UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Università degli Studi di Padova

Dipartimento di Scienze Chimiche

CORSO DI DOTTORATO DI RICERCA IN SCIENZE
MOLECOLARI
INDIRIZZO SCIENZE FARMACEUTICHE
XXIX CICLO

LIGAND–RECEPTOR RECOGNITION EVENTS DECODED AT MOLECULAR SCALE BY MEANS OF MOLECULAR DYNAMICS SIMULATIONS

Coordinatore: Ch.mo Prof. Antonino Polimeno
Supervisore: Ch.mo Prof. Stefano Moro

Dottorando: Giuseppe Deganutti

..
ii

..
ii

TABLE OF CONTENTS

ABSTRACT

During the last decades, the technological evolution has been very fast and has paved the way to a wide set of theoretical approaches able to support and stimulate the experimental component of biological sciences. From this standpoint, in drug discovery and design, the ability to make working hypothesis on how a small molecule interacts with its biological target can lead to rational approaches for the developing of new drug candidates.

Nowadays is possible to model the behaviour of chemical systems up to the atomistic scale, allowing retrieve insights on mechanisms behinds the ligand binding and unbinding from a receptor. Among the computational techniques available, molecular dynamics is able to take in account fundamental aspects linked to the time evolution of a biological system, such as structural flexibility and the dynamic role of water molecules in the protein binding sites.

During this Ph.D. project we employed molecular dynamics in order to disclose putative binding mechanisms of several ligands: more precisely, we applied the supervised molecular dynamics (SuMD) technique to decipher the binding pathways of both allosteric modulators and agonists to the adenosine receptors subtypes (belonging to class A of G protein-coupled receptors). Interestingly, findings highlights the coexistence of different potential recognition pathways that anticipate the formation of the orthosteric intermolecular complexes, as well as the crucial role of residues located at the extracellular portion of the protein.

SOMMARIO

Nel corso degli ultimi decenni, l'evoluzione tecnologica è stata così rapida da aprire la strada a una vasta gamma di approcci teorici in grado di supportare e stimolare la componente sperimentale delle scienze biologiche. Da questo punto di vista, in drug design, la capacità di fornire ipotesi di lavoro su come una piccola molecola interagisce con il suo bersaglio biologico può portare ad approcci razionali per lo sviluppo di nuovi candidati farmaci.

Oggigiorno è possibile modellare il comportamento di sistemi chimici fino alla scala atomica, consentendo di avere informazioni sul meccanismo che guida associazione e dissociazione da un recettore. Tra le tecniche di calcolo disponibili, la dinamica molecolare è in grado di prendere in considerazione fondamentali aspetti legati all'evoluzione temporale di un sistema biologico, quali la flessibilità strutturale e il ruolo dinamico delle molecole d'acqua nei siti di legame proteici.

Durante questo progetto di dottorato di ricerca abbiamo impiegato la dinamica molecolare al fine di rivelare i possibili meccanismi di riconoscimento di diversi ligandi, soprattutto nei confronti dei sottotipi recettoriali dell'adenosina (appartenenti alla classe A dei recettori accoppiati a proteine G): più precisamente, abbiamo applicato tecniche di dinamica molecolare supervisionata (SuMD) sia a modulatori allosterici che ad agonisti. È interessante notare che i risultati evidenziano la co-esistenza di diversi possibili meccanismi di riconoscimento, che anticipano la formazione dei complessi intermolecolari ortosterici, oltre ad il ruolo cruciale di residui localizzati nella porzione proteica extracellulare.

1 INTRODUCTION

1.1 BINDING AND UNBINDING: AN OVERVIEW

Binding between organic molecules has ever been the motor of life since its outset[1]. Inside biological systems, intermolecular recognitions often are the first step for transduction of the chemical signals required for physiological functionalities: gene regulation[2], hormonal signal cascades[3] and enzymatic metabolism[4] are just a few examples of biological events, subjected to binding with endogenous or exogenous chemicals. Likewise for all the natural phenomena, thermodynamics laws are able to explain and describe the physics and the driving forces that stand behind the formation of intermolecular complexes, as well as the kinetics behaviour of the entire binding and unbinding phenomena.

In this first section some fundamental aspects of binding, both from kinetics and thermodynamics point of view are shortly introduced to the reader, then the main forces that drive and influence binding are briefly elucidated.

1.1.1 BINDING THERMODYNAMICS AND KINETICS: TWO FACES OF THE SAME COIN

Is difficult to divide the binding thermodynamics characterization (quantified by the equilibrium association constant K_A and the dissociation constant K_D) from its kinetic description (reflected by the rate constants k_{on} and k_{off}). Considering a single step mechanism, at the steady state the ligand (L) and its receptor (R) form an intermolecular complex according to the law of mass action:

$$L + R \rightleftarrows LR \qquad (1)$$

The dynamic proprieties of the equilibrium allow to relate the thermodynamics dissociation constant K_D (which in turn is the reciprocal of K_A) to the binding kinetics rate constants k_{on} and k_{off} (equation 2):

$$K_D = \frac{[L][R]}{[LR]} = \frac{k_{off}}{k_{on}} = K_A^{-1} \qquad (2)$$

The recognition event, which depends on the concentration of both ligand and receptor, is a second order reaction (k_{on} is expressed in $M^{-1}s^{-1}$), while the dissociation is function only of the intermolecular complex concentration (k_{off} is measured in unit of s^{-1}). This important difference is caused by the intrinsic two steps nature of intermolecular recognition[5]: indeed, while the dissociation take place in one stride regulated only by the interactions between ligand and receptor, the association can be distinguished in two different phases. During the first stage (before any interaction occurs) the ligand diffuses through the media and the probability of a productive collision with the protein rely on its own concentration (we can assume a kinetics rate constant k_{diff} for this step, measured in M^{-1}). Upon the first interactions with the receptor are established, recognition no more relies on concentration but only on the binding mechanism and its energetic progress, leading to measure a proper interaction kinetics rate constant (namely k_{int}) in unit of s^{-1}. It follows that the total binding kinetics rate k_{on} can be considered the combination of the two phases (equation 3):

$$k_{on} = k_{diff} \cdot k_{int} \qquad (3)$$

Generally, for a drug–like compound, kinetics k_{on} value is in the range of 10^3 $M^{-1}s^{-1}$–10^9 $M^{-1}s^{-1}$ (the latter is about the rate limit of free diffusion in solution), while the k_{off} can spread from around 10^{-7} s^{-1} to approximately 1 s^{-1} and usually is the main contributor to the K_D value[6] as expressed in equation 2. Interestingly, super–fast binders (characterized by k_{on} values higher than 10^9 M^{-1} s^{-1}) have been evolutionary selected as effectors of physiologic processes that

need instant regulation, like in the case of acetylcholine on the enzyme acetylcholinesterase (AChE)[7] in the central nervous system (CNS).

During the last decade, an increasing attention has been addressed to the kinetics concept of residence time (t_r), first introduced in 2006 [8] and defined as the reciprocal of the k_{off} value (so is expressed in unity of seconds):

$$t_r = \frac{1}{k_{off}} \qquad (4)$$

According to experimental evidences, t_r is related to the *in vivo* biological effects triggered by the ligand[6,9]. Usually, a high t_r value is considered useful for the pharmacodynamics profile, especially if it is longer than the pharmacokinetics elimination lifetime, otherwise is the mass–balance effect, drove by elevated local drug concentrations, to sustain the effect of the drug[10] (Figure 1). Nevertheless, also adverse effects can be linked to t_r, both off–target and target mediated[11].

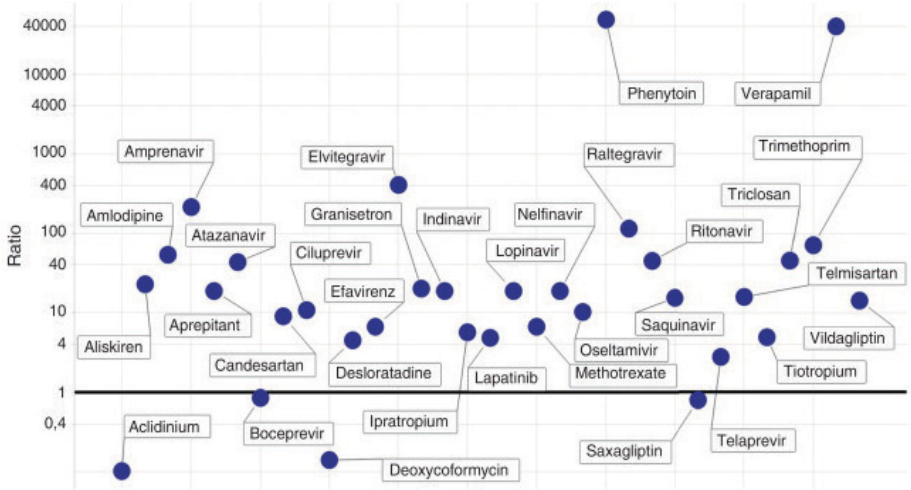

Figure 1. Plot showing the ratio between plasma half-lives and dissociation half-lives on a logarithmic scale (*y* axis) for some commercialized drugs. Dissociation half-lives is defined as $Dt_{1/2} = \frac{\ln 2}{k_{off}}$.(Adapted from ref. 10).

On the other hand, the k_{on} has a crucial role in protocols for binding measurements[12]: indeed, the kinetics on–rate of ligands involved in the experimental procedures drive the timing needed for the achievement of equilibrium conditions.

The free energy profile that characterizes binding and unbinding events is usually modelled as reported in Figure 2. According to the transition state theory (TST)[13], the highest point on the pathway ($LR^{\ddagger}$) represents the transition state between the unbound (L + R) and the bound configuration (LR). As stated in Eyring's equation (equation 4), the kinetics constants ($k_{on/off}$) are proportional to the exponential of the energies of activation of the respective transition states ($\Delta G^{\ddagger}_{on/off}$), through a pre-exponential factor that combines the Boltzmann's constant k_B, the Planck's constant h and the absolute temperature T:

$$K_{on/off} = \frac{k_B T}{h} e^{-\frac{\Delta G^{\ddagger} on/off}{RT}} \qquad (4)$$

Thermodynamics constants K_D and K_A, instead, are related to the $\Delta G_{binding}$, according to the equation 5:

$$K_D = e^{\frac{\Delta G_{binding}}{RT}} = K_A^{-1} \qquad (5)$$

Eyring's theory is the elegant combination of different thermodynamics, statistics and kinetics mathematical approaches developed until 1935[14]. Its power resides in its versatility, as is able to describe the kinetics behaviour of almost all the chemical systems. What allows us to apply this model (in principle developed for chemical reaction and two body interacting systems) to complex systems (as a fully solvated protein–ligand complexes can be considered) is that, along a single path, the dynamics of the transition depends only on the slowest steps involved (the one associated to the highest energy). Importantly, this assumption does not imply that the binding or unbinding kinetics are defined by only one

transition state, but more realistically, different pathways (therefore different transition states) can contribute with different weights to the observable k_{on} and k_{off} values.

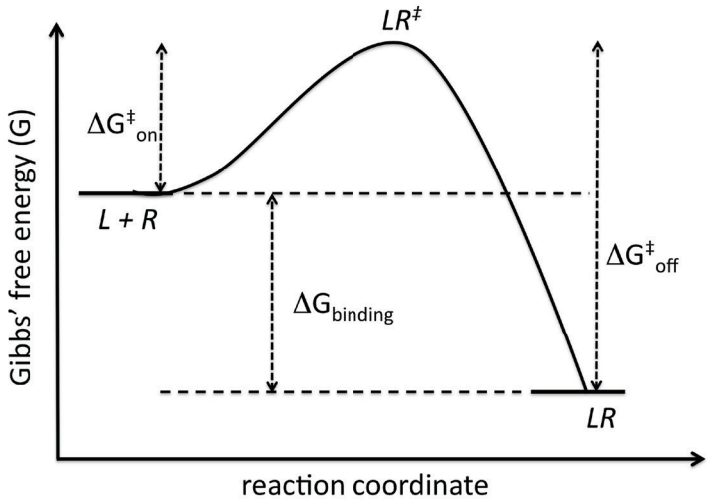

Figure 2. Simplified binding and unbinding energy profile. The highest energy point ($LR^{\ddagger}$) is the transition state between the unbound (L + R) and the bound state (LR). The energy difference between the two states is $\Delta G_{binding}$, while ΔG_{on} and ΔG_{off} quantifies the energy variation between $LR^{\ddagger}$ and the two energy basins.

Considering one of these possible energetic pathways (Figure 2), the definition of the reaction coordinate able to describe the free energy profile during the systems' transition become tricky, passing from a two body interacting system (which evolution is well depicted by the distance between the components of the system) to a biological system, where transitions may be better outlined by multiple descriptors[15–17] (e.g the combination of intermolecular distances, dihedral torsional angles, degree of solvation, angles formed by different centres of mass *etc etc*). To complicate the picture, there is the difficulty to determine which step, among all in the transition pathway, may represent the limiting one, or if two or more transition states are energetically degenerate, therefore equally contribute to the overall kinetics. In Figure 3, as an example, is reported the comparison between two reaction coordinates (r_{uni} and

r_{opt}) monitored during the molecular dynamics simulation of a 47–residues protein[16]: the distribution of the conformations follows a bimodal trend (Figure 3, panel B and E), which is better described by the reaction coordinate r_{opt}, able, indeed, to avoid overlaps (Figure 3, panel B) and to highlight the protein conformation associated to the transition state (the conformation associated to the lower probability).

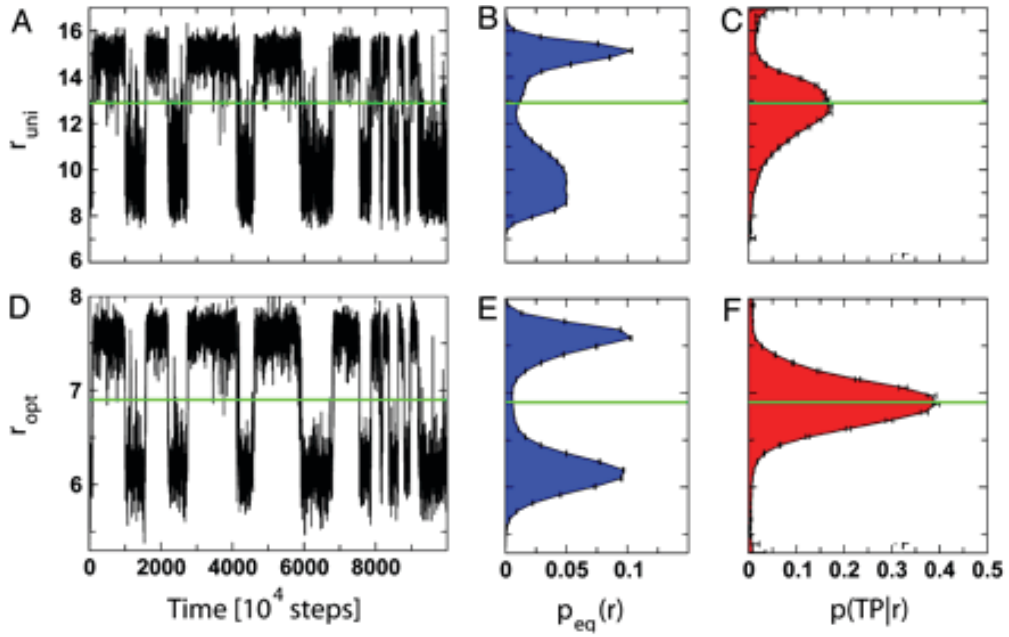

Figure 3. Comparison between reaction coordinates. (A and D) Reaction coordinates values during the same molecular dynamics simulation; (B and E) distribution probability of the two reaction coordinates; (C and F) the probability of being on a transition path. The green horizontal lines indicate transition states. (Adapted from ref [16]).

Ligand binding and unbinding events are multi steps processes, usually characterized by stable states that anticipate the final bound (or fully solvated unbound in the case oh unbinding) configuration[18–21] as depicted in Figure 4 for the case of alprenolol on the β_2 adrenoreceptor (β_2 AR)[20].

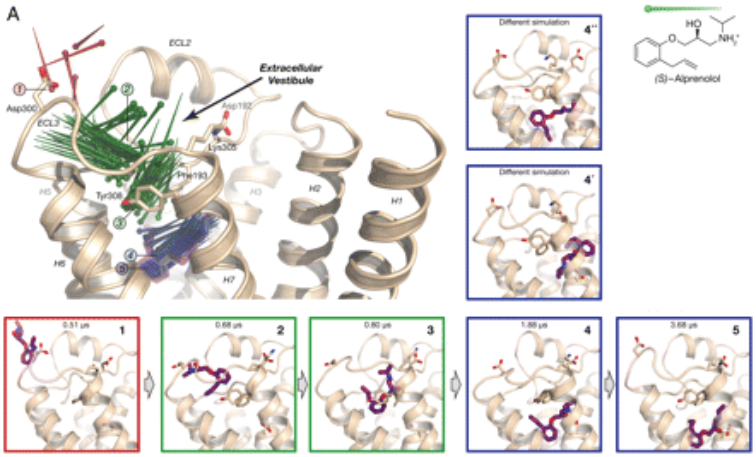

Figure 4. Alprenolol–β_2 AR binding metastable states. Pins indicate successive positions of alprenolol (A); the molecules moves from bulk (1), to the extracellular vestibule (2 and 3), and finally into the binding pocket (4 and 5). (Adapted from ref [20]).

From this standpoint, if we consider a two steps ligand–receptor recognition (or dissociation) event (equation 6), the formation of a metastable intermediate complex LR^{*M} has to be considered thermodynamically and kinetically related to the unbound state ($L + R$) and the final bound state (LR^{B}):

$$L + R \rightleftarrows LR^{*M} \rightleftarrows LR^{B} \qquad (6)$$

Again, the free energy profile can be simplified as reported in Figure 5. In this model the formation of an initial drug–receptor intermediate complex (LR^{*M}) anticipate the slower (because of the highest energy barrier $\Delta G^{\ddagger}_{B}$) formation of the final bound state (LR^{B}). In this case, compared to the single step binding model (equation 2), the definition of the thermodynamics dissociation constant K_D can be computed through a more complicate combination of the kinetics constants from each step (equation 7)[22,23]:

$$K_D = \frac{k_2}{k_1 + \frac{k_1 + k_1 k_3}{k_4}} \qquad (7)$$

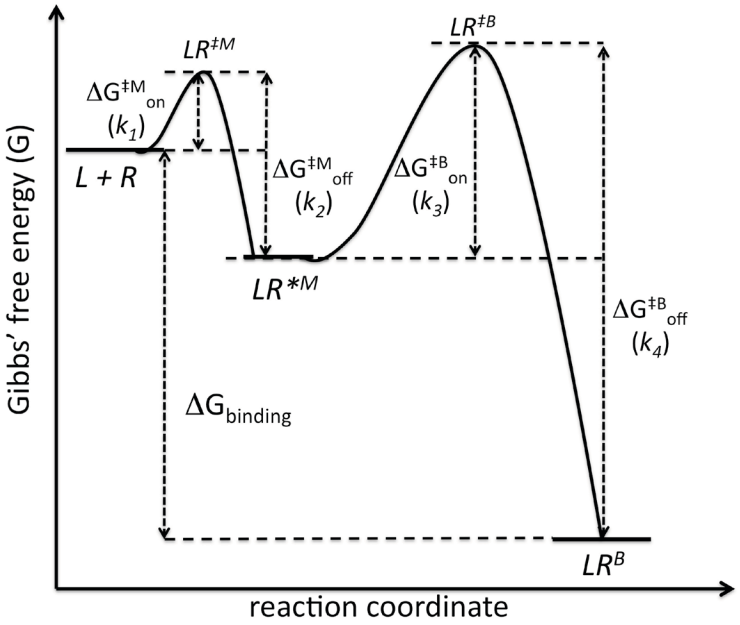

Figure 5. Energy profile of a two steps binding mechanism, characterized by a slower second step. Between the unbound state $L + R$ and the bound state LR^B the system experiences a stable intermediate complex LR^{*M}. The kinetics is governed by four constants (k_1, k_2, k_3 and k_4), while the thermodynamics is determined by the $\Delta G_{binding}$ value.

At the steady state, $\Delta G_{binding}$ can be quantified as the variation, from the initial unbound to the final bound state (and *vice versa* in the case of unbinding), of two state functions of the system: enthalpy (H) and entropy (S) (equation 8):

$$\Delta G_{binding} = G_{bound} - G_{unbound} = \Delta H_{binding} - T\Delta S_{binding} \qquad (8)$$

A negative $\Delta H_{binding}$ value is considered favourable for binding[24] and reflects a positive change in the number and/or strength of intermolecular interactions. In the same way, a negative value for $-T\Delta S_{binding}$ indicates an increase in entropy, ascribable to favourable changes[24] in translational, rotational or conformational entropy of the system. Studies carried out on sets of different intermolecular complexes highlighted the general tendency of enthalpy–entropy compensation: ligand binding appears to be energetically driven by either a

favourable enthalpy variation or a favourable entropy variation, but usually not by a concordant effect of the two[24,25].

An alternative point of view on binding and unbinding is offered by statistical mechanics[26,27]. According to this treatise, in equilibrium conditions, a system can be described as the sum of the energy of all its possible microstates, each one associated to a probability to exist quantified from its own energy. As a consequence, microstates with low energy are more populated and, if we label the i_{th} microstate by its energy E_i, the probability of a specific microstate is given by equation 9:

$$p_i = \frac{1}{Z} e^{-\beta E_i} \qquad (9)$$

where $-\beta E_i$ is called the Boltzmann factor ($\beta = 1/K_B T$) and Z is the partition function, defined as the sum of all the possible microstates. Intuitively, the probability associated to a ligand–receptor bound configuration increase exponentially with the energy of stabilization generated by the intermolecular complex formation.

1.1.2 BINDING DRIVING FORCES

During an intermolecular recognition, different events occur and in a general view, both the ligand and the receptor respectively: *i*) partially desolvate; *ii*) rotationally rearrange, in order to reach the right conformation for recognition and *iii*) establish new interactions. Each step needs a certain amount of energy for its activation and can be more related to a change in enthalpy or in entropy.

1.1.2.1 Intermolecular interactions

Like protein–protein association[28], also the recognition of small molecules is strongly affected by electrostatics interactions. In physiological pH conditions, molecules can be in ionized if their chemical structure includes basic or acidic functional groups, and when two charges of opposite signs establish a favourable ionic interaction, the enthalpy of the stabilization is determined by: *i*) the charge densities present on the respective atoms; *ii*) the interatomic distance and *iii*) the dielectric constant ε of the microenvironment. Notably, charges located at the protein surface may experience a different dielectric constant than charges buried in the protein core, with ε values that can diverge from ≈20-80 (at the bulk solvent interface) to ≈ 1-4 (inside buried protein cavities)[29,30]. During a salt bridge formation, the total free energy change ΔG_{tot} can be generalized as the algebraic sum of different terms[31] (equation 10):

$$\Delta G_{tot} = \Delta G_{dslv} + \Delta G_{brd} + \Delta G_{prt} + \Delta G_{rot} \qquad (10)$$

More precisely, ΔG_{dslv} takes in account the desolvation penalties that individual charged atoms have to overcome in their passage from the water to the protein environment; *ii*) ΔG_{brd} is the favourable enthalpy due to the interaction of oppositely charged functional groups; *iii*) ΔG_{prt} represents the

change in interactions between the new formed salt–bridge and the rest of the protein. Moreover, especially if long side chain amino acid residues are involved (as lysine or arginine), a loss of rotational entropy ΔG_{rot} should be considered for the overall ΔG_{tot}.

Moving to other classes of intermolecular interactions, hydrogen bonding probably represents the most common polar interaction in nature, being the intermolecular force characterizing the main constituent of the biosphere: the water. From a molecular point of view, at small separation distances between atoms involved, hydrogen bonds are governed by a balance between attractive and repulsive electrostatics, charge transfers and polarizations phenomena, while, at larger distances, are almost completely electrostatics[32]. In pure water, the strength of an hydrogen bond is about 3.7 kcal/mol[33], but the stability provided by a single interaction depends on the degree of solvent exposure[34] as well as the nature of the heteroatoms involved and their relative positions[35]. Moreover, the formation of a hydrogen bond, beside the enthalpy stabilization, implies also a minor entropy loss[36]. Recently, an increasing attention has been addressed to the understanding of how hydrogen bonds influence the binding and unbinding kinetics: hydrogen bonds buried inside protein cavities are generally shielded from water molecules and their rupture (and formation) are associated to energetically penalized transition states, as we can consider dehydration and rehydration occurring asynchronously[37] (Figure 6). This scenario is well exemplified by the biotin–streptavidin complex, where the water entrance in the binding pocket (Figure 7) during the unbinding is highly unfavourable due to the presence of a net of hidden hydrogen bonds[38,39].

Despite huge efforts made by the scientific community, there is still lack of a univocal model able to describe the contribution to intermolecular recognition[40] generated by hydrophobic interactions. From a computational point of view, hydrophobic interactions are estimated through the Lennard–Jones potential[41] or on the basis of the surfaces involved: the stabilization generated is generally quantified in the range of 25–75 cal/Å^2[42]. Unsaturated rings have been extensively reviewed[43] and in a recent study[44] Pi-Pi stacking

interactions, have been compared with hydrophobic interactions involving sp3 carbon atoms rings: interestingly, has emerged that in drug-like molecules the replacement of phenyl groups with saturated rings seems not to negatively affect the affinity.

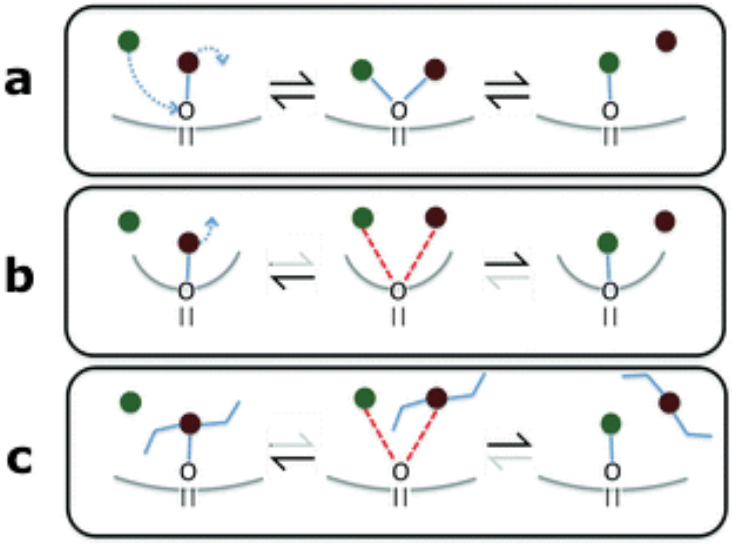

Figure 6. Representation of the hydrogen bond exchange processes between water and ligand (a) with a solvent-exposed polar atom; (b) for an almost buried polar atom; (c) with a solvent-exposed polar atom with a bulkier ligand. (Adapted from ref [37]).

