFRKME-EDANS. IC_{50} and EC_{50} calculations

and statistical analysis for all assays were performed with GraphPad Prism 9 (see statistical analysis section.

4.17. Nanopore sequencing of recombinant SARS-CoV-2 (rWA1) expressing mCherry

To validate the sequence of the recombinant SARS-CoV-2 (rWA1) expressing mCherry, we used the Nanopore sequencing "Midnight protocol", version 6 [57]. Primer pools generating 1200 bp overlapping amplicons were purchased from Integrated DNA Technologies, as referenced in the abovementioned protocol. The sequencing reactions were prepared using the Rapid Barcoding Kit SQK-RBK110.96 (Oxford Nanopore Technologies) and were performed in a sequencer (MinION Mk1B) using a proprietary flowcell (R9.4.1, Oxford Nanopore Technologies). Electrical signals are translated into nucleotide sequences (base-calling). Sequenced reads were sorted into separate files for each sample (demultiplexing). Demultiplexing was done using the super high accuracy model in Guppy 6.1.5. Output sequences generated so called fasta files and sequences below 200 and above 1200 bp were removed. Sequences between 200 and 1200 bp were assembled with the algorithm epi2me-labs/wf-artic v0.3.18 in Nextflow 22.04.4. The SARS-CoV-2 lineage pangolin 4.1.1 was used to map the sequences. A visualization application (Nextclade 2.4.0) was used to check mutations.

4.18. Protein structure preparation for molecular modelling

The three-dimensional structure of the SARS-CoV-2 $3CL^{pro}$ complexed with nirmatrelvir was retrieved from the Protein Data Bank (PDB code: 7RFW, method: X-ray diffraction, resolution: 1.73 Å [31]) and prepared for molecular modeling evaluations exploiting several tools implemented in the Molecular Operating Environment (MOE) 2022.02 suite [16]. Specifically, the "Structure Preparation" tool was used to assign each protein residue with alternative conformations to the one characterized by the highest occupancy value, and the "Protonate 3D" program was exploited to assign each titratable amino-acid to the most appropriate protonation state at a pH of 7.4. Finally, the coordinates of hydrogen atoms were energy minimized using the AMBER10:EHT force field [58] until a gradient of 0.1 kcal mol^{-1} $Å^{-2}$ was reached.

4.19. In silico alanine and resistance mutation scanning

An in silico evaluation of the impact of SARS-CoV-2 $3CL^{pro}$ mutations on both the stability of the protein and the affinity towards nirmatrelvir was conducted using the "Protein Design" module of MOE, using the previously mentioned complex structure. Particularly, the "Alanine Scan" and

"Resistance Scan" tools were used to perform two virtual mutagenesis experiments. First, we applied the "Alanine Scan" interface, in which each of the 612 amino acids composing the dimeric SARS-CoV-2 $3CL^{pro}$ was mutated into an alanine residue, calculating at each given time the energy difference between the mutated protein and the wild-type form concerning both the potential energy of the protein itself (dStability) and the affinity towards nirmatrelvir (dAffinity). Values were then color plotted on the crystal complex of nirmatrelvir and SARS-CoV-2 $3CL^{pro}$ using UCSF Chimera [59]. Then we used the "Resistance Scan" interface to investigate the impact of a selected pool of mutations: Y54C, G138S, L141F, L167F, and Q192R. For both types of calculations, the conformational sampling was carried out through LowModeMD [60], using the AMBER10:- EHT forced field coupled with the Generalized Born implicit solvent model [61]; the dAffinity value was determined through the GBVI/VSA [62] method.

4.20. Statistical analysis

Raw and normalized data are provided in Data file S4. Dose response data points of $3CL^{pro}$ -On, Off and biochemical assays were fitted using a four-parameter logistic regression (sigmoid, 4PL, X is concentration). IC_{50} values were extrapolated as the concentration value at which the signal was 50% between the top and bottom plateaus of each sub-dataset. Dose responses curves of the Src-$3CL^{pro}$ -Tat-Luc-based assay were fitted with the same regression, setting X as 2 for IC_{50} extrapolation.

$$Y = Bottom + \frac{Top-Bottom}{1+(\frac{IC_{50}}{X})HillSlope}$$

Data obtained with flow cytometry were normalized and fitbased using the non-linear regression function "[Agonist] vs. normalized response". The EC_{50} values were extrapolated as the medium value between the top and bottom plateaus of each subdataset.

$$Y = \frac{100*X}{EC_{50}+X}$$

Nirmatrelvir dose response curves of recombinant SARS-CoV-2 expressing mCherry were normalized and fitted using the nonlinear regression function "log(inhibitor) vs. normalized response - variable slope".

$$Y = \frac{100}{1+10^{\log(IC_{50}-X)*HillSlope}}$$

Kinetic parameters and catalytic activity of wild-type and mutant $3CL^{pro}$ enzymes were calculated as described in the corresponding method section. 95% confidence intervals were generated by the

described fittings and are provided in supplementary tables together with IC_{50} and EC_{50} values. All statistical analyses were performed with GraphPad Prism 9.

References

[1] P. Zhou *et al.*, “A pneumonia outbreak associated with a new coronavirus of probable bat origin,” *Nature*, vol. 579, no. 7798, pp. 270–273, Mar. 2020, doi: 10.1038/s41586-020-2012-7.

[2] R. Voelker, “Combination Drug for HCV Infection,” *JAMA*, vol. 318, no. 9, p. 790, Sep. 2017, doi: 10.1001/jama.2017.10930.

[3] M. Harrington and C. C. Carpenter, “Hit HIV-1 hard, but only when necessary,” *Lancet*, vol. 355, no. 9221, pp. 2147–2152, Jun. 2000, doi: 10.1016/S0140-6736(00)02388-6.

[4] K. Fan *et al.*, “Biosynthesis, Purification, and Substrate Specificity of Severe Acute Respiratory Syndrome Coronavirus 3C-like Proteinase,” *J. Biol. Chem.*, vol. 279, no. 3, pp. 1637–1642, Jan. 2004, doi: 10.1074/jbc.M310875200.

[5] B. H. Harcourt *et al.*, “Identification of Severe Acute Respiratory Syndrome Coronavirus Replicase Products and Characterization of Papain-Like Protease Activity,” *J. Virol.*, vol. 78, no. 24, pp. 13600–13612, Dec. 2004, doi: 10.1128/JVI.78.24.13600-13612.2004.

[6] Y. N. Lamb, “Nirmatrelvir Plus Ritonavir: First Approval,” *Drugs*, vol. 82, no. 5, pp. 585–591, Apr. 2022, doi: 10.1007/s40265-022-01692-5.

[7] S. Brogi *et al.*, “Covalent Reversible Inhibitors of Cysteine Proteases Containing the Nitrile Warhead: Recent Advancement in the Field of Viral and Parasitic Diseases,” *Molecules*, vol. 27, no. 8, p. 2561, Apr. 2022, doi: 10.3390/molecules27082561.

[8] S. E. Greasley *et al.*, “Structural basis for the in vitro efficacy of nirmatrelvir against SARS-CoV-2 variants,” *J. Biol. Chem.*, vol. 298, no. 6, p. 101972, Jun. 2022, doi: 10.1016/j.jbc.2022.101972.

[9] M. Sho Iketani, Hiroshi Mohri, Bruce Culbertson, Seo Jung Hong, Yinkai Duan, M. K. A. I. Luck1, A.-C. U. , Yicheng Guo, Zizhang Sheng, H. Y. Stephen P. Goff, Yosef Sabo, A. Chavez, and D. D. Ho1, “Multiple pathways for SARS-CoV-2 resistance to nirmatrelvir,” doi: https://doi.org/10.1101/2022.08.07.499047.

[10] K. V. Dirk Jochmans, Cheng Liu, Kim Donckers, Antitsa Stoycheva, Sandro Boland, Sarah K Stevens, Chloe De Vita, Bert Vanmechelen, Piet Maes, Bettina Trüeb, Nadine Ebert, Volker Thiel, Steven De Jonghe, Laura Vangeel, Dorothée Bardiot, Adreas Jekle, Lawrence M B, “The substitutions L50F, E166A and L167F in SARS-CoV-2 3CLpro are selected by a protease inhibitor in vitro and confer resistance to nirmatrelvir,” doi: https://doi.org/10.1101/2022.06.07.495116.

[11] R. S. H. Seyed Arad Moghadasi, Emmanuel Heilmann, Sofia N. Moraes, Fiona L. Kearns, Dorothee von Laer, Rommie E. Amaro, “Transmissible SARS-CoV-2 variants with resistance to clinical protease inhibitors,” doi: https://doi.org/10.1101/2022.08.07.503099.

[12] J. W. Yanmei Hu, Eric M. Lewandowski, Haozhou Tan, Ryan T. Morgan, Xiujun Zhang, Lian M. C. Jacobs, Shane G. Butler, Maura V. Mongora, John Choy, Yu Chen, “Naturally occurring mutations of SARS-CoV-2 main protease confer drug resistance to nirmatrelvir,” doi: https://doi.org/10.1101/2022.06.28.497978.

[13] A. K. Padhi and T. Tripathi, “Hotspot residues and resistance mutations in the nirmatrelvir-binding site of SARS-CoV-2 main protease: Design, identification, and correlation with globally circulating viral genomes,” *Biochem. Biophys. Res. Commun.*, vol. 629, pp. 54–60, Nov. 2022, doi: 10.1016/j.bbrc.2022.09.010.

[14] A. M. Kaufer, T. Theis, K. A. Lau, J. L. Gray, and W. D. Rawlinson, “Laboratory biosafety measures involving SARS-CoV-2 and the classification as a Risk Group 3 biological agent,” *Pathology*, vol. 52, no. 7, pp. 790–795, Dec. 2020, doi: 10.1016/j.pathol.2020.09.006.

[15] M. Baek *et al.*, “Accurate prediction of protein structures and interactions using a three-track neural network,” *Science (80-.).*, vol. 373, no. 6557, pp. 871–876, Aug. 2021, doi: 10.1126/science.abj8754.

[16] Chemical Computing Group ULC, “Molecular Operating Environment (MOE).” .

[17] E. Heilmann *et al.*, “A VSV-based assay quantifies coronavirus Mpro/3CLpro/Nsp5 main protease activity and chemical inhibition,” *Commun. Biol.*, vol. 5, no. 1, p. 391, Dec. 2022, doi: 10.1038/s42003-022-03277-0.

[18] M. Jaskolski *et al.*, “Crystallographic models of SARS-CoV-2 3CL pro : in-depth assessment of structure quality and validation,” *IUCrJ*, vol. 8, no. 2, pp. 238–256, Mar. 2021, doi: 10.1107/S2052252521001159.

[19] L. Silvestrini *et al.*, “The dimer-monomer equilibrium of SARS-CoV-2 main protease is affected by small molecule inhibitors,” *Sci. Rep.*, vol. 11, no. 1, p. 9283, Dec. 2021, doi: 10.1038/s41598-021-88630-9.

[20] J. W. Drake and J. J. Holland, “Mutation rates among RNA viruses,” *Proc. Natl. Acad. Sci.*, vol. 96, no. 24, pp. 13910–13913, Nov. 1999, doi: 10.1073/pnas.96.24.13910.

[21] D. A. Steinhauer, E. Domingo, and J. J. Holland, “Lack of evidence for proofreading mechanisms associated with an RNA virus polymerase,” *Gene*, vol. 122, no. 2, pp. 281–288, Dec. 1992, doi: 10.1016/0378-1119(92)90216-C.

[22] Y. Shu and J. McCauley, “GISAID: Global initiative on sharing all influenza data – from vision to reality,” *Eurosurveillance*, vol. 22, no. 13, Mar. 2017, doi: 10.2807/1560-7917.ES.2017.22.13.30494.

[23] S. Elbe and G. Buckland-Merrett, “Data, disease and diplomacy: GISAID’s innovative contribution to global health,” *Glob. Challenges*, vol. 1, no. 1, pp. 33–46, Jan. 2017, doi: 10.1002/gch2.1018.

[24] S. Khare *et al.*, “GISAID’s Role in Pandemic Response,” *China CDC Wkly.*, vol. 3, no. 49, pp. 1049–1051, 2021, doi: 10.46234/ccdcw2021.255.

[25] E. L. Hatcher *et al.*, “Virus Variation Resource – improved response to emergent viral outbreaks,” *Nucleic Acids Res.*, vol. 45, no. D1, pp. D482–D490, Jan. 2017, doi: 10.1093/nar/gkw1065.

[26] “U.S. Food and Drug Administration, Fact sheet for healthcare providers Emergency Use Authorization (EUA) for Paxlovid (2021).” .

[27] R. S. Moghadasi, Seyed Arad; Esler, Morgan A; Otsuka, Yuka; Becker, Jordan T; Moraes, Sofia N; Anderson, Constance B; Chamakuri, Srinivas; Belica, Christopher; Wick, Chloe; Harki, Daniel A; Young, Damian W; Scampavia, Louis; Spicer, Timothy P; Shi, Ke; Aihara, , “Gain-of-Signal Assays for Probing Inhibition of SARS-CoV-2 Mpro/3CLpro in Living Cells,” *MBio*, vol. 13, 2022, doi: mBio ; 13(3): e0078422, 2022 06 28.

[28] H. Mukae *et al.*, “A Randomized Phase 2/3 Study of Ensitrelvir, a Novel Oral SARS-CoV-2 3C-Like Protease Inhibitor, in Japanese Patients with Mild-to-Moderate COVID-19 or Asymptomatic SARS-CoV-2 Infection: Results of the Phase 2a Part,” *Antimicrob. Agents Chemother.*, vol. 66, no. 10, Oct. 2022, doi: 10.1128/aac.00697-22.

[29] C. Ye *et al.*, “Analysis of SARS-CoV-2 infection dynamic in vivo using reporter-expressing viruses,” *Proc. Natl. Acad. Sci.*, vol. 118, no. 41, Oct. 2021, doi: 10.1073/pnas.2111593118.

[30] C. Ye *et al.*, “Rescue of SARS-CoV-2 from a Single Bacterial Artificial Chromosome,” *MBio*, vol. 11, no. 5, Oct. 2020, doi: 10.1128/mBio.02168-20.

[31] D. R. Owen *et al.*, “An oral SARS-CoV-2 M pro inhibitor clinical candidate for the treatment of COVID-19,” *Science (80-.).*, vol. 374, no. 6575, pp. 1586–1593, Dec. 2021, doi: 10.1126/science.abl4784.

[32] Y. Zhao *et al.*, “Crystal structure of SARS-CoV-2 main protease in complex with protease inhibitor PF-

07321332," *Protein Cell*, vol. 13, no. 9, pp. 689–693, Sep. 2022, doi: 10.1007/s13238-021-00883-2.

[33] K. L. Morrison and G. A. Weiss, "Combinatorial alanine-scanning," *Curr. Opin. Chem. Biol.*, vol. 5, no. 3, pp. 302–307, Jun. 2001, doi: 10.1016/S1367-5931(00)00206-4.

[34] J. Whittaker, A. V. Groth, D. C. Mynarcik, L. Pluzek, V. L. Gadsbøll, and L. J. Whittaker, "Alanine Scanning Mutagenesis of a Type 1 Insulin-like Growth Factor Receptor Ligand Binding Site," *J. Biol. Chem.*, vol. 276, no. 47, pp. 43980–43986, Nov. 2001, doi: 10.1074/jbc.M102863200.

[35] X. Liu, L. Peng, Y. Zhou, Y. Zhang, and J. Z. H. Zhang, "Computational Alanine Scanning with Interaction Entropy for Protein–Ligand Binding Free Energies," *J. Chem. Theory Comput.*, vol. 14, no. 3, pp. 1772–1780, Mar. 2018, doi: 10.1021/acs.jctc.7b01295.

[36] P. Anand, D. Nagarajan, S. Mukherjee, and N. Chandra, "ABS–Scan: In silico alanine scanning mutagenesis for binding site residues in protein–ligand complex," *F1000Research*, vol. 3, p. 214, Dec. 2014, doi: 10.12688/f1000research.5165.2.

[37] M. Pavan, G. Bolcato, D. Bassani, M. Sturlese, and S. Moro, "Supervised Molecular Dynamics (SuMD) Insights into the mechanism of action of SARS-CoV-2 main protease inhibitor PF-07321332," *J. Enzyme Inhib. Med. Chem.*, vol. 36, no. 1, pp. 1646–1650, Jan. 2021, doi: 10.1080/14756366.2021.1954919.

[38] D. E. Kim, D. Chivian, and D. Baker, "Protein structure prediction and analysis using the Robetta server," *Nucleic Acids Res.*, vol. 32, no. Web Server, pp. W526–W531, Jul. 2004, doi: 10.1093/nar/gkh468.

[39] Z. Liu *et al.*, "Identification of SARS-CoV-2 spike mutations that attenuate monoclonal and serum antibody neutralization," *Cell Host Microbe*, vol. 29, no. 3, pp. 477-488.e4, Mar. 2021, doi: 10.1016/j.chom.2021.01.014.

[40] F. Schmidt *et al.*, "High genetic barrier to SARS-CoV-2 polyclonal neutralizing antibody escape," *Nature*, vol. 600, no. 7889, pp. 512–516, Dec. 2021, doi: 10.1038/s41586-021-04005-0.

[41] Y. Yahalom-Ronen *et al.*, "Neutralization of SARS-CoV-2 Variants by rVSV-ΔG-Spike-Elicited Human Sera," *Vaccines*, vol. 10, no. 2, p. 291, Feb. 2022, doi: 10.3390/vaccines10020291.

[42] S. Ullrich and C. Nitsche, "The SARS-CoV-2 main protease as drug target," *Bioorg. Med. Chem. Lett.*, vol. 30, no. 17, p. 127377, Sep. 2020, doi: 10.1016/j.bmcl.2020.127377.

[43] S. Ullrich, K. B. Ekanayake, G. Otting, and C. Nitsche, "Main protease mutants of SARS-CoV-2 variants remain susceptible to nirmatrelvir," *Bioorg. Med. Chem. Lett.*, vol. 62, p. 128629, Apr. 2022, doi: 10.1016/j.bmcl.2022.128629.

[44] W. Rut *et al.*, "SARS-CoV-2 Mpro inhibitors and activity-based probes for patient-sample imaging," *Nat. Chem. Biol.*, vol. 17, no. 2, pp. 222–228, Feb. 2021, doi: 10.1038/s41589-020-00689-z.

[45] K. M. Manandhar, Anjela; Srinivasulu, Vunnam; Hamad, Mohamad; Tarazi, Hamadeh; Omar, Hany; Colussi, Dennis J; Gordon, John; Childers, Wayne; Klein, Michael L; Al-Tel, Taleb H; Abou-Gharbia, Magid; Elokely, "Discovery of Novel Small-Molecule Inhibitors of SARS-CoV-2 Main Protease as Potential Leads for COVID-19 Treatment.," *J Chem Inf Model*, vol. 61, no. 9, pp. 4745–4757, 2021.

[46] D. R. Owen *et al.*, "An oral SARS-CoV-2 M pro inhibitor clinical candidate for the treatment of COVID-19.," *Science (80-.).*, vol. 374, no. 6575, pp. 1586–1593, Dec. 2021, doi: 10.1126/science.abl4784.

[47] A. Manandhar *et al.*, "Targeting SARS-CoV-2 M3CLpro by HCV NS3/4a Inhibitors: In Silico Modeling and In Vitro Screening," *J. Chem. Inf. Model.*, vol. 61, no. 2, pp. 1020–1032, Feb. 2021, doi: 10.1021/acs.jcim.0c01457.

[48] M.-F. Hsu *et al.*, "Mechanism of the Maturation Process of SARS-CoV 3CL Protease," *J. Biol. Chem.*, vol. 280, no. 35, pp. 31257–31266, Sep. 2005, doi: 10.1074/jbc.M502577200.

[49] T. Muramatsu *et al.*, “SARS-CoV 3CL protease cleaves its C-terminal autoprocessing site by novel subsite cooperativity,” *Proc. Natl. Acad. Sci.*, vol. 113, no. 46, pp. 12997–13002, Nov. 2016, doi: 10.1073/pnas.1601327113.

[50] D. G. Gibson, L. Young, R.-Y. Chuang, J. C. Venter, C. A. Hutchison, and H. O. Smith, “Enzymatic assembly of DNA molecules up to several hundred kilobases,” *Nat. Methods*, vol. 6, no. 5, pp. 343–345, May 2009, doi: 10.1038/nmeth.1318.

[51] M. J. Schnell, L. Buonocore, M. A. Whitt, and J. K. Rose, “The minimal conserved transcription stop-start signal promotes stable expression of a foreign gene in vesicular stomatitis virus,” *J. Virol.*, vol. 70, no. 4, pp. 2318–2323, Apr. 1996, doi: 10.1128/jvi.70.4.2318-2323.1996.

[52] D. Panda, P. X. Dinh, L. K. Beura, and A. K. Pattnaik, “Induction of Interferon and Interferon Signaling Pathways by Replication of Defective Interfering Particle RNA in Cells Constitutively Expressing Vesicular Stomatitis Virus Replication Proteins,” *J. Virol.*, vol. 84, no. 9, pp. 4826–4831, May 2010, doi: 10.1128/JVI.02701-09.

[53] S. E. Witko *et al.*, “An efficient helper-virus-free method for rescue of recombinant paramyxoviruses and rhadoviruses from a cell line suitable for vaccine development,” *J. Virol. Methods*, vol. 135, no. 1, pp. 91–101, Jul. 2006, doi: 10.1016/j.jviromet.2006.02.006.

[54] C. Dold *et al.*, “Application of interferon modulators to overcome partial resistance of human ovarian cancers to VSV-GP oncolytic viral therapy,” *Mol. Ther. - Oncolytics*, vol. 3, p. 16021, 2016, doi: 10.1038/mto.2016.21.

[55] G. Kärber, “Beitrag zur kollektiven Behandlung pharmakologischer Reihenversuche,” *Naunyn. Schmiedebergs. Arch. Exp. Pathol. Pharmakol.*

[56] L. Zhang *et al.*, “Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors,” *Science (80-.).*, vol. 368, no. 6489, pp. 409–412, Apr. 2020, doi: 10.1126/science.abb3405.

[57] N. E. Freed, M. Vlková, M. B. Faisal, and O. K. Silander, “Rapid and inexpensive whole-genome sequencing of SARS-CoV-2 using 1200 bp tiled amplicons and Oxford Nanopore Rapid Barcoding,” *Biol. Methods Protoc.*, vol. 5, no. 1, Jan. 2020, doi: 10.1093/biomethods/bpaa014.

[58] D.A. Case et al., “Amber 10, University of California, San Francisco.” .

[59] E. F. Pettersen *et al.*, “UCSF Chimera?A visualization system for exploratory research and analysis,” *J. Comput. Chem.*, vol. 25, no. 13, pp. 1605–1612, Oct. 2004, doi: 10.1002/jcc.20084.

[60] P. Labute, “LowModeMD—Implicit Low-Mode Velocity Filtering Applied to Conformational Search of Macrocycles and Protein Loops,” *J. Chem. Inf. Model.*, vol. 50, no. 5, pp. 792–800, May 2010, doi: 10.1021/ci900508k.

[61] A. Onufriev, D. A. Case, and D. Bashford, “Effective Born radii in the generalized Born approximation: The importance of being perfect,” *J. Comput. Chem.*, vol. 23, no. 14, pp. 1297–1304, Nov. 2002, doi: 10.1002/jcc.10126.

[62] P. Labute, “The generalized Born/volume integral implicit solvent model: Estimation of the free energy of hydration using London dispersion instead of atomic surface area,” *J. Comput. Chem.*, vol. 29, no. 10, pp. 1693–1698, Jul. 2008, doi: 10.1002/jcc.20933.

Qualitative Estimation of Protein-Ligand Complex Stability through Thermal Titration Molecular Dynamics (TTMD) Simulations

Matteo Pavan, Silvia Menin, **Davide Bassani**, Mattia Sturlese and Stefano Moro

Pavan, M., Menin, S., Bassani, D., Sturlese, M., & Moro, S. (2022). Qualitative Estimation of Protein–Ligand Complex Stability through Thermal Titration Molecular Dynamics Simulations. *Journal of Chemical Information and Modeling*. https://doi.org/10.1021/ACS.JCIM.2C00995

Abstract

The prediction of ligand efficacy has long been linked to thermodynamic properties such as the equilibrium dissociation constant, which considers both the association and the dissociation rate of a defined protein-ligand complex. In the last fifteen years, there has been a paradigm shift, with an increased interest in the determination of kinetic properties such as the drug-target residence time, since they better correlate with ligand efficacy compared to other parameters. In this article, we present Thermal Titration Molecular Dynamics (TTMD), an alternative computational method that combines a series of molecular dynamics simulations performed at progressively increasing temperatures with a scoring function based on protein-ligand interaction fingerprints for the qualitative estimation of protein-ligand binding stability. The protocol has been applied to four different pharmaceutically relevant test cases, including protein kinase CK1δ, protein kinase CK2, pyruvate dehydrogenase kinase 2, and SARS-CoV-2 main protease, on a variety of ligands with different sizes, structures, and experimentally determined affinity values. In all four cases, TTMD was successfully able to distinguish between high-affinity compounds (low nanomolar range) and low-affinity ones (micromolar), proving to be a useful screening tool for the prioritization of compounds in a drug discovery campaign.

1. Introduction

At the beginning of the 20th century, Paul Ehrlich's famous quote "*Corpora non agunt nisi fixata*" marked a pivotal moment in the history of modern pharmacology, molecular medicine, and drug development[1]. His statement, combined with independent observations about "receptive substances" by John Newport Langley[2], defined the birth of the receptor theory of drug action, which postulates that a drug can only work as long as it is bound to its target receptor[3][4].

Although the basic ideas of this cornerstone principle were formulated more than one hundred years ago[5], it was only in the 1970s that molecular receptors could be successfully isolated and purified[6][7]. This allowed for the development of different biochemical and cellular assays for the direct determination of the extent to which a drug is bound to its receptor under thermodynamic equilibrium conditions, i.e., the binding affinity[8][9]. Traditionally, this parameter is quantified either through the equilibrium dissociation constant (K_d) or through other proxy metrics such as the drug concentration responsible for the half-maximal inhibition/effect (IC_{50}/EC_{50}), and the inhibition constant (K_i)[10].

Although in principle, these measurements are all adequate predictors for *in vivo* efficacy, i.e., the capability of the drug to induce the desired response, they are all related to *in vitro* assays portrayed under closed system conditions[11]. Since in an open, *in vivo*, system the drug concentration is not a fixed variable and indeed varies over time because of various physiological processes, several authors thus suggested that the observables related to drug-receptor binding kinetics, such as the association (k_{on}) and dissociation (k_{off}) constant, could be better descriptors for drug efficacy[12][13][14][15]. Accordingly, while the binding affinity only depends on the free energy difference between the bound and unbound state, which can be directly correlated to the K_d, association and dissociation rates depend on the energy barriers that separate those states[10].