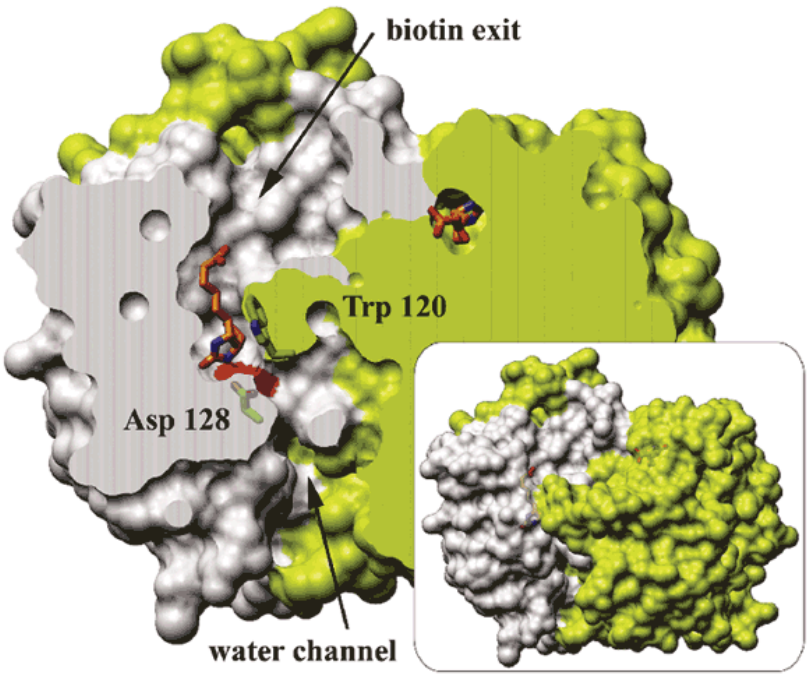

Figure 7. Biotin binding site location in the crystal structure of streptavidin. (Adapted from [38]).

1.1.2.2 Role of water molecules

Since the dawning of modern sciences, water has been considered a cornerstone of biological processes, because of its fundamental impact in protein plasticity, protein–protein interactions and small molecules' binding[45].

Nowadays, the role of water during ligand association and dissociation represents a challenging field for medicinal chemists[46], mainly due to the dichotomy between enthalpy and the entropy, which in water molecules probably finds its maximum expression within biological systems[47,48]. As a rule of thumb, water molecules in binding sites are partially restrained by irregular shapes of the protein surface. Moreover, cavities are frequently hydrophobic environments, whose solvation requires an energetic cost that can be regained thanks to the replacement of unfavourable water molecules with lipophilic moieties belonging to bound ligands[49–51]. From this point of view, the so-called "magic methyl effect"[52,53] is probably the most famous manifestation, in medicinal chemistry, of the effect mediated by the displacement of instable water molecules from the receptor surface during the binding. The analysis of several thousands crystal structures revealed that at least a water molecules mediates contacts between protein and ligand in almost two–thirds of all complex considered[54,55]. As demonstrated by the rational design of non-peptide HIV protease inhibitors[56], medicinal chemists have to take in account water and its importance in intermolecular recognition[57]. A number of different computational methods have been developed with the aim of predict the presence of water molecules in protein–ligand complexes[58], including Grand Canonical Monte Carlo (GCMC)[59] and WaterDock[60]. In general, MD–based techniques show relevant potentiality for this application, allowing to quantify the energetic contribution produced by water mediated interactions[61], or to evaluate the water accessibility to binding sites during unbinding events[62]. One

of these methods is WaterMap[63]: it consists of an all-atom MD simulation carried out in explicit solvent. Output trajectory thermodynamics is statistically–analysed to define potential hydration sites according to water clusters detected. WaterMap drove the development of a series of platelet-derived growth factor receptor β (PDGFR-β) inhibitors, thanks to the design of a structural modification of the ligands able to target a weakly bound water molecules, leading to a 25-fold improvement in IC_{50}[64] (Figure 8).

Interestingly, by means of long classic MD simulations, water molecules have been indicated as responsible for early kinetic barrier during ligand approach to G-protein coupled receptors (GPCRs), due to the dehydration that takes place as the drug associates with the extracellular vestibule of the receptors[20,65].

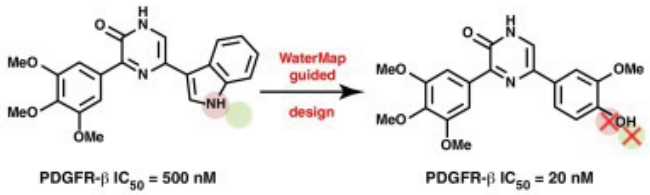

Figure 8. Development of PDGFR-β inhibitors using WatertMap. The replacement of an indole group is able to displace an unfavourable water molecule (in red) and improve the overall interactions through a water bridge molecule (in green). (Adapted from [58]).

1.1.2.3 Conformational changes

Almost thirty years ago a new concept was introduced in the field of protein biophysics[66,67], the energy landscape, which is the free energy associated to all the possible protein conformations. Each protein conformation is, in turn, a function of all the degree of freedom (mainly dihedral angles of the backbone). In this representation, valleys are stable and favourable conformations the protein can explore with kinetics and probabilities that depends on the shape of the surface that separate two or more minimum. This scenario suggests the

existence of an ensemble of possible pathways, intermediate states and transition states connecting two stable points.

In physiological conditions, the presence of a ligand in the surrounding environment of a protein can produce a shift in the protein conformational distribution[68]. Generally, two models of binding mechanism are considered (Figure 9): *a*) the "conformational selection" (equation 9), where the ligand *L* only binds one of the macromolecule conformational states *R',R",R'"* (usually presents only in small amounts) and *b*) the "induced fit" mechanism (equation 10), characterized by the ligand binding to the predominant protein conformation, followed by an isomerization of the complex (*RL**) to yield the final form (*RL*):

$$\langle R',R'',R''',\ldots\rangle + L \rightleftarrows LR \qquad (9)$$

$$R + L \rightleftarrows RL^* \rightleftarrows RL \qquad (10)$$

Even if is well established that the mechanistic model followed is influenced by the protein and ligand concentrations, it has been proposed a major role for the conformational selection[69–71].Considering the induced fit model, both the first and the second step can involve a conformational change of the target and/or the ligand, with a relative kinetics usually slower if is the target to experience a conformational isomerization, due to the higher number of degree of freedom implied[72]. This high number of degrees of freedom and the complexity of the free energy surface make difficult to estimate its overall thermodynamic influence on binding, especially for the entropic component[73].

Ligand and protein side chains conformational changes, instead, are linked to the exploration of a relatively low number of rotatable bonds and, therefore, are more easily addressable using a range of both experimental and computational techniques [74–76]. Notably, can exist a substantial difference between the more stable ligand conformations in the bulk solvent and the binding conformation[77,78] (bioactive in the case of an agonist) and it is important

to note that binding brings an unfavourable entropy contribution to the total free energy [79,80].

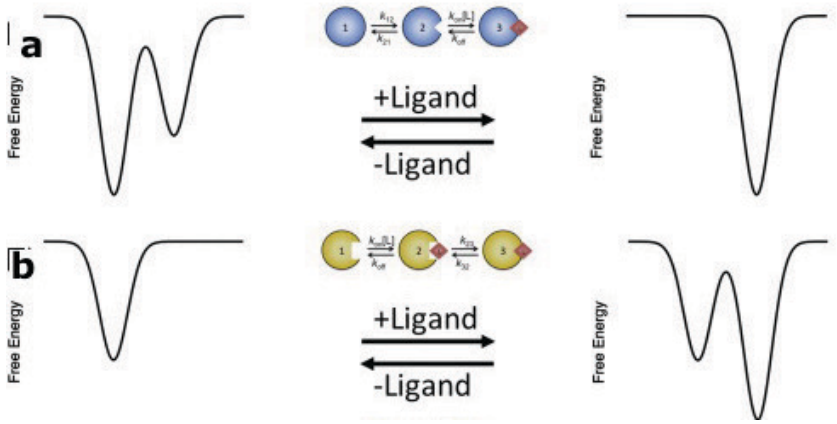

Figure 9. Comparison between conformational selection (a), where the ligand selectively binds a single conformation (stabilizing it) and induced fit (b), where the ligand induces conformational rearrangements in the macromolecule to optimize the interaction. (Adapted from ref [69]).

1.2 MOLECULAR DYNAMICS AS A POWERFUL TOOL FOR INVESTIGATE BINDING AND UNBINDING EVENTS

Structure based drug design (SBDD) has its roots in the middle of the last century, when the first pioneering results were obtained from X-ray crystallography techniques[81]. During the following decades the possibility to investigate biological macromolecules at atomistic resolution, as well as the constantly increasing computational performances, paved the way to molecular simulations in the pharmaceutical field. Docking simulations[82] rapidly became a fundamental tool in drug discovery, due to the opportunity of efficiently perform virtual screenings of vast virtual libraries of compounds and rationalise the structure-activity relationship (SAR) able to drive the optimization of hit and lead compounds. Main purpose of docking is to explore possible ligands' conformations in their binding site, semi-quantitatively evaluating the stability of intermolecular complexes generated on the basis of shape and electrostatics complementarity.

More recently, other force field–based techniques have been developed: alchemical perturbation (AP) methods[83] and molecular mechanics energies combined with the Poisson-Boltzmann or generalized Born and surface area continuum solvation (MM/PBSA and MM/PBSA) represent common tools for evaluate ligand binding affinities[84]. These approaches, whose results strongly depend on the chemical system considered, take advantage of short molecular dynamics (MD) simulations performed on both ligand and receptor conformations extracted from docking, quantifying the energetic stability of a binding mode and neglecting how the binder should reach the putative bound configuration.

In the last years increasing efforts have been addressed to decipher the ligand-receptor binding phenomena at an atomistic scale, with the aim of characterize the driving forces that move a small molecule from the solvated unbound state toward the stable final intermolecular complex, and *vice versa*. From this standpoint, MD is a promising tool, able to take in account protein

and ligand flexibility as well as the dynamic role of water molecules. Disclosure of a more detailed picture about binding mechanisms may allow the development of alternative and innovative drug discovery strategies, able to bridge the chemical modification introduced with the experimentally–determined kinetics. In this scenario, beside the structure-activity relationship (SAR) approach, the structure–kinetic relationship (SKR) is constantly gaining popularity[85]. The introduction of the time as a variable in computational simulations allow consider the systems' evolution according to how their components can dynamically and reciprocally interact: starting from an initial configuration on the energy landscape is possible to sample possible evolutions and pathways from a local minimum to another, passing through saddle points. In other words, by means of classic MD, which nowadays is already able to reach the millisecond time scale[86,87] thanks to the recent introduction of the graphic processing unit (GPU) technology and dedicated MD codes, is possible to reconstruct the ligand transition from the unbound state to the bound intermolecular complex, which represents a minimum of the system energy landscape. From this point of view, a crucial task is to sampling ligand binding or unbinding events an adequate number of times, as any observable retrieved from MD (e.g. energy values or kinetics rate) should derive from a statistical approach (and may consider the systematic error introduced by using rude molecular mechanics force fields).

On the other hand, the dissociation of a intermolecular complex can not be simulated using a classic MD approach, because of the high energy barrier of the transition state associated, that impose timescales still unreachable for the methodology. Indeed, in order to sample possible unbinding pathways and mechanisms, the so-called "enhanced sampling" methods are employed.

This section introduces the applicability of classic MD to the study of binding events as well as some enhanced methods that have been successfully used to characterize unbinding mechanisms.

1.2.1 CLASSIC MOLECULAR DYNAMICS FOR THE MODELING OF BINDING EVENTS

In drug design, classic MD is often combined with molecular docking methods with the aim of evaluate the quality of docking solutions as well as to their refinement and tuning[88,89]. However, classic MD represents a powerful method for disclosing the entire mechanisms of binding and has gained increasing visibility in the last years: in 2011 Buch *et al.* reconstructed for the first time at atomistic detail a ligand–protein recognition event[18], retrieving kinetic and thermodynamics parameters in good accordance with experimental data. A pioneering study in the G protein-coupled receptors (GPCRs) field is represented by the work of Dror *et al.*[20]. The authors performed classic MD simulations on both the β_2 adrenergic receptor (β_2 AR) and the β_1 adrenergic receptor (β_1 AR): interestingly, the antagonist alprenolol bound to β_2 AR exploring a predominant pathway, characterized by different metastable sites located between extracellular loop 2 (EL2) and 3 (ECL3). Decherchi *et al.*[90], instead, combined classic MD simulations with machine learning algorithms in order to reconstruct the binding of an inhibitor to the purine nucleoside phosphorylase enzyme. After microseconds of simulation, a final intermolecular complex in agreement with the experimentally determined one was gained, highlighting three different main binding pathways to the active site and therefore corroborating the experimental slow kinetics measured.

Usually, the concept of binding is associated to the initial ligand diffusion from the bulk solvent to the receptor surface, however it is important to consider that about 60% of drugs targets are located at the cell surface[91] and that membrane proteins are immersed in a peculiar environment. It follows that lipophilic binders may reach their recognition sites also from the lipid bilayer, as described by of Stanley *et al.*[19]. In this work, the binding of the inhibitor ML056 to the sphingosine-1-phosphate receptor 1 ($S1P_1R$) was reconstructed using classic MD simulations and Markov state models (MSMs)[92]: the aggregate

sampling time of over 800 μs allowed to highlight four different stages during the binding pathway.

MSMs[92] are a powerful statistical approach recently introduced for the computation of kinetics parameters from MD simulation. Briefly, assumed that short simulations allow observe system transitions between states that are kinetically near, a number of short parallel MD simulations are performed starting from different initial configurations. A structural descriptor is followed during the simulations and when its value become stationary a new simulation is started with the goal of to avoid the over sampling of metastable configurations (therefor the quality of the data set collected is improved). After the short MD trajectories are collected, a set of microstates is defined performing a clustering based on the structural descriptor (clusters are represented by the number of states that are comprised) and a microstate transition matrix is constructed. Configurations in the MD trajectories are, one by one, compared to microstates to find out to which microstate they can be assigned: the result is a translation of the overall MD trajectory from a series of structures over time to a series of microstates over time[93]. The aim is to count how many transitions are observed between each pair of microstates *i* and *j* during a lag time τ in order to estimate their probability in that time window[93]. MSMs constructions have been applied to chemical system of increasing complexity: recently Plattner *et al.* reconstructed the binding of benzamidine to β-Trypsin[94], revealing that the enzyme can exist in six main conformations and that the overall mechanism exhibits features of both induced fit and conformational selection. Gu *et al.*, instead, employed MSMs to study the choline binding to the endogenous protein ChoX[95], ascribing the recognition mainly to the conformational selection model. *Kohlhoff et al.*[96] carried out a cloud-based MSM approach to simulate a total time of 2.15 ms on the β_2 adrenergic receptor (β_2 AR). During this study, the authors started from both inactive and active crystal structures of the receptor, in presence of two different ligands (the partial inverse agonist carazolol and the full agonist BI-167107): findings revealed that binding to both of the two isomers may take place through multiple pathways, that comprises metastable intermediate sites.

Recently we have proposed an alternative unbiased MD based method for the characterization of possible pathways during recognition events: the supervised molecular dynamics[97,98] (SuMD). According to this approach, the timescale needed to reproduce complete intermolecular complex formations results in the range of nanoseconds, instead of hundreds of nanoseconds or microseconds usually necessary with classic MD. Sampling is gained without the introduction of bias, but just by applying a tabu–like algorithm (Figure 10) to monitor the distance between the center of masses of the ligand and the binding site, during short classic MD simulations.

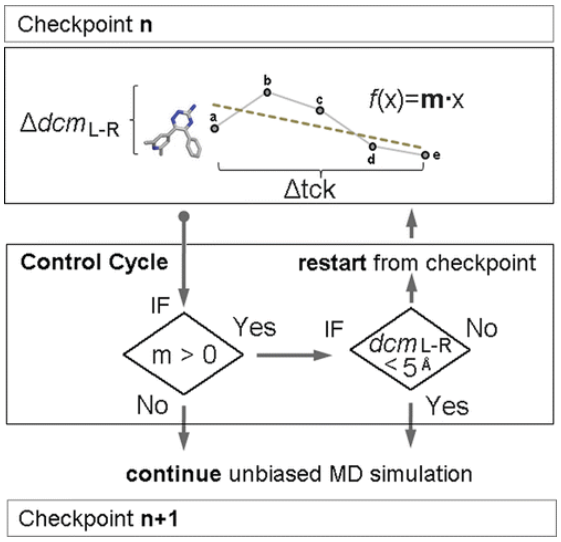

Figure 10. SuMD algorithm scheme. The ligand–receptor distance vector (dcm_{L-R}) is collected at regular time intervals and an interpolating linear function is calculated after each time window.

During each SuMD time window, an arbitrary number of distance points is collected at regular time intervals and fitted into a linear function. If the slope of the resulting straight line is negative, the ligand-receptor distance is likely to be shortened and the simulation is restarted using coordinates and velocities from the very last SuMD production step. Otherwise, the simulation is restored from the set of coordinates from the last productive step, by reassigning atoms velocities according to Boltzmann distribution. When the ligand–binding site distance reaches a predetermined value, the supervision is stopped and the

system can freely evolve without supervision. Some recent SuMD applications comprise the binding reconstructions of adenosine to its endogenous receptor A_{2A}[99], as well as of the positive allosteric modulator (PAM) LUF6000 on the adenosine receptor subtype A_3[100].

1.2.2 METHODS THAT INTRODUCE A SCALAR ENERGY BIAS IN THE SYSTEM POTENTIAL ENERGY

As already stated, dissociations of intermolecular complexes require the overcome of high-energy barriers in order to activate the process. This, in turn, affect the timescale needed to observe a spontaneous unbinding, which can reach up to days or weeks in the case of very slow-off ligands. In order to facilitate the systems' transition to the dissociate configuration, some MD–based approaches have specifically developed in order to modify the energy landscape. More precisely, the introduction of a new scalar value in the mathematical function that rule the potential of the system allow to flatten the energy surface and therefore to increase the probability to simulate kinetically–unfavourable events.

Recently, the scaled MD approach[101] (Figure 11), originally introduced as a tool for protein folding sampling, was applied to the unbinding study of seven modulators of the enzyme glucokinase 1 (GK1)[102,103]. Biasing the electrostatic and the Lennard-Jones potentials, the authors were able to simulate different ligands' transitions from the bound complex to the unbound state and, after some simplifications and by using a statistical approach, compared the computed residence times to the experimental values. Interestingly, even if this method implies a loss of details of the unbinding path, findings suggested that compounds characterized by a T-shaped geometry exhibits a longer residence time, ascribed to an induced fit binding mechanism inside the linearly shaped recognition cavity of the enzyme.

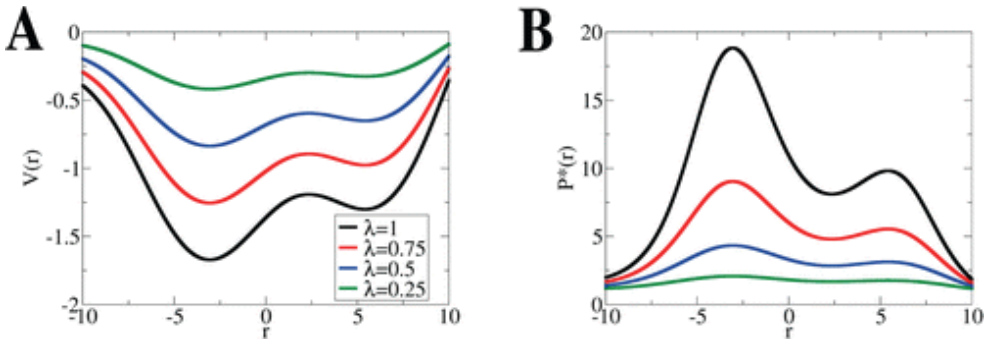

Figure 11. Schematic illustration of scaled MD. (A) Different scaling values lead to distinct energy surfaces: the higher the scaling factor the flatterer is the surface; (B) changes in probability distributions for the different scaling factors. (Adapted from ref [101]).

A further enhanced sampling approach is the accelerated molecular dynamics (AMD)[104]. During an AMD simulation, the system energy potential V*(r) changes continually, according to a function block that defines when to add the bias (equation 11):

$$V^*(r) = \begin{cases} V(r) \Leftrightarrow V(r) \geq E \\ V(r) + \Delta V(r) \Leftrightarrow V(r) < E \end{cases} \qquad (11)$$

V(r) represents the unbiased potential energy of the system and E is a predefined threshold energetic parameter. It follows that potential energy landscape minima that lie below E are raised, therefore reducing the neighbouring energy barriers[105]. The AMD usually bias the potential associated to the dihedral energy of the system, as recent employed in the GPCRs field to reconstruct the binding to the octopamine receptor[106] and the M_3 muscarinic receptor[107].

The possibility to modify the temperature in the simulation inspired the development of the temperature replica exchange algorithm (T–REMD), first applied to MD simulations by Sugita *et al* in 1999[108]. During T–REMD several parallel simulations are performed at different temperatures and periodically pairs of replicas are exchanged according to a statistical criterion: the resulting sampling in the temperature space increases the probability to escape from local

energy minima[109]. Differently from T–REMD, during the Hamiltonian replica exchange (H–REMD) are the force field parameters to be exchanged between parallel MD simulations. This approach allow to tune the bias only for a subset of the overall potential energy component of the system, instead of the total energy as happens in the case of T–REMD[110]. It follows that transitions are promoted thanks to a fewer number of MD replicas. In the work of Wang *et al.*[111] H–REMD technique was able to enhance the sampling of putative binding modes of three known ligands to T4 lysozyme L99A.

1.2.3 METHODS THAT REQUIRE THE PRELIMINARY DEFINITION OF COLLECTIVE VARIABLES

In physics the degrees of freedom embody all the variables that describe the possible configurations of a system, allowing describe it from a mechanic point of view. In a chemical system, modelled by means of molecular mechanics, these degrees of freedom can be represented by the force field equation parameters. Moreover, by mathematically combine more degrees of freedom is possible to define order parameters, which are reduced descriptor for a subset of degrees of freedom. Some examples of order parameters are angles formed by non-bonded atoms or distances between groups of atoms, but more complex, experimentally-derived order parameters are normally employed[112,113].

Before to perform a MD simulation is possible to define one or more order parameters, usually named as collective variables (CVs), able to describe the system transition the user is interested in: during the simulation, if an energy bias is added to the system along these CVs, the probability of observe the transition results enhanced. Important MD techniques that employ the bias of a set of CVs are metadynamics, steered MD, umbrella sampling (US) and temperature accelerated MD (TAMD).

After its first definition and application on simply chemical systems[114] (the dissociation of NaCl and the isomerization of the Alanine dipeptide in water), metadynamics soon demonstrated its potential usefulness in reconstructing the binding of chemical fragments on the enzyme β-Trypsine[115]. The algorithmic scheme behind metadynamics can be recapped as the seeding, at discrete time intervals, of a history–dependent energetic term centred along the predefined set of CVs (Figure 12). When the energetic bias is added at a certain instant, the probability that the system will revisit that specific configuration is decreased according to the shape of the energetic Gaussian function that is supplied. Thanks to this theoretical approach is possible to fill energy minimum on the energy surface defined by the CVs, therefore increasing the transition

probability from a basin to another one (e.g. the ligand bound and unbound states)[116].

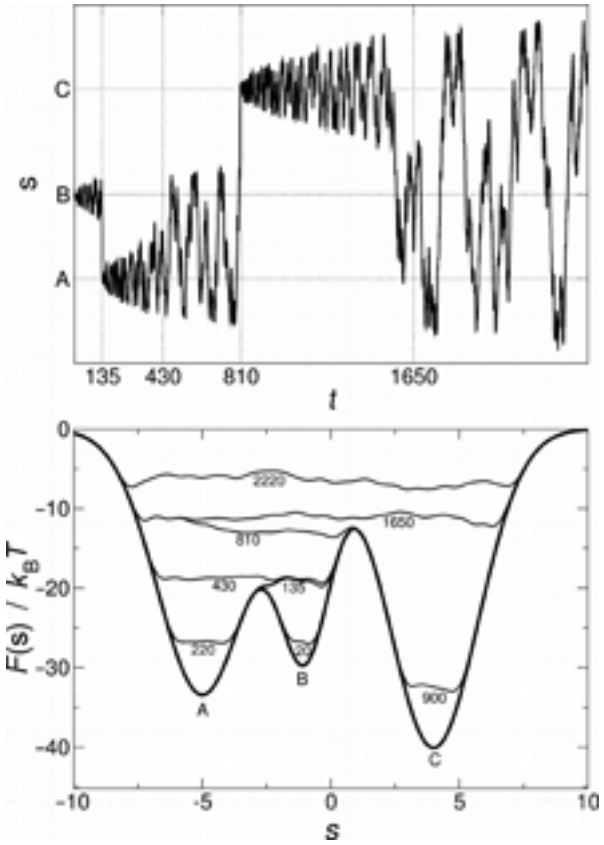

Figure 12. Example of metadynamics simulation that bias one CV (denoted as *s*). The time *t* is measured by counting the number of Gaussians deposited. At the beginning of the simulation the system is in the minimum B, then, after 135 energy depositions, it experiences a transition to the minimum A, which, in turns, is filled after 810 bias cycles; the system is then allowed to explore minimum C. (Top) Time evolution of the collective variables during the simulation. (Bottom) Representation of the progressive filling of the potential energy minimums by means of the Gaussians deposited along the trajectory. (Adapted from ref [116]).

In general, the definition of the order parameters that well describe the binding or unbinding events is not trivial. Crucial aspects to consider are the height and the width of the Gaussian energetic term added: as a rule of thumb, the height of the function should be much smaller than the highest energy barrier characterizing the process, while the width should be small enough to

track FES containing many basins connected by narrow paths[117]. Regarding this latter point, Branduardi *et al.* proposed the adaptive Gaussians metadynamics approach, able to adjust the Gaussians variance on the fly, on the basis of the local properties of the energy surface[118]. Last but not least, as the error introduced during metadynamics is inversely proportional to the square root of the time deposition[119], the energetic bias addition should not be too much frequent.

More recently, issues related to the overfilling effect of metadynamics (an excessive energy seeding that may lead to the exploration of regions of the space that are too high in energy and therefore associated to protein unfolding) have inspired the development of the well-tempered metadynamics[116,120] (WTmetaD): by tuning a new threshold parameter is possible to automatically limit the energy surface exploration to the physically interesting region, minimizing the bias added according to the energy already provided to the system. An elegant application of WTmetaD was the unbinding simulation of the selective COX-2 inhibitor SC-558[121]. Interestingly, findings highlighted the possibility that SC-558 engages COX-2 with two different binding poses, providing an explanation for the increased residence time in COX-2 for ligands structurally-related to SC-558 (Figure 13).

Binding of benzamidine on β-Trypsine and of SC-558 to COX-2 were also reconstructed implementing the so called funnel restraint potential in combination with WTmetaD: the core of this alternative MD methods is the introduction of a funnel-shaped restraint potential that limit the sampling of the ligand in the bulk solvent, speeding up the overall binding and unbinding simulation time[122].

Recently, Pratyush *et al.* provided a first attempt of to derivate kinetics rate values using WTmetaD by simulating the isomerization of the alanine dipeptide[123,124]. Salvalaglio *et al.* later introduced a posteriori statistical analysis (based on Poisson statistic) that establishes the reliability of the kinetics values thus generated[124]. In a very recent application of the former approach with a Markov State Model construction, the benzamidine unbinding mechanism from

β-Trypsin was completely disclosed[125]. Interestingly, beside the starting crystallographic binding pose, one alternative bound conformation and one metastable intermediate state were detected. Furthermore, the apo-form of the enzyme was suggested as existing in two different conformations: one characterized by the loop L in its crystallographic form (available for further binding), the other instead presenting a distorted loop L (protein is temporarily inactive).

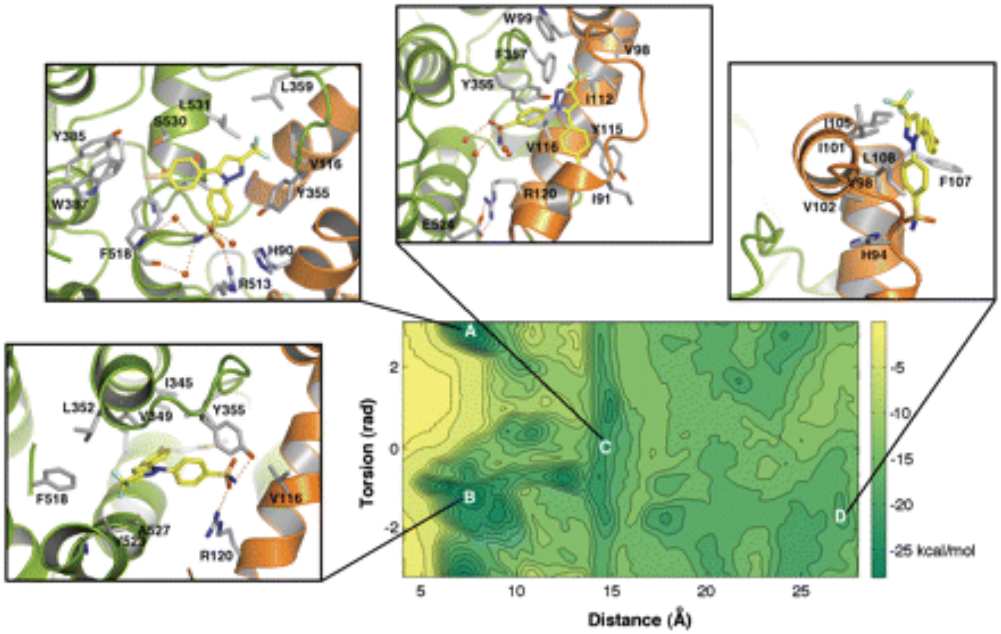

Figure 13. Dissociation energy surface of the COX-2 inhibitor SC-558. It was reconstructed as a function of the distance between the ligand and the protein and the value of a dihedral angle of the ligand. Four main energy basins A–D were found during the metadynamics (corresponding to the four displayed snapshots of the complex SC-558/COX-2): basin A, crystallographic pose; basin B, alternative pose; basin C, poses at gate site; basin D, external pose. (Adapted from ref [121]).

In general, binding and unbinding are influenced by multiple complex order parameters, such as the formation of intramolecular interactions, the rotations of entire protein domains or variations in the number of water molecules solvating the binding site. This complex scenario inspired Piana *et al.* to develop the so called bias–exchange metadynamics (BE-META)[126]: the heart

of this method is to exchange the CVs to bias between different parallel MD simulations, in order to sample a large number of different degrees of freedom of the system and, therefore, to explore energetic barriers orthogonal to the classic distance–based reaction coordinates. One of the first pharmaceutical application of BE-META was performed on the binding mechanism of the peptide substrate to HIV-1 protease[127].

Moving to further CVs–biasing methods, during a steered MD (SMD) simulation the transition between bound and unbound states is achieved by biasing the system energy with a harmonic potential, usually added on the protein–ligand distance. During the transition, is possible to compute the external work performed on the system, keeping track of energetic barriers along the unbinding pathway [128,129]. While the first SMD application on unbinding simulations was performed on the dissociation of the complex biotin–streptavidin[39] by Grubmuller *et al.*, more recent works employed this technique to study both binding [130]and unbinding mechanisms. Among these latter fall the study of Favia *et al.*, that sequentially applied SMD and metadynamics in order to reconstruct the dissociation mechanism of the endogenous hormone cortisone from the catalytic site of the enzyme 11β-HSD-1[131]: the authors first applied the SMD approach to sample different exit pathways, then tuned the selection of the CVs to bias on the basis of the trajectory that required the lower amount of work for the dissociation. SMD was also used to compare the unbinding force profiles of etoposide and some its analogs from the type II topoisomerase F14512[132], as well as for discriminate active cyclin-dependent kinase 5 (CDK5) inhibitors from inactive ligands[133].

A SMD–related methodology is the random expulsion MD[134] (REMD), during which the direction of the additional force is chosen randomly every time the ligand velocity falls below a predefined value (likely due to the encounter of relatively rigid parts of the protein). In a REMD application, a total of 360 short unbinding event of dihydroxyvitamin D3 from its receptor were reproduced[135], highlighting different putative pathways. REMD and SMD together, indeed,

were employed to study binding and unbinding mechanisms of the inhibitor GS-461203 on hepatitis C virus (HCV) non-structural protein 5B (NS5B)[136].

Umbrella sampling (US) represents the first CV–biasing method, developed[137] with the aim of compute the free energy change implied in a thermodynamic transition. Chosen a CV to bias (for pharmaceutical applications usually is the distance between the ligand and the binding site) the ligand is moved in a discrete way along the CV by performing several parallel short MD simulations (called "windows", Figure 14).

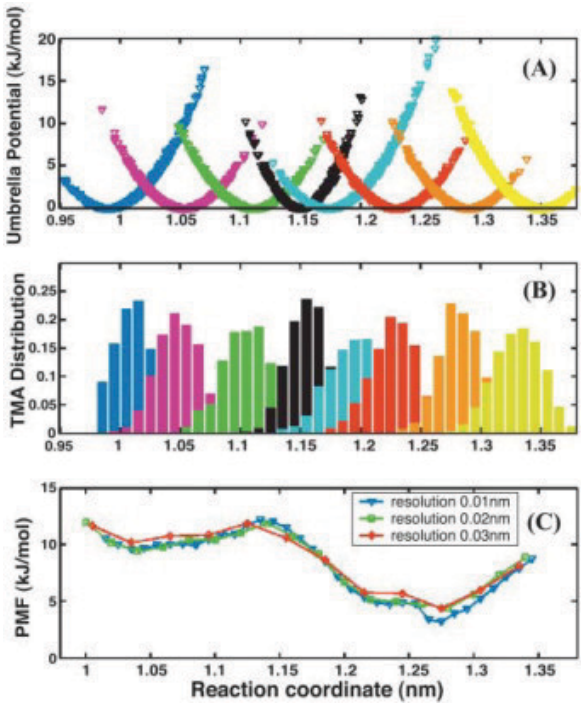

Figure 14. Example of an umbrella sampling workflow. (A) During different windows the position of the restrained ligand is sampled along the CV (reaction coordinate); (B) the probability distribution for each window is computed (using a 0.1-Å bin width in this case); (C) the energy surface along the CV (PMF) is calculated considering different resolutions of the grid elements. (Adapted from [138]).

During each window the ligand position x is subjected to an harmonic potential that allow to quantify the energy characterizing the regions along the pathway. The probability distribution of the ligand positions from each window is used to reconstruct the energy surface along the CV, by means of a post

processing analysis. From this point of view, the most common methods employed is the weighted histogram analysis method (WHAM)[139], which integrates the probability distribution considering both the bin chosen for its construction and the bias applied during the simulation. It follows that the distance between consecutive ligand positioning from a window to the consecutive one, as well as the value of the force constant used, can heavily influence the final result of the this approach. As a rule of thumb, the force constant applied should be large enough to allow the system to overcome energetic barriers, but at the same time not too large, in order to avoid very narrow positional distributions[140]. Notably, US was applied to a number of biological systems in order to energetically and kinetically characterize the putative binding of both small molecules and peptides[138,141].

A peculiar CV–biasing technique is represented by the temperature accelerated molecular dynamics (TAMD)[142]. TAMD consists in sampling along a set of CVs by applying a harmonic force coupled to a fictitious particle thermally kept at higher temperature. The idea is to increase the Brownian motion only on few degrees of freedom, while the rest of the system remain distributed at the real temperature according to Boltzmann distribution. Dong *et al.* applied TAMD to the unbinding of the adenosine receptor inhibitor ZM241385[21]. The authors simulated 10 dissociation events and detected a set of key protein residues later investigated in accurate mutation studies, as well as multiple distinct consecutive steps along the pathways. Experimental outcomes highlighted the importance of the hydrogen bond network at the extracellular vestibula involving the ligand, His264 and Glu169, responsible of slowing down the dissociation rate, due to the energy required for breaking these favourable interactions.

1.3 G PROTEIN-COUPLED RECEPTORS AND ADENOSINE RECEPTORS

G protein–coupled receptors (GPCRs) represent the most numerous family of membrane proteins (encoded by more than 800 genes in the human genome[143]) and mediate cellular responses to a wide range of stimuli, which comprise hormones, neurotransmitters and photons[144]. According to the GRAFS [143] nomenclature system, GPCRs can be classified into five families[145]:

- rhodopsin-like: is the largest GPCRs family. Members' diversity is mainly within the transmembrane (TMs) regions;
- secretin-like: is a small family of GPCRs characterized by an extracellular hormone-binding domain. Members share between 21 and 67% of sequence identity;
- adhesion: is the second largest GPCR family in humans. Members display a GPCR proteolytic (GPS) domain;
- metabotropic glutammate: includes metabotropic glutamate receptors (GRMs), GABABRs, the calcium-sensing receptor (CASR) and taste receptors;
- class F (Frizzled/Smoothened): consists of frizzled receptors, smoothened receptor (SMO) and teste2 receptors;

GPCRs are topologically characterized by three elements (Figure 15): *1*) the extracellular region, consisting in the N terminus and three extracellular loops (EL1–EL3) and responsive for ligands' access modulation; *2*) the transmembrane region (TM), which spams the membrane with seven α-helices (TM1–TM7) and shape the orthosteric binding site; *3*) the intracellular C terminus, comprising three intracellular loops (IL1–IL3) and interfacing with the effector proteins.

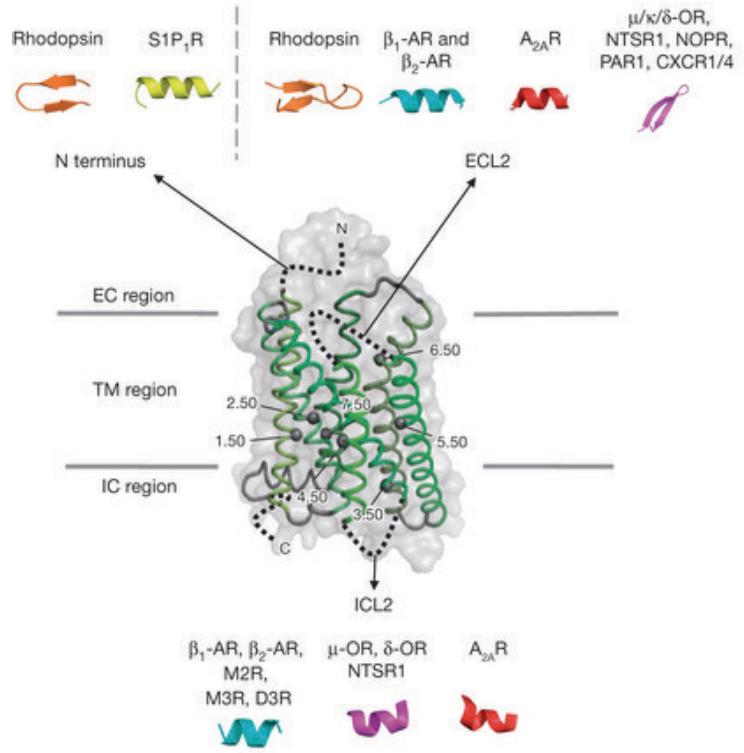

Figure 15. Generic structure of GPCRs. N terminus and extracellular loops are different, depending on the family. TM helices are shown as green cartoon. Amino acids are numbered using the Ballesteros/Weinstein numbering system. (Adapted from ref [146]).

Notably, in the extracellular region a conserved disulphide bridge links TM3 and EL2, stabilizing the overall protein structure limiting the degree of conformational changes of this region during receptor activation[147]. Further significant structural convergence is the 'ionic lock' between the highly conserved E/DRY motif on TM3 and a glutamate residue on TM6 (Figure 16): these amino acids form a network of interactions that stabilize the inactive-state conformation[148]. At the end of TM7 there is another important structural element, which participates in conformational changes associated with GPCR activation: the conserved NPXXY motif[149]: as evinced from X-ray crystal structures, a proline causes a distortion in the helical structure, allowing a tyrosine side chain to lie into a pocket formed by TM2, TM3, TM6 and TM7.

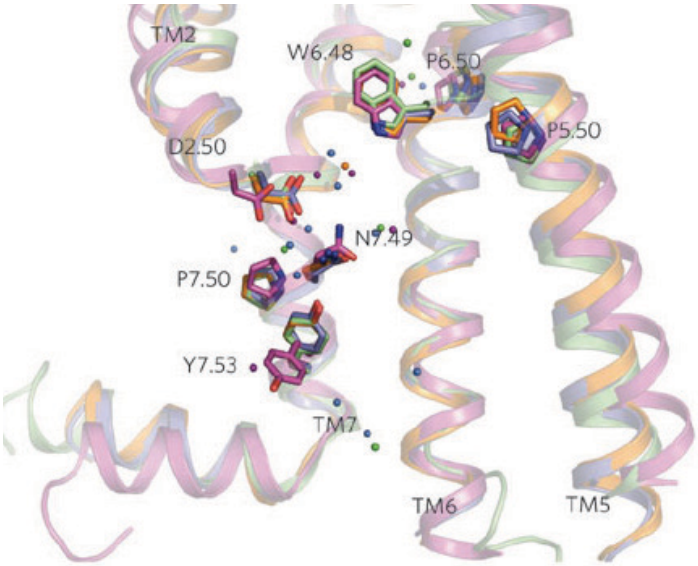

Figure 16. Location of some important conserved residues, probably part of a common pathway for propagating conformational changes from the orthosteric pocket to the G-protein coupling domains. Cluster of water molecules from (coloured spheres) are shown as spheres. Amino acids are numbered using the Ballesteros/Weinstein numbering system. (Adapted from ref [146]).

Variations in ELs, TMs helices, and side chains create a variety of orthosteric binding sites in the different GPCR subfamilies, reflecting the diversity of their corresponding ligands, as exemplified in Figure 17[150]. In general, binding sites are characterised by the presence of at least one key H-bonding or ionic interaction, as well as lipophilic hotspots, which are an important driving force of ligand interactions[151].

Experimental evidences outline that GPCRs exist in a dynamic equilibrium between inactive and active states (Figure 18) and the presence of the heterotrimeric G protein allow the receptors to reach the signaling state[150].

The energy difference between two states drives the transition probability between conformations (is a function of the energy barrier height between the two states). In absence of agonists, GPCRs have a basal activity[152]: energy changes occurring after ligand binding can modify the energy barrier between the two states or their stability (Figure 18). From this standpoint, inverse agonists shift the equilibrium toward inactive states, decreasing the level of basal

activity, whereas neutral antagonists do not affect the basal equilibrium. Binding of agonists shifts the equilibrium toward the activated states, which are characterized by large-scale conformational changes at the receptor's intracellular side[153].

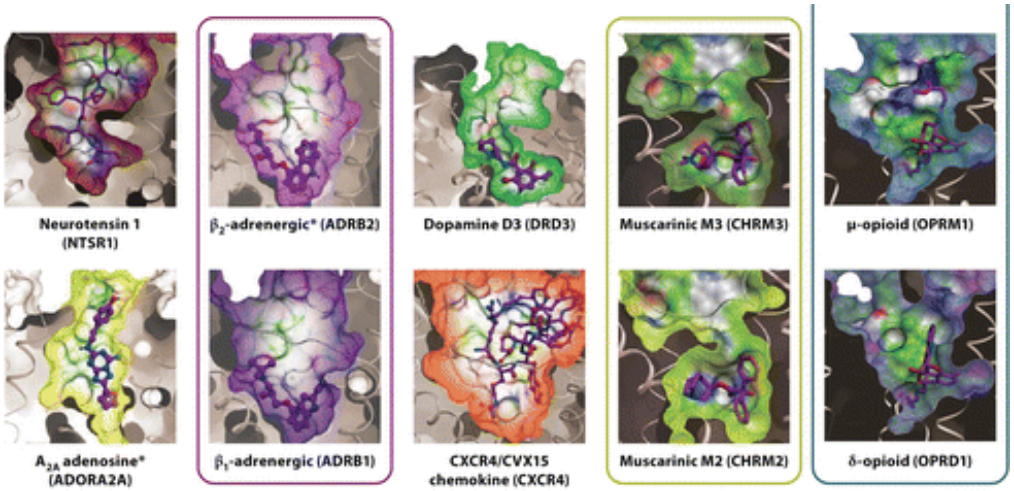

Figure 17. Examples of different orthosteric binding pockets in GPCRs. Complexes are shown as molecular surfaces; GPCRs subtypes with similar pockets are highlighted. (Adapted from ref [150]).

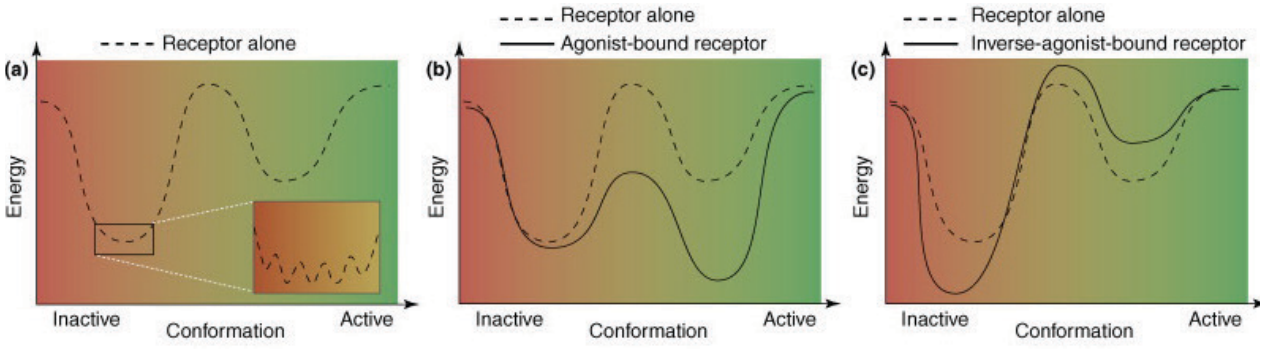

Figure 18. Simplified GPCRs energy landscape. (a) Conformational states of an apo GPCR; the inactive state is more stable (therefore more populated). (b) Agonist (or partial agonist) binding reduce the energy barrier and/or the energy of the active conformation. (c) Binding of an inverse agonist increases the energy barrier and/or reduces the energy of the inactive state conformation. (Adapted from ref [147]).

As shown in Figure 19, agonist–bound active conformation of the GPCRs can couples to heterotrimeric G proteins, which are formed by Gα, Gβ and Gγ subunits. Subsequently, GTP–bound Gα subunit separates from βγ and interacts with an effector, such as phospholipase C (PLC) or adenylyl cyclase (AC), initiating a second messenger cascade (usually cyclic AMP, inositol phosphates or Ca^{2+}) that triggers different cellular responses. The phosphorylation of activated receptors, primarily by GPCR kinases, enables GPCRs to bind to multifunctional scaffold proteins, called β-arrestins and able to promote internalization of the receptor. The GPCR can then be recycled back to the cell surface without ligand, restarting the cycle[154,155].

From a molecular point of view, GPCRs activation driven by agonist binding mainly involves structural rearrangements in TM6. In the adenosine A_{2A} subtype receptor (A_{2A} AR)[156] and the β_2 adrenoreceptor (β_2 AR)[157] the interactions between $Trp^{6.48}$ side chain and the agonist produce a shift of $Trp^{6.48}$ that stabilizes a swinging movement of helix VI (Figure 20, panel b) and allow the binding of the G protein (Figure 20, panel a). Moreover, at the cytoplasmic end of TM3, the $Arg^{3.50}$ of the DRY motif directly interacts with a the C terminus of the G protein[157].

More recently the pharmacologic concept of biased agonism has become increasingly popular. After agonist binding, GPCRs manly operate by recruiting the G proteins, however, interaction with other proteins can influence signalling events[158,159]. Indeed, ligands can preferentially trigger some particular effects, influencing therapeutic signalling pathways or, on the contrary, triggering averse effects. A structural and mechanistic understanding of these phenomena is not yet well understood, as crystallographic methods are not able to probe the equilibrium between different conformations. From this standpoint, molecular modelling approaches can lead to new insights into GPCR function[160] by means of molecular dynamics simulations, as for the observation of the β_2 AR active-to-inactive transitions[161].

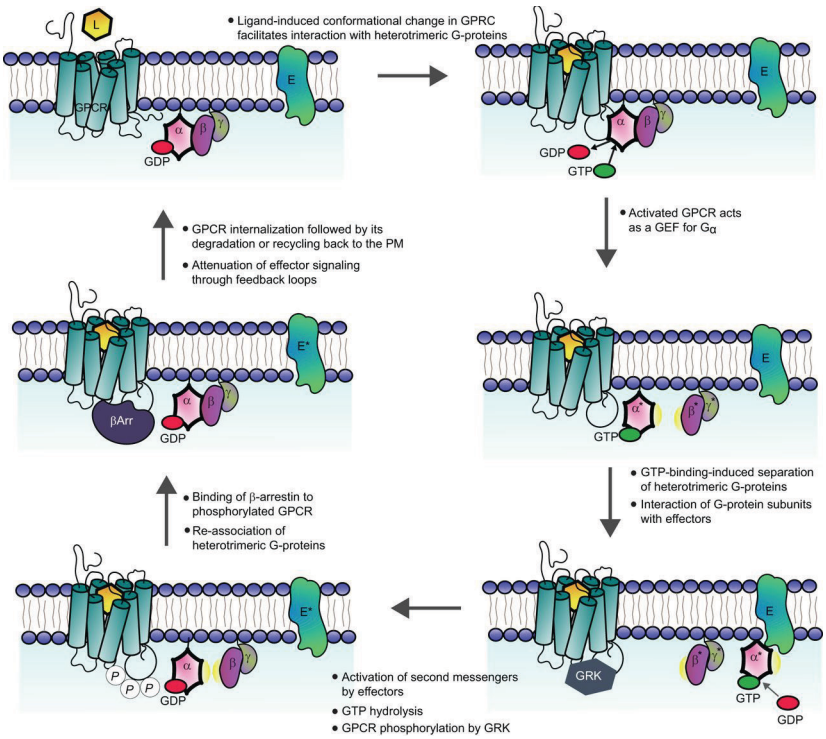

Figure 19. Activation cycle of GPCRs. Upon agonist binding, the receptor in the active conformation is able to bind the G protein, whose subunit Gα engage the effector protein triggering the second messenger cascade. β-arrestins are then responsible for GPCR internalization. (Adapted from ref [162]).

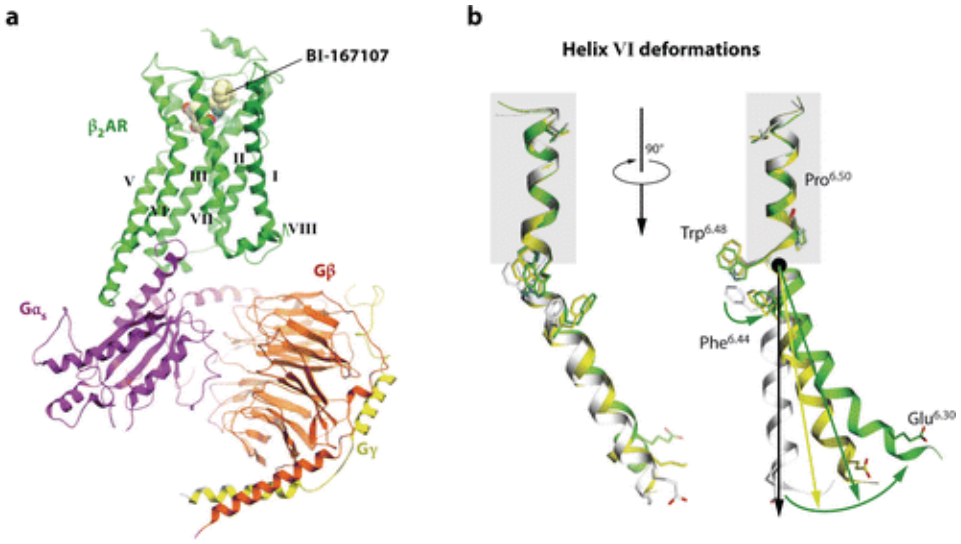

Figure 20. (a) Structure of β_2 AR in complex with agonist BI-167107 and the G protein heterotrimer (PDB code 3SN6 [157]). (b) Conformational changes in helix VI upon β_2AR activation. (Adapted from ref [150])

1.3.1 ADENOSINE RECEPTORS

The importance of the adenosine receptors' (ARs) pharmacology is daily experienced by million of coffee drinkers worldwide. Indeed, is well established that caffeine is able to non-selectively inhibit ARs subtypes[163] (named A_1, A_{2A}, A_{2B} and A_3 respectively) leading to a range of different biological responses and suggesting the potential usefulness of ARs agonists or blocking agents[164–166].

Each of these four receptors plays an essential role in responding to adenosine in the central nervous system, regulating pain, cerebral blood flow, basal ganglia functions, respiration and sleep. From a structural point of view, the sequence similarity and identity between human ARs is relatively high and can be ascribed mainly to the transmembrane regions (TMs) and the orthosteric binding sites (Table 1), while extracellular loops characterized by an higher degree of variability.

	Percentage of identity				Percentage of similarity			
Overall structure		A1	A2a	A2b	A3	A1	A2a	A2b	A3
	A1		38.1	44.0	49.4		48.5	59.6	70.4
	A2a	48.2		59.0	39.6	61.3		74.1	60.1
	A2b	44.8	47.6		37.4	60.7	59.7		56.0
	A3	48.2	30.6	35.8		68.7	46.4	53.6	
Orthosteric site		A1	A2a	A2b	A3	A1	A2a	A2b	A3
	A1		76.0	68.0	44.0		84.0	84.0	72.0
	A2a	76.0		76.0	52.0	84.0		88.0	72.0
	A2b	68.0	76.0		40.0	84.0	88.0		72.0
	A3	44.0	52.0	40.0		72.0	72.0	72.0	

Table 1. ARs identity and similarity of sequence, comparison among subtypes. In general, orthosteric binding sites (lower chart) have a higher degree of conservation.

Extracellular sources of adenosine are mainly linked to the release through a membrane transporter, a cell damage or the nucleotidase–mediated hydrolysis of extracellular adenine nucleotide[167] (Figure 21).

Adenosine-mediated activation of the A_1 AR inhibits adenylyl cyclase (AC) activity through G_i proteins[168] and results in increased activity of phospholipase C (PLC)[169]. In cardiac muscle and neurons, A_1 ARs activate pertussis toxin-sensitive K^+ channels (responsible for bradycardic effect[170]), as well as K_{ATP} and Ca^{2+} channels.

Adenosine binding to the A_{2A} AR increases AC activity through a stimulatory G_s protein, which seems to be associated with A_{2A}ARs in the peripheral systems but not in the striatum, where A_{2A}AR density is the highest but it mediate their effects mainly by activation of G_{olf}, which also activate the AC[171].

A_{2B} AR is coupled to both AC and PLC[172,173], which is responsible of many of the important functions of A_{2B} ARs[174]. Moreover, the arachidonic acid pathway was also demonstrated to be involved in A_{2B} AR activation[175].
Finally, the A_3AR pathway lead to the inhibition of AC[176], stimulation of PLC[177] and calcium mobilization. In cardiac cells, A_3 AR induce protection through the activation of K_{ATP} channels[178], while inhibition of proliferation was observed in human melanoma cells (via the ERK1/2 pathway)[179].