Thermodynamics and kinetics of bindings are interlinked by the equation $K_d = \frac{k_{off}}{k_{on}}$[10]. While, in theory, both kinetic constants should equally contribute to the determination of the K_d, physicochemical and pharmacological limitations on the k_{on} value[16] render the in vivo duration of a receptor-ligand complex almost entirely dependent on the k_{off} value[11]. Based on this observation, Copeland et al. first suggested that the key determinant of in vivo pharmacological activity and duration is not the binding affinity but, instead, the lifetime of the receptor-ligand complex, defined

as the residence time[12]. Furthermore, Copeland et al. proposed a mathematical formulation for the quantification of this parameter, defining it as the reciprocal of the k_{off} ($\tau=1/k_{off}$)[12].

From an experimental perspective, a plethora of methods for the determination of binding kinetics is available[17][18][19][20]. Each of them relies on monitoring the time-dependent evolution of a signal in response to the binding event[21]. The first strategy revolves around the radio-[22] and spectroscopic[20][23] labeling of ligands and includes techniques such as fluorescent resonance energy transfer (FRET)[24] and bioluminescence resonance energy transfer (BRET)[25]. An alternative approach revolves around the exploitation of label-free approaches such as surface plasmon resonance (SPR)[26][27], nuclear magnetic resonance (NMR)[28], surface acoustic wave method[29], and various declinations of isothermal titration calorimetry (ITC)[30][31]. Finally, another possible method is based on following enzymatic reactions, usually through the monitoring of spectroscopic parameters[32].

Alongside the aforementioned experimental protocols, various computational approaches exist than can flank and expand on the information that they provide by showcasing mechanistic information about the underlying process at an atomic level of detail[21][33][34]. Particularly, Molecular Dynamics (MD) simulations have been exploited to estimate thermodynamic properties such as the binding affinity for protein-ligand complexes, and due to the growing interest in the study of kinetics for drug discovery, they have recently been applied also to the estimation of kinetic properties[35]. Although it would theoretically be possible to exploit unbiased MD simulations for the determination of kinetics observables, biologically relevant events such as drug-target unbinding occur at much longer timescales than that of typical MD simulations, heavily restricting its limitations in terms of computational resources' availability[36]. and neglecting any real-world application of the technique[37]. For this reason, several different methods have been developed throughout the years that implement smart sampling strategies to reduce the required computational effort, such as various instances of metadynamics[38][39][40][41], which are based on repeatedly 'filling' the potential energy of the system by a sum of Gaussians centered along the trajectory followed by an appropriately chosen ensemble of collective variables (CVs)[42], scaled Molecular Dynamics (sMD), which relies on smoothing the potential energy surface (PES) by applying an appropriate scaling factor[43][44][45], and τ-Random Acceleration Molecular Dynamics (τ-RAMD), in which a small randomly oriented force vector is applied to the ligand[46][47][48].

In the present study, we present the first application of Thermal Titration Molecular Dynamics (TTMD), an alternative MD-based approach for the qualitative estimation of protein-ligand complex stability. The method relies on evaluating the conservation of the native binding mode for a ligand of interest throughout a series of molecular dynamics trajectories performed at progressively increasing temperature values. For validation purposes, the protocol has been applied to four different biomolecular targets of pharmaceutical interest: casein kinase 1δ (CK1δ), casein kinase 2 (CK2), pyruvate dehydrogenase kinase 2 (PDK2), and SARS-CoV-2 main protease (M^{pro}).

2. Materials and Methods

2.1. Hardware Overview

Each general molecular modeling operation, such as the preparation of protein-ligand complex structures, the setup for molecular dynamics simulations, and trajectory analyses were conducted on a 20 CPU Linux workstation equipped with an Intel Core i9-9820X 3.3 GHz processor. All molecular dynamics simulations were carried out on a GPU cluster composed of 20 NVIDIA drivers ranging from GTX980 to RTX2080Ti.

2.2. Structures Preparation

The three-dimensional coordinates of the protein-ligand complexes used in this study were retrieved from the Protein Data Bank (PDB)[49] and processed before molecular dynamics simulations through several tools provided by the Molecular Operating Environment (MOE) 2019.01 suite[50]. Four different macromolecular targets were considered in this work: casein kinase 1δ (CK1δ), casein kinase 2 (CK2), pyruvate dehydrogenase kinase 2 (PDK2), and SARS-CoV-2 main protease (M^{pro}). For each macromolecular target, the considered structures are reported in Table 1.

CK1δ	3UZP [51]	4TN6 [52]	5IH5 [53]	5IH6 [53]	5MQV [54]
CK2	2ZJW [55]	3H30 [56]	3PE1 [57]	3PE2 [57]	6HOU [58]
PDK2	4MP2 [59]	4V25 [60]	5J71 [61]	5M4M [62]	7EA0 [63]
M^{pro}	6M2N [64]	7LTJ [65]	7M8P [66]	7M91 [66]	7N44 [67]

***Table 1**: List of the protein-ligand complex structures used in this work. Complexes are grouped by macromolecule target.*

Each protein-ligand system was simulated in the monomeric form, except for SARS-CoV-2 M^{pro}, which was simulated in the dimeric form by applying a symmetric crystallographic transformation to each asymmetric unit. Firstly, all structures were pre-processed using the "Structure Preparation" tool, assigning alternates to the highest occupancy conformation, rebuilding missing loops through homology modeling, and correcting inconsistencies between the primary sequence and the tertiary structure. Secondly, the "Protonate3D" tool was exploited to add missing hydrogens to the system and to determine the most probable protonation state of titratable residues at pH = 7.4. Finally, every non-protein and non-ligand atom of the system was removed before saving the structure for further calculations, except for water molecules within 4.5 Å of the ligand that were not removed and were indeed considered in the simulations. Concerning the protonation state of the ligand, the most abundant protomer at pH 7.4 according to the "Protomers" tool was considered in the calculations, besides CK2 complex 2ZJW where two different protonation states were considered. Particularly, in the case of 2ZJW, the predominant form at pH 7.4 should be the neutral, non-charged one. However, in the context of the binding pocket, the interaction network of the hydroxyl in position 3 (the one facing Lys68 and the conserved water molecule W1) suggests the prevalence of a monocharged, ionized form. Since experimental data published in the literature does not clarify the correct protonation state for ellagic acid in the context of CK2 recognition[55][68][69], we opted to consider both hypotheses equally relevant (50/50).

2.3. System Setup for Molecular Dynamics Simulations and Equilibration Protocol

Each protein-ligand complex prepared as described before was further processed through various tools from Visual Molecular Dynamics (VMD) 1.9.2[70] and the Ambertools14[71] suite. Protein atoms were parametrized through the ff14SB[72] force field, while the General Amber Force Field (GAFF)[73] was utilized to parametrize the ligands. Partial charges were attributed to the ligand through the AM1-BCC method[74]. Each investigated system was solvated in a cubic box with a padding of 15 Å utilizing the TIP3P[75] model for water molecules. The proper number of sodium and chloride ions were added to neutralize the system and reach a salt concentration of 0.154 M. Before undergoing Molecular Dynamics (MD) simulations, each system was energy minimized for a total of 500 steps with the conjugate-gradient algorithm to remove clashes and bad contacts.

Afterward, each minimized system was subjected to a two-step equilibration protocol. During the first stage, a 0.1 ns simulation in the canonical ensemble (NVT) was performed, with harmonic positional restraints (5 Kcal $mol^{-1}Å^{-2}$ force constant) applied on both protein and ligand atoms. The second stage, instead, consisted of a 0.5 ns simulation carried out in the isothermal-isobaric ensemble (NPT), applying the same restraints only to the ligand position and the protein backbone.

Each MD simulation presented in this work, both in the equilibration and the production stage, was performed using an integration timestep of 2 fs, keeping the temperature at a constant value of 310 K through a Langevin thermostat[76], constraining the length of bonds involving hydrogen bonds through M-SHAKE algorithm[77], exploiting the particle-mesh Ewald (PME)[78] method to compute electrostatic interactions using cubic spline interpolation and a 1 Å grid spacing, and setting a 9.0 Å cutoff for the calculation of Lennard-Jones interactions. Simulations in the NPT ensemble were carried out keeping the pressure at a constant 1 atm value making use of a Monte Carlo barostat[79]. All MD simulations were run through the ACEMD 3[80] engine, which is based upon the open-source library for molecular simulations OpenMM 7[81].

2.4. Thermal Titration Molecular Dynamics (TTMD) simulations

Thermal Titration Molecular Dynamics (TTMD) is an alternative enhanced sampling molecular dynamics approach for the qualitative estimation of protein-ligand complex stability. The method relies on evaluating the conservation of the native binding mode for a ligand of interest throughout a series of molecular dynamics trajectories performed at progressively increasing temperature values. The protocol described herein is implemented as a Python 3.10 code, which relies on the Numpy, MDAnalysis[82][83], Open Drug Discovery Toolkit[84], and Scikit-learn libraries.

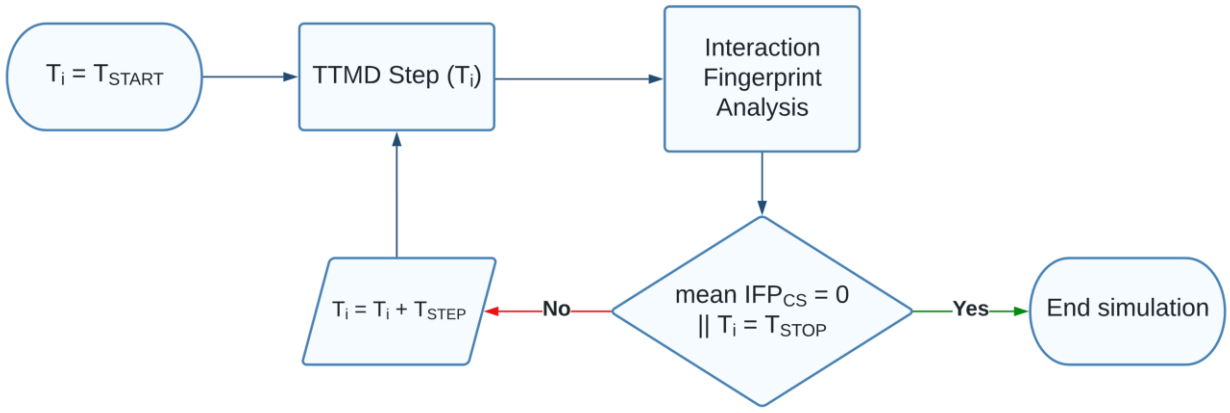


***Figure 1**: Computational workflow for a Thermal Titration Molecular Dynamics (TTMD) simulation.*

The workflow for a TTMD simulation is reported in Figure 1 and detailed hereafter. In detail, the task is accomplished through a series of short, unbiased, MD simulations performed at different, progressively increasing, temperatures in the NVT ensemble with the ACEMD3 engine. For each TTMD run, the duration of each simulation window (defined as "TTMD-step") is fixed and user-defined (10ns, in this case). The starting and final temperature values as well as the temperature increase between each "TTMD-step" are also defined by the user based on prior knowledge of the target, particularly regarding the conservation of the protein fold at higher simulation temperatures (which, in the context of this article, is carried out by monitoring the protein backbone RMSD throughout the simulation). In this work, the starting temperature was set to 300 K, the ending temperature was set to 450 K, while the temperature increase between each "TTMD-step" was set to 10 K.

The progress of the simulation is monitored through a scoring function based on protein-ligand interaction fingerprints. The scoring function, defined as IFP_{CS} and originally described in previous scientific work from our laboratory[85], exploits the Open Drug Discovery Toolkit Python library to calculate protein-ligand interaction fingerprints for each frame of the TTMD trajectory and compare it through the cosine similarity metric as implemented in the Scikit-learn Python module to a reference fingerprint, based on the last trajectory frame extracted from the second and last equilibration stage.

Specifically, each protein-ligand interaction fingerprint is an integer vector composed of rx8 elements, where r is the number of protein residues. Each protein residue is encoded into eight bits of information, one for each type of intermolecular interaction considered (hydrophobic contacts, aromatic face to face, aromatic edge to face, hydrogen bond with protein acting as donor, hydrogen bond with protein acting as acceptor, salt bridge with protein acting as the positively charged member, salt bridge with protein acting as the positively negative member, and ionic bond with a metal ion respectively). The mathematic formulation of the IFP_{CS} scoring function is reported in Equation 1:

$$IFP_{CS} = \frac{A \cdot B}{\|A\|\|B\|} * -1 \quad (1)$$

The IFP_{CS} value ranges from -1, indicating a total superposition between the reference and the query fingerprint, to 0, which indicates that every interaction determinant of the reference fingerprint is lost in the query.

At the end of each "TTMD-step", the average IFP_{CS} score for the step is calculated: if the value is null, indicating that for the whole duration of the step the original binding mode was not sampled, the TTMD trajectory is terminated, while if the value is not null the simulation proceeds to the next"TTMD-step".

2.5. TTMD Trajectory Analyses

Each TTMD trajectory is analyzed by making use of an in-house Python 3.10 script. The root-mean-squared deviation of atomic coordinates (RMSD) for both the ligand and the protein backbone is calculated for each frame through the MDAnalysis package. The per-residue decomposition of the protein-ligand interaction energy is computed for each frame by exploiting the NAMD Energy plugin (version 1.4)[86] for VMD. Three different plots are then generated, making use of the Matplotlib and seaborn Python packages. The first plot ("Titration Profile") reports the average IFP_{CS} value for each TTMD step as a function of the step temperature. A straight line joining the start and final state of the simulation is also drawn in the graph, and its slope is reported in the legend and stored for further analysis. The second graph illustrates the time-dependent per-residue decomposition of the interaction energy, with the 25 most contacted residues alongside the TTMD trajectory being

considered. The third and final plot reports the time-dependent evolution of the ligand and protein backbone RMSD and the IFP_{CS} value.

2.6. MS coefficient determination

For each TTMD simulation, a proxy value for the protein-ligand complex stability based on the conservation of the binding mode throughout the trajectory is calculated as reported in Equation 2:

$$MS = \frac{mean\ IFP_{CS}{}^{T^{end}} - (-1)}{T^{end} - T^{start}} \quad (2)$$

The MS coefficient is the slope of the straight line that interpolates the first and the last point of the "Titration Profile" plot described in the previous paragraph. In Equation 2, $mean\ IFP_{CS}{}^{T^{end}}$ is the average IFP_{CS} value for the last temperature explored in the TTMD trajectory, -1 is the IFP_{CS} value for the initial state of the simulation, T^{end} and T^{start} are the final and starting temperatures of the simulation. Values are positive and can vary between 0 (indicative of a strong binding) and 1 (related to a weak binding).

For each ligand, five independent TTMD simulations are performed, and the average MS coefficient is then calculated upon three of them, discarding the highest and the lowest value.

3. Results

To test and validate the applicability of the Thermal Titration Molecular Dynamics (TTMD) protocol, we performed four different case studies on three different pharmaceutically relevant targets of interest for our laboratory, specifically casein kinase 1δ (CK1δ), casein kinase 2 (CK2), pyruvate dehydrogenase kinase 2 (PDK2), and SARS-CoV-2 main protease (M^{pro}). For each protein target, five different protein-ligand complexes were chosen, based on the availability of binding affinity data. A list of all protein-ligand complexes used in the present work can be found in Table 1 at the beginning of the Materials and Methods section, while detailed information about the ligands utilized in each test case can be found in Tables S1-S4 (Supplementary Materials). For each protein-ligand complex investigated in the article, five independent TTMD simulations were carried out. The results for each test case are reported hereafter in separate paragraphs and discussed aggregately in the Discussion section of the manuscript. For each target, the conservation of the protein fold throughout the simulation is carried out by monitoring the time-dependent evolution of the protein backbone RMSD, as reported in the detailed analysis for each representative replicate illustrated in Figures S1-S21 (Supplementary Materials).

3.1. Protein Kinase CK1δ

Protein Kinase CK1δ is a serine-threonine kinase that belongs to the family of CK1 Kinases (Casein Kinase 1)[87]. Due to its pleiotropic nature (about 140 substrates have been reported so far), this kinase is involved in the regulation of several different cellular pathways[87][88]. Particularly relevant from a medicinal chemistry perspective is its involvement in several neurodegenerative diseases such as Alzheimer's Disease (AD), Parkinson's Disease (PD), and amyotrophic lateral sclerosis (ALS) by phosphorylating protein targets such as the Tau protein, α-synuclein, and TDP-43 (TransActivate Response DNA Binding Protein 43)[89]. 34 crystal structures of CK1δ, among which several protein-ligand complexes can be found, are deposited in the Protein Data Bank, with the affinity of co-crystallized inhibitors ranging over three orders of magnitude, making it a suitable target for the application of our computational protocol. The results of TTMD simulations performed on CK1δ crystal complexes are summarized in Table 2 and Figure 2, while a detailed analysis for a representative trajectory for each protein-ligand complex (the one highlighted in green in Table 2) is reported in Figures S1-S5 in the Supplementary Materials.

PDB ID	LIG ID	IC_{50} (nM)	MS_{MD1}	MS_{MD2}	MS_{MD3}	MS_{MD4}	MS_{MD5}	$MS_{average}$
3UZP	0CK	13	0.00203	0.00333	0.00228	0.00212	0.00275	0.0024
4TN6	PFO	3.9	0.00300	0.00345	0.00307	0.00214	0.00309	0.0031
5IH5	AUE	500	0.00539	0.00446	0.00357	0.00909	0.00347	0.0045
5IH6	AUG	2500	0.00380	0.00769	0.00420	0.00557	0.00594	0.0052
5MQV	D5Q	9	0.00283	0.00295	0.00317	0.00256	0.00340	0.0030

***Table 2**. Results for the Thermal Titration Molecular Dynamics (TTMD) simulations performed on the five investigated CK1δ complexes. For each protein-ligand complex, the PDB accession code, the ligand three-letter code, the experimentally determined affinity value, the MS coefficient for each simulation, and the average MS coefficient are reported. In each row, the lowest MS value is highlighted in red, while the highest value is highlighted in blue: both values were discarded for the calculation of the average MS coefficient reported in the last column. The most representative replicates, the one with the nearest MS coefficient to the average MS, is highlighted in green.*

As can be deduced by the analysis of the data extracted from the various TTMD simulations, the ligands respond differently to the protocol based on the experimental affinity value. As can be noticed in Figure 2, complexes 5IH5 and 5IH6, which are characterized by the lowest affinity values (500 and 2500 μM, respectively), are the ones with the highest MS coefficient value, indicating a loss of the native binding mode throughout the simulations. On the contrary, ligands with a good experimental affinity towards the target (in the low nanomolar range), are associated with good conservation of the native binding mode, as highlighted by the lower MS coefficient value. Based on this observation, a cutoff MS value of 0.004, able to distinguish between the tight and weak binders for CK1δ, can be determined. The detailed trajectory analyses provided in Figures S1-S5 illustrate how the loss of the native binding mode is primarily driven by the loss of crucial hydrogen bond interactions with the hinge region, particularly with Leu85 and Glu83. This evidence is in agreement with previously published works from our laboratory, which indicates how using an appropriate pharmacophore filter that takes into account the crucial hydrogen bond with the backbone of Leu85 leads to good results in virtual screening[85][90][91]. For visual reference, a comparison between the representative replicate for the 3UZP and 5IH6 is reported in Video S1.

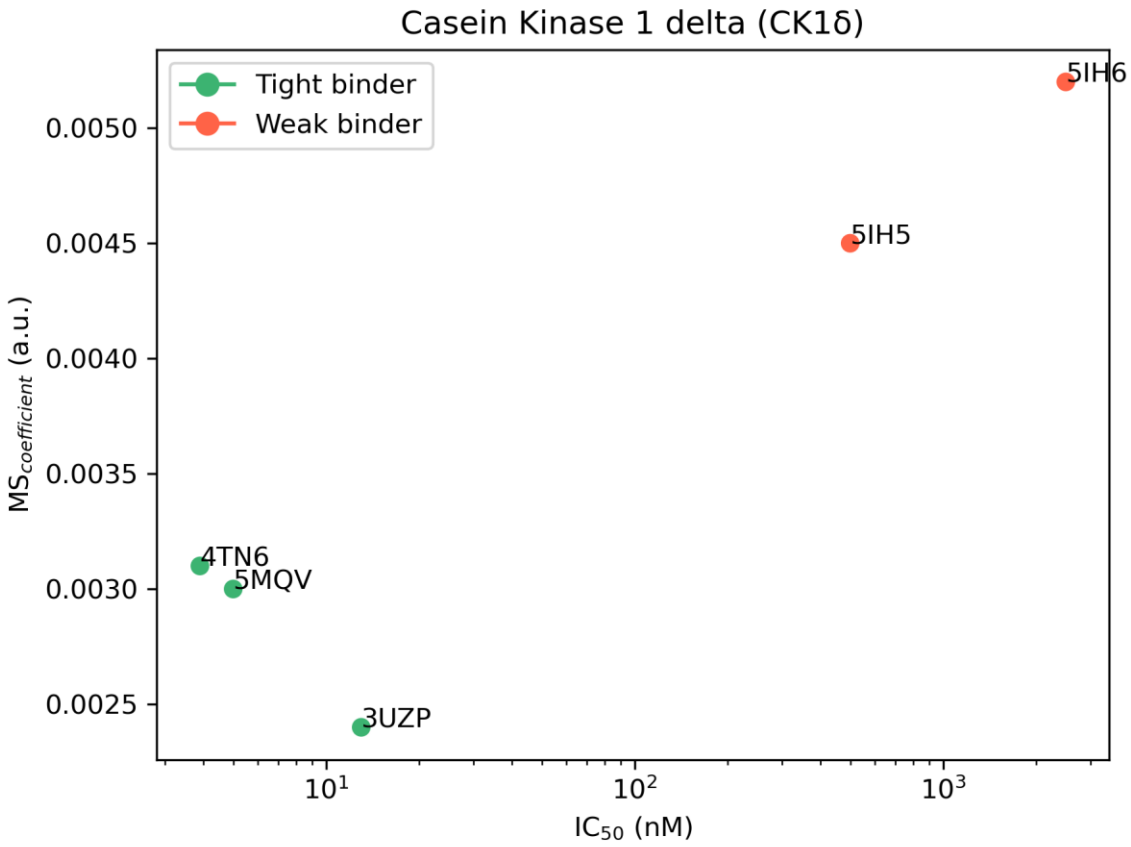


Figure 2. *Aggregate results of the Thermal Titration Molecular Dynamics (TTMD) simulations performed on the five investigated CK1δ complexes. On the horizontal axis, the experimentally determined affinity value (expressed as IC_{50}) is reported, while on the vertical axis the average MS coefficient is indicated. Each dot is color-coded as green or red and classified as a tight or weak binder based on the MS cutoff value of 0.004.*

3.2. Protein Kinase CK2

Protein Kinase CK2 is a serine-threonine kinase and represents one of the first identified protein kinases[92]. Similar to CK1δ, CK2 can phosphorylate a plethora of different substrates, being therefore involved in the regulation of several biological pathways[93]. The variety of biologically relevant scenarios in which CK2 is involved makes it a hot target from a pharmaceutical perspective, being related to several types of cancer, different neurodegenerative diseases (similarly to CK1δ), and also viral infections[94]. As of today, 214 crystal structures of CK2, among which several protein-ligand complexes can be found, are deposited in the Protein Data Bank, with the affinity of co-crystallized inhibitors ranging over five orders of magnitude, making it also a suitable target for the application of our computational protocol. The results of TTMD simulations carried out on CK2 crystal complexes are summarized in Table 3 and Figure 3, while a detailed analysis for a representative

trajectory for each protein-ligand complex (the one highlighted in green in Table 3) is reported in Figures S6-S11 in the Supplementary Materials. As explained in the Materials and Methods section, two different ligand protonation states are considered for the 2ZJW complex, the neutral and the negatively charged one. Although they are reported separately in both Table 3 and Figures S6-S7, they are considered as a single entity in Figure 3, where the average MS value between the two different protonation states is reported.

PDB ID	LIG ID	K_d (nM)	MS_{MD1}	MS_{MD2}	MS_{MD3}	MS_{MD4}	MS_{MD5}	$MS_{average}$
2ZJW (0)	REF	40	0.00318	0.00282	0.00269	0.00391	0.00205	0.0029
2ZJW (-)	REF	40	0.00192	0.00189	0.00227	0.00205	0.00133	0.0020
3H30	RFZ	13000	0.00504	0.00484	0.00395	0.00450	0.00453	0.0046
3PE1	3NG	1.5	0.00178	0.00231	0.00198	0.00219	0.00373	0.0022
3PE2	E1B	2.3	0.00179	0.00179	0.00171	0.00169	0.00170	0.0017
6HOU	V55	53400	0.0035	0.00667	0.00769	0.00364	0.00667	0.0057

***Table 3**. Results for the Thermal Titration Molecular Dynamics (TTMD) simulations performed on the five investigated CK2 complexes. For each protein-ligand complex, the PDB accession code, the ligand three-letter code, the experimentally determined affinity value, the MS coefficient for each simulation, and the average MS coefficient are reported. In each row, the lowest MS value is highlighted in red, while the highest value is highlighted in blue: both values were discarded for the calculation of the average MS coefficient reported in the last column. The most representative replicates, the one with the nearest MS coefficient to the average MS, is highlighted in green. For the complex 2ZJW, two different protonation states were independently considered in the simulations and are reported separately.*

As in the case of CK1δ, the investigated ligands show a different behavior during the TTMD simulations based on their experimental affinity value. The complexes characterized by a lower protein-ligand binding affinity (13 μM for 3H30, 53,4 μM for 6HOU), are also the ones characterized by the highest MS coefficient value (0.0046 and 0.0057, respectively). On the contrary, as observed for CK1δ, ligands with a binding affinity in the low nanomolar range are characterized by a conservation of the native binding mode throughout the simulation, resulting in a lower MS coefficient (below 0.003). Once again, an empirical threshold MS value of 0.004 can be extracted from this test set and utilized to distinguish between tight and weak binders.