Among all the G-protein coupled receptors (GPCRs) superfamily members, A_{2A} AR represents a fortunate starting point for structure-based drug design (SBDD) as, to date, despite the intrinsic difficulties in GPCRs crystallography[180] 20 X-ray structures have already been resolved: more precisely, A_{2A} was disclosed in complex with both agonists[156,181,182] (included a recent structure bound to an engineered G protein[183]) and antagonists[184–189]. Development of more selective compounds for adenosine receptor subtypes could provide a class of therapeutics for treating numerous human diseases, such as pain, Parkinson's disease, asthma, seizures, and many other neurological disorders (Figure 22).

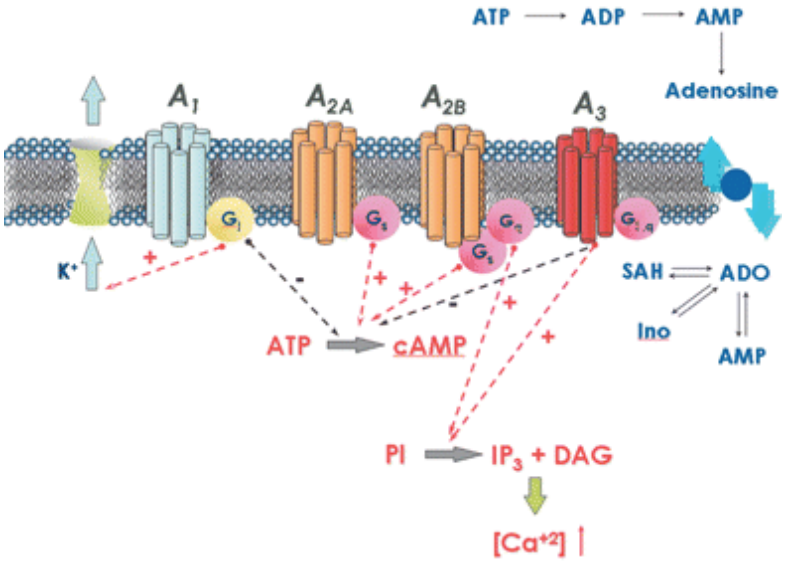

Figure 21. Schematic representation of ARs activation pathways, adenosine extracellular formation and adenosine intracellular metabolism. (Adapted from ref [190]).

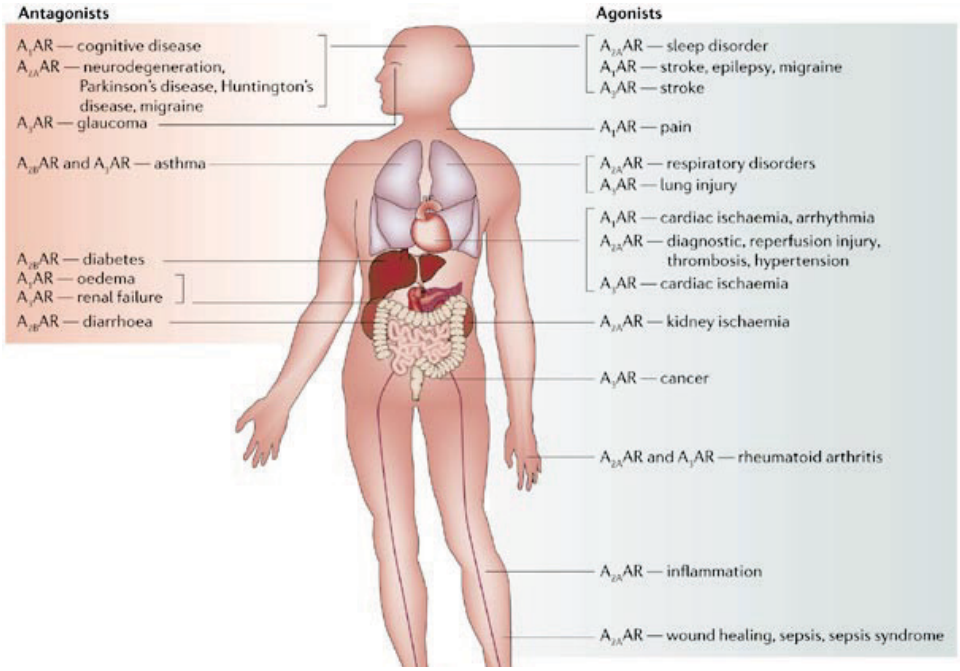

Fig. 22. Therapeutic potential applications of adenosine agonists and antagonists. A_1 AR is expressed in the brain, heart, adipose tissue, stomach, vas deferens, testis, spleen, kidney, aorta, liver, eye and bladder. A_{2A}AR is highly expressed in the central nervous system and in the immune cells, heart, lung and blood vessels. A_{2B}AR is located at low levels in almost all tissues. A_3ARs have been detected in various tissues including testis, lung, kidney, placenta, heart, brain, spleen and liver. (Adapted from ref [166]).

ARs, likewise many other GPCRs[191,192], can be selectively targeted with allosteric modulators. At molecular level, an allosteric modulator amplifies the action of agonists by stabilizing the intermolecular complex formed by adenosine and AR increasing the tissue specificity of the drug, that would act in concert with locally increased adenosine under pathological conditions, having little effect on sites where there are low basal adenosine levels. Recently an allosteric regulatory site for sodium was identified [189,193] and constant efforts are focused on the development of methods able to detect allosteric sites in ARs[100,194–196].

Bibliography

(1) Gilbert, W. Origin of Life: The RNA World. *Nature* **1986**, *319*, 618–618.

(2) Hippel, P. H. von. From "Simple" DNA-Protein Interactions to the Macromolecular Machines of Gene Expression. *Annu Rev Biophys Biomol Struct* **2007**, *36*, 79–105.

(3) Gether, U. Uncovering Molecular Mechanisms Involved in Activation of G Protein-Coupled Receptors. *Endocr Rev* **2000**, *21*, 90–113.

(4) Garcia-Viloca, M.; Gao, J.; Karplus, M.; Truhlar, D. G. How Enzymes Work: Analysis by Modern Rate Theory and Computer Simulations. *Science* **2004**, *303*, 186–195.

(5) Held, M.; Noé, F. Calculating Kinetics and Pathways of Protein-Ligand Association. *Eur J Cell Biol* **2012**, *91*, 357–364.

(6) Copeland, R. A. The Drug-Target Residence Time Model: A 10-Year Retrospective. *Nat Rev Drug Discov* **2016**, *15*, 87–95.

(7) Radić, Z.; Kirchhoff, P. D.; Quinn, D. M.; McCammon, J. A.; Taylor, P. Electrostatic Influence on the Kinetics of Ligand Binding to Acetylcholinesterase. Distinctions between Active Center Ligands and Fasciculin. *J Biol Chem* **1997**, *272*, 23265–23277.

(8) 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.

(9) Hothersall, J. D.; Brown, A. J.; Dale, I.; Rawlins, P. Can Residence Time Offer a Useful Strategy to Target Agonist Drugs for Sustained GPCR Responses? *Drug Discov Today* **2016**, *21*, 90–96.

(10) Dahl, G.; Akerud, T. Pharmacokinetics and the Drug-Target Residence Time Concept. *Drug Discov Today* **2013**, *18*, 697–707.

(11) Vauquelin, G.; Bostoen, S.; Vanderheyden, P.; Seeman, P. Clozapine, Atypical Antipsychotics, and the Benefits of Fast-off D2 Dopamine Receptor Antagonism. *Naunyn Schmiedebergs Arch Pharmacol* **2012**, *385*, 337–372.

(12) Hulme, E. C.; Trevethick, M. A. Ligand Binding Assays at Equilibrium: Validation and Interpretation. *Br J Pharmacol* **2010**, *161*, 1219–1237.

(13) Eyring, H. The Activated Complex in Chemical Reactions. *J. Chem. Phys.* **1935**, *3*, 107.

(14) Laidler, K. J.; King, M. C. Development of Transition-State Theory. *J Phys Chem* **1983**, *87*, 2657–2664.

(15) Tiwary, P.; Berne, B. J. Spectral Gap Optimization of Order Parameters for Sampling Complex Molecular Systems. *Proc Natl Acad Sci U S A* **2016**, *113*, 2839–2844.

(16) Best, R. B.; Hummer, G. Reaction Coordinates and Rates from Transition Paths. *Proc Natl Acad Sci U S A* **2005**, *102*, 6732–6737.

(17) Tiwary, P.; Berne, B. J. How Wet Should Be the Reaction Coordinate for Ligand Unbinding? *J Chem Phys* **2016**, *145*, 054113.

(18) Buch, I.; Giorgino, T.; De Fabritiis, G. Complete Reconstruction of an Enzyme-Inhibitor Binding Process by Molecular Dynamics Simulations. *Proc Natl Acad Sci U S A* **2011**, *108*, 10184–10189.

(19) Stanley, N.; Pardo, L.; Fabritiis, G. D. The Pathway of Ligand Entry from the Membrane Bilayer to a Lipid G Protein-Coupled Receptor. *Sci Rep* **2016**, *6*, 22639.

(20) Dror, R. O.; Pan, A. C.; Arlow, D. H.; Borhani, D. W.; Maragakis, P.; Shan, Y.; Xu, H.; Shaw, D. E. Pathway and Mechanism of Drug Binding to G-Protein-Coupled Receptors. *Proc Natl Acad Sci U S A* **2011**, *108*, 13118–13123.

(21) Guo, D.; Pan, A. C.; Dror, R. O.; Mocking, T.; Liu, R.; Heitman, L. H.; Shaw, D. E.; IJzerman, A. P. Molecular Basis of Ligand Dissociation from the Adenosine A2A Receptor. *Mol Pharmacol* **2016**, *89*, 485–491.

(22) Lu, H.; Tonge, P. J. Drug-Target Residence Time: Critical Information for Lead Optimization. *Curr Opin Chem Biol* **2010**, *14*, 467–474.

(23) Hoffmann, C.; Castro, M.; Rinken, A.; Leurs, R.; Hill, S. J.; Vischer, H. F. Ligand Residence Time at G-Protein-Coupled Receptors-Why We Should Take Our Time To Study It. *Mol Pharmacol* **2015**, *88*, 552–560.

(24) Geschwindner, S.; Ulander, J.; Johansson, P. Ligand Binding Thermodynamics in Drug Discovery: Still a Hot Tip? *J Med Chem* **2015**, *58*, 6321–6335.

(25) Reynolds, C. H.; Holloway, M. K. Thermodynamics of Ligand Binding and Efficiency. *ACS Med Chem Lett* **2011**, *2*, 433–437.

(26) Joshi, R. R. Statistical Mechanics of Antibody-Antigen Binding: Affinity Analysis. *Physica A: Statistical Mechanics and its Applications* **1995**, *218*, 214–228.

(27) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. The Statistical-Thermodynamic Basis for Computation of Binding Affinities: A Critical Review. *Biophys J* **1997**, *72*, 1047–1069.

(28) Sheinerman, F. B.; Norel, R.; Honig, B. Electrostatic Aspects of Protein-Protein Interactions. *Curr Opin Struct Biol* **2000**, *10*, 153–159.

(29) Schutz, C. N.; Warshel, A. What Are the Dielectric□ ?constants? Of Proteins and How to Validate Electrostatic Models? *Proteins* **2001**, *44*, 400–417.

(30) Li, L.; Li, C.; Zhang, Z.; Alexov, E. On the Dielectric "Constant" of Proteins: Smooth Dielectric Function for Macromolecular Modeling and Its Implementation in DelPhi. *J Chem Theory Comput* **2013**, *9*, 2126–2136.

(31) Kumar, S.; Nussinov, R. Close-Range Electrostatic Interactions in Proteins. *ChemBioChem* **2002**.

(32) Boobbyer, D. N.; Goodford, P. J.; McWhinnie, P. M.; Wade, R. C. New Hydrogen-Bond Potentials for Use in Determining Energetically Favorable Binding Sites on Molecules of Known Structure. *J Med Chem* **1989**, *32*, 1083–1094.

(33) Collins, K. D. Why Continuum Electrostatics Theories Cannot Explain Biological Structure, Polyelectrolytes or Ionic Strength Effects in Ion-Protein Interactions. *Biophys Chem* **2012**, *167*, 43–59.

(34) Nilsson, L. M.; Thomas, W. E.; Sokurenko, E. V.; Vogel, V. Beyond Induced-Fit Receptor-Ligand Interactions: Structural Changes That Can Significantly Extend Bond Lifetimes. *Structure* **2008**, *16*, 1047–1058.

(35) Gilli, G.; Gilli, P. Towards an Unified Hydrogen-Bond Theory. *J Mol Struct* **2000**, *552*, 1–15.

(36) Silverstein, K. A. T.; Haymet, A. D. J.; Dill, K. A. The Strength of Hydrogen Bonds in Liquid Water and Around Nonpolar Solutes. *J Am Chem Soc* **2000**, *122*, 8037–8041.

(37) Schmidtke, P.; Luque, F. J.; Murray, J. B.; Barril, X. Shielded Hydrogen Bonds as Structural Determinants of Binding Kinetics: Application in Drug Design. *J Am Chem Soc* **2011**, *133*, 18903–18910.

(38) Hyre, D. E.; Amon, L. M.; Penzotti, J. E.; Le Trong, I.; Stenkamp, R. E.; Lybrand, T. P.; Stayton, P. S. Early Mechanistic Events in Biotin Dissociation

from Streptavidin. *Nat Struct Biol* **2002**, *9*, 582–585.

(39) Grubmüller, H.; Heymann, B.; Tavan, P. Ligand Binding: Molecular Mechanics Calculation of the Streptavidin-Biotin Rupture Force. *Science* **1996**, *271*, 997–999.

(40) Meyer, E. E.; Rosenberg, K. J.; Israelachvili, J. Recent Progress in Understanding Hydrophobic Interactions. *Proc Natl Acad Sci U S A* **2006**, *103*, 15739–15746.

(41) Lennard-Jones, J. E. Cohesion. *Proceedings of the Physical Society* **1931**, *43*, 461–482.

(42) Kuntz, I. D.; Chen, K.; Sharp, K. A.; Kollman, P. A. The Maximal Affinity of Ligands. *Proc Natl Acad Sci U S A* **1999**, *96*, 9997–10002.

(43) Salonen, L. M.; Ellermann, M.; Diederich, F. Aromatic Rings in Chemical and Biological Recognition: Energetics and Structures. *Angew Chem Int Ed Engl* **2011**, *50*, 4808–4842.

(44) Gunaydin, H.; Bartberger, M. D. Stacking with No Planarity? *ACS Med Chem Lett* **2016**, *7*, 341–344.

(45) Ball, P. Water as an Active Constituent in Cell Biology. *Chem Rev* **2008**, *108*, 74–108.

(46) de Beer, S. B. A.; Vermeulen, N. P. E.; Oostenbrink, C. The Role of Water Molecules in Computational Drug Design. *Curr Top Med Chem* **2010**, *10*, 55–66.

(47) Ladbury, J. E. Just Add Water! The Effect of Water on the Specificity of Protein-Ligand Binding Sites and Its Potential Application to Drug Design. *Chem Biol* **1996**, *3*, 973–980.

(48) Dunitz, J. D. The Entropic Cost of Bound Water in Crystals and Biomolecules. *Science* **1994**, *264*, 670.

(49) Higgs, C.; Beuming, T.; Sherman, W. Hydration Site Thermodynamics Explain SARs for Triazolylpurines Analogues Binding to the A2A Receptor. *ACS Med Chem Lett* **2010**, *1*, 160–164.

(50) Mason, J. S.; Bortolato, A.; Weiss, D. R.; Deflorian, F.; Tehan, B.; Marshall, F. H. High End GPCR Design: Crafted Ligand Design and Druggability Analysis Using Protein Structure, Lipophilic Hotspots and Explicit Water Networks. *In silico pharmacology* **2013**, *1*, 23.

(51) Breiten, B.; Lockett, M. R.; Sherman, W.; Fujita, S.; Sayah, M. Al-; Lange, H.; Bowers, C. M.; Heroux, A.; Krilov, G.; Whitesides, G. M. Water Networks Contribute to Enthalpy/entropy Compensation in Protein-Ligand Binding. *J Am Chem Soc* **2013**, *135*, 15579–15584.

(52) Schönherr, H.; Cernak, T. Profound Methyl Effects in Drug Discovery and a Call for New C-H Methylation Reactions. *Angew Chem Int Ed Engl* **2013**, *52*, 12256–12267.

(53) Leung, C. S.; Leung, S. S. F.; Tirado-Rives, J.; Jorgensen, W. L. Methyl Effects on Protein-Ligand Binding. *J Med Chem* **2012**, *55*, 4489–4500.

(54) Klebe, G. Virtual Ligand Screening: Strategies, Perspectives and Limitations. *Drug Discov Today* **2006**, *11*, 580–594.

(55) Günther, J.; Bergner, A.; Hendlich, M.; Klebe, G. Utilising Structural Knowledge in Drug Design Strategies: Applications Using Relibase. *J Mol Biol* **2003**, *326*, 621–636.

(56) Lam, P. Y.; Jadhav, P. K.; Eyermann, C. J.; Hodge, C. N.; Ru, Y.; Bacheler, L. T.; Meek, J. L.; Otto, M. J.; Rayner, M. M.; Wong, Y. N. Rational Design of Potent, Bioavailable, Nonpeptide Cyclic Ureas as HIV Protease Inhibitors. *Science* **1994**, *263*, 380–384.

(57) Battistutta, R.; Mazzorana, M.; Cendron, L.; Bortolato, A.; Sarno, S.; Kazimierczuk, Z.; Zanotti, G.; Moro, S.; Pinna, L. A. The ATP-Binding Site of Protein Kinase CK2 Holds a Positive Electrostatic Area and Conserved Water Molecules. *Chembiochem* **2007**, *8*, 1804–1809.

(58) Bodnarchuk, M. S. Water, Water, Everywhere… It's Time to Stop and Think. *Drug Discov Today* **2016**, *21*, 1139–1146.

(59) Bortolato, A.; Tehan, B. G.; Bodnarchuk, M. S.; Essex, J. W.; Mason, J. S. Water Network Perturbation in Ligand Binding: Adenosine A(2A) Antagonists as a Case Study. *J Chem Inf Model* **2013**, *53*, 1700–1713.

(60) Ross, G. A.; Morris, G. M.; Biggin, P. C. Rapid and Accurate Prediction and Scoring of Water Molecules in Protein Binding Sites. *PLoS ONE* **2012**, *7*, e32036.

(61) Limongelli, V.; Marinelli, L.; Cosconati, S.; La Motta, C.; Sartini, S.; Mugnaini, L.; Da Settimo, F.; Novellino, E.; Parrinello, M. Sampling Protein Motion and Solvent Effect during Ligand Binding. *Proc Natl Acad Sci U S A* **2012**, *109*, 1467–1472.

(62) 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.

(63) Abel, R.; Young, T.; Farid, R.; Berne, B. J.; Friesner, R. A. Role of the Active-Site Solvent in the Thermodynamics of Factor Xa Ligand Binding. *J Am Chem Soc* **2008**, *130*, 2817–2831.

(64) Horbert, R.; Pinchuk, B.; Johannes, E.; Schlosser, J.; Schmidt, D.; Cappel, D.; Totzke, F.; Schächtele, C.; Peifer, C. Optimization of Potent DFG-in Inhibitors of Platelet Derived Growth Factor Receptorβ (PDGF-Rβ) Guided by Water Thermodynamics. *J Med Chem* **2015**, *58*, 170–182.

(65) 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.

(66) Dill, K. A.; Chan, H. S. From Levinthal to Pathways to Funnels. *Nat Struct Biol* **1997**, *4*, 10–19.

(67) Bryngelson, J. D.; Wolynes, P. G. Intermediates and Barrier Crossing in a Random Energy Model (with Applications to Protein Folding). *J Phys Chem* **1989**, *93*, 6902–6915.

(68) Kumar, S.; Ma, B.; Tsai, C. J.; Sinha, N.; Nussinov, R. Folding and Binding Cascades: Dynamic Landscapes and Population Shifts. *Protein Sci* **2000**, *9*, 10–19.

(69) Vogt, A. D.; Pozzi, N.; Chen, Z.; Di Cera, E. Essential Role of Conformational Selection in Ligand Binding. *Biophys Chem* **2014**, *186*, 13–21.

(70) Hammes, G. G.; Chang, Y.-C.; Oas, T. G. Conformational Selection or Induced Fit: A Flux Description of Reaction Mechanism. *Proc Natl Acad Sci U S A* **2009**, *106*, 13737–13741.

(71) Vogt, A. D.; Di Cera, E. Conformational Selection or Induced Fit? A Critical Appraisal of the Kinetic Mechanism. *Biochemistry* **2012**, *51*, 5894–5902.

(72) Pargellis, C.; Tong, L.; Churchill, L.; Cirillo, P. F.; Gilmore, T.; Graham, A. G.; Grob, P. M.; Hickey, E. R.; Moss, N.; Pav, S.; Regan, J. Inhibition of p38 MAP Kinase by Utilizing a Novel Allosteric Binding Site. *Nat Struct Biol* **2002**, *9*, 268–272.

(73) Frederick, K. K.; Marlow, M. S.; Valentine, K. G.; Wand, A. J. Conformational Entropy in Molecular Recognition by Proteins. *Nature* **2007**, *448*, 325–329.

(74) Igumenova, T. I.; Frederick, K. K.; Wand, A. J. Characterization of the Fast Dynamics of Protein Amino Acid Side Chains Using NMR Relaxation in

Solution. *Chem Rev* **2006**, *106*, 1672–1699.

(75) Karplus, M.; Kushick, J. N. Method for Estimating the Configurational Entropy of Macromolecules. *Macromolecules* **1981**, *14*, 325–332.

(76) DuBay, K. H.; Geissler, P. L. Calculation of Proteins' Total Side-Chain Torsional Entropy and Its Influence on Protein-Ligand Interactions. *J Mol Biol* **2009**, *391*, 484–497.

(77) Hruby, V. J. Designing Peptide Receptor Agonists and Antagonists. *Nat Rev Drug Discov* **2002**, *1*, 847–858.

(78) Anighoro, A.; la Vega de León, A. de; Bajorath, J. Predicting Bioactive Conformations and Binding Modes of Macrocycles. *J Comput Aided Mol Des* **2016**.

(79) Ryde, U. A Fundamental View of Enthalpy–entropy Compensation. *Medchemcomm* **2014**, *5*, 1324.

(80) Murray, C. W.; Verdonk, M. L. The Consequences of Translational and Rotational Entropy Lost by Small Molecules on Binding to Proteins. *J Comput Aided Mol Des*.

(81) Perutz, M. F.; Rossmann, M. G.; Cullis, A. F.; Muirhead, H.; Will, G.; North, A. C. T. Structure of Hæmoglobin: A Three-Dimensional Fourier Synthesis at 5.5-Å. Resolution, Obtained by X-Ray Analysis. *Nature* **1960**, *185*, 416–422.

(82) Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. Docking and Scoring in Virtual Screening for Drug Discovery: Methods and Applications. *Nat Rev Drug Discov* **2004**, *3*, 935–949.

(83) Chodera, J. D.; Mobley, D. L.; Shirts, M. R.; Dixon, R. W.; Branson, K.; Pande, V. S. Alchemical Free Energy Methods for Drug Discovery: Progress and Challenges. *Curr Opin Struct Biol* **2011**, *21*, 150–160.

(84) Genheden, S.; Ryde, U. The MM/PBSA and MM/GBSA Methods to Estimate Ligand-Binding Affinities. *Expert Opin Drug Discov* **2015**, *10*, 449–461.

(85) Tautermann, C. S. Impact, Determination and Prediction of Drug-Receptor Residence Times for GPCRs. *Curr Opin Pharmacol* **2016**, *30*, 22–26.

(86) Shaw, D. E.; Bowers, K. J.; Chow, E.; Eastwood, M. P.; Ierardi, D. J.; Klepeis, J. L.; Kuskin, J. S.; Larson, R. H.; Lindorff-Larsen, K.; Maragakis, P.; Moraes, M. A.; Dror, R. O.; Piana, S.; Shan, Y.; Towles, B.; Salmon, J. K.; Grossman, J. P.; Mackenzie, K. M.; Bank, J. A.; Young, C.; Deneroff, M. M.; Batson, B. Millisecond-Scale Molecular Dynamics Simulations on Anton. In *Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis - SC '09*; ACM Press: New York, New York, USA, 2009; p. 1.

(87) Shaw, D. E. Millisecond-Long Molecular Dynamics Simulations of Proteins on a Special-Purpose Machine. *Biophys J* **2013**, *104*, 45a.

(88) Sabbadin, D.; Ciancetta, A.; Moro, S. Bridging Molecular Docking to Membrane Molecular Dynamics to Investigate GPCR-Ligand Recognition: The Human A_{2A} Adenosine Receptor as a Key Study. *J Chem Inf Model* **2014**, *54*, 169–183.

(89) Alonso, H.; Bliznyuk, A. A.; Gready, J. E. Combining Docking and Molecular Dynamic Simulations in Drug Design. *Med Res Rev* **2006**, *26*, 531–568.

(90) Decherchi, S.; Berteotti, A.; Bottegoni, G.; Rocchia, W.; Cavalli, A. The Ligand Binding Mechanism to Purine Nucleoside Phosphorylase Elucidated via Molecular Dynamics and Machine Learning. *Nat Commun* **2015**, *6*, 6155.

(91) Overington, J. P.; Lazikani, B. Al-; Hopkins, A. L. How Many Drug Targets Are There? *Nat Rev Drug Discov* **2006**, *5*, 993–996.

(92) Chodera, J. D.; Noé, F. Markov State Models of Biomolecular Conformational

Dynamics. *Curr Opin Struct Biol* **2014**, *25*, 135–144.

(93) Pande, V. S.; Beauchamp, K.; Bowman, G. R. Everything You Wanted to Know about Markov State Models but Were Afraid to Ask. *Methods* **2010**, *52*, 99–105.

(94) Plattner, N.; Noé, F. Protein Conformational Plasticity and Complex Ligand-Binding Kinetics Explored by Atomistic Simulations and Markov Models. *Nat Commun* **2015**, *6*, 7653.

(95) Gu, S.; Silva, D.-A.; Meng, L.; Yue, A.; Huang, X. Quantitatively Characterizing the Ligand Binding Mechanisms of Choline Binding Protein Using Markov State Model Analysis. *PLoS Comput Biol* **2014**, *10*, e1003767.

(96) Kohlhoff, K. J.; Shukla, D.; Lawrenz, M.; Bowman, G. R.; Konerding, D. E.; Belov, D.; Altman, R. B.; Pande, V. S. Cloud-Based Simulations on Google Exacycle Reveal Ligand Modulation of GPCR Activation Pathways. *Nat Chem* **2014**, *6*, 15–21.

(97) 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* **2014**, *54*, 372–376.

(98) Cuzzolin, A.; Sturlese, M.; Deganutti, G.; Salmaso, V.; Sabbadin, D.; Ciancetta, A.; Moro, S. Deciphering the Complexity of Ligand-Protein Recognition Pathways Using Supervised Molecular Dynamics (SuMD) Simulations. *J Chem Inf Model* **2016**, *56*, 687–705.

(99) Sabbadin, D.; Ciancetta, A.; Deganutti, G.; Cuzzolin, A.; Moro, S. Exploring the Recognition Pathway at the Human A2A Adenosine Receptor of the Endogenous Agonist Adenosine Using Supervised Molecular Dynamics Simulations. *Medchemcomm* **2015**, *6*, 1081–1085.

(100) 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* **2015**, *23*, 4065–4071.

(101) Sinko, W.; Miao, Y.; de Oliveira, C. A. F.; McCammon, J. A. Population Based Reweighting of Scaled Molecular Dynamics. *J Phys Chem B* **2013**, *117*, 12759–12768.

(102) Mollica, L.; Theret, I.; Antoine, M.; Perron-Sierra, F.; Charton, Y.; Fourquez, J.-M.; Wierzbicki, M.; Boutin, J. A.; Ferry, G.; Decherchi, S.; Bottegoni, G.; Ducrot, P.; Cavalli, A. Molecular Dynamics Simulations and Kinetic Measurements to Estimate and Predict Protein-Ligand Residence Times. *J Med Chem* **2016**, *59*, 7167–7176.

(103) Mollica, L.; Decherchi, S.; Zia, S. R.; Gaspari, R.; Cavalli, A.; Rocchia, W. Kinetics of Protein-Ligand Unbinding via Smoothed Potential Molecular Dynamics Simulations. *Sci Rep* **2015**, *5*, 11539.

(104) Hamelberg, D.; Mongan, J.; McCammon, J. A. Accelerated Molecular Dynamics: A Promising and Efficient Simulation Method for Biomolecules. *J Chem Phys* **2004**, *120*, 11919–11929.

(105) Pierce, L. C. T.; Salomon-Ferrer, R.; Augusto F de Oliveira, C.; McCammon, J. A.; Walker, R. C. Routine Access to Millisecond Time Scale Events with Accelerated Molecular Dynamics. *J Chem Theory Comput* **2012**, *8*, 2997–3002.

(106) Kastner, K. W.; Izaguirre, J. A. Accelerated Molecular Dynamics Simulations of the Octopamine Receptor Using GPUs: Discovery of an Alternate Agonist-Binding Position. *Proteins* **2016**, *84*, 1480–1489.

(107) Kappel, K.; Miao, Y.; McCammon, J. A. Accelerated Molecular Dynamics Simulations of Ligand Binding to a Muscarinic G-Protein-Coupled Receptor. *Q Rev Biophys* **2015**, *48*, 479–487.

(108) Sugita, Y.; Okamoto, Y. Replica-Exchange Molecular Dynamics Method for Protein Folding. *Chem Phys Lett* **1999**, *314*, 141–151.

(109) Zhang, W.; Chen, J. Efficiency of Adaptive Temperature-Based Replica Exchange for Sampling Large-Scale Protein Conformational Transitions. *J Chem Theory Comput* **2013**, *9*, 2849–2856.

(110) Fukunishi, H.; Watanabe, O.; Takada, S. On the Hamiltonian Replica Exchange Method for Efficient Sampling of Biomolecular Systems: Application to Protein Structure Prediction. *J. Chem. Phys.* **2002**, *116*, 9058.

(111) Wang, K.; Chodera, J. D.; Yang, Y.; Shirts, M. R. Identifying Ligand Binding Sites and Poses Using GPU-Accelerated Hamiltonian Replica Exchange Molecular Dynamics. *J Comput Aided Mol Des* **2013**, *27*, 989–1007.

(112) Perilla, J. R.; Woolf, T. B. Towards the Prediction of Order Parameters from Molecular Dynamics Simulations in Proteins. *J Chem Phys* **2012**, *136*, 164101.

(113) Trbovic, N.; Kim, B.; Friesner, R. A.; Palmer, A. G. Structural Analysis of Protein Dynamics by MD Simulations and NMR Spin-Relaxation. *Proteins* **2008**, *71*, 684–694.

(114) Laio, A.; Parrinello, M. Escaping Free-Energy Minima. *Proc Natl Acad Sci U S A* **2002**, *99*, 12562–12566.

(115) Gervasio, F. L.; Laio, A.; Parrinello, M. Flexible Docking in Solution Using Metadynamics. *J Am Chem Soc* **2005**, *127*, 2600–2607.

(116) Barducci, A.; Bonomi, M.; Parrinello, M. Metadynamics. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2011**, *1*, 826–843.

(117) Laio, A.; Rodriguez-Fortea, A.; Gervasio, F. L.; Ceccarelli, M.; Parrinello, M. Assessing the Accuracy of Metadynamics. *J Phys Chem B* **2005**, *109*, 6714–6721.

(118) Branduardi, D.; Bussi, G.; Parrinello, M. Metadynamics with Adaptive Gaussians. *J Chem Theory Comput* **2012**, *8*, 2247–2254.

(119) Laio, A.; Gervasio, F. L. Metadynamics: A Method to Simulate Rare Events and Reconstruct the Free Energy in Biophysics, Chemistry and Material Science. *Reports on Progress in Physics* **2008**, *71*, 126601.

(120) Barducci, A.; Bussi, G.; Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys Rev Lett* **2008**, *100*, 020603.

(121) Limongelli, V.; Bonomi, M.; Marinelli, L.; Gervasio, F. L.; Cavalli, A.; Novellino, E.; Parrinello, M. Molecular Basis of Cyclooxygenase Enzymes (COXs) Selective Inhibition. *Proc Natl Acad Sci U S A* **2010**, *107*, 5411–5416.

(122) Limongelli, V.; Bonomi, M.; Parrinello, M. Funnel Metadynamics as Accurate Binding Free-Energy Method. *Proc Natl Acad Sci U S A* **2013**, *110*, 6358–6363.

(123) Tiwary, P.; Parrinello, M. From Metadynamics to Dynamics. *Phys Rev Lett* **2013**, *111*, 230602.

(124) Salvalaglio, M.; Tiwary, P.; Parrinello, M. Assessing the Reliability of the Dynamics Reconstructed from Metadynamics. *J Chem Theory Comput* **2014**, *10*, 1420–1425.

(125) 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.

(126) Piana, S.; Laio, A. A Bias-Exchange Approach to Protein Folding. *J Phys Chem B* **2007**, *111*, 4553–4559.

(127) 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.

(128) Abstract. Steered Molecular Dynamics (SMD) Induces Unbinding of Ligands and.

(129) Isralewitz, B.; Gao, M.; Schulten, K. Steered Molecular Dynamics and Mechanical Functions of Proteins. *Curr Opin Struct Biol* **2001**, *11*, 224–230.

(130) Shen, L.; Shen, J.; Luo, X.; Cheng, F.; Xu, Y.; Chen, K.; Arnold, E.; Ding, J.; Jiang, H. Steered Molecular Dynamics Simulation on the Binding of NNRTI to HIV-1 RT. *Biophys J* **2003**, *84*, 3547–3563.

(131) Favia, A. D.; Masetti, M.; Recanatini, M.; Cavalli, A. Substrate Binding Process and Mechanistic Functioning of Type 1 11β-Hydroxysteroid Dehydrogenase from Enhanced Sampling Methods. *PLoS ONE* **2011**, *6*, e25375.

(132) Palermo, G.; Minniti, E.; Greco, M. L.; Riccardi, L.; Simoni, E.; Convertino, M.; Marchetti, C.; Rosini, M.; Sissi, C.; Minarini, A.; De Vivo, M. An Optimized Polyamine Moiety Boosts the Potency of Human Type II Topoisomerase Poisons as Quantified by Comparative Analysis Centered on the Clinical Candidate F14512. *Chem Commun (Camb)* **2015**, *51*, 14310–14313.

(133) Patel, J. S.; Berteotti, A.; Ronsisvalle, S.; Rocchia, W.; Cavalli, A. Steered Molecular Dynamics Simulations for Studying Protein-Ligand Interaction in Cyclin-Dependent Kinase 5. *J Chem Inf Model* **2014**, *54*, 470–480.

(134) Lüdemann, S. K.; Lounnas, V.; Wade, R. C. How Do Substrates Enter and Products Exit the Buried Active Site of Cytochrome P450cam? 1. Random Expulsion Molecular Dynamics Investigation of Ligand Access Channels and Mechanisms. *J Mol Biol* **2000**, *303*, 797–811.

(135) Peräkylä, M. Ligand Unbinding Pathways from the Vitamin D Receptor Studied by Molecular Dynamics Simulations. *Eur Biophys J* **2009**, *38*, 185–198.

(136) Pan, D.; Niu, Y.; Ning, L.; Zhang, Y.; Liu, H.; Yao, X. Computational Study on the Binding and Unbinding Mechanism of HCV NS5B with the Inhibitor GS-461203 and Substrate Using Conventional and Steered Molecular Dynamics Simulations. *Chemometrics and Intelligent Laboratory Systems* **2016**, *156*, 72–80.

(137) Torrie, G. M.; Valleau, J. P. Nonphysical Sampling Distributions in Monte Carlo Free-Energy Estimation: Umbrella Sampling. *J Comput Phys* **1977**, *23*, 187–199.

(138) Bui, J. M.; Henchman, R. H.; McCammon, J. A. The Dynamics of Ligand Barrier Crossing inside the Acetylcholinesterase Gorge. *Biophys J* **2003**, *85*, 2267–2272.

(139) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. THE Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules. I. The Method. *J Comput Chem* **1992**, *13*, 1011–1021.

(140) Kästner, J. Umbrella Sampling. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2011**, *1*, 932–942.

(141) Yu, R.; Tabassum, N.; Jiang, T. Investigation of α-Conotoxin Unbinding Using Umbrella Sampling. *Bioorg Med Chem Lett* **2016**, *26*, 1296–1300.

(142) Maragliano, L.; Vanden-Eijnden, E. A Temperature Accelerated Method for Sampling Free Energy and Determining Reaction Pathways in Rare Events Simulations. *Chem Phys Lett* **2006**, *426*, 168–175.

(143) Fredriksson, R.; Lagerström, M. C.; Lundin, L.-G.; Schiöth, H. B. The G-Protein-Coupled Receptors in the Human Genome Form Five Main Families. Phylogenetic Analysis, Paralogon Groups, and Fingerprints. *Mol Pharmacol* **2003**, *63*, 1256–1272.

(144) Rosenbaum, D. M.; Rasmussen, S. G. F.; Kobilka, B. K. The Structure and Function of G-Protein-Coupled Receptors. *Nature* **2009**, *459*, 356–363.

(145) Lagerström, M. C.; Schiöth, H. B. Structural Diversity of G Protein-Coupled Receptors and Significance for Drug Discovery. *Nat Rev Drug Discov* **2008**, *7*, 339–357.

(146) Ballesteros, J. A.; Weinstein, H. [19] Integrated Methods for the Construction of Three-Dimensional Models and Computational Probing of Structure-Function Relations in G Protein-Coupled Receptors. In *Receptor Molecular Biology*; Methods in Neurosciences; Elsevier, 1995; Vol. 25, pp. 366–428.

(147) Venkatakrishnan, A. J.; Deupi, X.; Lebon, G.; Tate, C. G.; Schertler, G. F.; Babu, M. M. Molecular Signatures of G-Protein-Coupled Receptors. *Nature* **2013**, *494*, 185–194.

(148) Vogel, R.; Mahalingam, M.; Lüdeke, S.; Huber, T.; Siebert, F.; Sakmar, T. P. Functional Role of the "Ionic Lock"--an Interhelical Hydrogen-Bond Network in Family A Heptahelical Receptors. *J Mol Biol* **2008**, *380*, 648–655.

(149) Barak, L. S.; Ménard, L.; Ferguson, S. S.; Colapietro, A. M.; Caron, M. G. The Conserved Seven-Transmembrane Sequence NP(X)2,3Y of the G-Protein-Coupled Receptor Superfamily Regulates Multiple Properties of the Beta 2-Adrenergic Receptor. *Biochemistry* **1995**, *34*, 15407–15414.

(150) Katritch, V.; Cherezov, V.; Stevens, R. C. Structure-Function of the G Protein-Coupled Receptor Superfamily. *Annu Rev Pharmacol Toxicol* **2013**, *53*, 531–556.

(151) Cooke, R. M.; Brown, A. J. H.; Marshall, F. H.; Mason, J. S. Structures of G Protein-Coupled Receptors Reveal New Opportunities for Drug Discovery. *Drug Discov Today* **2015**, *20*, 1355–1364.

(152) Kobilka, B. K.; Deupi, X. Conformational Complexity of G-Protein-Coupled Receptors. *Trends Pharmacol Sci* **2007**, *28*, 397–406.

(153) Nygaard, R.; Frimurer, T. M.; Holst, B.; Rosenkilde, M. M.; Schwartz, T. W.

Ligand Binding and Micro-Switches in 7TM Receptor Structures. *Trends Pharmacol Sci* **2009**, *30*, 249–259.

(154) Kang, D. S.; Tian, X.; Benovic, J. L. Role of β-Arrestins and Arrestin Domain-Containing Proteins in G Protein-Coupled Receptor Trafficking. *Curr Opin Cell Biol* **2014**, *27*, 63–71.

(155) Tian, X.; Kang, D. S.; Benovic, J. L. β-Arrestins and G Protein-Coupled Receptor Trafficking. *Handb Exp Pharmacol* **2014**, *219*, 173–186.

(156) Xu, F.; Wu, H.; Katritch, V.; Han, G. W.; Jacobson, K. A.; Gao, Z.-G.; Cherezov, V.; Stevens, R. C. Structure of an Agonist-Bound Human A2A Adenosine Receptor. *Science* **2011**, *332*, 322–327.

(157) Rasmussen, S. G. F.; DeVree, B. T.; Zou, Y.; Kruse, A. C.; Chung, K. Y.; Kobilka, T. S.; Thian, F. S.; Chae, P. S.; Pardon, E.; Calinski, D.; Mathiesen, J. M.; Shah, S. T. A.; Lyons, J. A.; Caffrey, M.; Gellman, S. H.; Steyaert, J.; Skiniotis, G.; Weis, W. I.; Sunahara, R. K.; Kobilka, B. K. Crystal Structure of the β2 Adrenergic Receptor-Gs Protein Complex. *Nature* **2011**, *477*, 549–555.

(158) Rajagopal, S.; Rajagopal, K.; Lefkowitz, R. J. Teaching Old Receptors New Tricks: Biasing Seven-Transmembrane Receptors. *Nat Rev Drug Discov* **2010**, *9*, 373–386.

(159) Kenakin, T. Collateral Efficacy in Drug Discovery: Taking Advantage of the Good (allosteric) Nature of 7TM Receptors. *Trends Pharmacol Sci* **2007**, *28*, 407–415.

(160) Vilar, S.; Karpiak, J.; Berk, B.; Costanzi, S. In Silico Analysis of the Binding of Agonists and Blockers to the β2-Adrenergic Receptor. *J Mol Graph Model* **2011**, *29*, 809–817.

(161) Rosenbaum, D. M.; Zhang, C.; Lyons, J. A.; Holl, R.; Aragao, D.; Arlow, D. H.; Rasmussen, S. G. F.; Choi, H.-J.; Devree, B. T.; Sunahara, R. K.; Chae, P.

S.; Gellman, S. H.; Dror, R. O.; Shaw, D. E.; Weis, W. I.; Caffrey, M.; Gmeiner, P.; Kobilka, B. K. Structure and Function of an Irreversible Agonist-β(2) Adrenoceptor Complex. *Nature* **2011**, *469*, 236–240.

(162) Hanlon, C. D.; Andrew, D. J. Outside-in Signaling--a Brief Review of GPCR Signaling with a Focus on the Drosophila GPCR Family. *J Cell Sci* **2015**, *128*, 3533–3542.

(163) Rivera-Oliver, M.; Díaz-Ríos, M. Using Caffeine and Other Adenosine Receptor Antagonists and Agonists as Therapeutic Tools against Neurodegenerative Diseases: A Review. *Life Sci* **2014**, *101*, 1–9.

(164) Polosa, R.; Blackburn, M. R. Adenosine Receptors as Targets for Therapeutic Intervention in Asthma and Chronic Obstructive Pulmonary Disease. *Trends Pharmacol Sci* **2009**, *30*, 528–535.

(165) Stone, T. W.; Ceruti, S.; Abbracchio, M. P. Adenosine Receptors and Neurological Disease: Neuroprotection and Neurodegeneration. *Handb Exp Pharmacol* **2009**, 535–587.

(166) Jacobson, K. A.; Gao, Z.-G. Adenosine Receptors as Therapeutic Targets. *Nat Rev Drug Discov* **2006**, *5*, 247–264.

(167) Zimmermann, H. Extracellular Metabolism of ATP and Other Nucleotides. *Naunyn Schmiedebergs Arch Pharmacol* **2000**, *362*, 299–309.

(168) van Calker, D.; Müller, M.; Hamprecht, B. Adenosine Regulates via Two Different Types of Receptors, the Accumulation of Cyclic AMP in Cultured Brain Cells. *J Neurochem* **1979**, *33*, 999–1005.

(169) Rogel, A.; Bromberg, Y.; Sperling, O.; Zoref-Shani, E. Phospholipase C Is Involved in the Adenosine-Activated Signal Transduction Pathway Conferring Protection against Iodoacetic Acid-Induced Injury in Primary Rat Neuronal Cultures. *Neurosci Lett* **2005**, *373*, 218–221.

(170) Belardinelli, L.; Shryock, J. C.; Song, Y.; Wang, D.; Srinivas, M. Ionic Basis of the Electrophysiological Actions of Adenosine on Cardiomyocytes. *FASEB J* **1995**, *9*, 359–365.

(171) Kull, B.; Svenningsson, P.; Fredholm, B. B. Adenosine A(2A) Receptors Are Colocalized with and Activate G(olf) in Rat Striatum. *Mol Pharmacol* **2000**, *58*, 771–777.

(172) Peakman, M. C.; Hill, S. J. Adenosine A2B-Receptor-Mediated Cyclic AMP Accumulation in Primary Rat Astrocytes. *Br J Pharmacol* **1994**, *111*, 191–198.

(173) Feoktistov, I.; Biaggioni, I. Adenosine A2b Receptors Evoke Interleukin-8 Secretion in Human Mast Cells. An Enprofylline-Sensitive Mechanism with Implications for Asthma. *J Clin Invest* **1995**, *96*, 1979–1986.

(174) Linden, J.; Thai, T.; Figler, H.; Jin, X.; Robeva, A. S. Characterization of Human A(2B) Adenosine Receptors: Radioligand Binding, Western Blotting, and Coupling to G(q) in Human Embryonic Kidney 293 Cells and HMC-1 Mast Cells. *Mol Pharmacol* **1999**, *56*, 705–713.

(175) Donoso, M. V.; López, R.; Miranda, R.; Briones, R.; Huidobro-Toro, J. P. A2B Adenosine Receptor Mediates Human Chorionic Vasoconstriction and Signals through Arachidonic Acid Cascade. *Am J Physiol Heart Circ Physiol* **2005**, *288*, H2439–H2449.

(176) Zhou, Q. Y.; Li, C.; Olah, M. E.; Johnson, R. A.; Stiles, G. L.; Civelli, O. Molecular Cloning and Characterization of an Adenosine Receptor: The A3 Adenosine Receptor. *Proc Natl Acad Sci U S A* **1992**, *89*, 7432–7436.

(177) Abbracchio, M. P.; Brambilla, R.; Ceruti, S.; Kim, H. O.; Lubitz, D. K. von; Jacobson, K. A.; Cattabeni, F. G Protein-Dependent Activation of Phospholipase C by Adenosine A3 Receptors in Rat Brain. *Mol Pharmacol* **1995**, *48*, 1038–1045.

(178) Tracey, W. R.; Magee, W.; Masamune, H.; Oleynek, J. J.; Hill, R. J. Selective Activation of Adenosine A3 Receptors with N6-(3-Chlorobenzyl)-5'-N-Methylcarboxamidoadenosine (CB-MECA) Provides Cardioprotection via KATP Channel Activation. *Cardiovasc Res* **1998**, *40*, 138–145.

(179) Merighi, S.; Benini, A.; Mirandola, P.; Gessi, S.; Varani, K.; Leung, E.; Maclennan, S.; Borea, P. A. A3 Adenosine Receptor Activation Inhibits Cell Proliferation via Phosphatidylinositol 3-kinase/Akt-Dependent Inhibition of the Extracellular Signal-Regulated Kinase 1/2 Phosphorylation in A375 Human Melanoma Cells. *J Biol Chem* **2005**, *280*, 19516–19526.

(180) Ghosh, E.; Kumari, P.; Jaiman, D.; Shukla, A. K. Methodological Advances: The Unsung Heroes of the GPCR Structural Revolution. *Nat Rev Mol Cell Biol* **2015**, *16*, 69–81.

(181) Lebon, G.; Warne, T.; Edwards, P. C.; Bennett, K.; Langmead, C. J.; Leslie, A. G. W.; Tate, C. G. Agonist-Bound Adenosine A2A Receptor Structures Reveal Common Features of GPCR Activation. *Nature* **2011**, *474*, 521–525.

(182) Lebon, G.; Edwards, P. C.; Leslie, A. G. W.; Tate, C. G. Molecular Determinants of CGS21680 Binding to the Human Adenosine A2A Receptor. *Mol Pharmacol* **2015**, *87*, 907–915.

(183) Carpenter, B.; Nehmé, R.; Warne, T.; Leslie, A. G.; Tate, C. G. Structure of the Adenosine A2A Receptor Bound to an Engineered G Protein. *Nature* **2016**.

(184) Doré, A. S.; Robertson, N.; Errey, J. C.; Ng, I.; Hollenstein, K.; Tehan, B.; Hurrell, E.; Bennett, K.; Congreve, M.; Magnani, F.; Tate, C. G.; Weir, M.; Marshall, F. H. Structure of the Adenosine A(2A) Receptor in Complex with ZM241385 and the Xanthines XAC and Caffeine. *Structure* **2011**, *19*, 1283–1293.

(185) Jaakola, V.-P.; Griffith, M. T.; Hanson, M. A.; Cherezov, V.; Chien, E. Y. T.; Lane, J. R.; Ijzerman, A. P.; Stevens, R. C. The 2.6 Angstrom Crystal Structure

of a Human A2A Adenosine Receptor Bound to an Antagonist. *Science* **2008**, *322*, 1211–1217.

(186) Hino, T.; Arakawa, T.; Iwanari, H.; Yurugi-Kobayashi, T.; Ikeda-Suno, C.; Nakada-Nakura, Y.; Kusano-Arai, O.; Weyand, S.; Shimamura, T.; Nomura, N.; Cameron, A. D.; Kobayashi, T.; Hamakubo, T.; Iwata, S.; Murata, T. G-Protein-Coupled Receptor Inactivation by an Allosteric Inverse-Agonist Antibody. *Nature* **2012**, *482*, 237–240.

(187) Congreve, M.; Andrews, S. P.; Doré, A. S.; Hollenstein, K.; Hurrell, E.; Langmead, C. J.; Mason, J. S.; Ng, I. W.; Tehan, B.; Zhukov, A.; Weir, M.; Marshall, F. H. Discovery of 1,2,4-Triazine Derivatives as Adenosine A(2A) Antagonists Using Structure Based Drug Design. *J Med Chem* **2012**, *55*, 1898–1903.

(188) 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.; Cooke, R. M. Controlling the Dissociation of Ligands from the Adenosine A2A Receptor through Modulation of Salt Bridge Strength. *J Med Chem* **2016**, *59*, 6470–6479.

(189) Liu, W.; Chun, E.; Thompson, A. A.; Chubukov, P.; Xu, F.; Katritch, V.; Han, G. W.; Roth, C. B.; Heitman, L. H.; IJzerman, A. P.; Cherezov, V.; Stevens, R. C. Structural Basis for Allosteric Regulation of GPCRs by Sodium Ions. *Science* **2012**, *337*, 232–236.

(190) Moro, S.; Deflorian, F.; Spalluto, G.; Pastorin, G.; Cacciari, B.; Kim, S.-K.; Jacobson, K. A. Demystifying the Three Dimensional Structure of G Protein-Coupled Receptors (GPCRs) with the Aid of Molecular Modeling. *Chem. Commun.* **2003**, 2949.

(191) Nickols, H. H.; Conn, P. J. Development of Allosteric Modulators of GPCRs for Treatment of CNS Disorders. *Neurobiol Dis* **2014**, *61*, 55–71.

(192) Chatzidaki, A.; Millar, N. S. Allosteric Modulation of Nicotinic Acetylcholine

Receptors. *Biochem Pharmacol* **2015**, *97*, 408–417.

(193) Katritch, V.; Cherezov, V.; Stevens, R. C. Diversity and Modularity of G Protein-Coupled Receptor Structures. *Trends Pharmacol Sci* **2012**, *33*, 17–27.

(194) Nguyen, A. T.; Vecchio, E. A.; Thomas, T.; Nguyen, T. D.; Aurelio, L.; Scammells, P. J.; White, P. J.; Sexton, P. M.; Gregory, K. J.; May, L. T.; Christopoulos, A. The Role of the Second Extracellular Loop of the Adenosine A1 Receptor on Allosteric Modulator Binding, Signaling and Cooperativity. *Mol Pharmacol* **2016**.

(195) Massink, A.; Louvel, J.; Adlere, I.; van Veen, C.; Huisman, B. J. H.; Dijksteel, G. S.; Guo, D.; Lenselink, E. B.; Buckley, B. J.; Matthews, H.; Ranson, M.; Kelso, M.; IJzerman, A. P. 5'-Substituted Amiloride Derivatives as Allosteric Modulators Binding in the Sodium Ion Pocket of the Adenosine A2A Receptor. *J Med Chem* **2016**, *59*, 4769–4777.

(196) Guo, D.; Heitman, L. H.; IJzerman, A. P. Kinetic Aspects of the Interaction between Ligand and G Protein-Coupled Receptor: The Case of the Adenosine Receptors. *Chem Rev* **2016**.

2 AIM OF THE PROJECT

Computational chemists are constantly looking for new methods able to furnish a rationale for experimental data or to supply new hypothesis for applicative fields. From this standpoint, we have recently developed the supervised molecular dynamics (SuMD), a computational too able to gain insights at molecular level about ligands binding to macromolecules. Aims of this Ph.D project were to extensively validate the SuMD method on a wide range of endogenous targets, as well as to reconstruct the binding mechanisms of several adenosine receptor ligands. Better knowledge of the dynamic intermolecular recognition processes can allow propose structural modification in order to improve the kinetic behaviour of ligands as well as to decipher possible role of protein mutations in intermolecular recognitions.

3 SCIENTIFIC PUBBLICATIONS

The overall Ph.D. project was divided in different stages. A preliminary phase was focused on the application of the supervised molecular dynamics (SuMD) methodology: we tested this computational approach on a wide number of both cytosolic and membrane proteins with the aim of validate its ability to reconstruct X-ray determined intermolecular complexes as well as to elucidate the pathways the ligands putatively follow. This first part of the work resulted in the publication:

- "Deciphering the Complexity of Ligand-protein Recognition Pathways using Supervised Molecular Dynamics (SuMD) Simulations".

Following SuMD applications concerned the binding mechanisms of the endogenous effector adenosine (to the receptors subtype A_{2A}) and the A_3 positive allosteric modulator (PAM) LUF6000. Scientific publications related to these exertions are the following:

- "Exploring the recognition pathway at the human A_{2A} adenosine receptor of the endogenous agonist adenosine using supervised molecular dynamics simulations";

- "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".

In the last section of the chapter we introduce an overview on how to link different computational techniques, creating a complete pipeline able to facilitate the "hit to lead" compound optimization process. The related publication is:

- "New Trends in Inspecting GPCR-ligand Recognition Process: the Contribution of the Molecular Modeling Section (MMS) at the University of Padova".