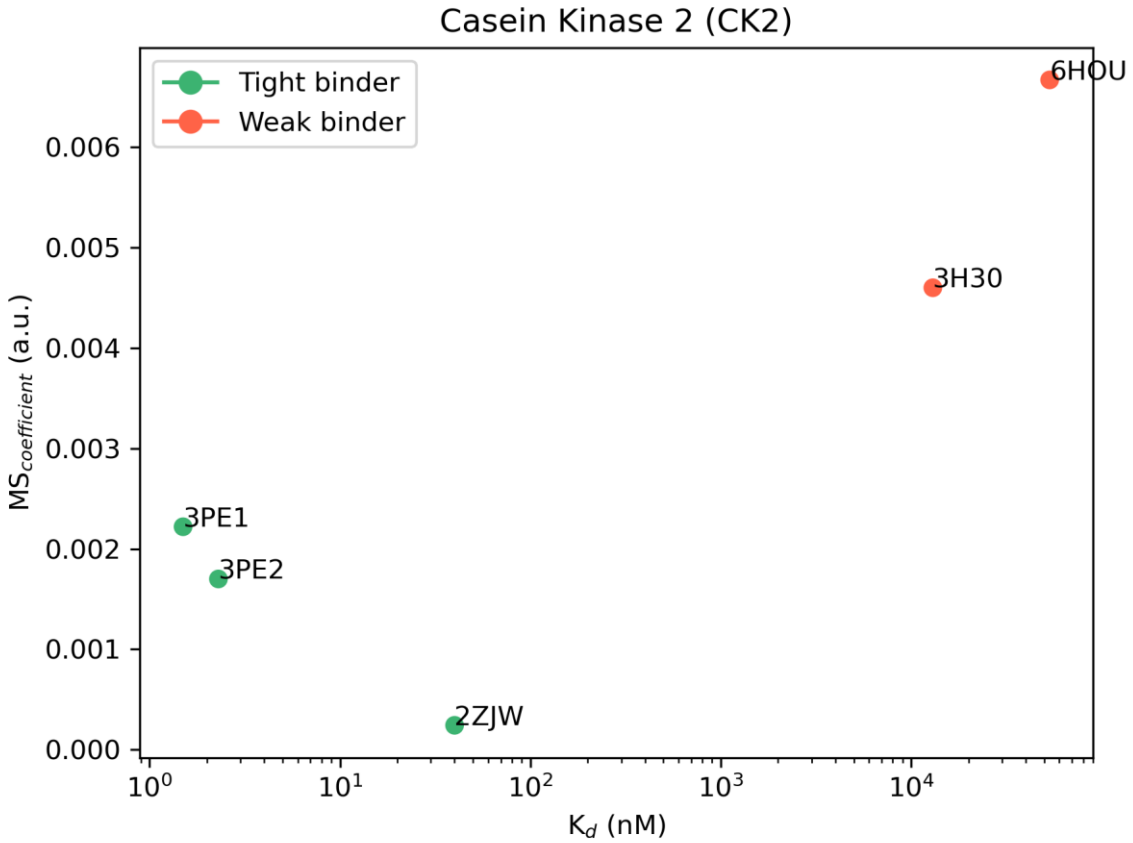


Figure 3. *Aggregate results of the Thermal Titration Molecular Dynamics (TTMD) simulations performed on the five investigated CK2 complexes. On the horizontal axis, the experimentally determined affinity value (expressed as K_d) is reported, while on the vertical axis the average MS coefficient is indicated. Each dot is color-coded as green or red and classified as a tight or weak binder based on the MS cutoff value of 0.004. For the complex 2ZJW, two different ligand protonation states were considered, but only one aggregate result (the average of the two states) is reported in the plot.*

As can be noticed by the evolution of the interaction pattern between ligands and the binding pocket throughout the trajectories (Figures S6-S11), the ligands with the most stable binding mode are the ones that tightly interact with Lys68 and Ile174: this is particularly noticeable in the case of ellagic acid (2ZJW), for which two different protonation states have been simulated. In the neutral form, the ellagic acid binding mode is less stable throughout the trajectory, while in the monocharged form, the ellagic acid binding mode is very stable even at high simulation temperatures, due to a favorable interaction with Lys68. A comparison between the representative replicate for complexes 3PE2 and 6HOU is shown in Video S2.

3.3. Pyruvate Dehydrogenase Kinase 2 (PDK2)

Pyruvate Dehydrogenase Kinase 2 (PDK2) is a pivotal enzyme in cellular energy metabolism that has previously been implicated in cancer[95]. PDK2 is a member of the GHKL ATPase/kinase superfamily and exerts its activity by phosphorylating and regulating the pyruvate dehydrogenase complex (PDC), which is a central control point in cellular energy metabolism since it links glycolysis with the tricarboxylic acid cycle[96][97]. Due to its involvement in the regulation of the energetic metabolism of cells, it is a drug target both from a metabolic and an antitumoral perspective. At the present moment, 33 crystal structures of PDK2, among which several protein-ligand complexes can be found, are deposited in the Protein Data Bank, with the affinity of co-crystallized inhibitors ranging over six orders of magnitude, making it a suitable target for the application of our computational protocol.

PDB ID	LIG ID	K_d(nM)	MS_{MD1}	MS_{MD2}	MS_{MD3}	MS_{MD4}	MS_{MD5}	$MS_{average}$
4MP2	PV1	3570	0.00603	0.01429	0.025	0.01111	0.00579	0.0105
4V25	SZ6	150	0.00216	0.00247	0.00194	0.00221	0.00332	0.0023
5J71	P35	110	0.00296	0.00296	0.00269	0.00243	0.00331	0.0029
5M4M	7FW	1	0.0028	0.0025	0.00286	0.00261	0.00318	0.0027
7EA0	W6P	958000	0.00347	0.00769	0.00667	0.00769	0.00665	0.0070

***Table 4**. Results for the Thermal Titration Molecular Dynamics (TTMD) simulations performed on the five investigated PDK2 complexes. For each protein-ligand complex, the PDB accession code, the ligand three-letter code, the experimentally determined affinity value, the MS coefficient for each simulation, and the average MS coefficient are reported. In each row, the lowest MS value is highlighted in red, while the highest value is highlighted in blue: both values were discarded for the calculation of the average MS coefficient reported in the last column. The most representative replicates, the one with the nearest MS coefficient to the average MS, is highlighted in green.*

The results of TTMD simulations performed on PDK2 crystal complexes are summarized in Table 4 and Figure 4, while a detailed analysis for a representative trajectory for each protein-ligand complex (the one highlighted in green in Table 4) is reported in Figures S12-S16 in the Supplementary Materials.

The analysis of results for the TTMD simulations performed on PDK2 protein-ligand complexes matches the ones already shown for CK1δ and CK2. Indeed, the ligands with the lowest binding affinity (3.57 µM for 4MP2, 958 µM for 7EA0), are the ones with the highest MS coefficient (0.0105 and 0.007, respectively), while on the contrary ligands characterized by a good binding affinity are

also the ones characterized by the lowest MS coefficient (below 0.003). The same threshold value used for previous cases (MS < 0.004) can also be utilized in this case to distinguish between the weak and the tight binders.

Looking at the evolution of the interaction pattern of various ligands throughout the simulations, it can be noticed that tight binders are characterized by persistent attractive interactions with Asp290 and Thr354. These residues are buried within the binding pocket, which contributes to the persistence of their interaction with the ligand compared to other more solvent-exposed residues such as Asn255, Arg258, and Glu262, that instead seem to be less relevant in retaining the ligand within the binding site. Noticeably, in the case of complex 4MP2, a repulsive interaction with Asp290 is present at the beginning of the simulation, and this could be a possible explanation for the low persistence of the native binding mode. Moreover, as can be noticed in Figures S14 and S15, in the case of complexes 5J71 and 5M4M, the fraction of the ligand which interacts with Asp290 and Thr354 barely moves from the starting position, fully retaining this interaction for the whole duration of the simulation, while partially losing the interactions with other more exposed residues which increases the ligand RMSD despite most of the binding determinants being conserved. A comparison between the representative replicate for complexes 4V25 and 4MP2 is illustrated in Video S3.

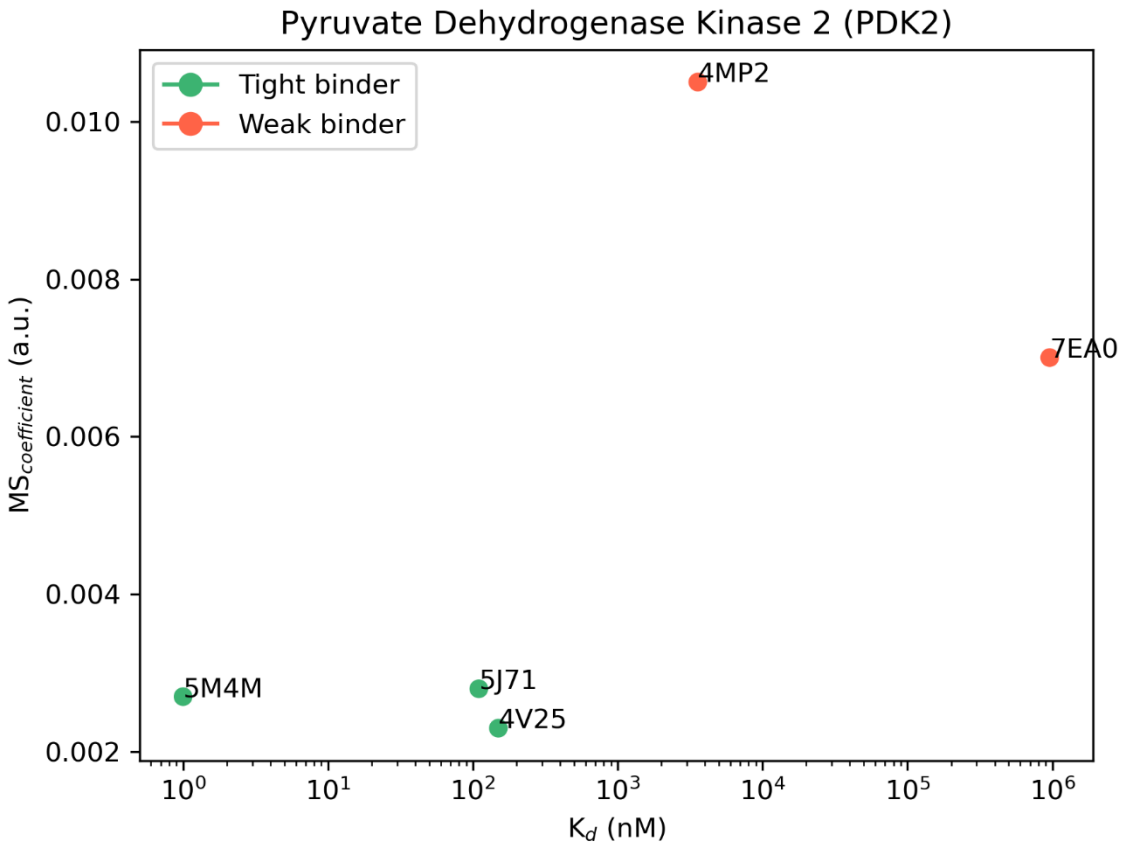


Figure 4. *Aggregate results of the Thermal Titration Molecular Dynamics (TTMD) simulations performed on the five investigated PDK2 complexes. On the horizontal axis, the experimentally determined affinity value (expressed as K_d) is reported, while on the vertical axis the average MS coefficient is indicated. Each dot is color-coded as green or red and classified as a tight or weak binder based on the MS cutoff value of 0.004.*

3.4. SARS-CoV-2 Main Protease (M^{pro})

SARS-CoV-2 is a betacoronavirus responsible for the COVID-19 pandemic which, to date, has caused the death of more than 6.5 million people around the world[98][99]. A pivotal enzyme in the virus' replication cycle is represented by its main protease (M^{pro}), a cysteine peptidase that is involved in the proteolytic cleavage of the pp1a/pp1ab polyproteins into several mature nonstructural proteins[100][101]. Due to its crucial role in the ability of the virus to replicate itself, the main protease is a validated antiviral target[102] and, as such, has become the focus of several different drug discovery campaigns[103][104][105], leading to 613 experimentally solved structures deposited on the Protein Data Bank, a marketed drug (Paxlovid, therapeutic association of Nirmatrelvir and Ritonavir)[106][107] and several inhibitors, with affinity values ranging from low nanomolar to micromolar and beyond. Its pharmaceutical relevance and the abundance of structural data make the

SARS-CoV-2 main protease a good target for the validation of the TTMD protocol. The results of TTMD simulations performed on SARS-CoV-2 M^{pro} crystal complexes are summarized in Table 5 and Figure 5, while a detailed analysis for a representative trajectory for each protein-ligand complex (the one highlighted in green in Table 5) is reported in Figures S17-S21 in the Supplementary Materials.

PDB ID	LIG ID	IC_{50}(nM)	MS_{MD1}	MS_{MD2}	MS_{MD3}	MS_{MD4}	MS_{MD5}	$MS_{average}$
6M2N	3WL	940	0.00714	0.00714	0.00557	0.00714	0.00909	0.0071
7LTJ	YD1	4200	0.00909	0.00642	0.00595	0.00667	0.00664	0.0066
7M8P	YSJ	20	0.00401	0.00336	0.00379	0.00317	0.00254	0.0034
7M91	YU4	25	0.00316	0.00437	0.00315	0.00313	0.00383	0.0034
7N44	06I	42	0.00325	0.00598	0.00398	0.00384	0.00383	0.0039

***Table 5**. Results for the Thermal Titration Molecular Dynamics (TTMD) simulations performed on the five investigated SARS-CoV-2 M^{pro} complexes. For each protein-ligand complex, the PDB accession code, the ligand three-letter code, the experimentally determined affinity value, the MS coefficient for each simulation, and the average MS coefficient are reported. In each row, the lowest MS value is highlighted in red, while the highest value is highlighted in blue: both values were discarded for the calculation of the average MS coefficient reported in the last column. The most representative replicates, the one with the nearest MS coefficient to the average MS, is highlighted in green.*

The analysis of the TTMD simulations performed on M^{pro} protein-ligand complexes matches the ones already shown for the previous cases. Once again, the ligands characterized by the lowest experimental binding affinity (0.94 µM for 6M2N, 4.2 µM for 7LTJ) are the ones associated with the highest MS coefficient (0.0071 and 0.0066, respectively). Accordingly, the ligands that present the highest binding affinity are associated with low MS coefficients, indicative of a persistent binding mode. Even for the SARS-CoV-2 M^{pro} it is possible to reutilize the previously determined threshold value ($MS < 0.004$) to separate the strong and the weak binders.

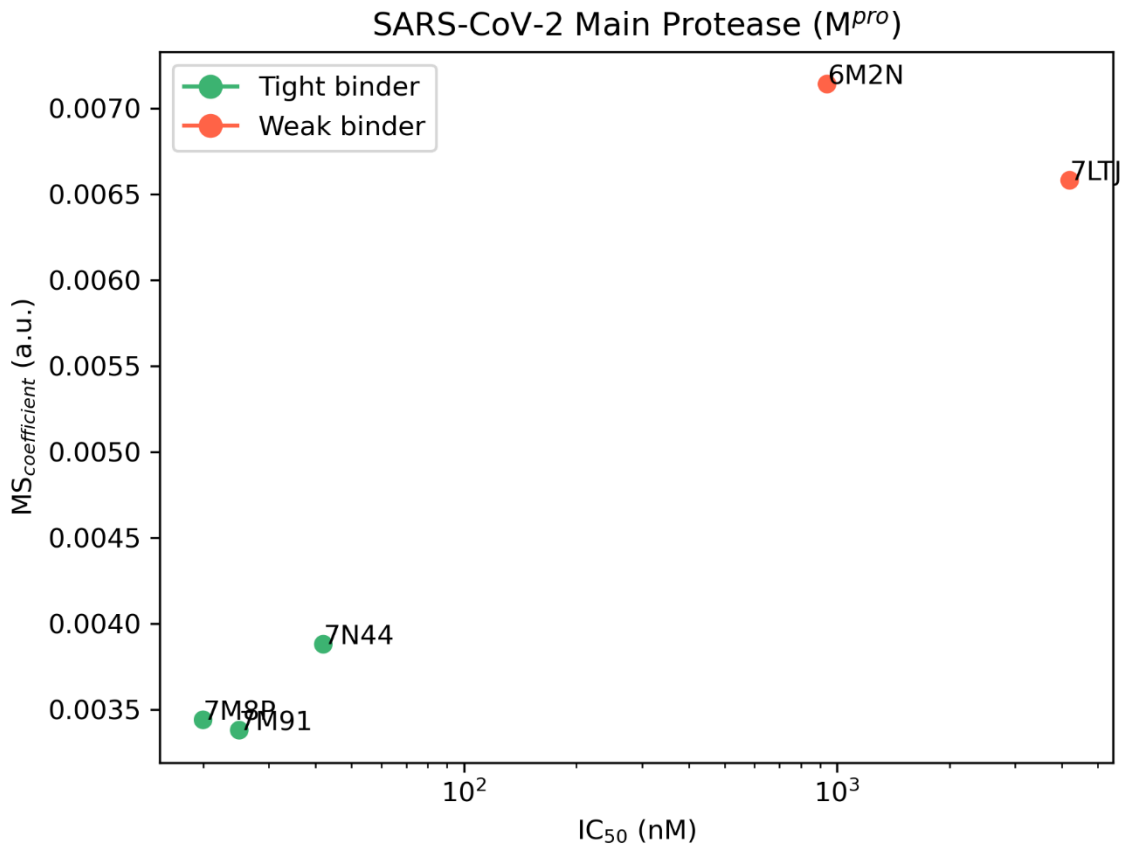


Figure 5. *Aggregate results of the Thermal Titration Molecular Dynamics (TTMD) simulations performed on the five investigated SARS-CoV-2 M^{pro} complexes. On the horizontal axis, the experimentally determined affinity value (expressed as IC$_{50}$) is reported, while on the vertical axis the average MS coefficient is indicated. Each dot is color-coded as green or red and classified as a tight or weak binder based on the MS cutoff value of 0.004.*

Regarding the evolution of the interaction pattern for the protein-ligand complexes that were investigated, it is possible to notice how the most persistent ligands are characterized by strong and stable interactions with key residues such as Met164-Glu166, located in a β-sheet that constitutes the central portion of the binding site lining several subpockets that precede the catalytic dyad such as S1, S2 and S3, and residues Leu141-Cys145, which line the S1 subpocket and constitute the so-called oxyanion loop, a structure that portrays a crucial role in the catalytic cycle of the protease[101]. On the contrary, interactions with residues lining the S2 and S4 subpocket seem to not be pivotal for the interaction with the catalytic site. As can be seen in Figures S20-S21, for example, the ligands partially lose their interactions with residues Asp187-Gln192: this causes a slight increase in the ligand RMSD towards the end of the simulation without overall altering the conservation of the native binding mode, as depicted by the interaction fingerprint analysis. Interestingly, in the case of complex 6M2N,

a repulsive interaction with Glu166 is present at the beginning of the simulation: considering the pivotal role that this residue portrays both in the dimerization process[108] (it forms a salt bridge through its side chain with the side chain of Ser1 of the second protomer) and in the binding of ligands, this could well explain the low persistence of the native binding mode for this ligand throughout the simulation. A comparison between the representative replicate for complexes 7LTJ and 7M91 is illustrated in Video S4.

4. Discussion

The Thermal Titration Molecular Dynamics (TTMD) method is an alternative protocol for the qualitative estimation of the protein-ligand complex stability based on the persistence of the native binding mode throughout a series of molecular dynamics simulations performed at progressively increasing temperatures. To evaluate the protocol capabilities, we performed four different case studies on an equal number of pharmaceutically relevant test cases, i.e. protein kinase CK1δ, protein kinase CK2, pyruvate dehydrogenase kinase 2, and SARS-CoV-2 main protease (M^{pro}). Despite its simplicity, the TTMD workflow was able to correctly discriminate between tight binders (with affinity values in the low nanomolar range) and weak binders (the ones with affinity values superior to the micromolar threshold) by applying an appropriate MS coefficient cutoff. This classification was performed on ligands with different scaffolds, and different interaction features, making its application interesting in real-world drug-discovery campaigns, where compounds from different chemical classes are usually identified in the first stages and are then subjected to an iterative optimization of their binding affinity towards the target of interest through chemical modification of their structures.

Contrary to most protocols that aim at predicting or estimating the drug-target residence time or other proxy values for the ligand affinity, TTMD does not require simulating the full unbinding event. Although this results in a rawer prediction compared to other similar protocols, this approach has two major advantages. The first one is that the simulation time is limited and can be accurately estimated right from the start. This is particularly useful in the case of a batch application of the protocol across a library of different compounds resulting from a screening campaign or an optimization process. Moreover, it facilitates the automatization of the process and its incorporation into existing drug-discovery pipelines. The second advantage is that by determining a relative metric (the MS coefficient) rather than an absolute one (i.e., the time required to observe the unbinding), there is no need for the definition of an arbitrary cutoff value for the detection of the unbinding event. Concerning this, most protocols exploit geometric descriptors such as the distance between the center of masses of the ligand and the binding site, or the distance between the ligand and the protein to define whether the ligand detached from the binding site[43,46]. This poses the problem of choosing the right distance because, in the case of deep and buried binding sites, the chosen cutoff value could not consider the whole unbinding process, leading once again to an underestimation of

the residence time. On the contrary, arbitrarily increasing the distance could elongate the simulation time without improving the prediction accuracy. Furthermore, using an interaction fingerprint-based metric instead of the standard RMSD for monitoring the evolution of the binding mode throughout the simulation results in lower sensitivity toward the chemical structure of the ligand: as highlighted in some of our trajectory analyses (Figures S14 and S15, for example), the presence of some ligand moieties that do not directly interact with the binding pocket or are slightly solvent exposed, leading to a less stable interaction with the target, could lead to an increase in the ligand RMSD without compromising the key binding determinants of the compounds. This could lead to a false perception of the unbinding event, causing errors in the evaluation of the persistence of the receptor-ligand complex, especially if very low cutoffs are utilized, such is the case in some studies in which classic MD simulations are used as a way to refine docking results and distinguish between native-like poses and decoys[109][110].

Another major advantage is related to the accessibility of the method. First, although the protocol in its current form exploits the ACEMD3 program to run molecular dynamics simulations, it can be easily and readily adapted to be utilized with any other major molecular dynamics engine such as OpenMM, GROMACS, or AMBER. Secondly, compared to other approaches TTMD is easier to implement. For example, contrary to metadynamics-based approaches, where the choice of the collective variable to monitor is not trivial[38], in the case of TTMD the user only needs to choose a temperature ramp where that ensures the conservation of the protein fold by monitoring a simple geometric descriptor such as the protein backbone RMSD. Although some attempts at optimizing the temperature ramp to decrease the simulation time without reducing the accuracy of the method are already going on in our laboratory, the temperature ramp proposed in this article should represent a good starting point for the third-party implementation of the method. Theoretically, increasing the simulation time for each “TTMD step” should provide an increase in the resolution of the technique, but would also result in an increased computational effort. On the contrary, reducing the simulation time for each TTMD step would reduce the computational effort, making the protocol more affordable, especially for those setups where a large number of different ligands is evaluated at a given time, but would also flatten the difference in MS coefficient between the ligands, thus decreasing the sensitivity of the technique. One possible solution could be to use different simulation times at different temperatures, for example simulating longer steps at lower temperatures and shorter ones at higher temperatures.

The pool of test cases provided in this article should, in principle, aid the user in the choice of a non-default temperature ramp, since the user could compare the results of its custom temperature ramp with the ones originally obtained and evaluate on its own the performances of a different ramp.

Other than estimating the protein-ligand binding affinity, the TTMD protocol could be easily adapted to perform mechanistic evaluations on the unbinding event, by appropriately tuning the temperature ramp and the simulation time to carry on the simulation until the native binding mode is completely lost. Although in the current form this protocol is not specifically designed for this purpose, it can be already used to discriminate different protein-ligand interactions based on their effect on the binding affinity. This indication could be very useful in the generation and refinement of pharmacophore filters, which are commonly used to reduce the false positive rates in docking-based virtual screening campaigns.

Another possible application of the TTMD protocol could be to distinguish the native binding pose from a set of decoy ones. This could be particularly useful in the case of fragment compounds, which can have several plausible binding modes, and are usually evolved into mature ligands by rationally modifying their scaffold to expand on their binding determinants without altering the existing interaction profile. Work in this sense is already going on in our laboratory and will be the scope of a future paper.

The last element that needs to be addressed in the nearest future is the applicability of the method to membrane proteins: so far, the protocol has been applied only to globular targets, but a wide variety of pharmaceutically relevant targets are membrane systems. Membrane systems are intrinsically more complicated to manage, because other than monitoring the protein fold throughout the simulations one has to decide how to manage the membrane. A possible solution could be to remove the membrane and treat the protein as soluble, possibly with the implementation of restraints on the atomic positions of atoms outside the binding site[43]. Evaluations in this sense are already going on in our laboratory to tune the protocol to be utilized also for this class of targets.

5. Conclusions

In this scientific work, we presented the first application of Thermal Titration Molecular Dynamics (TTMD), an alternative protocol for the qualitative estimation of protein-ligand complex stability by monitoring the conservation of the native ligand binding mode throughout a series of classic molecular dynamics simulations performed at progressively increasing temperatures through a scoring function based on protein-ligand interaction fingerprints. Four different test cases regarding the application of the technique on three different pharmaceutically relevant targets were presented. For each case, TTMD simulations were able to distinguish between tight (low nanomolar) from weak (micromolar) binders. The simplicity of the protocol, particularly regarding the choice of user-defined parameters to run the simulations, the agnosticism concerning the selection of the molecular dynamics engine, and the limited simulation time make it a viable choice for various medicinal chemistry projects, especially as a screening tool in the early stages of drug discovery campaigns. Further work is needed to extend the applicability domain of the technique to membrane proteins, and evaluations in this sense are already going on in our laboratory.

References

[1] Ehrlich, P. Address in Pathology, ON CHEMIOTHERAPY. *Br Med J* **1913**, *2*, 353–359. https://doi.org/10.1136/BMJ.2.2746.353.

[2] Langley, J. N. On the Reaction of Cells and of Nerve-Endings to Certain Poisons, Chiefly as Regards the Reaction of Striated Muscle to Nicotine and to Curari. *J Physiol* **1905**, *33*, 374. https://doi.org/10.1113/JPHYSIOL.1905.SP001128.

[3] Kaufmann, S. H. E. Paul Ehrlich: Founder of Chemotherapy. *Nat Rev Drug Discov* **2008**, *7*, 373–373. https://doi.org/10.1038/nrd2582.

[4] Maehle, A. H. "Receptive Substances": John Newport Langley (1852–1925) and His Path to a Receptor Theory of Drug Action. *Med Hist* **2004**, *48*, 153. https://doi.org/10.1017/S0025727300000090.

[5] Bennett, M. R. The Concept of Transmitter Receptors: 100 Years On. *Neuropharmacology* **2000**, *39*, 523–546. https://doi.org/10.1016/S0028-3908(99)00137-9.

[6] Gourley, D. R. Isolation and Characterization of Membrane Drug Receptors. *Prog Drug Res* **1976**, *20*, 323–346. https://doi.org/10.1007/978-3-0348-7094-8_10/COVER/.

[7] Rang, H. P. The Receptor Concept: Pharmacology's Big Idea. *Br J Pharmacol* **2006**, *147*, S9–S16. https://doi.org/10.1038/SJ.BJP.0706457.

[8] Borea, P. A.; Varani, K.; Gessi, S.; Gilli, P.; Dalpiaz, A. Receptor Binding Thermodynamics as a Tool for Linking Drug Efficacy and Affinity. *Farmaco* **1998**, *53*, 249–254. https://doi.org/10.1016/S0014-827X(98)00017-2.

[9] Kairys, V.; Baranauskiene, L.; Kazlauskiene, M.; Matulis, D.; Kazlauskas, E. Binding Affinity in Drug Design: Experimental and Computational Techniques. *https://doi.org/10.1080/17460441.2019.1623202* **2019**, *14*, 755–768. https://doi.org/10.1080/17460441.2019.1623202.