3.1 Deciphering the Complexity of Ligand-protein Recognition Pathways using Supervised Molecular Dynamics (SuMD) Simulations.

Alberto Cuzzolin, Mattia Sturlese, Giuseppe Deganutti, Veronica Salmaso, Davide Sabbadin, Antonella Ciancetta and Stefano Moro*

Abstract

Molecular recognition is a crucial issue in interpreting the mechanism of known active substances as well as in the development of novel active candidates, since both thermodynamic and kinetic aspects greatly affect the understanding of ligand-mediated signal transmission in living organisms or whether a chemical compound can be transformed in a drug candidate. The physicochemical bases governing the optimization of thermodynamic aspects of ligand binding are relatively well understood, but they remain still poorly comprehend for binding kinetics. Unfortunately, simulating this binding process is still a challenging task because it requires classical MD experiments in a long microsecond time scale that is affordable only with a high-level computational capacity. In order to overcome this limiting factor, we have recently implemented an alternative MD approach, named supervised molecular dynamics (SuMD) specifically in the field of G protein-coupled receptors (GPCRs). SuMD enables the investigation of ligand-receptor binding events independently from the starting position, chemical structure of the ligand, and also from its receptor binding affinity.

In this Article, we would like to present an extension of SuMD application domain including different types of proteins compared to GPCRs. In particular, we decided to deeply analyze the ligand-protein recognition pathways of six different case studies that we grouped into two different classes: globular and membrane proteins. Moreover, we would like to introduce the SuMD-Analyzer tool that we have specifically implemented to help the user in the analysis of the SuMD trajectories. Finally, we will emphasize the limit of SuMD

Published - Cuzzolin A, Sturlese M, Deganutti G, Salmaso V, Sabbadin D, Ciancetta A, Moro S. *J Chem Inf Model* 56, 687-705 (2016)

applicability domain as well as its strengths in analyzing the complexity of ligand-protein recognition pathways.

Introduction

The essential features of ligand-protein interaction are very often summarized under the expression "molecular recognition" incorporating both thermodynamic aspects (quantified by the K_d, the equilibrium dissociation constant) and kinetic aspects of ligand binding (reflected by the rate constants k_{on} and k_{off}). Consequently, molecular recognition is thus a crucial issue in interpreting the mechanism of known active substances as well as in the development of novel active candidates, since both thermodynamic and kinetic aspects greatly affect the understanding of ligand-mediated signal transmission in living organisms or whether a chemical compound can be transformed in a drug candidate[1].

The physico-chemical bases governing the optimization of thermodynamic aspects of ligand binding are relatively well understood but, unluckily, they remain still poorly comprehend for binding kinetics. In fact, the equilibrium dissociation constant value depends on the free energy difference between the ligand-protein bound and unbound states, both of which are chemically stable and generally experimentally observable. On the contrary, k_{on} and k_{off} rate constants depend on the height of the free energy barrier separating those states and, in particular, the highest free energy barrier defined as transition state characterized only by a fleeting existence[2]. Consequently, the major challenge in the optimization of the kinetics parameters is the complexity in characterizing all plausible approaching pathways of the ligand to its protein. In fact, different approaching pathways can be characterized by different metastable intermediate states (referred also as meta-binding sites)[3] connected to each other, and to the final bound state, by different transition states.

Understanding the molecular interactions between ligand and protein during the approaching pathways is thus central to the deep understanding and to the rational control of ligand binding kinetics. Even though experimental

techniques for measuring the kinetic parameters of ligand binding have existed for decades, all of them only provide indirect evidence about transient structures visited along a ligand-binding pathway[2]. Alternatively, computational methods, and in particular molecular dynamics (MD) simulations, can provide detailed structural information on metastable intermediate states (meta-binding sites) and transition states at the atomistic level of detail[4]. Due to increases in computational power, it has recently become possible to simulate the full process of spontaneous ligand-protein association which typically occurs on the microsecond timescale, providing direct access to detailed information on binding mechanisms that have been difficult to access experimentally[4,5]. Unfortunately, simulating this binding process is still a challenging task because it requires classical MD experiments in a long microsecond time scale that is affordable only with a high-level computational capacity. However, the probability of reproduce ligand- protein binding or unbinding event on an accessible timescale can be enhanced through the introduction of biased potentials that facilitate the crossing of energy barriers or the application of external forces on the ligand, respectively[6]. An alternative strategy that does not require the introduction of biases or external forces and enables to explore the ligand-protein approaching path in nanosecond simulation time scale has been recently proposed by us specifically in the field of G protein-coupled receptors (GPCRs)[7,8] The "supervised molecular dynamics" (SuMD) approach exploit a tabu-like algorithm to monitor the distance between the center of masses of the ligand atoms and the protein binding site in short (600 ps) standard MD simulations (Figure 1, left panel). According to this strategy, an arbitrary number of distance points is collected "on the flight" at regular intervals and fitted into a linear function $f(x)=mx$. If the slope (m) is negative, the ligand-receptor distance is likely to be shortened and the simulation is restarted from the last set of coordinates. Otherwise, the simulation is restored from the original set of coordinates and started over. The supervision is repeated until the ligand-receptor distance is less than 5 Å. The results of a SuMD simulation are displayed in a graph reporting the interaction energy toward the distance

between the ligand and the binding site (Figure 1, right panel). We have recently applied the SuMD approach to interpret at the molecular level: *i)* the binding of different antagonists at the human A_{2A} adenosine receptor (hA_{2A} AR) by detecting and characterizing a possible energetically stable meta-binding site[7] , *ii)* the binding of the natural agonist adenosine at the hA_{2A} AR by detecting and characterizing a possible energetically stable meta-binding site[9], *iii)* the positive allosteric modulation mediated by LUF6000 toward the human A_3 adenosine receptor (hA_3 AR) by suggesting at least two possible mechanisms to explain the available experimental data[10], and *iv)* the binding of different ligands at the human P2Y12 receptor by detecting and characterizing again possible energetically stable meta-binding site[11].

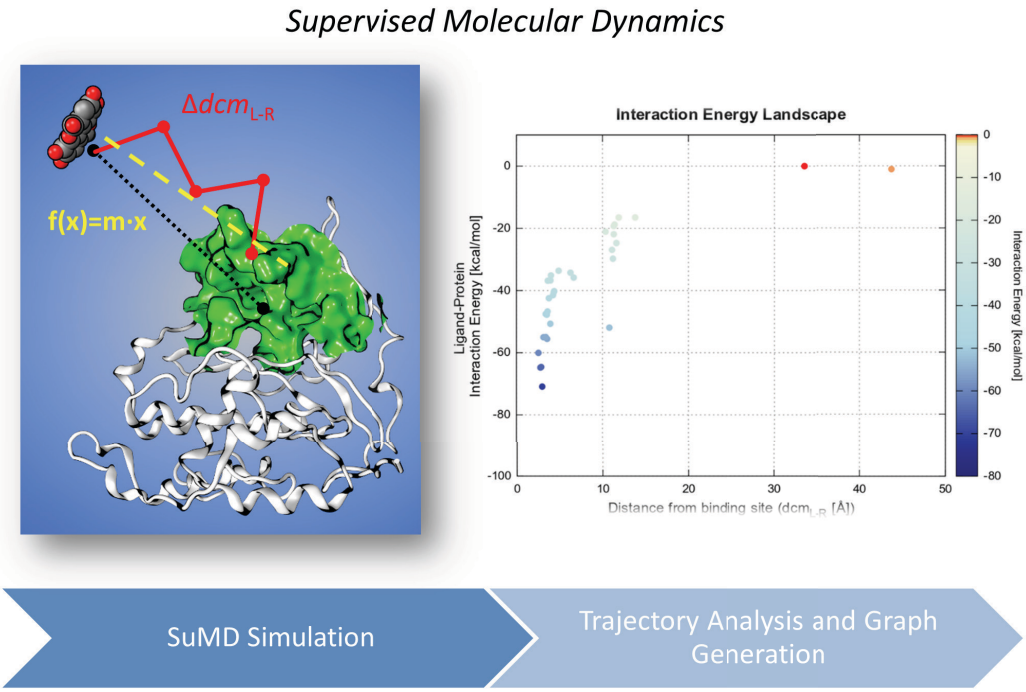

Figure 1. a) Schematic representation of Supervised Molecular Dynamics (SuMD) algorithm (left) and the outcoming ligand–protein interaction energy landscape. Interaction Energy values: kcal mol^{-1}.

In the present work, we would like to present an extension of SuMD application domain including different types of proteins compared to GPCRs. In particular, we decided to deeply analyzed the ligand-protein recognition

pathways of six different case studies that we grouped into two different classes of proteins: globular and membrane proteins, as summarized in Table 1. Moreover, we would like to introduce the SuMD-Analyzer tool that we have specifically implemented to help, also a non-expert user, in the analysis of the SuMD trajectories.

Table 1 - Structural summary of the selected ligand-protein PDB ID are reported

Globular System						
PDB	**Protein**	**Ligand**	**Resolution [Å]**	**Affinity**	**Ligand MW**	**Ref.**
2ZJW	**CK2**	**Ellagic Acid**	**2.40**	**K_i=0.04 µM**	**302.197**	**41**
13GS	**GSTP1-1**	**SASP**	**1.90**	**K_i=24 µM**	**398.39**	**44**
4K7I	**PRDX5**	**Benzen-1,2-diol**	**2.25**	**K_i=1500µM**	**110.11**	**45**
2VDB	**HSA**	**(S)-naproxen**	**2.25**	**K_i=1.2-1.8µM[1]**	**230.25**	**49,58**
Transmembrane Systems						
PDB	**Protein**	**Ligand**	**Resolution [Å]**	**Affinity**	**Ligand MW**	**Ref.**
3GWW	**LeuT**	**(S)-dluoxetine**	**2.46**	**IC_{50}=355mM**	**345.79**	**51**
2YDV	**$hA_{2A}AR$**	**NECA**	**2.60**	**K_i=13.8nM**	**308.29**	**55**

Materials and Methods

General.

All computations were performed on a hybrid CPU/GPU cluster. MD simulations were carried out with the ACEMD engine[12] on a GPU cluster

equipped with four NVIDIA GTX 580, two NVIDIA GTX 680, three NVIDIA GTX 780, and four NVIDIA GTX 980. Before running SuMD simulations, the following preliminary phases were carried out: *i)* protein-ligand system preparation; *ii)* ligand parameterization; *iii)* solvated system setup and equilibration. Two different protocols based on AMBER12[13]/General Amber Force Filed (GAFF)[14] and the CHARM27[15]/CHARMM General Force Field (CGenFF), force fields combinations were adopted for globular and transmembrane systems, respectively[16,17].

Systems Preparation.

Protein-ligand complexes were retrieved from the RCSB PDB database[18]. Proteins structures were prepared with the protein preparation tool as implemented in MOE[19]: hydrogen atoms were added to the complex and appropriate ionization states were assigned by means of the Protonate-3D tool[20]. Missing atoms in protein side chains were built according to either the AMBER12[13] or the CHARM27[15] force field topology. Missing loops were modeled by the default homology modelling protocol implemented in the MOE protein preparation tool. Non- natural N-terminal and C-terminal were capped to mimic the previous residue. For each considered system, the conformer with highest occupancy was selected whenever available. To avoid protein-ligand long range interactions in the starting geometry, the ligand was then moved at least 15 Å from any protein atom.

Ligand Parameterization.

Globular systems.

For the MD simulations based on the AMBER12 force field[13], the ligands were subjected to two energy minimization steps with MOPAC2012[21] using PM6 method[22] and Gaussian 09[23] (basis set: HF/6-31G*). After geometry minimization, ligand parameters were derived with GAFF[14] as implemented in ambertools2014[13] by using antechamber and parmchk tools. RESP partial

charges where calculated with Gaussian 09[23] following the procedure suggested by antechamber.

Transmembrane systems.

For the MD simulation based on the CHARMM27 force field[24], initial parameters for the ligands were retrieved from the paramchem service and subsequently optimized consistently to CGenFF[16,25] at the MP2/6-31G* level of theory[26] by using Gaussian 09[23] and the Force Field Toolkit[27] implemented in the VMD engine[28].

Solvated System Setup and Equilibration

Globular Systems.

Protein-ligand complexes were assembled with tleap tool using AMBER14SB[29] as force field for the protein[29]. The systems were explicitly solvated by a cubic water box with cell borders placed at least 12 Å away from any protein or ligand atom using TIP3P as water model[30]. To neutralize the total charge Na^+/Cl^- counter-ions were added to a final salt concentration of 0.150 M. The systems were energy minimized by 2000 step with conjugate-gradient method, then 50000 step of NVE (100 ps) followed by 1 ns of NPT simulation were carried out, both using 2 fs as time step and applying an harmonic positional constrain on protein and ligand atoms gradually reduced with a scaling factor of 0.1. Pressure was maintained at 1 atm using a Berendsen barostat[31]. The Langevin thermostat was set with a low damping constant of 1 ps^{-1} [32]. Bond lengths involving hydrogen atoms were constrained using the M-SHAKE algorithm[33]. The MD productive runs were conducted in a NVT ensemble. Long-range Coulomb interactions were handled using the particle mesh Ewald summation method (PME) setting the mesh spacing to 1.0 Å[34]. A non-bonded cut-off distance of 9 Å with a switching distance of 7.5 Å was used.

Transmembrane Systems.

Transmembrane proteins were embedded in a 1-palmitoyl-2- oleoyl-snglycero-3-phosphocholine (POPC) lipid bilayer according to the suggested orientation reported in the Orientations of Proteins in Membranes (OPM) database[35]. The systems were solvated with TIP3P[30] water using the program Solvate 1.0[36] and neutralized by Na^+/Cl^- counterions to a final concentration of 0.154 M. The systems were then equilibrated through a two steps procedure: in the first stage, after 2000 cycles of conjugate-gradient minimization algorithm (in order to reduce steric clashes produced by the system manual setting), 10 ns of MD simulation were performed in the NPT ensemble, restraining ligand and protein atoms by a force constant of 1 Kcal mol^{-1} $Å^{-2}$. The temperature was maintained at 298 K using a Langevin thermostat with a low damping constant of 1 ps^{-1} [32] pressure was maintained at 1 atm using a Berendsen barostat[31], bond lengths involving hydrogen atoms were constrained using the M-SHAKE algorithm[33] with an integration timestep of 2 fs. In the second stage, once water molecules diffused inside the protein cavity and the lipid bilayer reached equilibrium, the force constant was gradually reduced to 0.1 Kcal mol^{-1} $Å^{-2}$ for the next 10 ns of MD simulation.

Supervised Molecular Dynamics (SuMD)

SuMD is a command line tool written in python, tcl, and bash that operates the supervision of MD trajectories according to the algorithm that has been previously described[7]. The program exploits Visual Molecular Dynamics (VMD) and Gnuplot functionalities[28,37]. In its current implementation, SuMD is interfaced with the ACEMD[12] engine and supports AMBER and CHARMM force fields.

SuMD Input files.

SuMD requires a configuration file (selection.dat, Figure S1) organized in three major sections containing information about: *i)* the system; *ii)* the supervision procedure; and *iii)* the simulation settings. In the system settings

section, the following details about the molecular system need to be provided: *i)* the pdb file name containing the starting coordinates; *ii)* the 3-letter code name of the ligand; and *iii)* the residues describing the target binding site. In the supervision settings section, the following values are declared: *i)* the slope threshold (default value: 0); and *ii)* the number of maximum consecutive failed steps (default value: 33) to stop the simulation. In the simulation settings section, the following details must be specified: *i)* the force field to use; *ii)* the parameter file; *iii)* the GPU device ID to which the calculation will be addressed. In this section, a Boolean operator manages the introduction of a randomization step that varies the position of the ligand through a 600 ps of non-supervised MD simulation. In the same directory where SuMD is launched, a file containing the cell dimension as well as a parameter file (prmtop/psf with the same name of the pdb) must also be provided.

SuMD Main Code.

The workflow of the SuMD main code is reported in Figure 2A. As depicted, at the beginning of the simulation SuMD detects the atoms that identify the ligand and the target binding site, to define the distance between their mass centers $dcm_{(L\text{-}R)}$ that will be monitored. Then, a series of 600 ps classical MD simulations are performed. After each simulation, five $dcm_{(L\text{-}R)}$ distance points are collected at regular intervals of 75ps. Using these points, the slope value (*m*) is derived by a linear fitting. As previously described, if the resulting slope *m* is negative or below the user selected threshold (*i.e.* the distance $dcm_{(L\text{-}R)}$ is decreasing), the next simulation step starts from the last set of coordinates produced, otherwise the simulation is restarted by randomly assigning the atomic velocities. To avoid problematic starting geometries (*i.e.* geometries prone to lead to dead-end pathway), in the first simulation step, SuMD supervises the distance $dcm_{(L\text{-}R)}$ with a maximum threshold of 31 failed attempts (Preliminary Run). In the case this threshold is reached, SuMD callbacks a randomization process on the set of coordinates supplied by the user by a classical 600 ps MD simulation.

During the following steps, the simulations are perpetuated under the supervision rules. In particular, the first time a slope value below the threshold is recorded, the program enters the so- called "SuMD Run". When the distance $dcm_{(L-R)}$ drops below 5 Å the supervision is disabled and the simulation proceeds though a classical MD simulation. At the end of the simulation, only the productive steps are saved, chronologically numbered and stored in a separate directory.

SuMD log file.

At each SuMD simulation step, a log file (Figure 1S) is updated collecting information about: *i)* the step number; *ii)* the $dcm_{(L-R)}$ distance; *iii)* the slope value (*m*); *iv)* the electrostatic and van der Waals potential energy contributions of the ligand-receptor interaction energy (IE). A counter keeps trace on how many times each SuMD step has been attempted. Furthermore, three counters corresponding to the $dcm_{(L-R)}$ distance ranges 0-2 Å, 2-5 Å, and 5-9 Å are reported. These distances monitors how many times the binding site is approached, *i.e.* how often the $dcm_{(L-R)}$ distance lies below the long-range interaction cutoff. These counters determine the program termination criteria (see following section) and, according to the binding site definition supplied by the user, they might represent: the target binding site, its neighbors, and putative allosteric/meta-binding sites, respectively.

SuMD Termination Criteria.

A SuMD simulation is terminated when one of the following criteria is satisfied: *i)* no negative slope (*m*) values are recorded for a user-selected number of steps (33 consecutive steps by default); *ii)* one of the distance counters reaches a maximum value of 19 (*i.e.* the $dcm_{(L-R)}$ lies in the same region for 11.5 nonconsecutive nanoseconds).

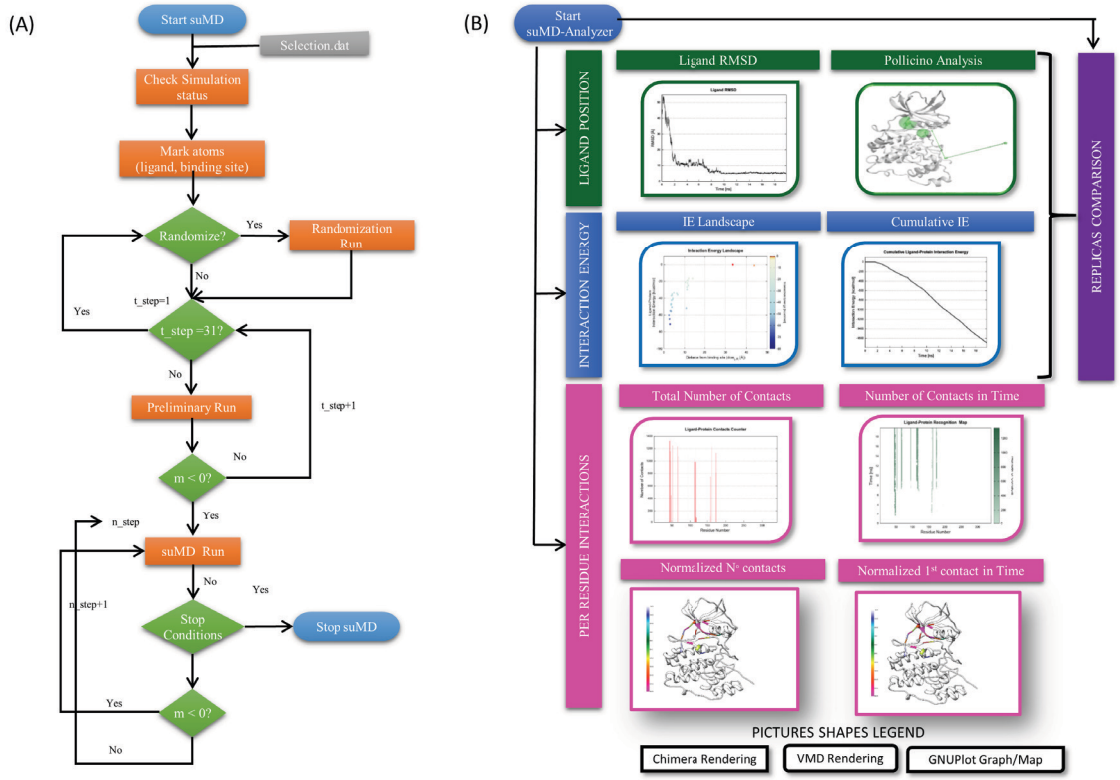

Figure 2. (A) Workflow of the SuMD main code (B) Workflow of the SuMD-Analyzer tool.

SuMD-Analyzer Tool

The SuMD-Analyzer is a standalone tool written in python, tcl, and bash to analyze the SuMD trajectories (Figure 2B). The tool is integrated with VMD[28] and UCSF Chimera[38] for the graphical visualization and exploits Wordom[39] and Gnuplot[37] functionalities. The provided analyses cross over four different aspects: *i)* the ligand position, *ii)* the IE, *iii)* the per residue interactions, and *iv)* the replicas comparison.

When the SuMD-Analyzer is launched, the trajectories produced by SuMD are merged and aligned to the starting reference structure using the RMSD tool in VMD by using alpha-carbon atoms for the superposition. The merged trajectory is subjected to a striding procedure picking one frame every 5 through the VMD animate module.

Ligand Position.

Two analyses follow the coordinates explored by the ligand during the SuMD trajectory (Figure 2B, green boxes): *i)* the Root Mean Square Deviation (RMSD); and *ii)* the so-called "Pollicino Analysis". If a reference complex structure is available, the RMSD between the ligand and the reference coordinates supplied is computed along the trajectory. The calculation is performed on the heavy atoms of the ligand using the measure rmsd function implemented in VMD and the data obtained are plotted against the time using Gnuplot[37] (Figure 2B, left green box).

The Pollicino Analysis is a representation that graphically renders the recognition pathway explored by the ligand. At the end of each SuMD step, the coordinates of the ligand mass center are collected and clustered according to their $dcm_{(L\text{-}R)}$ using a threshold value of 2 Å. The coordinates belonging to the same cluster are averaged and represented by a sphere which radius depends on the population of the cluster. Arrows indicate the chronological order onto which the regions where the sphere reside are approached by the ligand mass center (Figure 2B, right green 22 box).

Interaction Energy.

The ligand-protein interaction is analyzed by means of the mdenergy function embedded into VMD. The electrostatic and van der Waals contributions to the potential energy are calculated for each frame and summed to obtain the total IE. With this value, two graphs are derived (Figure 2B, blue boxes): *i)* the "Interaction Energy Landscape", and *ii)* the "Cumulative Interaction Energy". The former chart displays the total IE profile with respect to the $dcm_{(L\text{-}R)}$ through a colorimetric scale representing the IE value. Each point displayed in the chart represents the last position of the corresponding SuMD step (Figure 2B, left blue box). The latter plot shows the cumulative sum of the total IE values for each frame against the time. Therefore, each point is the sum of all previous IE values. Changes in the observed trend highlight how

the variation of ligand conformation/position affects the IE (Figure 2B, right blue box).

Per Residue Interactions.

A further set of analyses was developed to highlight the most important residues involved in the ligand recognition pathway (Figure 2B, upper magenta boxes): *i)* the "Protein-ligand Contacts Count", and the *ii)* "Ligand-Protein Recognition Map". In the first graph (Figure 2B, upper left magenta box), the residues more frequently approached by the ligand during the trajectory are reported and for each residue the total number of established contacts is rendered as histograms. In this representation, at each SuMD frame only the residues lying within a distance of 4 Å from any ligand atoms are considered. In the second graph (Figure 2B, upper right magenta box), the residue approached by the ligand are depicted with respect to the simulation time. In particular, each dot in the map represents a trajectory frame colored according to the total number of contacts the ligand has established with a particular residue. White dots means that, at the considered frame, the residue atoms are farther then 4 Å from ligand atoms, while green dots correspond to a contact event and the sum of the contact is coded by the light-green to dark-green scale.

To support the user in the topological localization of the residues mainly interacting with the ligand during the trajectory, molecular 3D representations of the protein are automatically set using USF Chimera[38] (Figure 2B, lower magenta boxes). In particular, the number of ligand- protein contacts is normalized and stored into the B-factor field of the involved residue in the protein pdb file. In the protein 3D representation "Chimera_count" (Figure 2B, lower left magenta box) the ribbons are colored according the so-derived B-factor values. A similar representation, "Chimera_time" (Figure 2B, lower right magenta box), is available with the color code (blue-to-violet) reflecting the chronological order onto which the residues have been approached by the ligand for the first time.

Replicas Analysis.

The "Replicas Analysis" (Figure 2B, violet box) is a manager that compares the molecular recognition event occurred in different SuMD replicas. The manager extracts from each trajectory the data regarding the ligand position and the IE, merges the data for each analysis in graphs colored according the replica number to better appreciate the differences.

Results and Discussion

Case Studies Selection

As already anticipated, in this work SuMD applicability domain has been extended using six different case studies, grouped into two major protein classes: *i)* globular systems, and *ii)* transmembrane systems (as summarized in Table 1). Specifically, considering the globular proteins we selected: *a)* the human Caseine Kinase 2 (CK2) in complex with Ellagic acid; *b)* the P1-1 isoform of Glutathione S-transferase (GSTP1-1) in complex with Sulphasalazine (2-hydroxy-(5-{[4-(2-pyridinylamino)sulfonyl]phenyl}azo) benzoic acid, SASP); *c)* the human Peroxiredoxin 5 (PRDX5) in complex with a benzen-1,2-diol; and *d)* the human Serum Albumin (HSA) in complex with (S)-naproxen. Considering the membrane proteins, we selected: *a)* the Leucine transporter (LeuT) from *Aquifex aeolicus* in complex with (S)-fluoxetine; and *b)* the human Adenosine A Receptor (hA_{2A} AR) in complex with the synthetic agonist 5'-N-Ethylcarboxamidoadenosine (NECA). An overview of the structural features of the considered ligand-protein complex is reported in Figure 3 and briefly described in the following.

CK2 is a ubiquitous and constitutively active serine/threonine kinase (PK) that phosphorylates more than 300 substrates. It is involved in the regulation of numerous cellular process such as cycle progression, apoptosis, transcription and viral infection[40]. The catalytic alpha subunit is composed by two lobes connected by a small loop called "hinge region". The N- terminal lobe presents five β-strands and the α-helix C involved in the substrate recognition,

whereas the C-terminal lobe is composed of α-helices. All PKs present a glycine rich loop (P-loop), an activation loop, and a catalytic loop[40]. The X-ray complex highlights that the inhibitor binds to Lys49, Ser51and His160 as shown in Figure 3A[41].