[10] Pan, A. C.; Borhani, D. W.; Dror, R. O.; Shaw, D. E. Molecular Determinants of Drug–Receptor Binding Kinetics. *Drug Discov Today* **2013**, *18*, 667–673. https://doi.org/10.1016/J.DRUDIS.2013.02.007.

[11] Copeland, R. A. The Drug–Target Residence Time Model: A 10-Year Retrospective. *Nature Reviews Drug Discovery 2015 15:2* **2015**, *15*, 87–95. https://doi.org/10.1038/nrd.2015.18.

[12] Copeland, R. A.; Pompliano, D. L.; Meek, T. D. Drug-Target Residence Time and Its Implications for Lead Optimization. *Nat Rev Drug Discov* **2006**, *5*, 730–739. https://doi.org/10.1038/nrd2082.

[13] Swinney, D. C. The Role of Binding Kinetics in Therapeutically Useful Drug Action. *Curr Opin Drug Discov Devel* **2009**, *12*, 31–39.

[14] Zhang, R.; Monsma, F. The Importance of Drug-Target Residence Time. *Curr Opin Drug Discov Devel* **2009**, *12*, 488–496.

[15] Lu, H.; Tonge, P. J. Drug–Target Residence Time: Critical Information for Lead Optimization. *Curr Opin Chem Biol* **2010**, *14*, 467–474. https://doi.org/10.1016/J.CBPA.2010.06.176.

[16] Hill, T. L. Effect of Rotation on the Diffusion-Controlled Rate of Ligand-Protein Association. *Proc Natl Acad Sci U S A* **1975**, *72*, 4918. https://doi.org/10.1073/PNAS.72.12.4918.

[17] Guo, D.; Hillger, J. M.; Ijzerman, A. P.; Heitman, L. H. Drug-Target Residence Time—A Case for G Protein-Coupled Receptors. *Med Res Rev* **2014**, *34*, 856–892. https://doi.org/10.1002/MED.21307.

[18] Cusack, K. P.; Wang, Y.; Hoemann, M. Z.; Marjanovic, J.; Heym, R. G.; Vasudevan, A. Design Strategies to Address Kinetics of Drug Binding and Residence Time. *Bioorg Med Chem Lett* **2015**, *25*, 2019–2027. https://doi.org/10.1016/J.BMCL.2015.02.027.

[19] Bernetti, M.; Cavalli, A.; Mollica, L. Protein–Ligand (Un)Binding Kinetics as a New Paradigm for Drug Discovery at the Crossroad between Experiments and Modelling. *Medchemcomm* **2017**, *8*, 534–550. https://doi.org/10.1039/C6MD00581K.

[20] Meyer-Almes, F. J. Kinetic Binding Assays for the Analysis of Protein–Ligand Interactions. *Drug Discov Today Technol* **2015**, *17*, 1–8. https://doi.org/10.1016/J.DDTEC.2015.08.004.

[21] Bernetti, M.; Masetti, M.; Rocchia, W.; Cavalli, A. Kinetics of Drug Binding and Residence Time. *https://doi.org/10.1146/annurev-physchem-042018-052340* **2019**, *70*, 143–171. https://doi.org/10.1146/ANNUREV-PHYSCHEM-042018-052340.

[22] Hulme, E. C.; Trevethick, M. A. Ligand Binding Assays at Equilibrium: Validation and Interpretation. *Br J Pharmacol* **2010**, *161*, 1219–1237. https://doi.org/10.1111/J.1476-5381.2009.00604.X.

[23] Sridharan, R.; Zuber, J.; Connelly, S. M.; Mathew, E.; Dumont, M. E. Fluorescent Approaches for Understanding Interactions of Ligands with G Protein Coupled Receptors. *Biochim Biophys Acta* **2014**, *1838*, 15–33. https://doi.org/10.1016/J.BBAMEM.2013.09.005.

[24] Ilien, B.; Franchet, C.; Bernard, P.; Morisset, S.; Weill, C. O.; Bourguignon, J. J.; Hibert, M.; Galzi, J. L. Fluorescence Resonance Energy Transfer to Probe Human M1 Muscarinic Receptor Structure and Drug Binding Properties. *J Neurochem* **2003**, *85*, 768–778. https://doi.org/10.1046/J.1471-4159.2003.01717.X.

[25] Stoddart, L. A.; Johnstone, E. K. M.; Wheal, A. J.; Goulding, J.; Robers, M. B.; MacHleidt, T.; Wood, K. V.; Hill, S. J.; Pfleger, K. D. G. Application of BRET to Monitor Ligand Binding to GPCRs. *Nat Methods* **2015**, *12*, 661–663. https://doi.org/10.1038/NMETH.3398.

[26] Núñez, S.; Venhorst, J.; Kruse, C. G. Target–Drug Interactions: First Principles and Their Application to Drug Discovery. *Drug Discov Today* **2012**, *17*, 10–22. https://doi.org/10.1016/J.DRUDIS.2011.06.013.

[27] Patching, S. G. Surface Plasmon Resonance Spectroscopy for Characterisation of Membrane Protein-Ligand Interactions and Its Potential for Drug Discovery. *Biochim Biophys Acta* **2014**, *1838*, 43–55. https://doi.org/10.1016/J.BBAMEM.2013.04.028.

[28] Millet, O.; Bernadó, P.; Garcia, J.; Rizo, J.; Pons, M. NMR Measurement of the off Rate from the First Calcium-Binding Site of the Synaptotagmin I C2A Domain. *FEBS Lett* **2002**, *516*, 93–96. https://doi.org/10.1016/S0014-5793(02)02508-5.

[29] Gronewold, T. M. A.; Baumgartner, A.; Hierer, J.; Sierra, S.; Blind, M.; Schäfer, F.; Blümer, J.; Tillmann, T.; Kiwitz, A.; Kaiser, R.; Zabe-Kühn, M.; Quandt, E.; Famulok, M. Kinetic Binding Analysis of Aptamers Targeting HIV-1 Proteins by a Combination of a Microbalance Array and Mass Spectrometry (MAMS). *J Proteome Res* **2009**, *8*, 3568–3577. https://doi.org/10.1021/PR900265R.

[30] Burnouf, D.; Ennifar, E.; Guedich, S.; Puffer, B.; Hoffmann, G.; Bec, G.; Disdier, F.; Baltzinger, M.; Dumas, P. KinITC: A New Method for Obtaining Joint Thermodynamic and Kinetic Data by Isothermal Titration Calorimetry. *J Am Chem Soc* **2012**, *134*, 559–565. https://doi.org/10.1021/JA209057D/SUPPL_FILE/JA209057D_SI_001.PDF.

[31] Li, D.; Chen, L.; Wang, R.; Liu, R.; Ge, G. Synergetic Determination of Thermodynamic and Kinetic Signatures Using Isothermal Titration Calorimetry: A Full-Curve-Fitting Approach. *Anal Chem* **2017**, *89*,

7130–7138. https://doi.org/10.1021/ACS.ANALCHEM.7B01091/ASSET/IMAGES/LARGE/AC-2017-01091U_0005.JPEG.

[32] Acker, M. G.; Auld, D. S. Considerations for the Design and Reporting of Enzyme Assays in High-Throughput Screening Applications. *Perspect Sci (Neth)* **2014**, *1*, 56–73. https://doi.org/10.1016/J.PISC.2013.12.001.

[33] Bruce, N. J.; Ganotra, G. K.; Kokh, D. B.; Sadiq, S. K.; Wade, R. C. New Approaches for Computing Ligand-Receptor Binding Kinetics. *Curr Opin Struct Biol* **2018**, *49*, 1–10. https://doi.org/10.1016/J.SBI.2017.10.001.

[34] Nunes-Alves, A.; Kokh, D. B.; Wade, R. C. Recent Progress in Molecular Simulation Methods for Drug Binding Kinetics. *Curr Opin Struct Biol* **2020**, *64*, 126–133. https://doi.org/10.1016/J.SBI.2020.06.022.

[35] De Vivo, M.; Masetti, M.; Bottegoni, G.; Cavalli, A. Role of Molecular Dynamics and Related Methods in Drug Discovery. *J Med Chem* **2016**, *59*. https://doi.org/10.1021/acs.jmedchem.5b01684.

[36] Pan, A. C.; Xu, H.; Palpant, T.; Shaw, D. E. Quantitative Characterization of the Binding and Unbinding of Millimolar Drug Fragments with Molecular Dynamics Simulations. *J Chem Theory Comput* **2017**, *13*, 3372–3377. https://doi.org/10.1021/ACS.JCTC.7B00172/SUPPL_FILE/CT7B00172_SI_002.ZIP.

[37] Schuetz, D. A.; de Witte, W. E. A.; Wong, Y. C.; Knasmueller, B.; Richter, L.; Kokh, D. B.; Sadiq, S. K.; Bosma, R.; Nederpelt, I.; Heitman, L. H.; Segala, E.; Amaral, M.; Guo, D.; Andres, D.; Georgi, V.; Stoddart, L. A.; Hill, S.; Cooke, R. M.; De Graaf, C.; Leurs, R.; Frech, M.; Wade, R. C.; de Lange, E. C. M.; IJzerman, A. P.; Müller-Fahrnow, A.; Ecker, G. F. Kinetics for Drug Discovery: An Industry-Driven Effort to Target Drug Residence Time. *Drug Discov Today* **2017**, *22*, 896–911. https://doi.org/10.1016/J.DRUDIS.2017.02.002.

[38] Tiwary, P.; Limongelli, V.; Salvalaglio, M.; Parrinello, M. Kinetics of Protein-Ligand Unbinding: Predicting Pathways, Rates, and Rate-Limiting Steps. *Proc Natl Acad Sci U S A* **2015**, *112*, E386–E391. https://doi.org/10.1073/PNAS.1424461112/SUPPL_FILE/PNAS.1424461112.SAPP.PDF.

[39] Bortolato, A.; Deflorian, F.; Weiss, D. R.; Mason, J. S. Decoding the Role of Water Dynamics in Ligand-Protein Unbinding: CRF1R as a Test Case. *J Chem Inf Model* **2015**, *55*, 1857–1866. https://doi.org/10.1021/ACS.JCIM.5B00440/ASSET/IMAGES/LARGE/CI-2015-00440A_0006.JPEG.

[40] Pietrucci, F.; Marinelli, F.; Carloni, P.; Laio, A. Substrate Binding Mechanism of HIV-1 Protease from Explicit-Solvent Atomistic Simulations. *J Am Chem Soc* **2009**, *131*, 11811–11818. https://doi.org/10.1021/JA903045Y/SUPPL_FILE/JA903045Y_SI_001.PDF.

[41] Sun, H.; Li, Y.; Shen, M.; Li, D.; Kang, Y.; Hou, T. Characterizing Drug-Target Residence Time with Metadynamics: How to Achieve Dissociation Rate Efficiently without Losing Accuracy against Time-Consuming Approaches. *J Chem Inf Model* **2017**, *57*, 1895–1906. https://doi.org/10.1021/ACS.JCIM.7B00075/ASSET/IMAGES/LARGE/CI-2017-00075P_0004.JPEG.

[42] Bussi, G.; Laio, A. Using Metadynamics to Explore Complex Free-Energy Landscapes. *Nature Reviews Physics 2020 2:4* **2020**, *2*, 200–212. https://doi.org/10.1038/s42254-020-0153-0.

[43] Mollica, L.; Decherchi, S.; Zia, S. R.; Gaspari, R.; Cavalli, A.; Rocchia, W. Kinetics of Protein-Ligand Unbinding via Smoothed Potential Molecular Dynamics Simulations. *Scientific Reports 2015 5:1* **2015**, *5*, 1–12. https://doi.org/10.1038/srep11539.

[44] Schuetz, D. A.; Bernetti, M.; Bertazzo, M.; Musil, D.; Eggenweiler, H. M.; Recanatini, M.; Masetti, M.; Ecker, G. F.; Cavalli, A. Predicting Residence Time and Drug Unbinding Pathway through Scaled Molecular

Dynamics. *J Chem Inf Model* **2019**, *59*, 535–549. https://doi.org/10.1021/ACS.JCIM.8B00614/ASSET/IMAGES/LARGE/CI-2018-00614Y_0008.JPEG.

[45] Bernetti, M.; Rosini, E.; Mollica, L.; Masetti, M.; Pollegioni, L.; Recanatini, M.; Cavalli, A. Binding Residence Time through Scaled Molecular Dynamics: A Prospective Application to HDAAO Inhibitors. *J Chem Inf Model* **2018**, *58*, 2255–2265. https://doi.org/10.1021/ACS.JCIM.8B00518/ASSET/IMAGES/MEDIUM/CI-2018-00518D_M001.GIF.

[46] Kokh, D. B.; Amaral, M.; Bomke, J.; Grädler, U.; Musil, D.; Buchstaller, H. P.; Dreyer, M. K.; Frech, M.; Lowinski, M.; Vallee, F.; Bianciotto, M.; Rak, A.; Wade, R. C. Estimation of Drug-Target Residence Times by τ-Random Acceleration Molecular Dynamics Simulations. *J Chem Theory Comput* **2018**, *14*, 3859–3869. https://doi.org/10.1021/ACS.JCTC.8B00230/ASSET/IMAGES/LARGE/CT-2018-00230J_0003.JPEG.

[47] Nunes-Alves, A.; Kokh, D. B.; Wade, R. C. Ligand Unbinding Mechanisms and Kinetics for T4 Lysozyme Mutants from TRAMD Simulations. *Curr Res Struct Biol* **2021**, *3*, 106–111. https://doi.org/10.1016/J.CRSTBI.2021.04.001.

[48] Kokh, D. B.; Wade, R. C. G Protein-Coupled Receptor-Ligand Dissociation Rates and Mechanisms from TrAMD Simulations. *J Chem Theory Comput* **2021**, *17*, 6610–6623. https://doi.org/10.1021/ACS.JCTC.1C00641/SUPPL_FILE/CT1C00641_SI_002.MP4.

[49] Berman, H. M. The Protein Data Bank. *Nucleic Acids Res* **2000**, *28*, 235–242. https://doi.org/10.1093/nar/28.1.235.

[50] *Molecular Operating Environment (MOE), 2019.01; Chemical Computing Group ULC, 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2021.* https://www.chemcomp.com/Research-Citing_MOE.htm (accessed 2021-01-19).

[51] Long, A.; Zhao, H.; Huang, X. Structural Basis for the Interaction between Casein Kinase 1 Delta and a Potent and Selective Inhibitor. *J Med Chem* **2012**, *55*, 956–960. https://doi.org/10.1021/JM201387S.

[52] *RCSB PDB - 4TN6: CK1d in complex with inhibitor*. https://www.rcsb.org/structure/4TN6 (accessed 2022-06-23).

[53] Ursu, A.; Illich, D. J.; Takemoto, Y.; Porfetye, A. T.; Zhang, M.; Brockmeyer, A.; Janning, P.; Watanabe, N.; Osada, H.; Vetter, I. R.; Ziegler, S.; Schöler, H. R.; Waldmann, H. Epiblastin A Induces Reprogramming of Epiblast Stem Cells into Embryonic Stem Cells by Inhibition of Casein Kinase 1. *Cell Chem Biol* **2016**, *23*, 494–507. https://doi.org/10.1016/j.chembiol.2016.02.015.

[54] Halekotte, J.; Witt, L.; Ianes, C.; Krüger, M.; Bührmann, M.; Rauh, D.; Pichlo, C.; Brunstein, E.; Luxenburger, A.; Baumann, U.; Knippschild, U.; Bischof, J.; Peifer, C.; Koch, P.; Laufer, S. Optimized 4,5-Diarylimidazoles as Potent/Selective Inhibitors of Protein Kinase CK1δ and Their Structural Relation to P38α MAPK. *Molecules* **2017**, *22*. https://doi.org/10.3390/MOLECULES22040522.

[55] Sekiguchi, Y.; Nakaniwa, T.; Kinoshita, T.; Nakanishi, I.; Kitaura, K.; Hirasawa, A.; Tsujimoto, G.; Tada, T. Structural Insight into Human CK2α in Complex with the Potent Inhibitor Ellagic Acid. *Bioorg Med Chem Lett* **2009**, *19*, 2920–2923. https://doi.org/10.1016/J.BMCL.2009.04.076.

[56] Raaf, J.; Brunstein, E.; Issinger, O. G.; Niefind, K. The CK2α/CK2β Interface of Human Protein Kinase CK2 Harbors a Binding Pocket for Small Molecules. *Chem Biol* **2008**, *15*, 111–117. https://doi.org/10.1016/J.CHEMBIOL.2007.12.012.

[57] Battistutta, R.; Cozza, G.; Pierre, F.; Papinutto, E.; Lolli, G.; Sarno, S.; Obrien, S. E.; Siddiqui-Jain, A.; Haddach, M.; Anderes, K.; Ryckman, D. M.; Meggio, F.; Pinna, L. A. Unprecedented Selectivity and

Structural Determinants of a New Class of Protein Kinase CK2 Inhibitors in Clinical Trials for the Treatment of Cancer. *Biochemistry* **2011**, *50*, 8478–8488. https://doi.org/10.1021/BI2008382.

[58] Cozza, G.; Zonta, F.; Dalle Vedove, A.; Venerando, A.; Dall'Acqua, S.; Battistutta, R.; Ruzzene, M.; Lolli, G. Biochemical and Cellular Mechanism of Protein Kinase CK2 Inhibition by Deceptive Curcumin. *FEBS J* **2020**, *287*, 1850–1864. https://doi.org/10.1111/febs.15111.

[59] Tso, S. C.; Qi, X.; Gui, W. J.; Wu, C. Y.; Chuang, J. L.; Wernstedt-Asterholm, I.; Morlock, L. K.; Owens, K. R.; Scherer, P. E.; Williams, N. S.; Tambar, U. K.; Wynn, R. M.; Chuang, D. T. Structure-Guided Development of Specific Pyruvate Dehydrogenase Kinase Inhibitors Targeting the ATP-Binding Pocket. *Journal of Biological Chemistry* **2014**, *289*, 4432–4443. https://doi.org/10.1074/JBC.M113.533885.

[60] Moore, J. D.; Staniszewska, A.; Shaw, T.; D'Alessandro, J.; Davis, B.; Surgenor, A.; Baker, L.; Matassova, N.; Murray, J.; Macias, A.; Brough, P.; Wood, M.; Mahon, P. C. VER-246608, a Novel Pan-Isoform ATP Competitive Inhibitor of Pyruvate Dehydrogenase Kinase, Disrupts Warburg Metabolism and Induces Context-Dependent Cytostasis in Cancer Cells. *Oncotarget* **2014**, *5*, 12862–12876. https://doi.org/10.18632/ONCOTARGET.2656.

[61] Tso, S. C.; Lou, M.; Wu, C. Y.; Gui, W. J.; Chuang, J. L.; Morlock, L. K.; Williams, N. S.; Wynn, R. M.; Qi, X.; Chuang, D. T. Development of Dihydroxyphenyl Sulfonylisoindoline Derivatives as Liver-Targeting Pyruvate Dehydrogenase Kinase Inhibitors. *J Med Chem* **2017**, *60*, 1142–1150. https://doi.org/10.1021/ACS.JMEDCHEM.6B01540.

[62] Brough, P. A.; Baker, L.; Bedford, S.; Brown, K.; Chavda, S.; Chell, V.; D'Alessandro, J.; Davies, N. G. M.; Davis, B.; Le Strat, L.; Macias, A. T.; Maddox, D.; Mahon, P. C.; Massey, A. J.; Matassova, N.; McKenna, S.; Meissner, J. W. G.; Moore, J. D.; Murray, J. B.; Northfield, C. J.; Parry, C.; Parsons, R.; Roughley, S. D.; Shaw, T.; Simmonite, H.; Stokes, S.; Surgenor, A.; Stefaniak, E.; Robertson, A.; Wang, Y.; Webb, P.; Whitehead, N.; Wood, M. Application of Off-Rate Screening in the Identification of Novel Pan-Isoform Inhibitors of Pyruvate Dehydrogenase Kinase. *J Med Chem* **2017**, *60*, 2271–2286. https://doi.org/10.1021/ACS.JMEDCHEM.6B01478.

[63] Akaki, T.; Bessho, Y.; Ito, T.; Fujioka, S.; Ubukata, M.; Mori, G.; Yamanaka, K.; Orita, T.; Doi, S.; Iwanaga, T.; Ikegashira, K.; Hantani, Y.; Nakanishi, I.; Adachi, T. Fragment-Based Lead Discovery to Identify Novel Inhibitors That Target the ATP Binding Site of Pyruvate Dehydrogenase Kinases. *Bioorg Med Chem* **2021**, *44*, 116283. https://doi.org/10.1016/j.bmc.2021.116283.

[64] Su, H. xia; Yao, S.; Zhao, W. feng; Li, M. jun; Liu, J.; Shang, W. juan; Xie, H.; Ke, C. qiang; Hu, H. chen; Gao, M. na; Yu, K. qian; Liu, H.; Shen, J. shan; Tang, W.; Zhang, L. ke; Xiao, G. fu; Ni, L.; Wang, D. wen; Zuo, J. ping; Jiang, H. liang; Bai, F.; Wu, Y.; Ye, Y.; Xu, Y. chun. Anti-SARS-CoV-2 Activities in Vitro of Shuanghuanglian Preparations and Bioactive Ingredients. *Acta Pharmacol Sin* **2020**, *41*, 1167–1177. https://doi.org/10.1038/S41401-020-0483-6.

[65] Clyde, A.; Galanie, S.; Kneller, D. W.; Ma, H.; Babuji, Y.; Blaiszik, B.; Brace, A.; Brettin, T.; Chard, K.; Chard, R.; Coates, L.; Foster, I.; Hauner, D.; Kertesz, V.; Kumar, N.; Lee, H.; Li, Z.; Merzky, A.; Schmidt, J. G.; Tan, L.; Titov, M.; Trifan, A.; Turilli, M.; van Dam, H.; Chennubhotla, S. C.; Jha, S.; Kovalevsky, A.; Ramanathan, A.; Head, M. S.; Stevens, R. High-Throughput Virtual Screening and Validation of a SARS-CoV-2 Main Protease Noncovalent Inhibitor. *J Chem Inf Model* **2022**, *62*, 116–128. https://doi.org/10.1021/ACS.JCIM.1C00851.

[66] Deshmukh, M. G.; Ippolito, J. A.; Zhang, C. H.; Stone, E. A.; Reilly, R. A.; Miller, S. J.; Jorgensen, W. L.; Anderson, K. S. Structure-Guided Design of a Perampanel-Derived Pharmacophore Targeting the SARS-CoV-2 Main Protease. *Structure* **2021**, *29*, 823-833.e5. https://doi.org/10.1016/J.STR.2021.06.002.

[67] Zhang, C. H.; Spasov, K. A.; Reilly, R. A.; Hollander, K.; Stone, E. A.; Ippolito, J. A.; Liosi, M. E.; Deshmukh, M. G.; Tirado-Rives, J.; Zhang, S.; Liang, Z.; Miller, S. J.; Isaacs, F.; Lindenbach, B. D.; Anderson, K. S.; Jorgensen, W. L. Optimization of Triarylpyridinone Inhibitors of the Main Protease of SARS-CoV-2 to Low-Nanomolar Antiviral Potency. *ACS Med Chem Lett* **2021**, *12*, 1325–1332. https://doi.org/10.1021/ACSMEDCHEMLETT.1C00326.

[68] Cozza, G.; Bonvini, P.; Zorzi, E.; Poletto, G.; Pagano, M. A.; Sarno, S.; Donella-Deana, A.; Zagotto, G.; Rosolen, A.; Pinna, L. A.; Meggio, F.; Moro, S. Identification of Ellagic Acid as Potent Inhibitor of Protein Kinase CK2: A Successful Example of a Virtual Screening Application. *J Med Chem* **2006**, *49*, 2363–2366. https://doi.org/10.1021/JM060112M/SUPPL_FILE/JM060112MSI20060306_104657.PDF.

[69] Cozza, G.; Gianoncelli, A.; Bonvini, P.; Zorzi, E.; Pasquale, R.; Rosolen, A.; Pinna, L. A.; Meggio, F.; Zagotto, G.; Moro, S. Urolithin as a Converging Scaffold Linking Ellagic Acid and Coumarin Analogues: Design of Potent Protein Kinase CK2 Inhibitors. *ChemMedChem* **2011**, *6*, 2273–2286. https://doi.org/10.1002/CMDC.201100338.

[70] Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J Mol Graph* **1996**, *14*, 33–38. https://doi.org/10.1016/0263-7855(96)00018-5.

[71] Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber Biomolecular Simulation Programs. *J Comput Chem* **2005**, *26*, 1668–1688. https://doi.org/10.1002/jcc.20290.

[72] Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. Ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from Ff99SB. *J Chem Theory Comput* **2015**, *11*, 3696–3713. https://doi.org/10.1021/ACS.JCTC.5B00255.

[73] Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J Comput Chem* **2004**, *25*, 1157–1174. https://doi.org/10.1002/JCC.20035.

[74] Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: II. Parameterization and Validation. *J Comput Chem* **2002**, *23*, 1623–1641. https://doi.org/10.1002/jcc.10128.

[75] Jorgensen, W. L.; Tirado-Rives, J. The OPLS Potential Functions for Proteins. Energy Minimizations for Crystals of Cyclic Peptides and Crambin. *J Am Chem Soc* **1988**, *110*, 1657–1666. https://doi.org/10.1021/JA00214A001/ASSET/JA00214A001.FP.PNG_V03.

[76] Davidchack, R. L.; Handel, R.; Tretyakov, M. V. Langevin Thermostat for Rigid Body Dynamics. *J Chem Phys* **2009**, *130*, 234101. https://doi.org/10.1063/1.3149788.

[77] Kräutler, V.; Van Gunsteren, W. F.; Hünenberger, P. H. A Fast SHAKE Algorithm to Solve Distance Constraint Equations for Small Molecules in Molecular Dynamics Simulations. *J Comput Chem* **2001**, *22*, 501–508. https://doi.org/10.1002/1096-987X.

[78] Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J Chem Phys* **1998**, *103*, 8577. https://doi.org/10.1063/1.470117.

[79] Faller, R.; De Pablo, J. J. Constant Pressure Hybrid Molecular Dynamics–Monte Carlo Simulations. *J Chem Phys* **2001**, *116*, 55. https://doi.org/10.1063/1.1420460.

[80] Harvey, M. J.; Giupponi, G.; De Fabritiis, G. ACEMD: Accelerating Biomolecular Dynamics in the Microsecond Time Scale. *J Chem Theory Comput* **2009**, *5*, 1632–1639. https://doi.org/10.1021/ct9000685.

[81] Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L.-P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D.; Wiewiora, R. P.; Brooks, B. R.; Pande, V. S. OpenMM 7: Rapid Development of High Performance Algorithms for Molecular Dynamics. *PLoS Comput Biol* **2017**, *13*, e1005659. https://doi.org/10.1371/journal.pcbi.1005659.

[82] Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. MDAnalysis: A Toolkit for the Analysis of Molecular Dynamics Simulations. *J Comput Chem* **2011**, *32*, 2319–2327. https://doi.org/10.1002/JCC.21787.

[83] Gowers, R. J.; Linke, M.; Barnoud, J.; Reddy, T. J. E.; Melo, M. N.; Seyler, S. L.; Domański, J.; Dotson, D. L.; Buchoux, S.; Kenney, I. M.; Beckstein, O. MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. *Proceedings of the 15th Python in Science Conference* **2016**, 98–105. https://doi.org/10.25080/MAJORA-629E541A-00E.

[84] Wójcikowski, M.; Zielenkiewicz, P.; Siedlecki, P. Open Drug Discovery Toolkit (ODDT): A New Open-Source Player in the Drug Discovery Field. *J Cheminform* **2015**, *7*, 26. https://doi.org/10.1186/s13321-015-0078-2.

[85] Pavan, M.; Menin, S.; Bassani, D.; Sturlese, M.; Moro, S. Implementing a Scoring Function Based on Interaction Fingerprint for Autogrow4: Protein Kinase CK1δ as a Case Study. *Front Mol Biosci* **2022**, *0*, 629. https://doi.org/10.3389/FMOLB.2022.909499.

[86] Phillips, J. C.; Hardy, D. J.; Maia, J. D. C.; Stone, J. E.; Ribeiro, J. V.; Bernardi, R. C.; Buch, R.; Fiorin, G.; Hénin, J.; Jiang, W.; McGreevy, R.; Melo, M. C. R.; Radak, B. K.; Skeel, R. D.; Singharoy, A.; Wang, Y.; Roux, B.; Aksimentiev, A.; Luthey-Schulten, Z.; Kalé, L. V.; Schulten, K.; Chipot, C.; Tajkhorshid, E. Scalable Molecular Dynamics on CPU and GPU Architectures with NAMD. *J Chem Phys* **2020**, *153*, 044130. https://doi.org/10.1063/5.0014475.

[87] Knippschild, U.; Gocht, A.; Wolff, S.; Huber, N.; Löhler, J.; Stöter, M. The Casein Kinase 1 Family: Participation in Multiple Cellular Processes in Eukaryotes. *Cell Signal* **2005**, *17*, 675–689. https://doi.org/10.1016/j.cellsig.2004.12.011.

[88] Xu, P.; Ianes, C.; Gärtner, F.; Liu, C.; Burster, T.; Bakulev, V.; Rachidi, N.; Knippschild, U.; Bischof, J. Structure, Regulation, and (Patho-)Physiological Functions of the Stress-Induced Protein Kinase CK1 Delta (CSNK1D). *Gene* **2019**, *715*.

[89] Perez, D. I.; Gil, C.; Martinez, A. Protein Kinases CK1 and CK2 as New Targets for Neurodegenerative Diseases. *Med Res Rev* **2011**, *31*, 924–954. https://doi.org/10.1002/MED.20207.

[90] Cescon, E.; Bolcato, G.; Federico, S.; Bissaro, M.; Valentini, A.; Ferlin, M. G.; Spalluto, G.; Sturlese, M.; Moro, S. Scaffold Repurposing of In-House Chemical Library toward the Identification of New Casein Kinase 1 δ Inhibitors. *ACS Med Chem Lett* **2020**, *11*. https://doi.org/10.1021/acsmedchemlett.0c00028.

[91] Bolcato, G.; Cescon, E.; Pavan, M.; Bissaro, M.; Bassani, D.; Federico, S.; Spalluto, G.; Sturlese, M.; Moro, S. A Computational Workflow for the Identification of Novel Fragments Acting as Inhibitors of the Activity of Protein Kinase CK1δ. *Int J Mol Sci* **2021**, *22*, 9741. https://doi.org/10.3390/ijms22189741.

[92] Venerando, A.; Ruzzene, M.; Pinna, L. A. Casein Kinase: The Triple Meaning of a Misnomer. *Biochem J* **2014**, *460*, 141–156. https://doi.org/10.1042/BJ20140178.

[93] Meggio, F.; Pinna, L. A. One-Thousand-and-One Substrates of Protein Kinase CK2? *FASEB J* **2003**, *17*, 349–368. https://doi.org/10.1096/FJ.02-0473REV.

[94] Borgo, C.; D'Amore, C.; Sarno, S.; Salvi, M.; Ruzzene, M. Protein Kinase CK2: A Potential Therapeutic Target for Diverse Human Diseases. *Signal Transduction and Targeted Therapy 2021 6:1* **2021**, *6*, 1–20. https://doi.org/10.1038/s41392-021-00567-7.

[95] Atas, E.; Oberhuber, M.; Kenner, L. The Implications of PDK1–4 on Tumor Energy Metabolism, Aggressiveness and Therapy Resistance. *Front Oncol* **2020**, *10*, 2835. https://doi.org/10.3389/FONC.2020.583217/BIBTEX.

[96] Zhang, S.; Hulver, M. W.; McMillan, R. P.; Cline, M. A.; Gilbert, E. R. The Pivotal Role of Pyruvate Dehydrogenase Kinases in Metabolic Flexibility. *Nutr Metab (Lond)* **2014**, *11*, 1–9. https://doi.org/10.1186/1743-7075-11-10/FIGURES/2.

[97] Patel, M. S.; Korotchkina, L. G. Regulation of the Pyruvate Dehydrogenase Complex. *Biochem Soc Trans* **2006**, *34*, 217–222. https://doi.org/10.1042/BST20060217.

[98] Wang, C.; Horby, P. W.; Hayden, F. G.; Gao, G. F. A Novel Coronavirus Outbreak of Global Health Concern. *The Lancet* **2020**, *395*, 470–473. https://doi.org/10.1016/S0140-6736(20)30185-9.

[99] *COVID Live - Coronavirus Statistics - Worldometer*. https://www.worldometers.info/coronavirus/ (accessed 2022-10-06).

[100] Dai, W.; Zhang, B.; Jiang, X. M.; Su, H.; Li, J.; Zhao, Y.; Xie, X.; Jin, Z.; Peng, J.; Liu, F.; Li, C.; Li, Y.; Bai, F.; Wang, H.; Cheng, X.; Cen, X.; Hu, S.; Yang, X.; Wang, J.; Liu, X.; Xiao, G.; Jiang, H.; Rao, Z.; Zhang, L. K.; Xu, Y.; Yang, H.; Liu, H. Structure-Based Design of Antiviral Drug Candidates Targeting the SARS-CoV-2 Main Protease. *Science (1979)* **2020**, *368*, 1331–1335. https://doi.org/10.1126/SCIENCE.ABB4489/SUPPL_FILE/ABB4489_DAI_SM.PDF.

[101] Fornasier, E.; Macchia, M. L.; Giachin, G.; Sosic, A.; Pavan, M.; Sturlese, M.; Salata, C.; Moro, S.; Gatto, B.; Bellanda, M.; Battistutta, R. A New Inactive Conformation of SARS-CoV-2 Main Protease. *Acta Crystallogr D Struct Biol* **2022**, *78*, 363–378. https://doi.org/10.1107/S2059798322000948.

[102] Ullrich, S.; Nitsche, C. The SARS-CoV-2 Main Protease as Drug Target. *Bioorg Med Chem Lett* **2020**, *30*, 127377. https://doi.org/10.1016/J.BMCL.2020.127377.

[103] Consortium, T. C. M.; Achdout, H.; Aimon, A.; Bar-David, E.; Barr, H.; Ben-Shmuel, A.; Bennett, J.; Bilenko, V. A.; Bilenko, V. A.; Boby, M. L.; Borden, B.; Bowman, G. R.; Brun, J.; BVNBS, S.; Calmiano, M.; Carbery, A.; Carney, D.; Cattermole, E.; Chang, E.; Chernyshenko, E.; Chodera, J. D.; Clyde, A.; Coffland, J. E.; Cohen, G.; Cole, J.; Contini, A.; Cox, L.; Cvitkovic, M.; Dias, A.; Donckers, K.; Dotson, D. L.; Douangamath, A.; Duberstein, S.; Dudgeon, T.; Dunnett, L.; Eastman, P. K.; Erez, N.; Eyermann, C. J.; Fairhead, M.; Fate, G.; Fearon, D.; Fedorov, O.; Ferla, M.; Fernandes, R. S.; Ferrins, L.; Foster, R.; Foster, H.; Gabizon, R.; Garcia-Sastre, A.; Gawriljuk, V. O.; Gehrtz, P.; Gileadi, C.; Giroud, C.; Glass, W. G.; Glen, R.; Glinert, I.; Godoy, A. S.; Gorichko, M.; Gorrie-Stone, T.; Griffen, E. J.; Hart, S. H.; Heer, J.; Henry, M.; Hill, M.; Horrell, S.; Huliak, V. D.; Hurley, M. F. D.; Israely, T.; Jajack, A.; Jansen, J.; Jnoff, E.; Jochmans, D.; John, T.; Jonghe, S. de; Kantsadi, A. L.; Kenny, P. W.; Kiappes, J. L.; Kinakh, S. O.; Koekemoer, L.; Kovar, B.; Krojer, T.; Lee, A.; Lefker, B. A.; Levy, H.; Logvinenko, I. G.; London, N.; Lukacik, P.; Macdonald, H. B.; MacLean, B.; Malla, T. R.; Matviiuk, T.; McCorkindale, W.; McGovern, B. L.; Melamed, S.; Melnykov, K. P.; Michurin, O.; Mikolajek, H.; Milne, B. F.; Morris, A.; Morris, G. M.; Morwitzer, M. J.; Moustakas, D.; Nakamura, A. M.; Neto, J. B.; Neyts, J.; Nguyen, L.; Noske, G. D.; Oleinikovas, V.; Oliva, G.; Overheul, G. J.; Owen, D.; Pai, R.; Pan, J.; Paran, N.; Perry, B.; Pingle, M.; Pinjari, J.; Politi, B.; Powell, A.; Psenak, V.; Puni, R.; Rangel, V. L.; Reddi, R. N.; Reid, S. P.; Resnick, E.; Ripka, E. G.; Robinson, M. C.; Robinson, R. P.; Rodriguez-Guerra, J.; Rosales, R.; Rufa, D.; Saar, K.; Saikatendu, K. S.; Schofield, C.; Shafeev, M.; Shaikh, A.; Shi, J.; Shurrush, K.; Singh, S.; Sittner, A.; Skyner, R.; Smalley, A.; Smeets, B.; Smilova, M. D.; Solmesky, L. J.; Spencer, J.; Strain-Damerell, C.; Swamy, V.; Tamir, H.; Tennant, R.; Thompson, W.; Thompson, A.; Tomasio, S.; Tsurupa, I. S.;

Tumber, A.; Vakonakis, I.; Rij, R. P. van; Vangeel, L.; Varghese, F. S.; Vaschetto, M.; Vitner, E. B.; Voelz, V.; Volkamer, A.; Delft, F. von; Delft, A. von; Walsh, M.; Ward, W.; Weatherall, C.; Weiss, S.; White, K. M.; Wild, C. F.; Wittmann, M.; Wright, N.; Yahalom-Ronen, Y.; Zaidmann, D.; Zidane, H.; Zitzmann, N. Open Science Discovery of Oral Non-Covalent SARS-CoV-2 Main Protease Inhibitor Therapeutics. *bioRxiv* **2022**, 2020.10.29.339317. https://doi.org/10.1101/2020.10.29.339317.

[104] Luttens, A.; Gullberg, H.; Abdurakhmanov, E.; Vo, D. D.; Akaberi, D.; Talibov, V. O.; Nekhotiaeva, N.; Vangeel, L.; Jonghe, S. de; Jochmans, D.; Krambrich, J.; Tas, A.; Lundgren, B.; Gravenfors, Y.; Craig, A. J.; Atilaw, Y.; Sandström, A.; Moodie, L. W. K.; Lundkvist, Å.; Hemert, M. J. van; Neyts, J.; Lennerstrand, J.; Kihlberg, J.; Sandberg, K.; Danielson, U. H.; Carlsson, J. Ultralarge Virtual Screening Identifies SARS-CoV-2 Main Protease Inhibitors with Broad-Spectrum Activity against Coronaviruses. *J Am Chem Soc* **2022**, *144*, 2905–2920. https://doi.org/10.1021/JACS.1C08402.

[105] Ghahremanpour, M. M.; Tirado-Rives, J.; Deshmukh, M.; Ippolito, J. A.; Zhang, C. H.; Cabeza De Vaca, I.; Liosi, M. E.; Anderson, K. S.; Jorgensen, W. L. Identification of 14 Known Drugs as Inhibitors of the Main Protease of SARS-CoV-2. *ACS Med Chem Lett* **2020**, *11*, 2526–2533. https://doi.org/10.1021/ACSMEDCHEMLETT.0C00521/SUPPL_FILE/ML0C00521_SI_002.XLSX.

[106] Owen, D. R.; Allerton, C. M. N.; Anderson, A. S.; Aschenbrenner, L.; Avery, M.; Berritt, S.; Boras, B.; Cardin, R. D.; Carlo, A.; Coffman, K. J.; Dantonio, A.; Di, L.; Eng, H.; Ferre, R.; Gajiwala, K. S.; Gibson, S. A.; Greasley, S. E.; Hurst, B. L.; Kadar, E. P.; Kalgutkar, A. S.; Lee, J. C.; Lee, J.; Liu, W.; Mason, S. W.; Noell, S.; Novak, J. J.; Obach, R. S.; Ogilvie, K.; Patel, N. C.; Pettersson, M.; Rai, D. K.; Reese, M. R.; Sammons, M. F.; Sathish, J. G.; Singh, R. S. P.; Steppan, C. M.; Stewart, A. E.; Tuttle, J. B.; Updyke, L.; Verhoest, P. R.; Wei, L.; Yang, Q.; Zhu, Y. An Oral SARS-CoV-2 M pro Inhibitor Clinical Candidate for the Treatment of COVID-19. *Science (1979)* **2021**, *374*, 1586–1593. https://doi.org/10.1126/science.abl4784.

[107] Pavan, M.; Bolcato, G.; Bassani, D.; Sturlese, M.; Moro, S. Supervised Molecular Dynamics (SuMD) Insights into the Mechanism of Action of SARS-CoV-2 Main Protease Inhibitor PF-07321332. *J Enzyme Inhib Med Chem* **2021**, *36*, 1646–1650. https://doi.org/10.1080/14756366.2021.1954919.

[108] Cheng, S. C.; Chang, G. G.; Chou, C. Y. Mutation of Glu-166 Blocks the Substrate-Induced Dimerization of SARS Coronavirus Main Protease. *Biophys J* **2010**, *98*, 1327–1336. https://doi.org/10.1016/j.bpj.2009.12.4272.

[109] Proctor, E. A.; Yin, S.; Tropsha, A.; Dokholyan, N. V. Discrete Molecular Dynamics Distinguishes Nativelike Binding Poses from Decoys in Difficult Targets. *Biophys J* **2012**, *102*, 144–151. https://doi.org/10.1016/J.BPJ.2011.11.4008.

[110] Liu, K.; Kokubo, H. Exploring the Stability of Ligand Binding Modes to Proteins by Molecular Dynamics Simulations: A Cross-Docking Study. *J Chem Inf Model* **2017**, *57*, 2514–2522. https://doi.org/10.1021/ACS.JCIM.7B00412/SUPPL_FILE/CI7B00412_SI_001.PDF.

Investigating RNA-Protein Recognition Mechanisms through Supervised Molecular Dynamics (SuMD) Simulations

Matteo Pavan, **Davide Bassani**, Mattia Sturlese and Stefano Moro

M. Pavan, D. Bassani, M. Sturlese, and S. Moro, "Investigating RNA–protein recognition mechanisms through supervised molecular dynamics (SuMD) simulations," *NAR Genomics Bioinforma.*, vol. 4, no. 4, Oct. 2022, doi: 10.1093/nargab/lqac088.

Abstract

Ribonucleic acid (RNA) plays a key regulatory role within the cell, cooperating with proteins to control the genome expression and several biological processes. Due to its characteristic structural features, this polymer can mold itself into different three-dimensional structures able to recognize target biomolecules with high affinity and specificity, thereby attracting the interest of drug developers and medicinal chemists. One successful example of the exploitation of RNA's structural and functional peculiarities is represented by aptamers, a class of therapeutic and diagnostic tools that can recognize and tightly bind several pharmaceutically relevant targets, ranging from small molecules to proteins, making use of the available structural and conformational freedom to maximize the complementarity with their interacting counterparts. In this scientific work, we present the first application of Supervised Molecular Dynamics (SuMD), an enhanced sampling Molecular Dynamics-based method for the study of receptor-ligand association processes in the nanoseconds timescale, to the study of recognition pathways between RNA aptamers and proteins, elucidating the main advantages and limitations of the technique while discussing its possible role in the rational design of RNA-based therapeutics.

1. Introduction

According to the central dogma of molecular biology, ribonucleic acid (RNA) is considered the functional link between deoxyribonucleic acid (DNA), which is involved in the storage of genetic information, and proteins, which are the effectors of most pivotal cell functions[1]. Complementing this ancestral and simplistic description of the biological role of RNA, in recent times this polymer has been linked with a variety of regulatory activities within the cell, cooperating with proteins to finely tune the genome expression and other biological processes[2]. The consideration that the vast majority of human RNA is not translated into proteins[3] in conjunction with the fact that a large number of newly discovered non-coding RNAs are associated with various pathologies and illnesses[4][5], caused an increase in the popularity of RNA among the scientific community, both from a biological and a therapeutic perspective[6][7].

From a structural point of view, RNA exists mainly as a double-stranded molecule with a chain length that can span from a few tens of nucleobases, as in small hairpins[8], to few thousand nucleotides, as in long non-coding sequences[9]. Compared to DNA, the higher conformational freedom of ribonucleic acids implies that they can assume a wide variety of three-dimensional structures in solution[10][11], organizing themselves in functional domains specifically designed to recognize other nucleic acids[12], proteins[13], glycated derivatives[14] or small organic molecules[15].

From a functional perspective, successful exploitation of RNA's ability to mold itself into different three-dimensional structures able to recognize target biomolecules with high affinity and specificity is represented by aptamers[16] This class of single-stranded oligonucleotides that fold into defined and complex architectures including stems, loops, bulges, hairpins, pseudoknots, triplexes, or quadruplexes, can bind several molecular targets, such as proteins, small organic molecules, and ions, thus classifying as a useful tool both for a therapeutic and diagnostic purpose[17][18].

As is the case for other nucleic acids, interactions between aptamers and proteins are characterized by a complex network of van der Waals, hydrogen bond, stacking, and general non-polar interactions that define the complementarity of shape and electrostatic properties at the surface between the two interactors and determine the specificity of binding to a certain target[19][20]. To the present date, the rising amount of experimentally

solved three-dimensional RNA structures, especially concerning their complexes with both macro and small molecules, has led to an increased interest by the scientific community in the investigation of RNA structures at an atomistic level of detail, to apply Structure-Based Drug Design (SBDD) strategies for the rational design of novel therapeutic entities[21][22].

Among the tools that are routinely used, both in the academic and industrial environment, for investigating the structural determinants of biological entities' recognition, molecular docking is by far the most widely and successfully adopted[23]. Originally developed for predicting the interaction between small organic molecules and proteins[24], throughout the years this computational technique has been also applied to the investigation of protein-protein[25], protein-peptides[26], and antigen-antibodies[27] complexes with various degrees of success. Contrary to proteins, the application of molecular docking to the study of complexes involving nucleic acids has been so far very limited, and mainly applied to the prediction of the binding mode of small molecules[28]. Compared to the aforementioned established docking protocols, a smaller number of methods are available for the investigation of nucleic acids-protein complexes, due to some intrinsic structural peculiarities of nucleic acids, particularly in the case of RNA[29]. The first limitation is represented by the distinctive charge distribution that characterizes the RNA surface compared to that of proteins[30], the second one is portrayed by the neglected treatment of the role of solvent[31] and the third one is the lack or limited consideration of the structural flexibility and dynamicity of ribonucleic acids[32].

One possible approach to overcome the limitations of molecular docking is represented by molecular dynamics (MD) simulations. Despite the enhanced description of the binding event derived from the explicit treatment of solvent molecules and the consideration of both receptor and ligand flexibility, MD simulations are rarely carried out to investigate the whole binding event due to the long simulation times and computational effort required to sample these infrequent events and are therefore mostly exploited for the refinement of docking results[33].

To mitigate the time constraints of classic molecular dynamics, one possible strategy is the exploitation of enhanced sampling techniques that allow increasing the frequency of observing desired events, such as the receptor-ligand association[34]. Among the plethora of enhanced sampling protocols that have been developed throughout the years, Supervised Molecular Dynamics (SuMD) has proven to be particularly successful in

investigating ligand-receptor recognition pathways at an atomistic level of detail without applying any energetic bias to the system while contemporarily reducing by orders of magnitude the simulation time compared to classic MD simulations[35]. Particularly, from a ligand perspective, SuMD simulations have proven useful to work with a variety of molecular entities, ranging from small organic molecules (both fragment[36][37][38][39][40][41] and mature, lead-like, compounds[42][43][44]), to more complex chemical species such as macrocycles[45] and peptide ligands[46]. From a receptor point of view, instead, SuMD was successfully applied to the study of both soluble[47][48] and membrane[49][50][51] systems, including both protein[52][53] and nucleic[54] targets.

In this scientific work, SuMD simulations were applied for the first time to the study of the recognition process between RNA macromolecules and proteins, to extend the applicability domain of the methodology. Particularly, we decided to focus our attention on RNA aptamers, due to their relevance as diagnostic and therapeutic tools and due to the variety of their structural landscape.

Briefly, we present four different applications of the SuMD methodology to RNA aptamer-protein complexes. Three test cases, involving systems in which the three-dimensional structure of the complex is known and deposited in the Protein Data Bank[55], are used to validate the ability of the SuMD protocol to correctly reproduce the experimental data while giving additional useful information that goes beyond the final state of the recognition process. The fourth and final case, concerning instead a complex whose structure has not yet been experimentally determined, is presented to show a possible prospective application of the SuMD protocol, discussing at the same time the advantages and limitations of the technique other than its possible role in a typical drug discovery pipeline.

2. Materials and Methods

2.1. Hardware Overview

General molecular modeling operations, such as the preparation of RNA aptamer-protein complex structures, the system setup for molecular dynamics simulations, and trajectory analysis were performed on an 8 CPU Linux workstation equipped with an Intel Xeon E5-1620 3.50 GHz processor. All molecular dynamics simulations were performed on a GPU cluster composed of 20 NVIDIA devices ranging from GTX980 to Titan V.

2.2. Structures Preparation

The three-dimensional coordinates of the three RNA aptamer-protein complex used as control cases in this study were retrieved from the Protein Data Bank[56] (PDB ID: 3DD2, 4PDB, 5VOE) and prepared for subsequent simulations exploiting several modules from the Molecular Operating Environment (MOE) 2019.01 suite[57]. At first, structures were pre-processed through the “Structure Preparation” tool, to assign each residue with alternate conformations to the one with the highest occupancy, build missing loops through homology modeling, and correct inconsistencies between the primary sequence and the tertiary structure. For compatibility with each piece of software used in this work and for consistency with previous studies involving the SuMD approach, structures were manually edited to mutate each non-natural nucleic residue (e.g.: fluorinated nucleotides) to the corresponding natural alternative. Afterward, titratable residues were assigned to the most probable protonation state at pH = 7.40 exploiting the “Protonate3D” tool. Finally, each non-protein or non-nucleic residue was removed, and the nucleic ligand was moved away from the binding site at a distance of at least 30 Å from the nearest receptor atom, in order to explore the conformational degree of freedom of the ligand throughout the recognition process.

For the investigation of the recognition process between the SARS-CoV-2 Spike RNA aptamer, the structure of the Spike Receptor Binding Domain (RBD) was retrieved from the Protein Data Bank (accession code: 6M0J[58]). Concerning the RNA aptamer, the experimental structure was not available, therefore the primary sequence was obtained from the Supplementary Material of the original work from Valero et al.[59] and sequentially submitted to the NUPACK[60] and 3DRNA[61] webservers to retrieve the

predicted secondary and tertiary structure respectively. In both cases, default parameters were chosen, and the lowest energy structure coming from the 3DRNA webserver was used for both docking calculations and MD simulations. Every other preparation passage is identical to the ones executed for the three control cases.

2.3. System Setup and Equilibration Protocol

Nucleic-protein systems coming from the preliminary preparation stage were then further processed making use of both Visual Molecular Dynamics (VMD) 1.9.2[62] and several tools from the Ambertools14 suite[63]. Each protein or nucleic atom was parameterized according to the ff14SB force field with χ modification tuned for RNA (χ_{OL3})[64][65][66]. At first, each system was solvated in a cubic box of TIP3P[67] water molecules with a padding of 35 Å. Afterward, each solvated system was neutralized through the addition of an appropriate number of Na^+ and Cl^- counterions until a salt concentration of 0.154 M was reached. Before Molecular Dynamics (MD) simulations, each system was subjected to a 1500-step energy minimization phase with the conjugate-gradient algorithm.

Each minimized system underwent then a two stages equilibration protocol. The first stage consisted of 1 ns of simulation in the canonical ensemble (NVT), applying a 5 kcal mol^{-1} $Å^{-2}$ harmonic positional restraint on each protein and nucleic atom. The second stage consisted instead of a 2 ns simulation in the isothermal-isobaric ensemble (NPT), with the same restraints applied only to the backbone atoms of both the protein and the nucleic acid. For each MD simulation performed in this scientific work, an integration step of 2 fs was used, the temperature was kept at a constant value of 310 K through a Langevin thermostat[68], and the M-SHAKE algorithm[69] was used to constrain the length of bonds involving hydrogen atoms, the particle-mesh Ewald (PME)[70] method was exploited to compute electrostatic interactions using cubic spline interpolation and a 1 Å grid spacing, while a 9.0 Å cutoff was set for calculation of Lennard-Jones interactions. For simulations in the NPT ensemble, the pressure was kept at a constant value of 1 atm through a Monte Carlo barostat[71]. Finally, all MD simulations were run through the ACEMD 3[72] engine, which is based upon OpenMM 7[73].

2.4. Supervised Molecular Dynamics (SuMD) Simulations

Supervised Molecular Dynamics (SuMD) is a well-established enhanced-sampling molecular dynamics approach that has been successfully applied to the study of the recognition process between various molecular entities at an atomic level of details on the nanosecond timescale[35].

The main advantage of the SuMD approach compared to traditional molecular dynamics simulations is the improved ability to sample infrequent events such as molecular association processes, thus reducing the timescale of the simulation that is required to spontaneously observe a binding event from the microseconds range to a few nanoseconds.