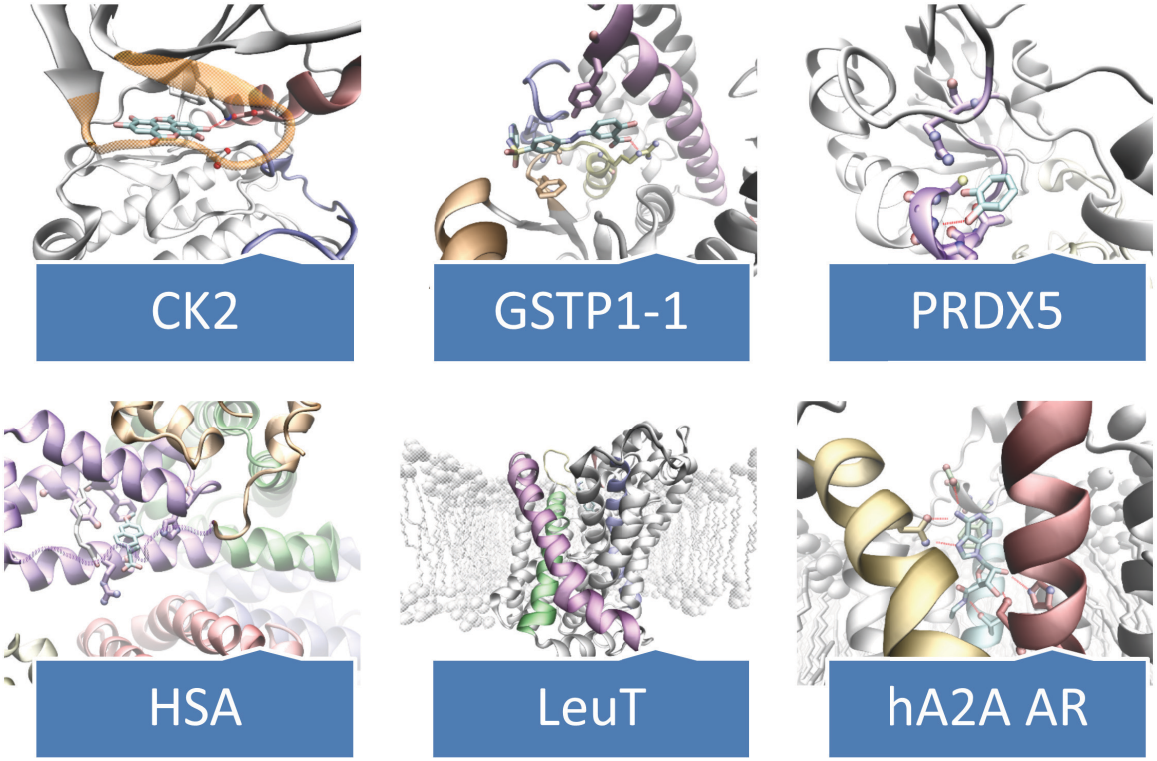

Figure 3. Overview of the X-ray protein-ligand complexes used as validation cases. On the top: Acid Ellagic-CK2, SASP-GSTP1-1, Benzen-1,2-diol-PDRX5; On the bottom: (S)-naproxen-HAS, (S)-fluoxetin-LeuT, NECA-hA_{2A}AR.

Glutathione S-transferases (GSTs) are homodimeric phase II detoxification enzymes, active in the bioconjugation of glutathione (GSH) to a wide range of both endogenous and exogenous molecules. The catalytic region of GSTs is topologically subdivided in two different site: *i)* the G-site, selective for GSH recognition and highly conserved crosswise GSTs isoforms, and *ii)* the H-site, less conserved and responsible for the binding of electrophilic molecules[42]. Isoform P1-1 probably represents the most studied GST and has been related to the development of tumors resistance towards numerous anti-cancer drugs[43]. SASP, which is able to inhibit GSTs without acting as a co-

substrate for the conjugation reaction with GSH, has been co-crystallized with GSTP1-1 and represents a starting point for structure-based design of new anticancer drugs[44]. The X-ray complex (Figure 3B) highlights that the inhibitor binds to a hydrophobic pocket formed by Phe8, Val10, Val35, Ile104 and Tyr108 side-chains. SASP phenyl ring and salicylic acid moiety are engaged in π–π stacking interactions with the aromatic side-chain of Phe8 and Tyr108, respectively, while the ligand carboxylate is involved in an electrostatic interaction with the Arg13 side chain.

To extend the SuMD capabilities on low affinity ligand we selected the recently solved structure of PRDX5 in complex with a benzen-1,2-diol[45]. PRDX5 belongs to the ubiquitary peroxiredoxin family which role relies on the hydrogen peroxide and alkyl hydroperoxides reduction. PRDX5 plays a remarkable role in post-ischemic inflammations in the brain[46,47]. The catechol was identified by a fragment based screening and the dissociation constant was estimated in the millimolar range (K_d=1.5 +/- 0.5mM). More interestingly, the system was extensively characterized by NMR spectroscopy both with structure-based experiments and ligand-based experiments, resulting in a solid model system for a low-affinity binding event[45]. In the X-ray complex (Figure 3C) the catechol ring is localized to the N-terminus of the second helix establishing a hydrogen-bond network with the backbone nitrogen of Gly46 and Cys47 residues. The sidechain of Arg127 is oriented towards the hydroxyl moiety and contributes to the binding with an additional hydrogen bond. Similarly, the thiol group of Cys47 is faced to the catechol. The Pro40, Leu116 and Phe120 establish hydrophobic interactions with the aromatic ring.

The Human Serum Albumin (HSA) is a deeply investigated protein for its ability in bind a wide range of different molecules in human plasma. (S)-naproxen strongly binds HSA and more interestingly in different sites depending on the presence of other small molecules (e.g. hormones, xenobiotic, fatty acids)[48,49]. The only structure available for this complex was obtained in presence of decanoic acid driving the accommodation of the naproxen molecule in the IB site, a vast and hydrophobic pocket where a multitude of different ligand can be

hosted[49]. In the IB site (S)-naproxen inserts its naphthalene scaffold within the hydrophobic pocket and interacts directly with the aliphatic tail of decanoic acid and the residues Ile142, Phe157 and Tyr161 (Figure 3D). The carboxylic group is partially exposed to the solvent but is surrounded by several charged residue forming the entrance of the pocket: Arg145, Lys 190 and in particular Arg186.

Neurotransmitter sodium symporter (NSS) family includes the human serotonin transporter (SERT), norepinephrine transporter (NET) and dopamine transporter (DAT)[50]. To date, there is a lack of focused information about the structure of these important therapeutic targets. In recent past, the crystallographic structure of the LeuT from Aquifex aeolicus (a NSS family member) has been disclosed with the aim of better understand the basis for selective serotonin re-uptake inhibitors (SSRIs) activity towards serotonin transporters[51]. LeuT-(S)-fluoxetine X-ray complex (Figure 3E) highlights hydrophobic contacts between the inhibitor and Leu29, Arg30, Tyr108 and Phe253 side chains. (S)-fluoxetine secondary amino group points towards the extracellular space and engage Asp401 in an electrostatic interaction, while the extracellular gate is locked by the salt bridge between Asp404 and Arg30.

Moving to the last key study, adenosine receptors (ARs) belong to the G protein-coupled receptors (GPCRs) superfamily. The known four subtypes, termed adenosine A_1, A_{2A}, A_{2B} and A_3 receptors, are widely distributed in human body, involved in several physio-pathological processes and represent potential targets for the treatment of several diseases[52]. In the last decade, X-ray structures of the hA_{2A} AR in complex with agonists and antagonists have been released thus offering the basis for molecular modeling investigation[53] including also SuMD simulations[7,54,10]. Here we focus on the complex with NECA[55] (Figure 3F) that features a strong polar interaction between the exocyclic amine group of NECA and the side chain of the conserved Asn253 residue; a hydrogen bond with the nitrogen atom of NECA acetamide moiety and the Thr88 side chain; and an aromatic π-π stacking with the conserved Phe168, located in the second extracellular loop (EL2), and hydrophobic contacts with, among others, the Leu249 side chain.

Globular Systems

Acid Ellagic-CK2 recognition pathway.

In the starting geometry the ligand was placed at a distance of 50 Å from the binding site. After the initial randomization step, the distance reduced to 43 Å. As depicted in Figure 4A and shown in Video S1, the first interaction between the ligand and the protein is established after 2 ns of productive trajectory and is mediated Lys49 that directs the ligand to the P-loop of the kinase. As shown by the Pollicino analysis (Figure 4B), the ellagic acid approaches the region of the P-loop and mostly interacts with the Arg47, Lys49, Glu53 and the Lys71 (Figure 4A). These residues describe an interaction site, at 10.5 Å where the ligand resides for about 6 ns. In facts, the ligand RMSD plot (Figure 4C) records stable values in the 2-8 ns time lapse. The IE with the protein in this site is about -20 kcal/mol (Figure 4D at $dcm_{(L-R)}$= 10 Å); the Per Residue Contacts Count graph (Figure 4E) highlights that the above mentioned residues are those establishing the greatest number of contact, whereas the corresponding 3D models helps in identifying their location (Figure 4F) and the chronological order at which they have been approached by the ligand (Figure 5A). Approximately after 7 ns of simulation the ligand moves toward the orthosteric site, where Leu45 stabilizes its conformation and the side-chain of His160 hampers its passage. Through an interaction mediated by Arg43 the ligand overcomes the His160 gate and reaches new interaction site described by Asp120, Arg47, and Met163. The permanency in this site is about of 2 ns with an interaction energy of -51 kcal/mol (Figure 4C-D). Consistently, the RMSD plot presents another plateau in the time range 8-10 ns (Figure 4C) that corresponds to the swarm of dots in the IE Landscape at $dcm_{(L-R)}$ = 11 Å (Figure 4D). A further stabilizing interaction with the Asn118 induces a shift in the ligand position that places the ring system parallel to the β7-β8 strands (Video S1).

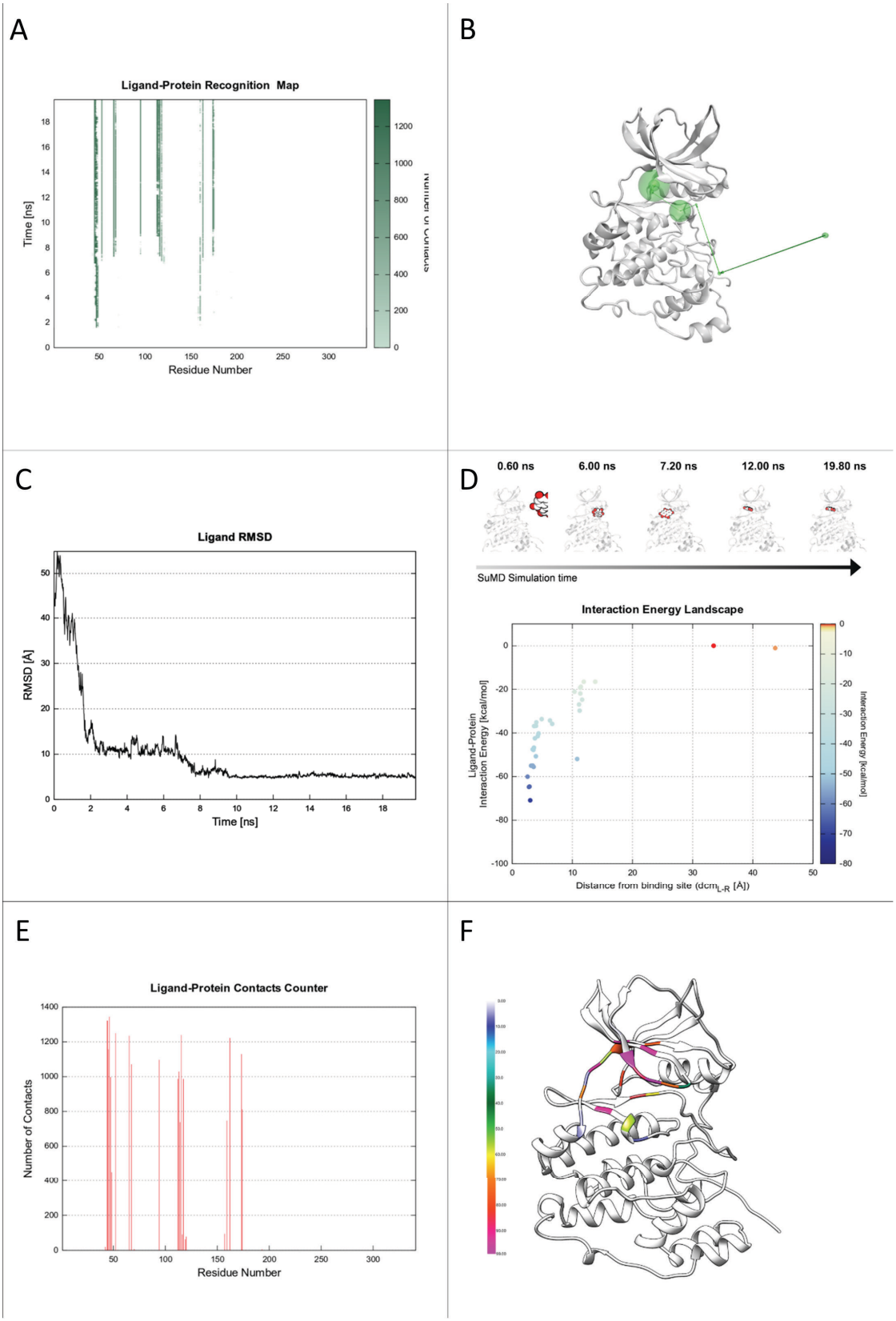

Figure 4. Acid Ellagic-CK2 recognition pathway. (A) Ligand-Protein Recognition Map (B) Pollicino Analysis (C) Ligand RMSD (D) IE Landscape (E) Ligand-Protein Contacts Count (F) Chimera contacts.

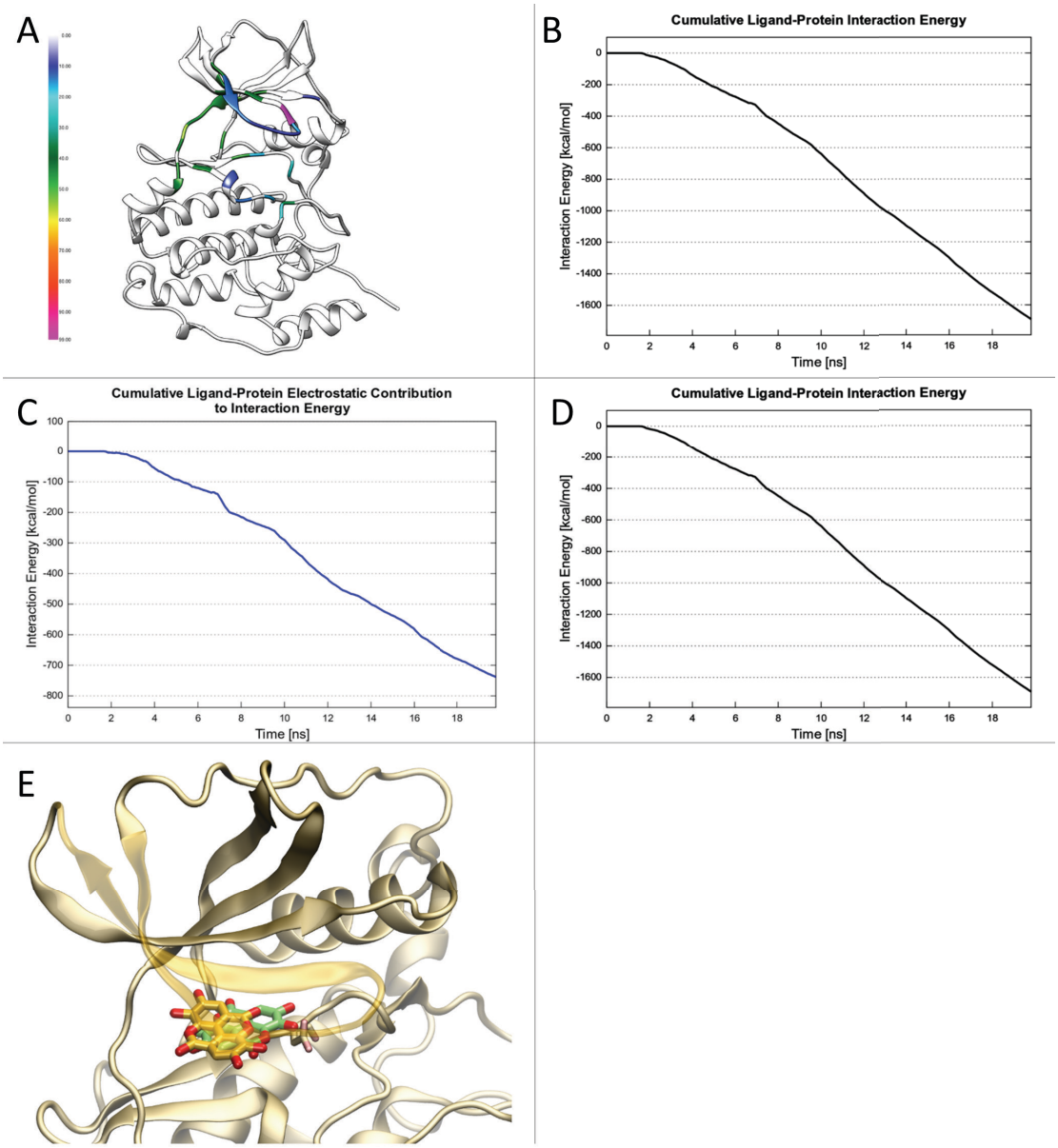

Figure 5. Acid Ellagic-CK2 recognition pathway. (A) Chimera time (B) Cumulative IE (C) Cumulative IE electrostatic contribution (D) Cumulative IE van der Waals contribution (E) Superimposition between SuMD endpoint conformation and X-ray binding mode.

As shown in the Cumulative Ligand-Protein IE (Figure. 5B) and its corresponding decomposition into electrostatic and van de Waal contribution, (Figure 5C and D, respectively) the change in the slope indicates that new conformation has a lower interaction energy than the previous one.

In particular, as highlighted by the comparison of the graphs relative to the electrostatic and van der Waals contribution (Figure 5C and D, respectively), the stabilization can be ascribed by the establishment on an electrostatic interaction with Asp175. As result of the new interaction the ligand moves into the orthosteric site (Figure 5E) and interacts with Lys159, Val66, Val117, Val53, His115 and Lys68 by maintaining the same position is maintained till the end of the SuMD simulation. The RMSD plot shows another plateau from 10 ns to the end, whereas the IE Landscape indicates that in this time lapse the ligand is at a distance around 2.5 Å with an IE between -40 to -70 kcal/mol.

The simulation was replicated three times and Replicas Analysis results are reported in Figure 6. In particular, the RMSD plot indicates that one replica does not reach the orthosteric site (Figure 6A, green line), whereas the others reach the same final RMSD value.

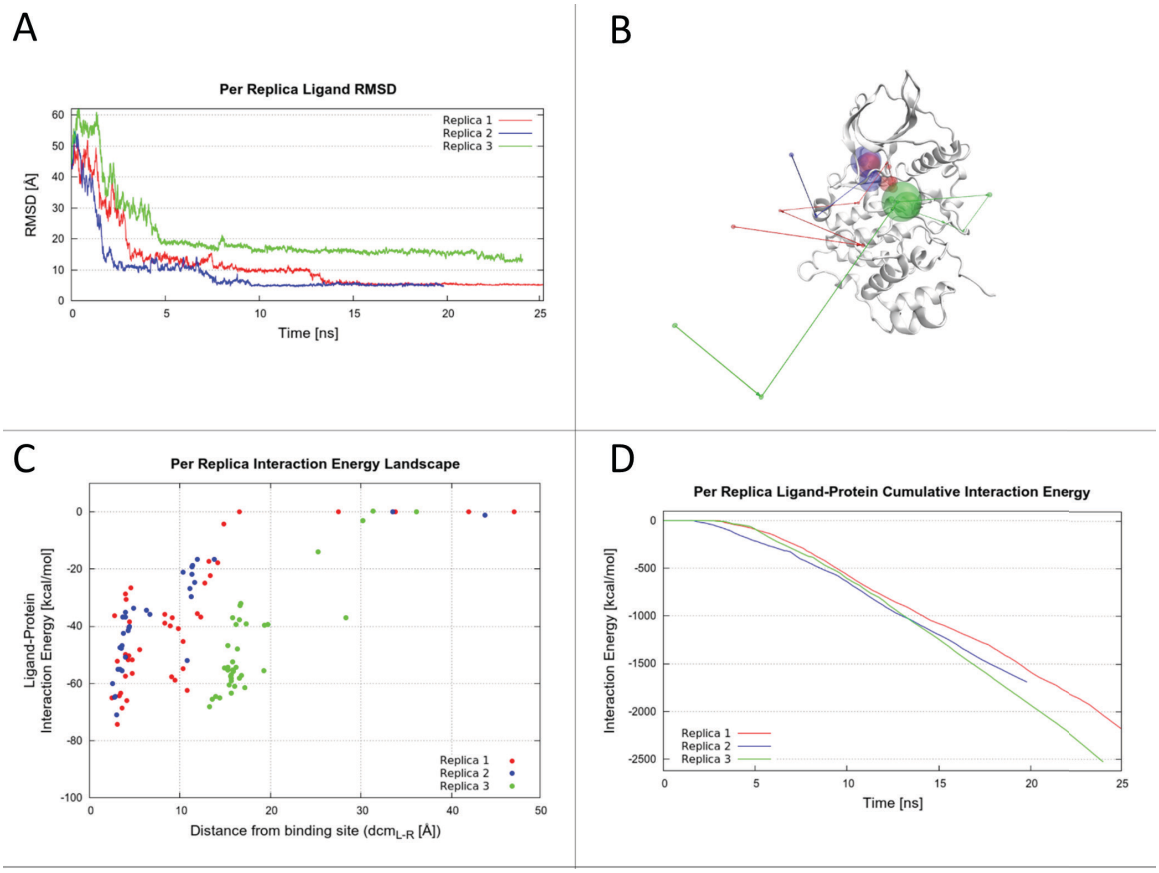

Figure 6. Acid Ellagic-CK2 recognition pathway (A) Per Replica Ligand RMSD (B) Per Replica Pollicino Analysis (C) Per Replica IE Landscape (D) Per Replica cumulative IE.

The same conclusion arises from the investigation of the Pollicino analysis where the ligand pathway of the two replicas converge in proximity of the protein (Figure 6B, red and blue spheres).

The Per Replica IE Landscape helps in explaining why the third replica does not reach the orthosteric site: as indicated by the green dots in Figure 5C the ligand reaches a different interaction site with an IE of -60 kcal/mol, a value close to the IE of the replicas that converge into in the orthosteric site (Figure 6C, red and blue dots). This consideration is confirmed by the trend of the Per Replica Cumulative IE that highlights a more negative slope for the third replica (Figure 6D, green line), indicating a very strong interaction.

SASP-GSTP1-1 recognition pathway.

During the SuMD simulation the SASP reaches the GSTP1-1 catalytic H site in less than 6 ns (Video S2). The IE landscape highlights the formation of the first protein-ligand stabilizing interaction when the ligand and protein H site distance is 15 Å (point a, Figure 7A and 7B).

In this preliminary complex, SASP engages the Gly205 backbone oxygen in a hydrogen bond interaction through its sulfamide nitrogen atom and establishes an aromatic π-π stacking interaction between the salicylic moiety and Tyr108 (interactions corresponding to the first continues lines in the Protein-Ligand Recognition Map, Figure 7C). This situation anticipates a ligand positional shift that allows the SASP salicylic carboxylate to approach the positively charged Arg13 side chain, while the benzene ring replaces the salicylic aromatic moiety in the π-π stacking interaction with Tyr108 (point b, Figure 7A). The energy stabilization of the complex increases and, after 8 ns of simulation, SASP proceeds toward a farther conformation, able to gain a more favourable electrostatic interaction geometry with Arg13 side chain, after the displacement of two water molecules from the solvation sphere of the positively charged residue. This new pose (point c, Figure 7A and Figure 7B) is retained until the end of SuMD simulation, with the exception of conformational changes occurring to thepyridylsulfamoyl moiety, able to fit in the hydrophobic pocket

delimited by Phe8, Val35 Trp38. During the SASP - GSTP1-1 recognition event GSH remain in the catalytic G site of the enzyme, not interacting with the inhibitor. Figure 7D highlights all the residues involved in the interaction with SASP during the SuMD simulations: the selective contacts towards only one enzymatic subunit, as well as the topologically restricted area interested, are well defined by the ribbon colorations.

Considering the SASP crystallographic conformation as geometrical reference, the ligand RMSD analysis (Figure 7E) reaches a minimum after 15 ns of simulation (Figure 7F), before level out at a value of about 5 Å. Figure S2 reports other ligand-protein interaction energy analysis.

The Replicas Analysis (Figure S3) depicts a recognition event with no meta-stable binding sites and characterized by almost a univocal pathway. Nevertheless, in one replica, in the final complex SASP is rotated by 180° (as highlighted by the higher RMSD value) and loses the electrostatic stabilization between its salicylic moiety and Arg13 side chain.

Benzen-1,2-diol–PDRX5 recognition pathway.

The simulation was repeated on both monomeric and dimeric form yielding similar results. However, here we will focus on the dimeric form according to solution NMR studies, in which the authors stated the protein as dimer[56]. At the beginning of randomization step the fragment was placed at 78 Å from PDRX5 binding site ($dcm_{(L-R)}$= 78 Å). As reported in figure 8A, 8B (point b) and 8C, after nearly 3 ns the fragment approaches the protein in a region located at around 30 Å from the primary binding site (Video S3).

This meta-binding site lies in the opposite monomeric subunit with respect to the primary binding site and it is defined by residues Leu62, Lys63, Val69, and Val70. As shown by the IE landscape and the Pollicino Analysis (Figure 8A and 8B, respectively), this site engages the ligand with favorable interactions for a couple of nanoseconds. In particular, the formation of a hydrogen bond between the hydroxyl groups of catechol and the carbonyl moiety of the backbone amide of residue Lys95 stabilizes this conformation.