In detail, this task is accomplished by performing a sequence of short, unbiased, MD simulations followed by an evaluation of the simulation progress by a tabu-like algorithm. In this case, each of these MD simulations, defined as the “SuMD-step”, is run in the canonical ensemble at a constant temperature of 310 K for 300 ps, but the length of the “SuMD-step” can vary and is chosen according to the system that is studied. At the end of each “SuMD-step”, the distance between the center of mass of the ligand and one of the user-defined binding sites is computed at each step of the simulation, and this data is then fitted into a linear function: if the slope of the resulting straight-line is negative, indicating that the ligand is approaching the binding site, the “SuMD-step” is considered productive and retained for the generation of the final trajectory, while the final state of the simulation is used as the starting point for the successive step. On the contrary, if the slope is positive, thereby indicating that the ligand is not approaching the binding site, the “SuMD-step” is considered not productive and therefore discarded: in this case, the step is repeated by randomly reassigning the particle velocities through the Langevin thermostat and retaining the final coordinates from the end of the previous “SuMD-step”. The supervision algorithm is switched off when the distance between the two centers of mass falls below a threshold value (10 Å, in this case): from that point on, the simulation proceeds for the other 10 ns of classic molecular dynamics, allowing the system to relax and reach the final state of the simulation without any external geometric biased imposed by the supervision. For each control case study, 10 SuMD simulations were collected: while every single one was visually inspected, only the best one according to the geometrical agreement with the reference (based on the RMSD between the ligand coordinates in the final step of the simulation and the ligand coordinate in the reference experimental structure after optimal superposition of

the protein backbone) was thoroughly analyzed and discussed in the manuscript. In the case where the structure of the complex was not available, the best replica was chosen based on the MMGBSA interaction energy instead. In the current implementation, the SuMD code is written in Python and exploits the Numpy and ProDy[74] modules to perform the aforementioned geometrical supervision throughout the simulation. A list of residues utilized to define the ligand and protein binding site for each case is provided in Table S4 (Supplementary Materials).

2.5. Trajectory Analysis

The SuMD trajectories were analyzed by making use of an-in house tool written in Python 3 which represents an evolution and customization of the original one which is described in the work of Salmaso et al.[46].

Initially, trajectories representing each single "SuMD-step" were merged into a single collective trajectory. Then, obtained trajectories were pre-processed by applying a stride and retaining one frame every 20 ps, superposing and aligning each frame on the protein backbone atoms of the first frame, and wrapping it into an image of the system simulated under periodic boundary conditions (PBC). Both geometric and energetic analyses were performed on the so-obtained SuMD trajectories.

Concerning the geometric properties of the system, regarding both the nucleic ligand and the protein receptor, the time-dependent evolution of both backbone RMSD and radius of gyration, a global and a time-dependent per-residue decomposition of the backbone RMSF were collected and reported in an aggregated panel. Furthermore, the geometric performance of the SuMD protocol in reproducing the experimental bound conformation of the ligand was evaluated by computing the ligand backbone RMSD compared to the experimental reference throughout the entire simulation. All these geometric analyses were performed making use of the appropriate functions of the MDAnalysis[75][76] Python library and plotted through the Matplotlib[77] module.

Regarding the energetic analysis, an estimation of the ligand-receptor interaction energy alongside the SuMD trajectory was obtained both through the MMGBSA protocol, as implemented in AMBER 14, and through the "NAMD Energy" plugin for VMD, which exploits the NAMD[78] package to retrieve an estimate of the interaction energy defined as the sum

of the van der Waals and electrostatic contribution calculated according to the user-defined force field (AMBER 14, in this case). The energy values were then plotted both as a function of the simulation time and of the RMSD to the reference pose, giving both a time-dependent and a geometry-dependent energetic profile of the trajectory.

Finally, a per-residue interaction energy decomposition analysis was carried out exploiting once again the "NAMD Energy" plugin for VMD: plots report a time-dependent per-residue decomposition of the interaction energy for both the receptor and the ligand and a bidimensional interaction energy matrix in which interacting residues on the ligand side are correlated with the corresponding interacting residues on the receptor side. For all these per-residue analyses, the 25 most frequently contacted residues throughout the trajectory are considered (25 for the receptor, as well as for the ligand), defining contacting residues as the ones that are at a maximum distance of 4.5 Å from the nearest atom of the counterpart, either the ligand or the receptor.

A movie representation of the trajectory alongside the dynamic evaluation of its geometric and energetic features is also provided by the same analysis tool, which exploits VMD for the visual rendering of the simulated system. For uniformity reasons, in each plot and video, residue numbering is related to the fasta sequence for the wild-type receptor, as retrieved from the UniProt database.

2.6. Docking

To evaluate SuMD's ability to reproduce the native conformation of the RNA aptamer-protein complex, we decided to compare its performance with the one of molecular docking. The program chosen to accomplish this task was HADDOCK[25] ("High Ambiguity Driven protein-protein DOCKing", version 2.4) since it has already been extensively used for dealing with protein-nucleic acid complexes[79] and it uses *a priori* information to steer the docking calculation in a similar way to how SuMD works.

Each one of the crystal structures used as control (the ones coming from PDB codes 3DD2, 4PDB, and 5VOE) was subjected to a docking run. For all these cases, the nucleic acid was treated as the ligand, while the protein was considered as the receptor. The binding site was defined based on residues at the contact surface in the crystallographic structures, both on the protein and nucleic acid sides.

Concerning the SARS-CoV-2 Spike RBD RNA aptamer, the selected protein residues were chosen instead based on the contact surface with human ACE2 in the structure 6M0J, while for the aptamer the residues were selected based on the information coming from the original paper by Valero et. al. The list of protein and aptamer residues used as input for each docking calculation is reported in Table S3 (Supplementary Material). All user-definable parameters for molecular docking were kept as default.

HADDOCK starts with a randomization stage, in which the docking partners are placed far in space from one another (about 150 Å) and randomly rotated around their centers of mass. The following step consists of a rigid body energy minimization, which is followed by the rigid-body docking of the ligand and the receptor, allowing to obtain 1000 complexes. The 200 best solutions in terms of intermolecular energies obtained at this stage are subjected to simulated annealing refinements. Both the intra- and inter-molecular energies are evaluated by HADDOCK using full electrostatic and van der Waals energy terms with an 8.5 Å distance cutoff using the OPLS[80] nonbonded parameters. The final complexes are then clustered based on the Fraction of Common Contacts[81] (FCCs) with a 0.6 cutoff similarity for clustering, and the clusters are then ranked for energetics.

2.7. RNA Aptamer-Protein Complexes Molecular Dynamics Simulations

To evaluate the dynamic behavior of aptamer-bound protein complexes, we performed several classic molecular dynamics simulations. At first, each system was subjected to a preparation step exploiting both AmberTools14 and VMD 1.9.2, as previously mentioned in the preparation stage for SuMD simulations. Specifically, each protein-nucleic acid complex was singularly solvated in an explicit TIP3P water box with a 40 Å padding. Each of these simulation boxes was then neutralized using Na^+/Cl^- ions until reaching a physiological salt concentration of 0.154 M. The preparation phase was followed by a two-step equilibration protocol. The first equilibration was carried out in the canonical ensemble (NVT) and was composed of 1500 steps of energy minimization with a conjugate-gradient algorithm followed by a 1 ns MD simulation. In this first passage, harmonic positional restraints of 5 kcal mol^{-1} $Å^{-2}$ were applied on both the protein and the nucleic acid, while the temperature was kept at the constant value of 310 K exploiting a Langevin thermostat (with friction coefficient set to 0.1 ps^{-1}). The second equilibration was performed in the isothermal-

isobaric ensemble (NPT), also this time for 1 ns of MD simulation in which the harmonic positional restraints of 5 kcal mol^{-1} $Å^{-2}$ were applied just on the protein and nucleic acid backbones. The pressure was kept constant at the value of 1.0 atm using a Monte Carlo barostat. In each of the equilibration steps, a 2 ps integration step was adopted, the bonds involving the hydrogen atoms were constrained through the M-SHAKE algorithm, and a 9.0 Å cutoff was used for the calculation of the Lennard-Jones interaction. For the electrostatic interaction, a particle-mesh Ewald method (PME) was used. After this preparation phase, three different 50 ns replicates of classic MD simulation in the NVT ensemble at 310 K were executed.

2.8. Free RNA-Aptamer Molecular Dynamics Simulations

To complement the investigation of the structural dynamicity of investigated RNA aptamer-protein complexes, we also performed a classic MD simulation of the free RNA aptamer. To accomplish this task, we retrieved the coordinates for each nucleic acid molecule from the previously mentioned complexes. Each of these aptamers was then subjected to the same protocol described before for RNA aptamer-protein complexes, except for the parts related to the protein which, in this case, was not part of the system.

3. Results

To assess the applicability domain and accuracy of Supervised Molecular Dynamics simulation in the context of the nucleic acids-protein recognition processes, we opted for a retrospective validation approach, evaluating the ability of the protocol to correctly reproduce the binding mode of nucleic ligands found in experimentally solved complex structures, focusing both on the sampling and ranking capabilities of the protocol. Particularly, we decided to focus our attention on the class of RNA aptamers, both for their therapeutical relevance and for their challenging nature due to their peculiar structural features such as intrinsically higher flexibility and density of negative charge compared to ligands considered in the past applications of the SuMD protocol. In the following paragraphs, we present the application of SuMD to three different case studies for which the experimental structure of the RNA aptamer-protein complex is available on the Protein Data Bank, focusing on the geometrical accuracy of the technique in reproducing the experimentally determined binding mode and monitoring both the geometric and energetic features of the recognition process, stretching beyond the final state of the simulation. The three test cases are reported in chronological order, starting from the oldest structure to the most recent one. Furthermore, we also present a prospective application of the SuMD protocol to the investigation of a complex whose structure has not yet been experimentally determined, to present and discuss the role, the advantages, and the limitations of implementing the SuMD protocol in a pipeline for the rational design of RNA-based therapeutics. Information about each SuMD simulation reported in the manuscript are encompassed in Table S2 (Supplementary Material).

3.1. RNA aptamer bound to human thrombin (PDB ID: 3DD2)

Due to its ability to process several proteins that are part of the coagulation cascade, including the cleavage of soluble fibrinogen into fibrin, which is responsible for the formation of clots, human thrombin is a serine protease that exerts a pivotal role in blood coagulation and is, therefore, a target of interest for anticoagulation therapy[82][83][84]. Two surface regions of thrombin (exosite-1 and exosite-2), which are located on opposite sides of the molecule and away from the catalytic site, are responsible for its ability to interact with various macromolecular substrates. Particularly, exosite-2 is responsible for

the binding of thrombin to heparin, a clinically used oligosaccharide that mediates its anticoagulant effect by facilitating the interaction of thrombin with its endogenous inhibitor antithrombin[85].

In 2001, White et al. reported the discovery of Toggle-25, an RNA aptamer that was developed to bind with a high affinity to both human and porcine thrombin, leading to the inhibition of both plasma clot formation and platelet activation[86]. In 2008, Long et. al. were able to solve the crystal structure of Toggle-25t, a 25 nucleotide truncated version of Toggle-25, bound to the exosite-2 of human thrombin at a resolution of 1.90 Å (PDB ID: 3DD2[87]), thereby allowing a structural characterization of the complex which nicely complements previous biochemical analysis[87]. As reported in the original publication, the aptamer recognizes the thrombin exosite-2 in its native state, since no conformational changes can be observed between the unbound and bound state. Despite a relatively simple secondary structure, defined by a stem-loop with an internal bulge, Toggle-25t can achieve a selective and high-affinity binding to human thrombin (K_d = 0.54±0.1 nM) thanks to the good complementarity of shape and electrostatic properties between the negatively charged aptamer and the basic protein region responsible for its recognition. The absence of significant structural alteration of the protein upon binding, the therapeutic relevance of the target, and the relatively modest size of the aptamer (25 residues), make this complex an ideal target for the application of the SuMD protocol to the study of RNA aptamer-protein interactions. As previously introduced, two simplifications were introduced in the system investigated through SuMD: firstly, each 2' fluoro substituted pyrimidine residue was retro-mutated to the correspondent naturally occurring nucleotide, secondly, the divalent Mg^{2+} ion was not included in the system. The choice of retro mutating the fluorine-containing nucleotides was done for compatibility reasons since some of the software used in this scientific work (e.g., HADDOCK) could not work with non-natural nucleic residues. According to the original publication by Long et al.[87], the introduction of fluorine mainly impacted the aptamer resistance to ribonuclease rather than the binding affinity, as also underlined by the fact that only one single 2' fluoro group is in direct contact with the protein, specifically at the level of the U17-Arg126 interaction. Concerning the presence of Mg^{2+} ions, despite their undisputable importance in the field of RNA folding, we opted not to include them at all in the simulations due to some intrinsic molecular mechanics limitations that hamper the possibility of fully and accurately describing their interaction with RNA[88],

other than the difficulties in accurately predicting their locations without hints from experimental data[89].

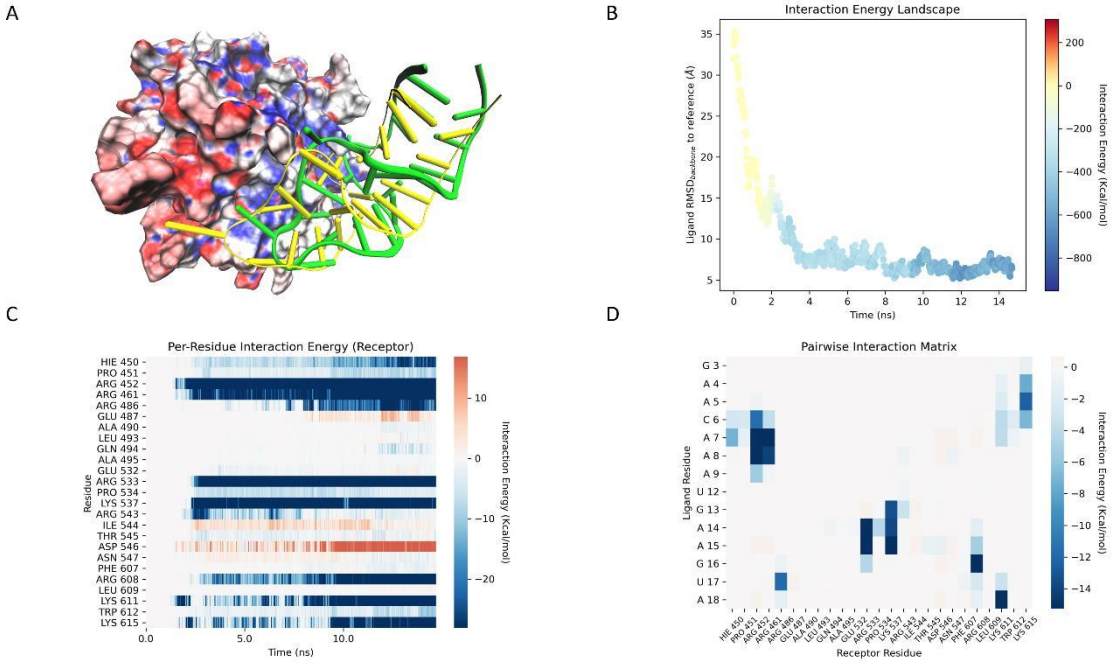


Figure 1. *This panel encompasses the putative recognition pathway between the RNA aptamer Toggle-25t and the exosite-2 of human thrombin described by the best trajectory obtained through the SuMD protocol (the one with the lowest RMSD to the crystal reference). (A) Visual representation of the Toggle-25t conformation sampled in the last frame of the SuMD trajectory (green) superposed with the native Toggle-25t conformation (yellow), as found in the crystal structure deposited in the Protein Data Bank with accession code 3DD2 (*$RMSD_{SuMD\text{-}Crystal}$*: 6.41 Å). The aptamer is represented as a ribbon, while the protein is represented as a Connolly surface colored according to the electrostatic potential as calculated with the APBS software[90], where red indicates a negatively charged area while blue indicates a positively charged one. (B) Profile of the ligand-receptor interaction energy (defined as the sum of the electrostatic and van der Waals contribution) throughout the recognition process as a function of both the simulation time and the RMSD between the ligand position during the trajectory and the ligand position in the crystal. (C) Receptor per-residue decomposition of the receptor-ligand interaction energy throughout the SuMD trajectory as a function of the simulation time: the 25 most-contacted residues are reported in the plot. (D) Per-residue interaction energy matrix: the 25 most-contacted residues for both the receptor and the ligand are considered, while each square composing the heatmap represents the average value of the interaction energy between the two paired residues alongside the trajectory.*

As can be seen in Video 1 (Supplementary Material), about 15 ns of simulation time was sufficient to sample a putative molecular recognition event between the Toggle-25t RNA aptamer and the human thrombin exosite-2. This is a quite remarkable result, considering the usual hundreds of nanoseconds that are necessary to spontaneously sample a binding event with classic MD simulations. As illustrated by Video 1 and summarized in Figure 1, the final state of the SuMD simulation converged quite well both from a geometrical and an interactive point of view toward the crystal reference.

Concerning the geometric accuracy, the RMSD between the ligand backbone conformation in the final state of the simulation and the native ligand backbone conformation observed in the crystal structure is 6.41 Å, which is quite impressive considering the intrinsic structural flexibility of these objects, which is also confirmed by the structural deviation than can be observed in classic MD simulations of both the crystal complex (average RMSD of the nucleic acid backbone in the final step of the simulation across three MD replicates: 2.87 Å) and the free aptamer (average RMSD of the nucleic acid backbone in the final step of the simulation across three MD replicates: 3.85 Å). It is not surprising, therefore, that SuMD performs worse compared to molecular docking from a geometrical point of view (RMSD between the best docking pose and the crystal binding mode of the aptamer: 2.51 Å).

Despite a lower geometrical accuracy of the method compared to the one of docking, that derives mainly from the high intrinsic flexibility of the nucleic ligand during the simulation (see also Figure S4, Supplementary Material, which reports the pairwise RMSD matrix of the free RNA aptamer during classic MD simulations), SuMD can correctly pose the negatively charged RNA-aptamer in a native-like conformation that maximizes the complementarity of shape and electrostatic features with the electropositive concave surface of the thrombin exosite-2, as highlighted in Figure 1 (panel A).

Concerning the capability of the protocol to correctly recapture the pivotal binding features despite a suboptimal geometric accuracy, SuMD can accurately describe the main interaction determinants, as illustrated by a comparison between the per-residue energy decomposition of the first 300 ps of the crystal complex classic MD simulation and the last 300 ps of the SuMD simulation (Figure S5, Supplementary Material). Particularly, as can be seen in Figure 1 (panels C and D), SuMD correctly captures the pivotal role of both Arg 101 (461) and Arg 233 (608) in driving the recognition mechanism, serving as electrostatic recruiters in the initial phases of the process and acting as anchors to stabilize the bound state in the final part of the simulation, in agreement with mutagenesis data that assess how mutation of each of this two residue completely abrogates any aptamer's effect[91]. Particularly, Arg 233 (608) along with Arg 165 (533) are responsible for the formation of a stacked interaction domain motif defined as an "A-Arg zipper" which involves five unpaired adenine residues on the ligand side, A4, A5, A7, A15, and A18 respectively. As can be seen in Figure 1 panel D, SuMD individuates all these five adenine residues as key interaction determinants, in agreement with the experimental data. Finally, SuMD also discriminates

the non-relevance of Gln 239 (614), which is not reported in the analysis as it is not one of the 25 most contacted residues during the trajectory, coherently with mutagenesis data that shows how a mutation of this residue does not affect the aptamer binding[91].

On the ligand side, SuMD once again correctly encompasses the different roles portrayed by different nucleotides. As can be seen in Figure 1 (panel D), the interaction between U17 and Arg 126 (486) is retrieved by the SuMD simulation: mutagenesis data shows how the mutation of U17, one of the flipped-out nucleotides, with adenine has essentially no effect on the aptamer's affinity for the target. Looking at our ligand-based interaction map it can be noticed how this interaction is the only one in which this nucleotide is involved, other than being less intense compared to more prominent interaction such as the aforementioned "A-Arg zipper", which suggest a non-pivotal contribution to the binding affinity. This can be related to the fact the U17 interacts with Arg126 through the backbone and not through its sidechain so that, as pointed out by the work of Jeter et al., this interaction would be maintained even when substituting the base[91]. Substitution of U12 with adenine results in a nearly three-orders-of-magnitude diminished binding affinity: this is due to the stabilizing role that U12 plays towards A15 through a non-Watson-Crick base pairing, one of the bases involved in the formation of the "A-Arg zipper"[91]. Once again, as depicted in Figure 1 panel D, SuMD correctly recognizes the pivotal role portrayed by A15, while contemporarily elucidating the indirect role of U12, which is not involved in any major interactions with protein residues. All other geometric and energetic analyses performed on the trajectory are summarized in Figures S1-S3 (Supplementary Materials).

To assess the predictive power of the method, we retrospectively analyzed all trajectories using two different metrics, i.e. the electrostatic interaction energy and the MMGBSA interaction energy of the final state of the simulation. The idea to use these two metrics stems from the consideration that RNA binding to proteins usually requires a good level of complementarity of steric and electrostatic properties at the binding interface. As reported in Figure S19, both metrics can successfully distinguish the native and native-like poses, i.e. the ones with a superimposable interaction pattern with the reference (measured through the Mean Signed Error and the Root-Mean Squared Error of the per-residue interaction energy decomposition, panels C and D respectively), from the decoys. This observation suggests that both metrics could be utilized in a prospective application of SuMD to rank

poses coming from different simulations prioritizing the ones that are most similar to the native binding mode.

3.2. RNA aptamer bound to Bacillus anthracis ribosomal protein S8 (PDB ID: 4PDB)

Thanks to its central function in the vital cycle of bacteria, the bacterial ribosome, complex machinery responsible for protein synthesis in prokaryotic organisms, has been extensively studied, both from a structural and a functional perspective, and has been validated as a target for multiple antibiotic drugs[92]. Protein-RNA interactions play a key role in the assembly, maturation, and function of the bacterial ribosome[93]. Among these, association processes involving ribosomal protein S8 are particularly relevant since it not only participates in the 30s subunit assembly by binding to 16S rRNA but, additionally, it also serves as a translational repressor of the *spc* operon mRNA, which encodes for 11 ribosomal proteins including S8 itself[94].

Due to its relevance, the complex formed between bacterial ribosomal protein S8 and 16S rRNA has been thoroughly characterized through different techniques, allowing us to establish that most of the protein–RNA contacts involve helices 21 and 25 and that a small RNA portion located in helix 21 is sufficient to confer specificity and high affinity to the S8-RNA interaction[95]. Furthermore, the interaction determinants between S8 and its RNA targets are largely conserved, and the same degree of conservation applies also to the overall fold of various S8 proteins[96][97][98]. Finally, the complementarity of shape and electrostatic properties that is required for the binding entails a high level of nucleotide sequence and secondary structure conservation, to impose an RNA shape that optimizes interaction properties with the protein surface[99].

To fetch RNA secondary structures that deviate from the conserved bacterial motif while retaining the ability to bind the S8 protein, in 2014 Davlieva et al. performed a SELEX experiment that led to the discovery of a 38-mer RNA aptamer that can bind the Bacillus anthracis S8 protein with high affinity (K_d = 110 ± 30 nM), determining at the same time the structure of the bound complex between the RNA aptamer and its protein target[99]. The selection process was based on an RNA stem-loop scaffold containing symmetric and asymmetric internal loops of 16 randomized nucleotides, with the resulting aptamer sequence forming a secondary structure with a symmetric internal loop[99]. The absence of

major structural rearrangements on the protein side of the interaction (0.65 Å $RMSD_{backbone}$ between the free and the bound form) and the relevance of this interaction from a mechanistic perspective led us to consider this complex as a suitable casestudy to validate the SuMD protocol.

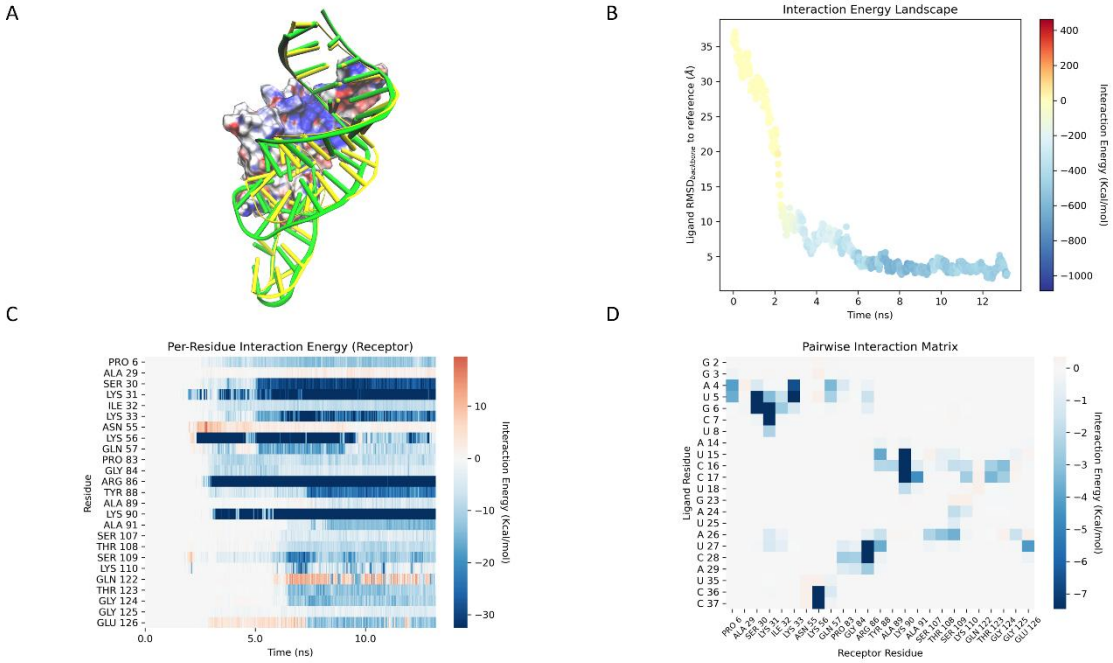


Figure 2. *This panel encompasses the putative recognition pathway between the RNA aptamer and the S8 ribosomal protein of Bacillus anthracis described by the best trajectory obtained through the SuMD protocol (the one with the lowest RMSD to the crystal reference). (A) Visual representation of the RNA aptamer conformation sampled in the last frame of the SuMD trajectory (green) superposed with the native RNA aptamer conformation (yellow), as found in the crystal structure deposited in the Protein Data Bank with accession code 4PDB ($RMSD_{SuMD\text{-}Crystal}$: 2.61 Å). The aptamer is represented as a ribbon, while the protein is represented as a Connolly surface colored according to the electrostatic potential as calculated with the APBS software[90], where red indicates a negatively charged area while blue indicates a positively charged one. (B) Profile of the ligand-receptor interaction energy (defined as the sum of the electrostatic and van der Waals contribution) throughout the recognition process as a function of both the simulation time and the RMSD between the ligand position during the trajectory and the ligand position in the crystal. (C) Receptor per-residue decomposition of the receptor-ligand interaction energy throughout the SuMD trajectory as a function of the simulation time: the 25 most-contacted residues are reported in the plot. (D) Per-residue interaction energy matrix: the 25 most-contacted residues for both the receptor and the ligand are considered, while each square composing the heatmap represents the average value of the interaction energy between the two paired residues alongside the trajectory.*

As can be deducted from Video 2 (Supplementary Materials), less than 15 ns of simulation time were needed to sample a putative recognition mechanism between the RNA aptamer and the ribosomal S8 protein. As underlined by both Video 2 and Figure 2, the final state of the SuMD simulation converged impressively well with the experimental data, both from a geometric and an interactive point of view.

Specifically, regarding the geometric accuracy of the method, the RMSD between the ligand backbone conformation in the final state of the simulation and the native ligand backbone conformation observed in the crystal structure is only 2.61 Å, a result comparable with the performance of molecular docking (RMSD between the best docking pose and the crystal binding mode of the aptamer: 2.20 Å). The higher geometric accuracy of the SuMD protocol, compared to 3DD2 case, can be partially explained by the lower degree of conformational freedom available to the aptamer, as can be seen in Figure S9 (Supplementary Materials, average RMSD of the nucleic acid backbone in the final step of the simulation across three MD replicates: 2.27 Å), despite similar stability of the bound state (average RMSD of the nucleic acid backbone in the final step of the simulation across three MD replicates: 2.80 Å).

Intriguingly, despite the impressive geometric convergence of the trajectory with the experimental data, the comparison between the per-residue energy decomposition from the last 300 ps of the SuMD simulation and the first 300 ps of the classic MD simulation of the crystal complex reveals a slightly lower congruence of the binding mode compared to the 3DD2 case (Figure S10, Supplementary Materials). It is important to notice that, in this case, the difference is not due to the interaction pattern, which for the most part is correctly depicted by the SuMD simulation analysis, but instead to the relative strength of the interactions. Indeed, as can be noticed in Figure 2 panel A, the final state of the SuMD simulation is slightly shifted compared to the crystal reference, considering that the predominant electrostatic component to the total interaction energy is proportional to the squared distance between the two interactors, even small differences in the relative position of interacting residues can alter the quantitative estimation of the interaction energy. Coherently with this interpretation of the data coming from the simulation analysis, our SuMD protocol can qualitatively describe the vast majority of the key interaction determinants.

In the crystal structure, the interaction between the S8 protein and the RNA aptamer involves a strip of electropositive charge (as is also visible in Video 2 and Figure 2, panel A) along which the phosphate backbone of the aptamer traverses from residues A4-C7 and U27-A29: this behavior is correctly captured by our simulation, as depicted in Figure 2 panel D. Moreover, as can be noticed in Figure 2 panels C and D, SuMD correctly intercepts the polar interactions that form between the backbone of aptamer residues C16, C17, A24, U25, U27, and A26 and their counterparts on the protein side such as Glu 126, Ser 107, Gly 124,

Lys 110, Ser 109, Ala 91 and Thr 123. Additionally, as depicted in Figure 2 panel D, the SuMD protocol is also able to spot some water-mediated interactions such as the contact between U27 to Glu 126. Finally, as can be depicted in Figure 2 panel D, SuMD can retrieve the stacked interaction between the peptide bond of highly conserved residues Ser 107 – Thr 108 – Ser 109 and the purine ring of A26: an analogous stacking interaction, involving A642, is present also in the complexes between the S8 protein and its natural RNA interactors, where it represents the only base-specific contacts of the complex[99]. All other geometric and energetic analyses performed on the trajectory are summarized in Figures S6-S8 (Supplementary Materials).

As for the previous case, we once again retrospectively analyzed all trajectories using the same metrics utilized before (electrostatic interaction energy and MMGBSA interaction energy), to assess the predictive power of the method. As reported in Figure S20, also in this case both metrics can successfully distinguish the native and native-like poses from the decoys. This observation further supports the idea that either of the two metrics could be utilized in a prospective application of SuMD to rank poses coming from different simulations to prioritize the most similar to the native binding mode.

3.3. RNA aptamer bound to human factor Xa (PDB ID: 5VOE)

One of the key events of the coagulation cascade is represented by the formation of the prothrombinase complex, macromolecular machinery formed by the serine protease factor Xa (FXa) and its cofactor factor Va (FVa)[100][101]: this membrane-mediated interaction enhances the catalytic activity of the FXa, leading to increased conversion of prothrombin into thrombin by a factor of approximately 10^5 and is, thereby, an interesting target for anticoagulation therapy[102][103]. Despite a great effort devoted to the development of both small-molecule inhibitors of the FXa catalytic site and peptide inhibitors directed at epitopes on the binding interface with FVa, both these approaches have led to disappointing therapeutic results since the interaction surface is large and involves multiple hotspots[104][105] and the inhibition at the catalytic site-level interferes with natural regulatory agents such as antithrombin III[106].

To avoid the limitations of traditional small molecule ligands, in 2010 Buddai et. al. developed RNA_{11f7t}, an RNA aptamer that exerts a potent anticoagulant effect by binding to

FXa with high affinity (Kd 1.1 ± 0.2 nM) and selectivity (~3000 fold over other coagulation proteases) and inhibiting its interaction with FVa[107].

In 2018, Gunaratne et al. were able to solve the crystal structure of 11F7t bound to FXa, allowing us to better comprehend the key structural features that characterize this interaction[108]. Specifically, in agreement with previous biochemical data, the analysis of the structure confirmed that the interaction occurs at a protein site which is implicated in the binding of both anticoagulant drug heparin and coagulation factor Factor Va (FVa)[109], with the interaction surface involving a central aptamer loop formed by residues C8, A10, A21, and C28-C30 and a protease exosite formed by Leu 59, Arg 64, Val 88, Ile 89, Asn 92, Arg 93, Lys 236, and Arg 240[108]. The absence of notable structural alterations of the protease upon aptamer binding, the therapeutical relevance of the target, and the relatively contained size of the aptamer (36 residues) induced us to consider it to validate the SuMD protocol. As in the case of structure 3DD2, 2' fluoro-modified nucleotides were retro-mutated to the corresponding natural alternatives, and the presence of the two Mg^{2+} ions was not considered in the simulations.

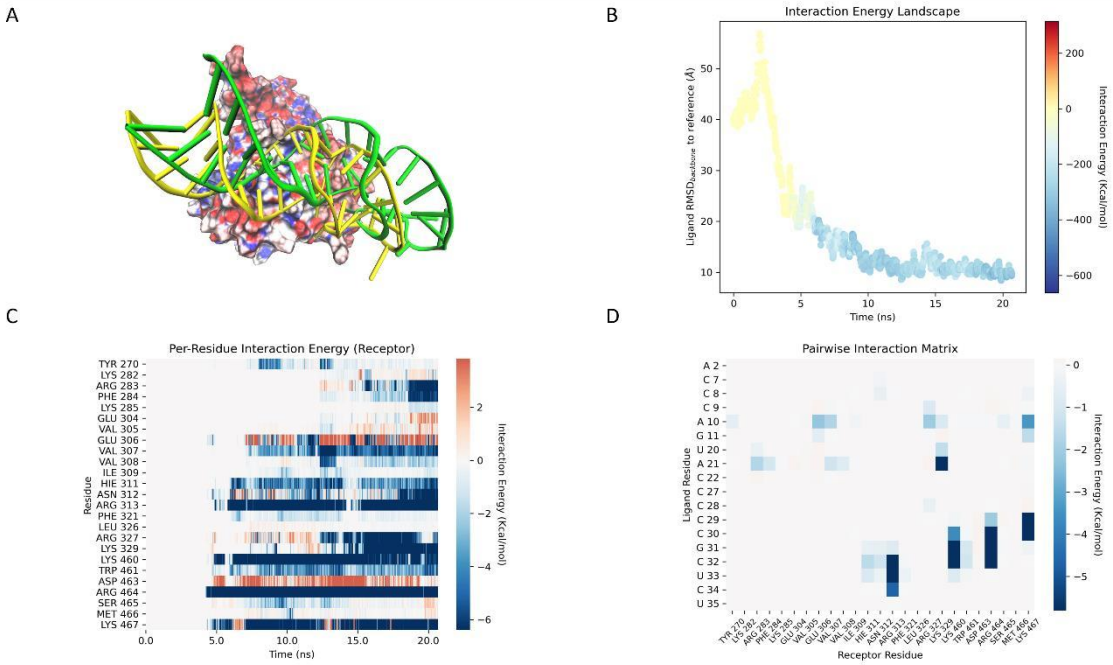


Figure 3. *This panel encompasses the putative recognition pathway between the 11F7t RNA aptamer and human factor Xa described by the best trajectory obtained through the SuMD protocol (the one with the lowest RMSD to the crystal reference). (A) Visual representation of the RNA aptamer conformation sampled in the last frame of the SuMD trajectory (green) superposed with the native RNA aptamer conformation (yellow), as found in the crystal structure deposited in the Protein Data Bank with accession code 5VOE (RMSD$_{SuMD\text{-}Crystal}$: 9.12 Å). The aptamer is represented as a ribbon, while the protein is represented as a Connolly surface colored according to the electrostatic potential as calculated with the APBS software[90], where red indicates a negatively charged area while blue indicates a positively charged one. (B) Profile of the ligand-receptor*

interaction energy (defined as the sum of the electrostatic and van der Waals contribution) throughout the recognition process as a function of both the simulation time and the RMSD between the ligand position during the trajectory and the ligand position in the crystal. (C) Receptor per-residue decomposition of the receptor-ligand interaction energy throughout the SuMD trajectory as a function of the simulation time: the 25 most-contacted residues are reported in the plot. (D) Per-residue interaction energy matrix: the 25 most-contacted residues for both the receptor and the ligand are considered, while each square composing the heatmap represents the average value of the interaction energy between the two paired residues alongside the trajectory.

As can be deemed by Video 3 (Supplementary Materials), in this case about 20 ns of simulation time were sufficient to sample a presumptive association pathway between the RNA aptamer 11F7t and human coagulation factor Xa. As can be noticed in Figure 3 (panel A), in this case, the geometric accuracy of the SuMD protocol in reproducing the crystal complex was worse compared to the first two cases, as also denoted by the 9.12 Å RMSD value between the final state of the simulation and the crystal reference. A first explanation of the lower geometric accuracy of the method can be found in the analysis of the classic MD simulation performed on both the crystal complex and the free aptamer: as highlighted by Figure S14 (Supplementary Materials) and by the RMSD of the nucleic backbone in the crystal complex MD (5.98 Å) and the free aptamer MD (4.93 Å), this aptamer has a significantly higher degree of conformational freedom compared to the previous two cases, which increase the difficulty of sampling the binding event due to the reduced amount of time that the aptamer spends in the binding competent conformation. Intriguingly, in this case, also, molecular docking compared worse than in the previous one, with the most correct pose having an RMSD of 3.17 Å to the crystal reference, indicating that the structural flexibility of the object (which is not considered by molecular docking) is not the only determinant of the performance of the protocol.

Contrary to the expectations, SuMD converged quite well from an interactive point of view with the experimental data: as can be seen in Figure S15 (Supplementary Materials), SuMD can qualitatively retrieve most of the native crystal interactions even if, as in the case of complex 4PDB, the estimation of the relative interaction strength is not always congruent.

Specifically, as can be noticed in Figure 3 panel D, the SuMD protocol can correctly retrieve the pivotal part played by residues A10, A21, and C29-C30, while slightly missing out on the importance of contacts with residues C8 and C28. On the protein side, as illustrated by Figure 3 panel C, SuMD precisely describes the central role portrayed by residues Arg 64 (283), Val 88 (308), Ile 89 (309), Asn 92 (312), Arg 93 (313), Lys 236 (460) and Arg 240 (464),

while only passing up on the interaction with Leu 59 (278). Interestingly, Arg 240 (464) and Lys 236 (460) are key residues for the binding of heparin, according to mutagenesis studies[110]. Finally, SuMD analysis highlights how Arg 165 (387) and Lys 169 (391), two critical residues in the recognition of factor FXa by either factor Va and/or prothrombin, are not contacted during the trajectory, in agreement with both the crystal structure and previous observations which pointed out to the possibility that the abrogation of factor Va binding happened through an indirect effect rather than through occlusion of the interaction surface[107][108]. All other geometric and energetic analyses performed on the trajectory are summarized in Figures S11-S13 (Supplementary Materials).

Finally, encouraged by the promising insights provided by the first two cases, we retrospectively analyzed all trajectories using the same scoring metrics defined before, to establish if they were once again able to distinguish the native and native-like poses from the decoys. As can be noticed in Figure S21, disappointingly both metrics fail to prioritize the most geometrically accurate solution, preferring instead the second-best one. Curiously, the reference crystal is also scored poorly by both metrics, indicating even if the SuMD protocol would have been able to sample it we would not have been able to prioritize it. As can be noticed in Table S1, this case also HADDOCK fails to rank the most geometrically accurate pose as the top solution, attributing to it a lower rank than a completely incorrect pose (RMSD: 19.79 Å), which ranks as the second best one. The observation that HADDOCK, despite incorporating information about native contacts in its scoring protocol, has trouble in correctly ranking poses for this case, combined with the intrinsic instability of the crystal complex, as indicated by the high RMSD value (5.98 Å) in the classic MD simulations and the low interaction energy values attributed by both metrics, indicate how this case might be an outlier, thus justifying the hypothesis of using the previously proposed metrics in prospective applications of the SuMD protocol.

3.4. RNA aptamer bound to SARS-CoV-2 Spike glycoprotein Receptor Binding Domain (RBD)

The outbreak of the COVID-19 pandemic in December 2019 caused an unprecedented worldwide public health crisis, leading to the death of more than six million people all over the world[111][112]. This illness is caused by the severe acute respiratory syndrome

coronavirus 2 (SARS-CoV-2), a betacoronavirus able to infect human cells by expressing a surface glycoprotein known as spike (S) glycoprotein, which interacts through its receptor-binding domain (RBD) with the human angiotensin-converting enzyme 2 (hACE2) that mediates the viral uptake process in conjunction with the associated transmembrane serine protease 2 (TMPRSS2)[113]. Due to the central role that this interaction plays in SARS-CoV-2 infectivity, the vast majority of therapeutic and prophylactic efforts in contrasting the COVID-19 pandemic have therefore been directed towards the inhibition of the spike-hACE2 interaction, either through vaccination or administration of monoclonal antibodies[114][115].

To overcome the main disadvantages of monoclonal antibodies, such as their high production costs, poor room temperature stability, and immunogenicity, in 2021 Valero et al. performed a SELEX experiment for the identification of a serum-stable RNA aptamer that could tightly bind the RBD of SARS-CoV-2 spike protein preventing the interaction with hACE2 thereby neutralizing viral infectivity[59]. This experiment led to the identification of RBD-PB6, an elongated stem-loop RNA aptamer that can selectively interact with the RBD with low nanomolar affinity ($K_D \approx 18$ nM), inhibiting the binding of RBD to hACE2 in a concentration-dependent manner[59].

Considering the encouraging results shown by the three applications of the SuMD protocol to the study of recognition processes between RNA aptamers and proteins and to illustrate a possible application of the protocol in a prospective scenario, we used SuMD simulations to shed light on the possible association pathway between the RBD-PB6 aptamer and the SARS-CoV-2 spike RBD. A model of the RNA aptamer structure was obtained through the 3dRNA webserver, based on the input primary sequence retrieved from the original publication and on the secondary structure prediction by the NUPACK webserver, while the structure of the SARS-CoV-2 spike RBD was retrieved from the crystal structure of the complex between the RBD and hACE2, deposited in the PDB with accession code 6M0J. The RBD structure that is present in this crystal comprises protein residues ranging from Thr 333 to Gly 526, slightly shorter than the construct used in the SELEX experiment which included residues from Arg 319 to Asn 532. However, residues that are not experimentally solved in the crystal structure are on the opposite side relative to the hACE2 interaction interface, so their absence should not impact the validity of the simulation. In agreement with experimental data that indicate how the RBD-PB6 RNA aptamer and hACE2 compete for the

same binding site on the RBD surface, SuMD simulations were carried out to sample a putative recognition mechanism between RBD-PB6 and RBD surface that is responsible for interaction with hACE2.

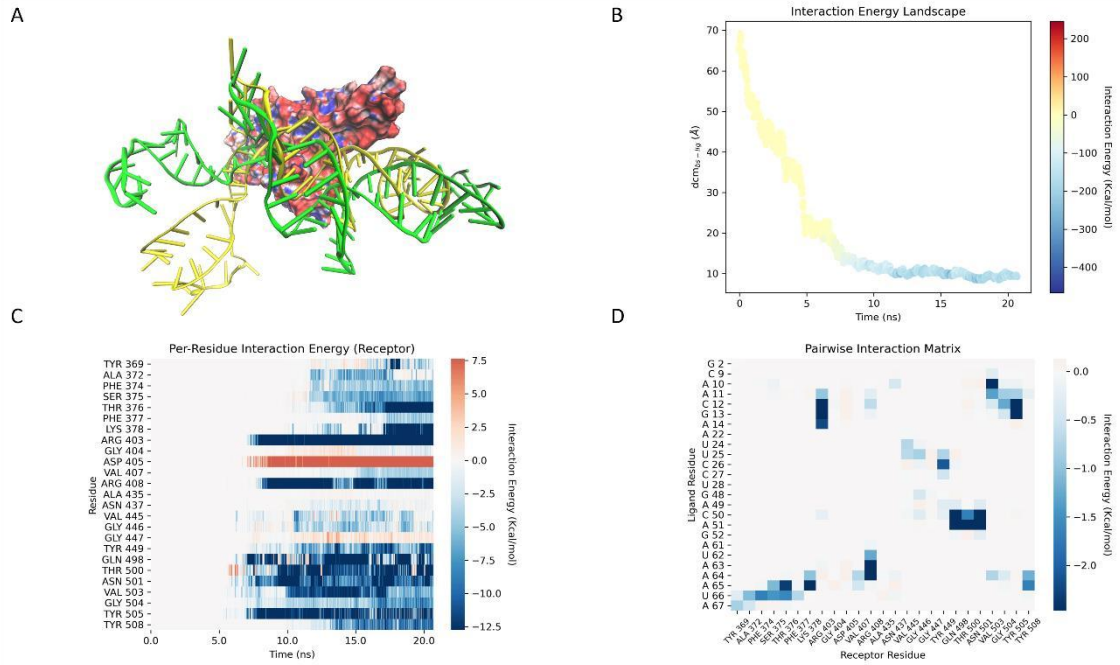


Figure 4. *This panel encompasses the putative recognition pathway between the RBD-PB6 RNA aptamer and the SARS-CoV-2 Spike RBD described by the best trajectory obtained through the SuMD protocol according to the MMGBSA interaction energy. (A) Visual representation of the RNA aptamer conformation sampled in the last frame of the SuMD trajectory (green) superposed with the best docking-predicted RNA aptamer conformation (yellow). The aptamer is represented as a ribbon, while the protein is represented as a Connolly surface colored according to the electrostatic potential as calculated with the APBS software[90], where red indicates a negatively charged area while blue indicates a positively charged one. (B) Profile of the ligand-receptor interaction energy (defined as the sum of the electrostatic and van der Waals contribution) throughout the recognition process as a function of both the simulation time and the RMSD between the ligand position during the trajectory and the ligand position in the crystal. (C) Receptor per-residue decomposition of the receptor-ligand interaction energy throughout the SuMD trajectory as a function of the simulation time: the 25 most-contacted residues are reported in the plot. (D) Per-residue interaction energy matrix: the 25 most-contacted residues for both the receptor and the ligand are considered, while each square composing the heatmap represents the average value of the interaction energy between the two paired residues alongside the trajectory.*

As can be observed in Video 4 (Supplementary Materials), about 20 ns of simulation time was enough to sample a putative recognition mechanism between the RBD-PB6 aptamer and the SARS-CoV-2 spike RBD. Interestingly, as can be noticed in Figure 4 panel A, there is a discrete level of convergence between the final state of the simulation and the docking-predicted binding mode, especially in the region ranging from A10 to A61, which is the minimal portion of the full-length aptamer which fully retains its binding capabilities to the RBD ($RMSD_{backbone\ SuMD\text{-}docking}$: 11.69 Å). Regarding the recognition mechanism proposed by

the SuMD protocol, an analysis of the interaction pattern reveals that the first contacts on the aptamer side involve residues A10-A14 and G48-G52, all of which fall into the conserved aptamer moiety that is required for binding to RBD. After these initial contacts take place and steer the binding event, other ancillary stabilizing interactions occur, such as the one with residues U24-U26 and the one involving residues U62-U66. On the protein side, instead, the main contacts involve polar and charged residues, such as Thr 376, Lys 378, Arg 403, Arg 408, Tyr 449, Gln 498, Thr 500, Asn 501, and Tyr 505. Interestingly, the most contacted residues during the trajectory do not include either Lys 417 or Glu 484, which are involved in the mutations K417N and E484K that characterize most viral variants with augmented infectivity compared to the wild-type virus. This evidence is in agreement with experimental data showing the RBD-PB6 affinity for the spike protein is practically unaffected by these mutations[59]. On the contrary, SuMD simulations indicate Asn 501 (involved in the N501Y mutation) as one of the most important residues in the recognition process: contrary to other previously mentioned residues, Asn 501 is surrounded by other interacting residues that can be found both in the complex between hACE2 and RBD and in the final state of the SuMD simulation, such as Gln 498, Thr 500 and Tyr 505, which could justify the contained impact of this mutation on the binding affinity of the aptamer compared to the other two. Interestingly, the most recent viral variants of concern include mutations such as Q498R, Y505H, and D405N that increase the positive charge on the spike surface: based on the analysis of the interaction pattern predicted by SuMD (Figure 4, panel C), all these mutations should increase the affinity with the negatively charged RBD-PB6 RNA aptamer, not only justifying its affinity towards the alpha and beta variant of SARS-CoV-2 but also towards the one from omicron on[116][117]. All other geometric and energetic analyses performed on the trajectory are summarized in Figures S16-S18 (Supplementary Materials).

4. Discussion

In this scientific work, we presented the first-ever application of Supervised Molecular Dynamics (SuMD) to the study of recognition processes between RNA macromolecules and proteins. Specifically, we concentrated our efforts on the aptamer class, due to their relevance as both therapeutic and diagnostic tools.

Three different retrospective case studies, where the structure of the RNA aptamer-protein complex was available, were presented to validate the ability of the SuMD protocol to reproduce the experimental data. In all three cases, despite the intrinsic limitations derived from the relatively high conformational freedom of the ribonucleic ligands, SuMD was able to converge quite well, both from a geometric and interactive point of view, with the experimentally solved complex structures. The lower geometric accuracy did not impair the ability to retrieve most if not all the binding determinants of the native complex, with the increased RMSD compared to docking being related to the portion of the ligand not directly involved in the binding interface. Despite the increased complexity of the considered system compared to the usually investigated protein-small molecule complexes, the simulation times required to sample a putative recognition mechanism between the RNA aptamers and their protein target were comparable: in all presented cases, indeed, 10-20 ns of simulation time were sufficient to capture the entire association pathway, from the unbound state to the final complex. The reduced computational effort that the SuMD platform provides compared to classic, unsupervised, molecular dynamics simulation, makes it more suitable for its implementation in a drug discovery pipeline, flanking and complementing the role of already established approaches such as molecular docking. Due to the limited sampling capability of molecular dynamics-based methods compared to molecular docking, now the optimal strategy would be to combine the two techniques rather than using them in a mutually exclusive fashion: the rapidity of molecular docking could be useful to generate a series of a reasonable binding hypothesis that could be then more thoroughly investigated through SuMD simulations. The rapid increase in computational power available to scientists will hopefully make it possible to solve the sampling issue of MD-based methods, allowing them to fully replace more physically approximate methods such as molecular docking.

Concerning the applicability of the SuMD protocol in a prospective scenario, we also presented a case study where the experimental structure of the RNA aptamer-protein complex was not experimentally solved. Particularly, due to its therapeutic relevance, we decided to investigate the recognition process between the RBD-PB6 RNA aptamer, developed by Valero et al. in 2021[59], and the SARS-CoV-2 spike protein receptor-binding domain RBD. We showed how, even in the absence of an experimentally solved structure of the aptamer, the SuMD protocol can be coupled with structure prediction tools to give a structural prediction congruent with experimental evidence.

Regarding the ability of the SuMD protocol to be applied in a prospective scenario, one crucial point regards the capability to discriminate between the native-like binding mode and decoys. In the three presented cases, the simulation that was carefully analyzed and discussed in the manuscript was the one that presented the best geometric agreement with the crystal reference but such a metric could not be applied to a prospective investigation of complexes whose structures have not already been solved. By retroactively analyzing the geometric and energetic profile of the generated trajectories, we noticed that the electrostatic component of the interaction energy plays a fundamental role in steering the association process. Particularly, as can be noticed in Figures S19-S21 (Supplementary Materials), both the MMGBSA interaction energy and the electrostatic component on its own can discriminate and prioritize the native complex and native-like poses from decoys in two out of three case studies. The only exception to this rule is represented by the complex between the RNA aptamer 11F7t and human factor Xa (PDB ID: 5VOE), for which both the electrostatic and MMGBSA scoring metrics indicate the second most geometrically accurate pose as the one with the most favorable energetic profile. Interestingly, in this case, molecular docking also fails to prioritize the most geometrically accurate pose, ranking it as the third-best one (Table S1, Supplementary Materials). The indication that docking and SuMD, despite using different scoring metrics, failed to prioritize the native-like conformation suggests that there is still room for improvement regarding the scoring of complexes involving nucleic acids. However, despite this, the use of MMGBSA and or electrostatic interaction energy as scoring metrics can still be relatively accurate in suggesting reasonable binding mode hypotheses that are congruent with experimental evidence, as also previously pointed out by a benchmark study by Chen et al. from 2018[118]. Due to these considerations, we opted for using MMGBSA as a scoring metric for

our prospective study of the interaction between the RNA aptamer RBD-PB6 and the SARS-CoV-2 spike RBD.

The last aspect that is worth addressing is related to the choice of the residues to consider for the supervision of the association process throughout the SuMD trajectory. As is the case for molecular docking, where the binding site is defined by the user through a sphere or a box wrapped around the area of interest, SuMD also requires the user to specify a residue selection on both the receptor and ligand sides that are used to compute the distance between the center of mass of the binding site and of the ligand that is fed to the supervision algorithm. The choice of residues is usually based on prior knowledge of the interaction site derived from experimental evidence, but there could be some cases where this choice is not obvious. A first possible solution to this problem is represented by the analysis of the electrostatic potential of the receptor surface: as can be seen in Video 1-4, the recognition between the RNA aptamers and their protein targets usually involves a high level of complementarity of electrostatic properties, with the negatively charged ribonucleic surface being nicely harbored by positively charged patches on the protein side. A second possible solution is to perform a docking calculation to have a first indication of the preferable binding mode of the object, followed by a more extensive characterization of the binding mode through SuMD simulations. This solution was used in the context of this article for the study of the interaction between the RBD-PB6 RNA aptamer and the SARS-CoV-2 spike RBD.

The possibility to investigate different binding sites and binding mode hypotheses can also be viewed as a strong point of the SuMD technique: for example, in the case of complex ribosomal protein S8, two different RNA recognition sites are available for the aptamer, specifically the site involved in the interaction with helix 21 and the site that mediates interaction with helix 25[99]. SuMD simulations would allow investigating both possibilities at the same time, elucidating the mechanistic details that determine the preferential recognition of the primary binding site thereby helping the rational development of selective binders.

Furthermore, concerning the exploration of different binding hypotheses, SuMD would allow deciphering the possibility of alternative stoichiometries. For example, in the case of aptamer 11F7t, Buddai et al. noticed a peculiar and difficult to rationalize binding stoichiometry, other than a strong Ca^{2+} dependence of the interaction[107]. The authors

discussed various possibilities, including a possible effect on the protein and/or aptamer structure, but also a possible calcium-induced aptamer dimerization[107]. In this case, the exploitation of the SuMD technique would have allowed the exploration of all these different hypotheses, which could not be investigated through static, time-independent techniques such as molecular docking.

References

[1] Gilbert, W. Origin of life: The RNA world. *Nature 1986 319:6055* **319**, 618–618 (1986).

[2] Breaker, R. R. & Joyce, G. F. The Expanding View of RNA and DNA Function. *Chem Biol* **21**, 1059–1065 (2014).

[3] Hangauer, M. J., Vaughn, I. W. & McManus, M. T. Pervasive Transcription of the Human Genome Produces Thousands of Previously Unidentified Long Intergenic Noncoding RNAs. *PLoS Genet* **9**, (2013).

[4] Cheetham, S. W., Gruhl, F., Mattick, J. S. & Dinger, M. E. Long noncoding RNAs and the genetics of cancer. *Br J Cancer* **108**, 2419–2425 (2013).

[5] Esteller, M. Non-coding RNAs in human disease. *Nat Rev Genet* **12**, 861–874 (2011).

[6] Morris, K. v. & Mattick, J. S. The rise of regulatory RNA. *Nature Reviews Genetics 2014 15:6* **15**, 423–437 (2014).

[7] Connelly, C. M., Moon, M. H. & Schneekloth, J. S. The Emerging Role of RNA as a Therapeutic Target for Small Molecules. *Cell Chem Biol* **23**, 1077–1090 (2016).

[8] Bartel, D. P. MicroRNAs: Genomics, Biogenesis, Mechanism, and Function. *Cell* **116**, 281–297 (2004).

[9] Ponting, C. P., Oliver, P. L. & Reik, W. Evolution and Functions of Long Noncoding RNAs. *Cell* **136**, 629–641 (2009).

[10] Salmon, L., Yang, S. & Al-Hashimi, H. M. Advances in the Determination of Nucleic Acid Conformational Ensembles. *http://dx.doi.org/10.1146/annurev-physchem-040412-110059* **65**, 293–316 (2014).

[11] Cruz, J. A. & Westhof, E. The dynamic landscapes of RNA architecture. *Cell* **136**, 604–609 (2009).

[12] Bartel, D. P. MicroRNAs: Target Recognition and Regulatory Functions. *Cell* **136**, 215–233 (2009).

[13] Draper, D. E. Protein-RNA recognition. *Annu Rev Biochem* **64**, 593–620 (1995).

[14] Lorger, M., Engstler, M., Homann, M. & Göringer, H. U. Targeting the variable surface of African trypanosomes with variant surface glycoprotein-specific, serum-stable RNA aptamers. *Eukaryot Cell* **2**, 84–94 (2003).

[15] Disney, M. D. Targeting RNA with Small Molecules To Capture Opportunities at the Intersection of Chemistry, Biology, and Medicine. *J Am Chem Soc* **141**, 6776–6790 (2019).

[16] Stoltenburg, R., Reinemann, C. & Strehlitz, B. SELEX—A (r)evolutionary method to generate high-affinity nucleic acid ligands. *Biomol Eng* **24**, 381–403 (2007).

[17] Keefe, A. D., Pai, S. & Ellington, A. Aptamers as therapeutics. *Nature Reviews Drug Discovery 2010 9:7* **9**, 537–550 (2010).

[18] Jayasena, S. D. Aptamers: An Emerging Class of Molecules That Rival Antibodies in Diagnostics. *Clin Chem* **45**, 1628–1650 (1999).

[19] Jones, S., Daley, D. T. A., Luscombe, N. M., Berman, H. M. & Thornton, J. M. Protein–RNA interactions: a structural analysis. *Nucleic Acids Res* **29**, 943–954 (2001).

[20] Morozova, N., Myers, J. & Shamoo, Y. Protein–RNA interactions: exploring binding patterns with a three-dimensional superposition analysis of high resolution structures. *Bioinformatics* **22**, 2746–2752 (2006).

[21] Reynolds, A. *et al.* Rational siRNA design for RNA interference. *Nat Biotechnol* **22**, 326–330 (2004).

[22] Boese, Q. *et al.* Mechanistic insights aid computational short interfering RNA design. *Methods Enzymol* **392**, 73–96 (2005).

[23] Meng, X.-Y., Zhang, H.-X., Mezei, M. & Cui, M. Molecular Docking: A Powerful Approach for Structure-Based Drug Discovery. *Current Computer Aided-Drug Design* **7**, 146–157 (2012).

[24] Kuntz, I. D., Blaney, J. M., Oatley, S. J., Langridge, R. & Ferrin, T. E. A geometric approach to macromolecule-ligand interactions. *J Mol Biol* **161**, 269–288 (1982).

[25] Dominguez, C., Boelens, R. & Bonvin, A. M. J. J. HADDOCK: A protein-protein docking approach based on biochemical or biophysical information. *J Am Chem Soc* **125**, 1731–1737 (2003).

[26] Ciemny, M. *et al.* Protein–peptide docking: opportunities and challenges. *Drug Discov Today* **23**, 1530–1537 (2018).

[27] Pedotti, M., Simonelli, L., Livoti, E. & Varani, L. Computational Docking of Antibody-Antigen Complexes, Opportunities and Pitfalls Illustrated by Influenza Hemagglutinin. *Int J Mol Sci* **12**, 226 (2011).

[28] David Morley, S. & Afshar, M. Validation of an empirical RNA-ligand scoring function for fast flexible docking using RiboDock. *J Comput Aided Mol Des* **18**, 189–208 (2004).

[29] Nithin, C., Ghosh, P. & Bujnicki, J. M. Bioinformatics Tools and Benchmarks for Computational Docking and 3D Structure Prediction of RNA-Protein Complexes. *Genes 2018, Vol. 9, Page 432* **9**, 432 (2018).

[30] Disney, M. D. Targeting RNA with Small Molecules To Capture Opportunities at the Intersection of Chemistry, Biology, and Medicine. *J Am Chem Soc* **141**, 6776 (2019).

[31] Fulle, S. & Gohlke, H. Molecular recognition of RNA: challenges for modelling interactions and plasticity. *J Mol Recognit* **23**, 220–231 (2010).

[32] Hermann, T. Rational ligand design for RNA: the role of static structure and conformational flexibility in target recognition. *Biochimie* **84**, 869–875 (2002).

[33] De Vivo, M., Masetti, M., Bottegoni, G. & Cavalli, A. Role of Molecular Dynamics and Related Methods in Drug Discovery. *J Med Chem* **59**, 4035–4061 (2016).

[34] Bernardi, R. C., Melo, M. C. R. & Schulten, K. Enhanced sampling techniques in molecular dynamics simulations of biological systems. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1850**, 872–877 (2015).

[35] Sabbadin, D. & Moro, S. Supervised molecular dynamics (SuMD) as a helpful tool to depict GPCR-ligand recognition pathway in a nanosecond time scale. *J Chem Inf Model* **54**, 372–376 (2014).

[36] Deganutti, G., Moro, S., Ciruela, F. & Sotelo, E. Supporting the Identification of Novel Fragment-Based Positive Allosteric Modulators Using a Supervised Molecular Dynamics Approach: A Retrospective Analysis Considering the Human A2A Adenosine Receptor as a Key Example. *Molecules 2017, Vol. 22, Page 818* **22**, 818 (2017).

[37] Bolcato, G., Bissaro, M., Sturlese, M. & Moro, S. Comparing Fragment Binding Poses Prediction Using HSP90 as a Key Study: When Bound Water Makes the Difference. *Molecules 2020, Vol. 25, Page 4651* **25**, 4651 (2020).

[38] Ferrari, F. *et al.* HT-SuMD: Making Molecular Dynamics Simulations Suitable for Fragment-Based Screening. a Comparative Study with NMR. (2020) doi:10.26434/CHEMRXIV.12582662.V1.

[39] Bissaro, M. *et al.* Inspecting the Mechanism of Fragment Hits Binding on SARS-CoV-2 M ^{pro} by Using Supervised Molecular Dynamics (SuMD) Simulations. *ChemMedChem* **16**, 2075–2081 (2021).

[40] Bolcato, G. *et al.* A Computational Workflow for the Identification of Novel Fragments Acting as Inhibitors of the Activity of Protein Kinase CK1δ. *Int J Mol Sci* **22**, 9741 (2021).

[41] Grosjean, H. *et al.* SAMPL7 protein-ligand challenge: A community-wide evaluation of computational methods against fragment screening and pose-prediction. *Journal of Computer-Aided Molecular Design 2022* 1–21 (2022) doi:10.1007/S10822-022-00452-7.

[42] Bolcato, G., Bissaro, M., Pavan, M., Sturlese, M. & Moro, S. Targeting the coronavirus SARS-CoV-2: computational insights into the mechanism of action of the protease inhibitors lopinavir, ritonavir and nelfinavir. *Sci Rep* **10**, 20927 (2020).

[43] Pavan, M., Bolcato, G., Bassani, D., Sturlese, M. & Moro, S. Supervised Molecular Dynamics (SuMD) Insights into the mechanism of action of SARS-CoV-2 main protease inhibitor PF-07321332. *J Enzyme Inhib Med Chem* **36**, 1646–1650 (2021).

[44] Bolcato, G., Pavan, M., Bassani, D., Sturlese, M. & Moro, S. Ribose and Non-Ribose A2A Adenosine Receptor Agonists: Do They Share the Same Receptor Recognition Mechanism? *Biomedicines 2022, Vol. 10, Page 515* **10**, 515 (2022).

[45] Hassankalhori, M., Bolcato, G., Bissaro, M., Sturlese, M. & Moro, S. Shedding Light on the Molecular Recognition of Sub-Kilodalton Macrocyclic Peptides on Thrombin by Supervised Molecular Dynamics. *Front Mol Biosci* **8**, 730 (2021).

[46] Salmaso, V., Sturlese, M., Cuzzolin, A. & Moro, S. Exploring Protein-Peptide Recognition Pathways Using a Supervised Molecular Dynamics Approach. *Structure* **25**, 655-662.e2 (2017).

[47] Bissaro, M. *et al.* Targeting Protein Kinase CK1δ with Riluzole: Could It Be One of the Possible Missing Bricks to Interpret Its Effect in the Treatment of ALS from a Molecular Point of View? *ChemMedChem* **13**, 2601–2605 (2018).

[48] Panday, S. K., Sturlese, M., Salmaso, V., Ghosh, I. & Moro, S. Coupling Supervised Molecular Dynamics (SuMD) with Entropy Estimations to Shine Light on the Stability of Multiple Binding Sites. *ACS Med Chem Lett* **10**, 444–449 (2019).

[49] Deganutti, G., Cuzzolin, A., Ciancetta, A. & Moro, S. Understanding allosteric interactions in G protein-coupled receptors using Supervised Molecular Dynamics: A prototype study analysing the human A3 adenosine receptor positive allosteric modulator LUF6000. *Bioorg Med Chem* **23**, 4065–4071 (2015).

[50] Paoletta, S. *et al.* Modeling ligand recognition at the P2Y12 receptor in light of X-ray structural information. *J Comput Aided Mol Des* **29**, 737–756 (2015).

[51] Palazzotti, D. *et al.* Deciphering the Molecular Recognition Mechanism of Multidrug Resistance Staphylococcus aureus NorA Efflux Pump Using a Supervised Molecular Dynamics

Approach. *International Journal of Molecular Sciences 2019, Vol. 20, Page 4041* **20**, 4041 (2019).

[52] Cuzzolin, A. *et al.* Deciphering the Complexity of Ligand-Protein Recognition Pathways Using Supervised Molecular Dynamics (SuMD) Simulations. *J Chem Inf Model* **56**, 687–705 (2016).

[53] Deganutti, G., Moro, S. & Reynolds, C. A. A Supervised Molecular Dynamics Approach to Unbiased Ligand-Protein Unbinding. *ACS Appl Mater Interfaces* **2020**, 1804–1817 (2020).

[54] Bissaro, M., Sturlese, M. & Moro, S. Exploring the RNA-Recognition Mechanism Using Supervised Molecular Dynamics (SuMD) Simulations: Toward a Rational Design for Ribonucleic-Targeting Molecules? *Front Chem* **8**, 107 (2020).

[55] Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res* **28**, 235 (2000).

[56] Berman, H. M. The Protein Data Bank. *Nucleic Acids Res* **28**, 235–242 (2000).

[57] Molecular Operating Environment (MOE), 2019.01; Chemical Computing Group ULC, 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2021. https://www.chemcomp.com/Research-Citing_MOE.htm.

[58] Lan, J. *et al.* Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature 2020 581:7807* **581**, 215–220 (2020).

[59] Valero, J. *et al.* A serum-stable RNA aptamer specific for SARS-CoV-2 neutralizes viral entry. *Proc Natl Acad Sci U S A* **118**, (2021).

[60] Zadeh, J. N. *et al.* NUPACK: Analysis and design of nucleic acid systems. *J Comput Chem* **32**, 170–173 (2011).

[61] Wang, J., Wang, J., Huang, Y. & Xiao, Y. 3dRNA v2.0: An Updated Web Server for RNA 3D Structure Prediction. *Int J Mol Sci* **20**, (2019).

[62] Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *J Mol Graph* **14**, 33–38 (1996).

[63] Case, D. A. *et al.* The Amber biomolecular simulation programs. *J Comput Chem* **26**, 1668–1688 (2005).

[64] Maier, J. A. *et al.* ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J Chem Theory Comput* **11**, 3696–3713 (2015).

[65] Pérez, A. *et al.* Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of α/γ Conformers. *Biophys J* **92**, 3817–3829 (2007).

[66] Zgarbová, M. *et al.* Refinement of the Cornell et al. Nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles. *J Chem Theory Comput* **7**, 2886–2902 (2011).

[67] Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J Chem Phys* **79**, 926–935 (1983).

[68] Davidchack, R. L., Handel, R. & Tretyakov, M. v. Langevin thermostat for rigid body dynamics. *J Chem Phys* **130**, 234101 (2009).

[69] Kräutler, V., van Gunsteren, W. F. & Hünenberger, P. H. A Fast SHAKE Algorithm to Solve Distance Constraint Equations for Small Molecules in Molecular Dynamics Simulations. *J Comput Chem* **22**, 501–508 (2001).

[70] Essmann, U. *et al.* A smooth particle mesh Ewald method. *J Chem Phys* **103**, 8577 (1998).

[71] Faller, R. & de Pablo, J. J. Constant pressure hybrid Molecular Dynamics–Monte Carlo simulations. *J Chem Phys* **116**, 55 (2001).

[72] Harvey, M. J., Giupponi, G. & de Fabritiis, G. ACEMD: Accelerating biomolecular dynamics in the microsecond time scale. *J Chem Theory Comput* **5**, 1632–1639 (2009).

[73] Eastman, P. *et al.* OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Comput Biol* **13**, e1005659 (2017).

[74] Bakan, A., Meireles, L. M. & Bahar, I. ProDy: Protein dynamics inferred from theory and experiments. *Bioinformatics* **27**, 1575–1577 (2011).

[75] Michaud-Agrawal, N., Denning, E. J., Woolf, T. B. & Beckstein, O. MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *J Comput Chem* **32**, 2319–2327 (2011).

[76] Gowers, R. J. *et al.* MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. *Proceedings of the 15th Python in Science Conference* 98–105 (2016) doi:10.25080/MAJORA-629E541A-00E.

[77] Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Comput Sci Eng* **9**, 90–95 (2007).

[78] Phillips, J. C. *et al.* Scalable molecular dynamics on CPU and GPU architectures with NAMD. *J Chem Phys* **153**, 044130 (2020).

[79] van Zundert, G. C. P. *et al.* The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J Mol Biol* **428**, 720–725 (2016).

[80] Jorgensen, W. L. & Tirado-Rives, J. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J Am Chem Soc* **110**, 1657–1666 (1988).

[81] Rodrigues, J. P. G. L. M. *et al.* Clustering biomolecular complexes by residue contacts similarity. *Proteins: Structure, Function, and Bioinformatics* n/a-n/a (2012) doi:10.1002/prot.24078.

[82] Stubbs, M. T. & Bode, W. The clot thickens: clues provided by thrombin structure. *Trends Biochem Sci* **20**, 23–28 (1995).

[83] Hoffman, M. & Monroe, D. M. A cell-based model of hemostasis. *Thromb Haemost* **85**, 958–965 (2001).

[84] di Cera, E. Thrombin Interactions. *Chest* **124**, 11S-17S (2003).

[85] Rau, J. C., Beaulieu, L. M., Huntington, J. A. & Church, F. C. Serpins in thrombosis, hemostasis and fibrinolysis. *Journal of Thrombosis and Haemostasis* **5**, 102–115 (2007).

[86] White, R. *et al.* Generation of Species Cross-reactive Aptamers Using "Toggle" SELEX. *Molecular Therapy* **4**, 567–573 (2001).

[87] Long, S. B., Long, M. B., White, R. R. & Sullenger, B. A. Crystal structure of an RNA aptamer bound to thrombin. *RNA* **14**, 2504–2512 (2008).

[88] Vangaveti, S., Ranganathan, S. v. & Chen, A. A. Advances in RNA molecular dynamics: a simulator's guide to RNA force fields. *Wiley Interdiscip Rev RNA* **8**, e1396 (2017).

[89] Giambaşu, G. M., Case, D. A. & York, D. M. Predicting Site-Binding Modes of Ions and Water to Nucleic Acids Using Molecular Solvation Theory. *J Am Chem Soc* (2019) doi:10.1021/JACS.8B11474/ASSET/IMAGES/MEDIUM/JA-2018-11474K_M001.GIF.

[90] Jurrus, E. *et al.* Improvements to the <scp>APBS</scp> biomolecular solvation software suite. *Protein Science* **27**, 112–128 (2018).

[91] Jeter, M. L. *et al.* RNA aptamer to thrombin binds anion-binding exosite-2 and alters protease inhibition by heparin-binding serpins. *FEBS Lett* **568**, 10–14 (2004).

[92] Poehlsgaard, J. & Douthwaite, S. The bacterial ribosome as a target for antibiotics. *Nature Reviews Microbiology 2005 3:11* **3**, 870–881 (2005).

[93] Nierhaus, K. H. The assembly of the prokaryotic ribosome. *Biosystems* **12**, 273–282 (1980).

[94] Nomura, M., Yates, J. L., Dean, D. & Post, L. E. Feedback regulation of ribosomal protein gene expression in Escherichia coli: Structural homology of ribosomal RNA and ribosomal protein mRNA. *Proc Natl Acad Sci U S A* **77**, 7084–7088 (1980).

[95] Wu, H., Jiang, L. & Zimmermann, R. A. The binding site for ribosomal protein S8 in 16S rRNA and spc mRNA from Escherichia coli: minimum structural requirements and the effects of single bulged bases on S8-RNA interaction. *Nucleic Acids Res* **22**, 1687 (1994).

[96] Merianos, H. J., Wang, J. & Moore, P. B. The structure of a ribosomal protein S8/spc operon mRNA complex. *RNA* **10**, 954–964 (2004).

[97] Tishchenko, S. *et al.* Detailed analysis of RNA-protein interactions within the ribosomal protein S8-rRNA complex from the archaeon Methanococcus jannaschii. *J Mol Biol* **311**, 311–324 (2001).

[98] Brodersen, D. E., Clemons, W. M., Carter, A. P., Wimberly, B. T. & Ramakrishnan, V. Crystal structure of the 30 S ribosomal subunit from Thermus thermophilus: structure of the proteins and their interactions with 16 S RNA. *J Mol Biol* **316**, 725–768 (2002).

[99] Davlieva, M., Donarski, J., Wang, J., Shamoo, Y. & Nikonowicz, E. P. Structure analysis of free and bound states of an RNA aptamer against ribosomal protein S8 from Bacillus anthracis. *Nucleic Acids Res* **42**, 10795–10808 (2014).

[100] Mann, K. G., Nesheim, M. E., Church, W. R., Haley, P. & Krishnaswamy, S. Surface-dependent reactions of the vitamin K-dependent enzyme complexes. *Blood* **76**, 1–16 (1990).

[101] Mann, K. G., Jenny, R. J. & Krishnaswamy, S. COFACTOR PROTEINS IN THE ASSEMBLY AND EXPRESSION OF BLOOD CLOTTING ENZYME COMPLEXES. *https://doi.org/10.1146/annurev.bi.57.070188.004411* **57**, 915–956 (2003).

[102] Gross, P. L. & Weitz, J. I. New anticoagulants for treatment of venous thromboembolism. *Arterioscler Thromb Vasc Biol* **28**, 380–386 (2008).

[103] Bauer, K. A. New anticoagulants. *Curr Opin Hematol* **15**, 509–515 (2008).

[104] Krishnaswamy, S. Prothrombinase complex assembly. Contributions of protein-protein and protein-membrane interactions toward complex formation. *Journal of Biological Chemistry* **265**, 3708–3718 (1990).

[105] Krishnaswamy, S. Exosite-driven substrate specificity and function in coagulation. *Journal of Thrombosis and Haemostasis* **3**, 54–67 (2005).

[106] Björk, I. & Olson, S. T. Antithrombin. *Adv Exp Med Biol* **425**, 17–33 (1997).

[107] Buddai, S. K. *et al.* An Anticoagulant RNA Aptamer That Inhibits Proteinase-Cofactor Interactions within Prothrombinase *. *Journal of Biological Chemistry* **285**, 5212–5223 (2010).

[108] Gunaratne, R. *et al.* Combination of aptamer and drug for reversible anticoagulation in cardiopulmonary bypass. *Nature Biotechnology 2018 36:7* **36**, 606–613 (2018).

[109] Rezaie, A. R. Identification of Basic Residues in the Heparin-binding Exosite of Factor Xa Critical for Heparin and Factor Va Binding *. *Journal of Biological Chemistry* **275**, 3320–3327 (2000).

[110] Rezaie, A. R. Identification of Basic Residues in the Heparin-binding Exosite of Factor Xa Critical for Heparin and Factor Va Binding. *Journal of Biological Chemistry* **275**, 3320–3327 (2000).

[111] Guarner, J. Three Emerging Coronaviruses in Two Decades: The Story of SARS, MERS, and Now COVID-19. *American Journal of Clinical Pathology* vol. 153 420–421 Preprint at https://doi.org/10.1093/ajcp/aqaa029 (2020).

[112] COVID Live - Coronavirus Statistics - Worldometer. https://www.worldometers.info/coronavirus/.

[113] Hoffmann, M. *et al.* SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* **181**, 271-280.e8 (2020).

[114] Thanh Le, T. *et al.* The COVID-19 vaccine development landscape. *Nat Rev Drug Discov* **19**, 305–306 (2020).

[115] Taylor, P. C. *et al.* Neutralizing monoclonal antibodies for treatment of COVID-19. *Nature Reviews Immunology 2021 21:6* **21**, 382–393 (2021).

[116] Sartore, G. *et al.* In silico evaluation of the interaction between ACE2 and SARS-CoV-2 Spike protein in a hyperglycemic environment. *Sci Rep* **11**, (2021).

[117] Bassani, D., Ragazzi, E., Lapolla, A., Sartore, G. & Moro, S. Omicron Variant of SARS-CoV-2 Virus: In Silico Evaluation of the Possible Impact on People Affected by Diabetes Mellitus. *Front Endocrinol (Lausanne)* **13**, (2022).

[118] Chen, F. *et al.* Assessing the performance of MM/PBSA and MM/GBSA methods. 8. Predicting binding free energies and poses of protein-RNA complexes. *RNA* **24**, 1183–1194 (2018).

CONCLUSION

In this Ph.D. work, we applied some of the most actual techniques in Computer-Aided Drug Design (CADD) to a wide variety of scenarios and case studies, every time analyzing their advantage and limitations, also providing useful improvements from a methodological perspective. Molecular Docking is for sure one of the most used techniques in the present work, and detailed reports of its performance have been reported for both SARS-CoV-2 main protease (M^{pro}) and GPCRs. Moreover, the consensus docking approach was beneficial in the drug discovery pipeline which led to novel low-micromolar fragments inhibiting Casein Kinase 1δ (CK1δ). This last target was also chosen as the case study for the application of a novel scoring function for *de novo* design based on the program Autogrow4. This scoring function, based on three-dimensional pharmacophore filtration and Protein-Ligand Interaction Fingerprint (PLIFs) similarity ranking, has been shown to virtually improve the results given by the algorithm in terms of the quality of the molecules produced. Experimental validation of the method will follow in the upcoming months.

Molecular Dynamics (MD) is another very powerful tool in the hands of computational drug designers, and in this Ph.D. this method has been used both with post-docking purposes for the prioritization of the already mentioned small CK1δ inhibitors and for stability assessment of the crystallographic binding modes of the non-covalent SARS-CoV-2 M^{pro} ligands. Both Molecular Docking and Molecular Dynamics have been also described among the main techniques which have been shown to be beneficial in the development of novel compounds for neuroinflammation treatment.

Remaining on the topic of Molecular Dynamics, we also reported the first validation of Thermal Titration Molecular Dynamics (TTMD) as a simple but powerful tool for the classification of ligands based on a correlation between potency and MD residence time. Further studies on this specific approach are now being carried out in our lab, enlarging the applicability domain of the method, and validating it in drug discovery pipelines.

On the other hand, Supervised Molecular Dynamics (SuMD) found an even more central role in the present work, demonstrating to be effective in describing and rationalizing several molecular scenarios. First, the technique was used to correctly predict the binding mode of Nirmatrelvir to SARS-CoV-2 M^{pro}, then we also attempted to evaluate its performance in the pose prediction of the small early fragments complexed with the same protein. Moreover, SuMD allowed exploring the differences between the recognition paths of ribose and non-ribose $A_{2A}R$ ligands at the molecular level. In this work, we also reported the first and successful application of SuMD for the description of the aptamer-protein recognition path.

CADD techniques are very proficient for the help that they can give in elucidating the molecular reasons beneath biological and physiological processes. In this Ph.D. work, molecular modeling was successfully exploited to propose a valid rationale for the different responses of diabetic patients to COVID-19 infection, both for WT and Omicron variants. Moreover, it also allowed analyzing the mutation patterns of the main SARS-CoV-2 lineages, bringing to the conclusion that the main protease remains valid (if not the most valid) molecular target for viral infection. Alanine scanning was also combined with classical molecular modeling to evaluate the basis for the resistance of different M^{pro} mutants to Nirmatrelvir. In addition to that, the importance of GPCRs in the development and progression of Amyotrophic Lateral Sclerosis (ALS) has been extensively described in a detailed review.