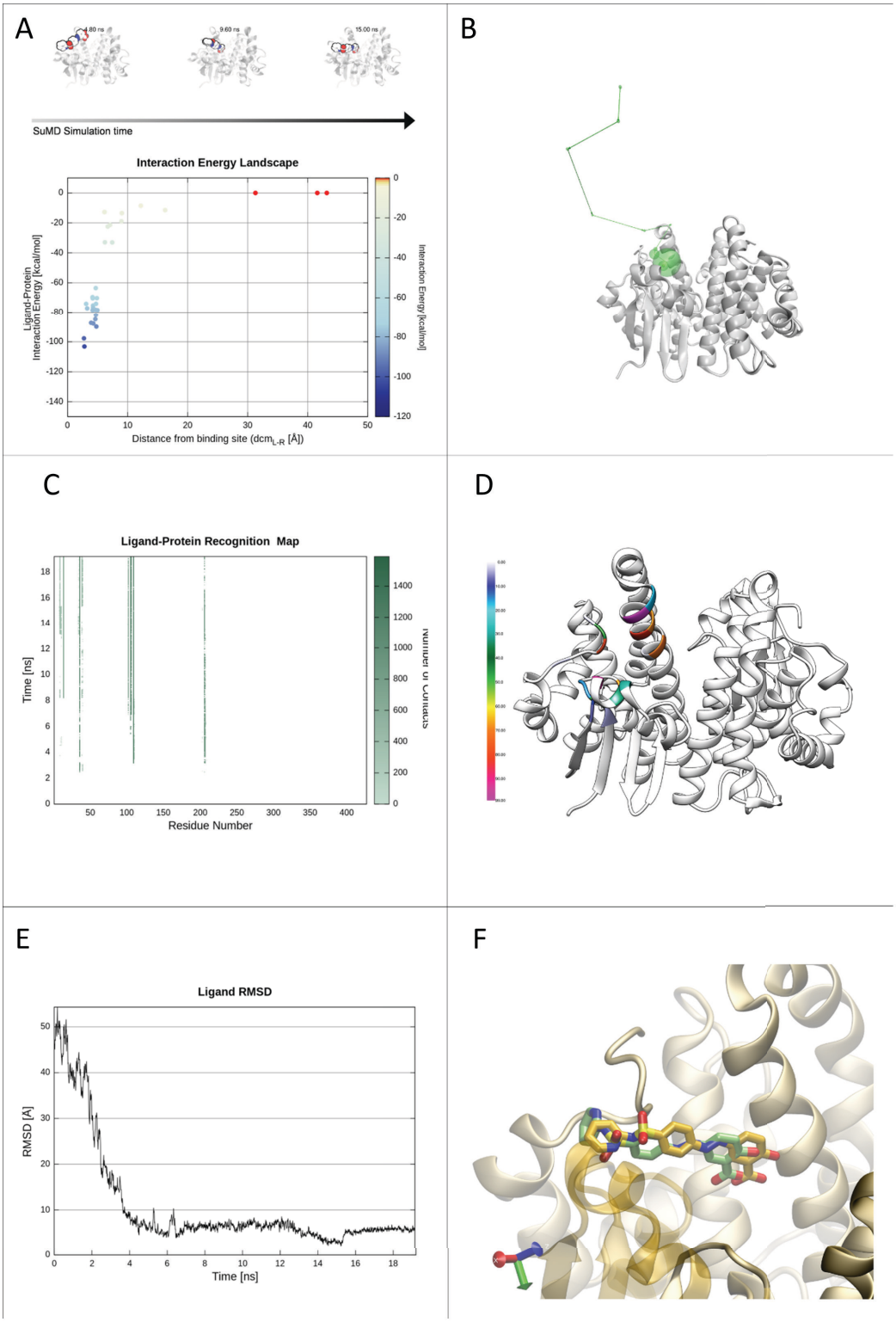

Figure 7 – SASP-GSTP1-1 recognition pathway. (A) IE Landscape (B) Pollicino Analysis (C) Ligand-Protein Recognition Map (D) Chimera contacts (E) Ligand RMSD (F) Superimposition between SuMD endpoint conformation and X-ray binding mode.

After nearly 6 ns the fragment is released by this site and fluctuates to finally reach the primary binding site thought a series of molecular interaction, including residues (chronologically sorted): Glu91, Glu16, Glu18, Phe79 belonging to the first monomer unit (SI figure S4). Finally, the fragment accesses to the binding site where fluctuates experimenting different conformations in accordance with its affinity in the millimolar range. The fluctuations of the fragment in the binding site are also evident in the Protein-Ligand Energy profiles, in which the energy wavers around the value of -20 kcal/mol (SI figure S4). During the fluctuation, the catechol enters in contact with most of the residue forming the site, in particular (sorted by number of molecular contacts during the trajectory): Thr146, Thr44, Arg127, Phe120, Leu116, Gly46 and Cys47 (Figure 8C, 8D). The main conformation observed corresponds to the crystallographic one, as reported in Figure 8E and 8F where the RMSD reaches a minimum value 0.69 Å at 17.3 ns.

The simulation was repeated in three times randomizing the position of the ligand. The Replicas Analysis is reported in Figure S5. Briefly, in each replica the fragment reached the primary binding site experiencing the conformation reported in the crystallographic data with the best RMSD respectively of 1.12 and 1.24 Å for the replica 2 and 3.

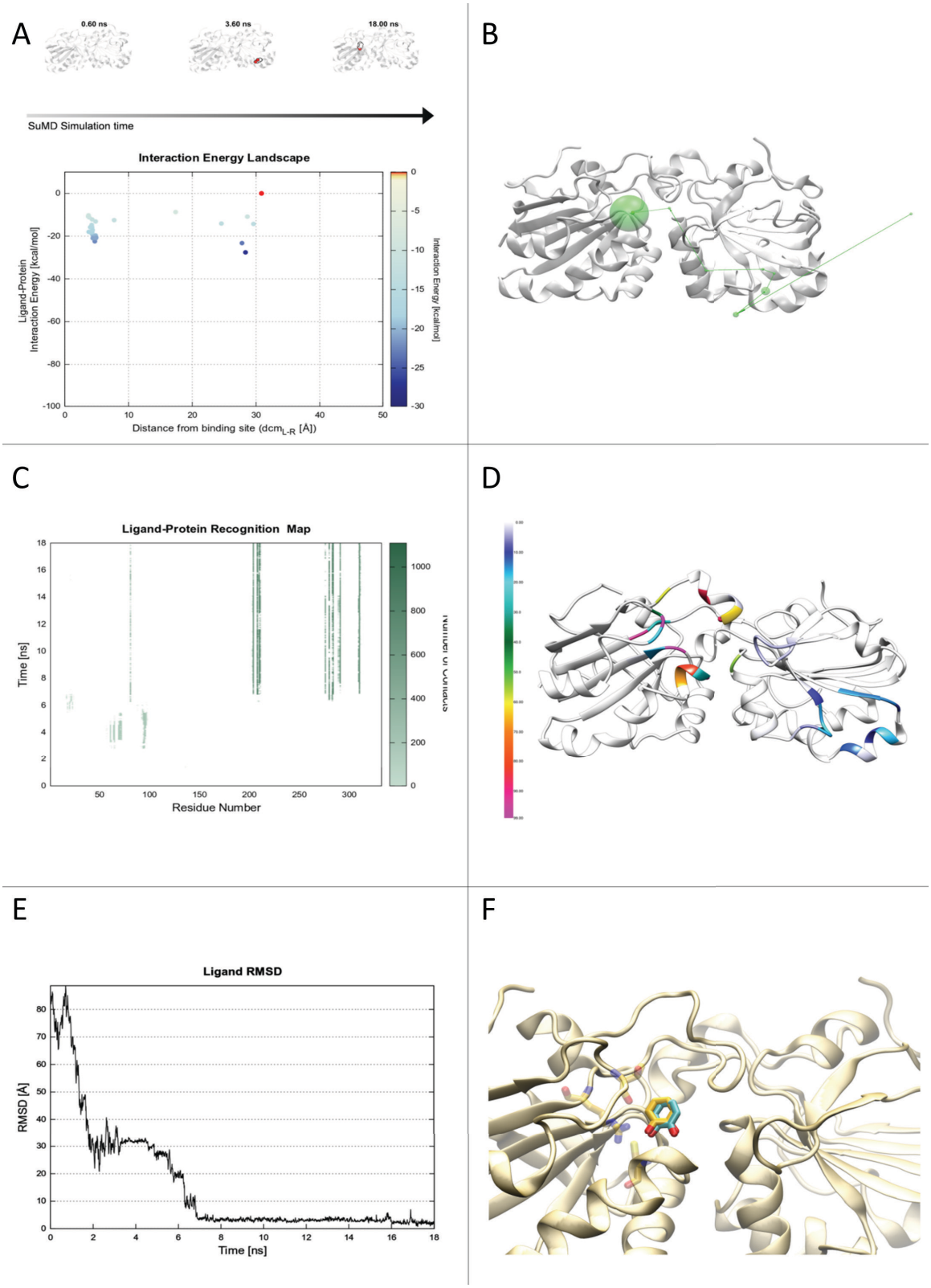

Figure 8 –Benzen-1,2-diol-PDRX5 recognition pathway. (A) IE Landscape (B) Pollicino Analysis (C) Ligand-Protein Recognition Map (D Chimera contacts. (E) Ligand RMSD (F) Superimposition between SuMD endpoint conformation and X-ray binding mode.

(S)-naproxen-HAS recognition pathway.

SuMD simulation was performed maintaining decanoic ligand in the IB site according the crystallographic geometries. (S)-naproxen was separated from HSA-decanoid acid complex by 32 Å from IB site (point a in Figure 9A and 9B). In the first SuMD step the ligand fluctuate till 50 Å from the IB site. As reported in Figure 9C after a couple of nanosecond the ligand approaches the first protein site in its trajectory by engaging Lys510 and Thr564 (Video S4). Shortly after, the ligand establishes a network of interaction for 1 ns (from 2.3 ns to 3.2) in a site located at around $dcm_{(L-R)}$= 20 Å (point b in Figure A), defined by residues: Val116, Pro118, Val122, Thr133 and Phe134. Then, the Naproxen molecule approach a second, where it fluctuates for about 3 ns by establishing strong interaction with residues Leu115, Pro118, Lys137, and Ile142 (as also evident from Protein-Ligand Interaction energy in figure S6 in SI).

This meta-binding site is located in front of the principal binding site to which is separated by the presence of a long extended loop (residue 106 to 119) that acts as a gate for the IB site. Finally after 6 ns, (S)-naproxen is able to pass behind the extended loop and reach the IB site (residues Leu115, Ile142, Phe157, Tyr161) as show by Figure 9B and 9E. Within the primary site, the ligand is able to place the methyl ether group in the proximity of Phe152 very similarly to the orientation of crystal structure. On the other hand, the naphthalene core and in particular the carboxylic group adopt a different orientation due to the presence of the extended loop. This different orientation abolishes the ionic interaction between the carboxyl group and the Arg112 observed in the crystallographic structure (Figure 9F). At the end of the simulation the RMSD fluctuates around 5 Å, reaching the lowest value of 4.76 at 12.70 ns (Figure 9E and 9F).

Interestingly, in the other replicas (Figure S7) the ligand reaches the IB site by approaching the extended loop from a different position and occupies a slightly different location in the vast IB site. This suggests the loop might have a crucial role in the recognition process (Figure S7).

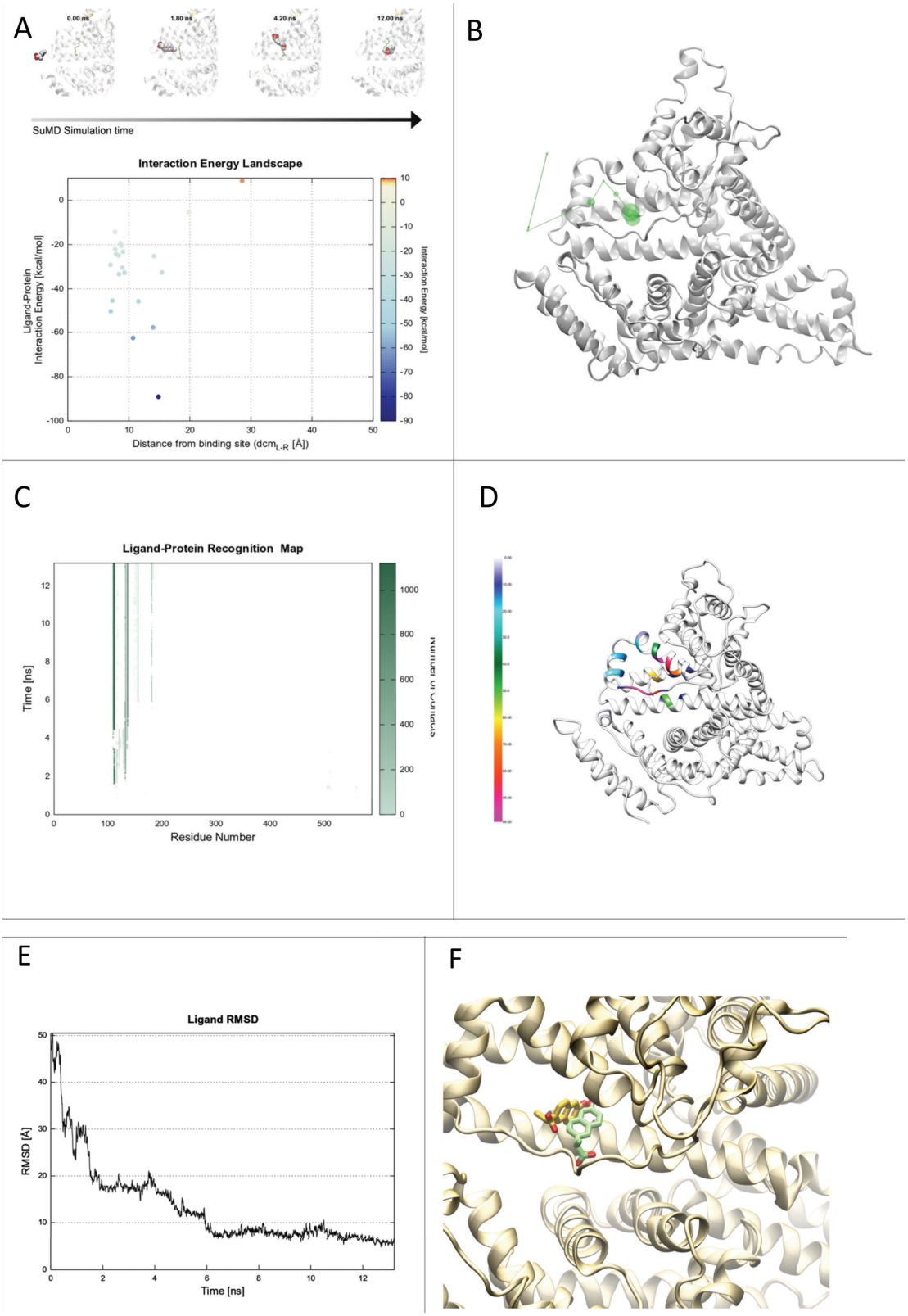

Figure 9. (S)-naproxen-HAS recognition pathway. (A) IE Landscape (B) Pollicino Analysis (C) Ligand-Protein Recognition Map (D Chimera contacts. (E) Ligand RMSD (F) Superimposition between SuMD endpoint conformation and X-ray binding mode.

Published - Cuzzolin A, Sturlese M, Deganutti G, Salmaso V, Sabbadin D, Ciancetta A, Moro S. *J Chem Inf Model* 56, 687-705 (2016)

Transmembrane Systems

(S)-fluoxetine-LeuT recognition pathway.

The (S)-fluoxetine recognition pathway highlights, after 1 ns of SuMD simulation, a first electrostatic interaction between Asp 158 side chain and the ligand charged secondary amine group (Video S5). The energetic stabilization characterizing this complex corresponds to the IE landscape minimum reported in the Figure 10A, point a and Figure 10B. This preliminary complex is able to favor the ligand approach towards an inner pocket of LeuT, topologically defined by Tyr 471 and the aliphatic chains of Lys 474 and Glu 478, reciprocally involved in an ionic interaction. Hydrophobic contacts stabilize this intermolecular complex for about 2 ns, before a conformational change allows (S)-fluoxetine to establish a more favorable electrostatic interaction whit Glu 402 side chain. This scenario anticipates the ligand repositioning inside an inner hydrophobic site, where the ligand engages for almost 7ns Tyr471, Trp406, Ile475 and Phe405 side chains in lipophilic interactions through its phenyl ring (point b, Figure 10A and Figure 10B). During the remaining simulation time, the inhibitor makes contacts with Ala319 (EL4) and the side chains of the key residues Asp404 and Arg30 (point c, Figure 10A and Figure 10B and continuous lines corresponding to the last 4ns of SuMD simulation in Figure 10C), both located at the protein extracellular gate and involved in a ionic lock that 10terically obstructs the SSRIs binding site disclosed by LeuT crystallographic structure. Figure 10D summarizes all the amino acids involved in the SFX recognition event during the SuMD simulation.

The RMSD plot (Figure 10E) outlines the inhibitor difficulty in reproducing the experimental pose (Figure 10F). Investigation on LeuT crystal structure without co-crystallized inhibitor reveale an alternative conformation of Arg30 side chain, and the absence of the gate ionic lock (Figure S12)[57] : it is possible to speculate that the LeuT extracellular gate, during SuMD simulation timescale, is able to remain in a stable conformation, previously induced by the inhibitor binding and retained even after the removal of the ligand during the system preparation for SuMD.

Replicas analysis (Figure S9) highlights two alternative recognition pathways through the extracellular vestibule, unable to enable SFX to reproduce the binding mode observed in the crystallographic complex and characterized by accentuated energy variations in proximity of the extracellular transporter gate.

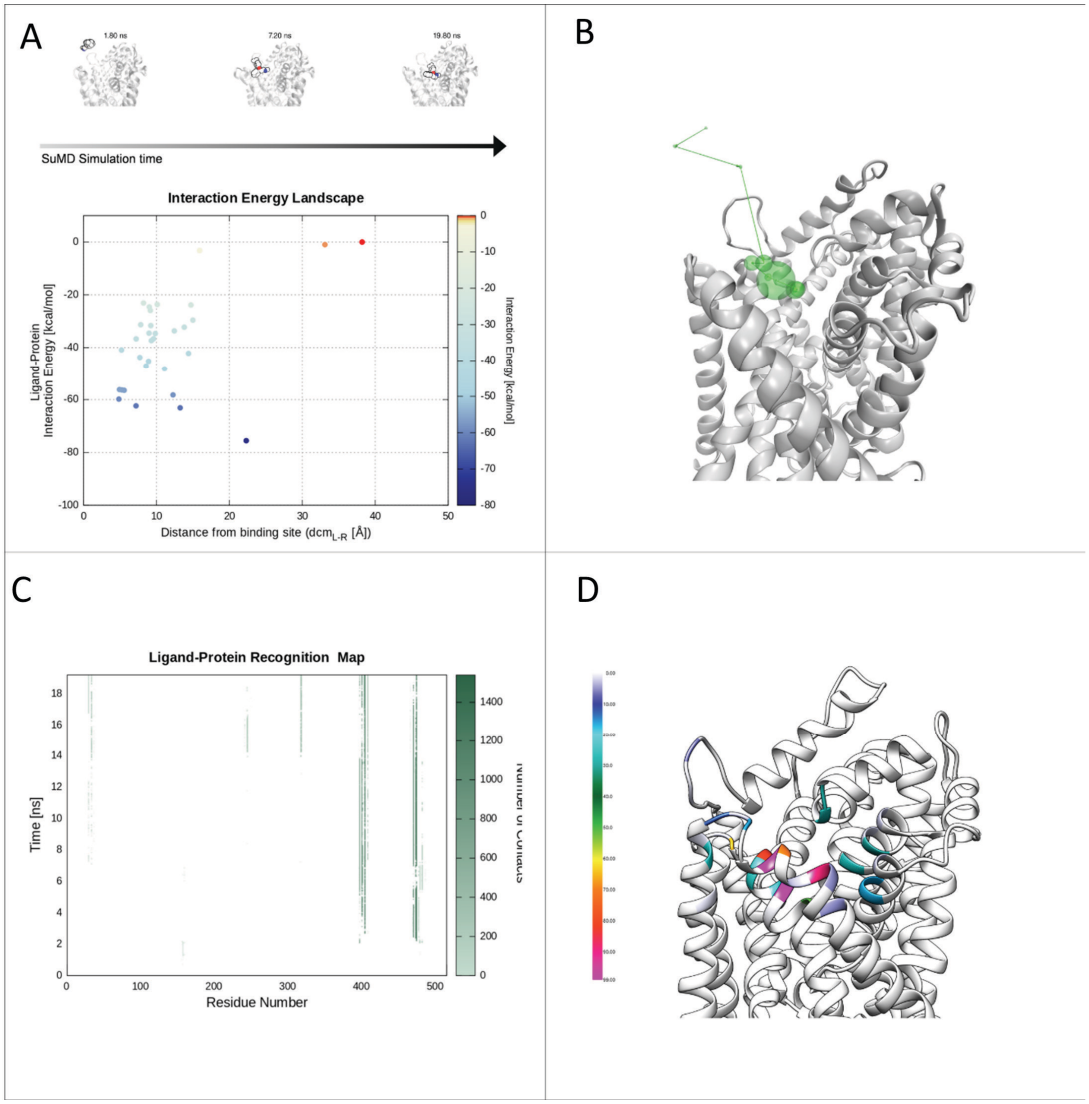

Figure 10. (S)-fluoxetin-LeuT recognition pathway. (A) IE Landscape (B) Pollicino Analysis (C) Ligand-Protein Recognition Map (D Chimera contacts.

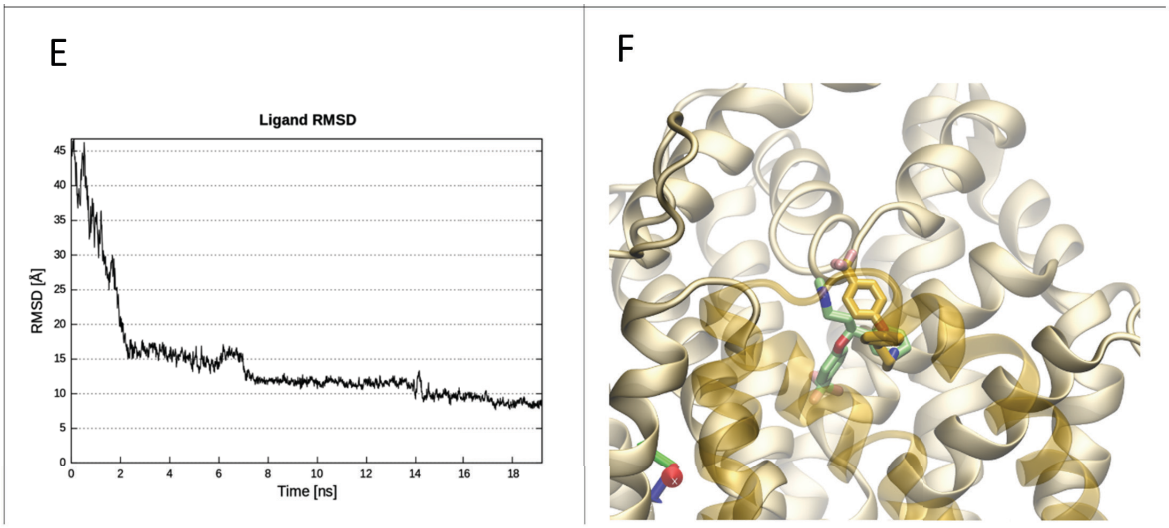

Figure 10 (continuation). (E) Ligand RMSD (F) Superimposition between SuMD endpoint conformation and X-ray binding mode.

NECA–hA$_{2A}$ AR recognition pathway.

NECA establishes the first stabilizing contacts with hA$_{2A}$ AR after about 4ns of SuMD simulation (Video S6). During this initial scenario (point a, Figure 11A and Figure 11B) the ligand approaches the protein topological structure defined by ECL2 N-terminus and the residues located at top of TM5 and TM6. More precisely, NECA engages Phe257 (6.59) side chain in a π-π stacking interaction through its purine scaffold and locates the N-ethylcarboxamido moiety towards a pocket delimited by Trp143 (ECL2), Pro173 (ECL2), and Asn175 (TM5) side chains, as highlighted by the first stripes in Figure 11C and the yellow and violet ribbon coloration in Figure 11D.

This complex anticipates a repositioning that allows the ligand to reach a meta-stabile binding site, mainly characterized by a π-π stacking interaction with His264 (EL3) side chain, an hydrophobic contact in the direction of Met174 (TM5) side chain, and an hydrogen bond interaction between its C2' hydroxide group and Asn253 (TM6) (point b, Figure 11A and Figure 11B).

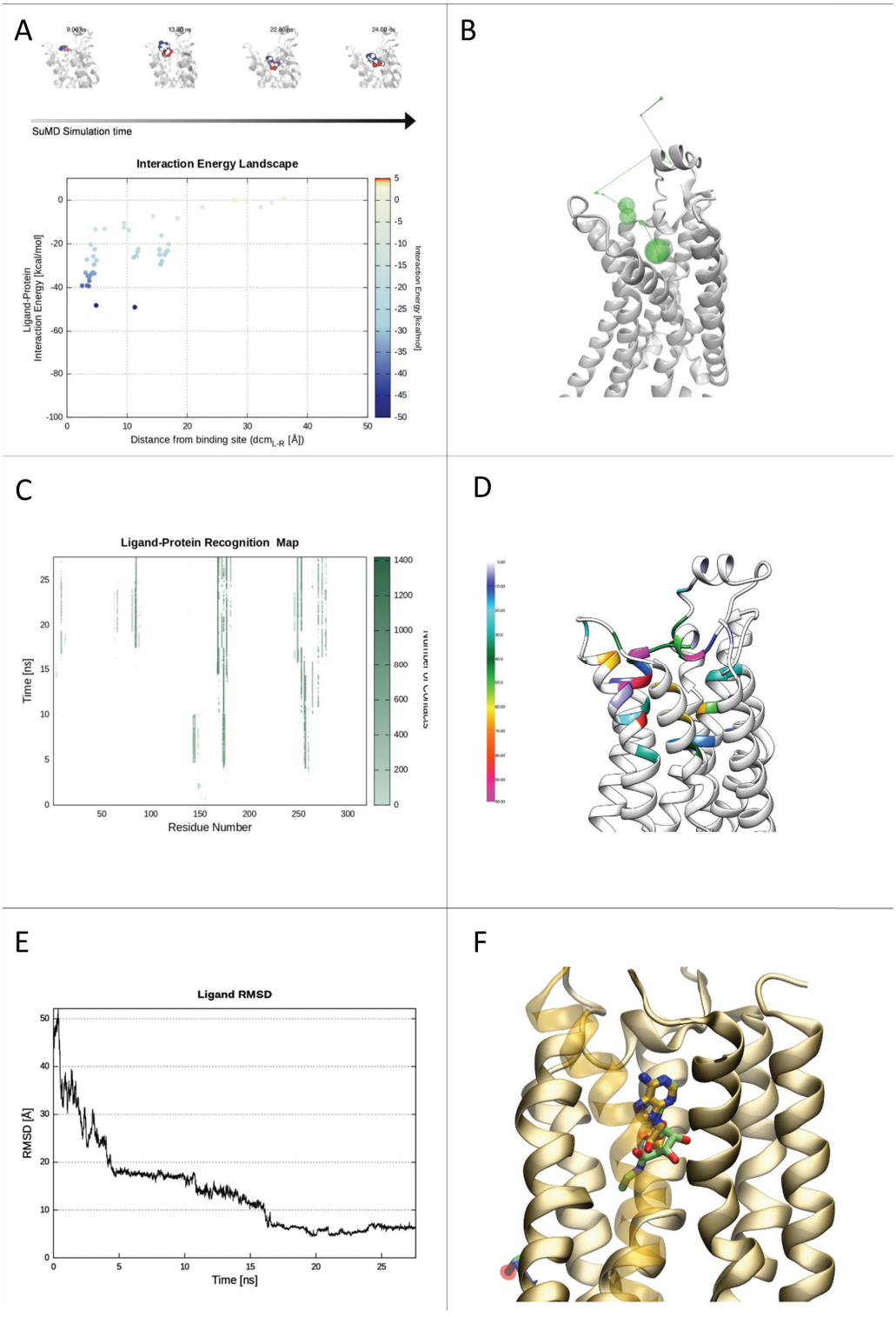

Figure 11. NECA-hA_{2A} AR recognition pathway. (A) IE Landscape (B) Pollicino Analysis (C) Ligand-Protein Recognition Map (D) Chimera contacts (E) Ligand RMSD (F) Superimposition between SuMD endpoint conformation and X-ray binding mode.

During the time slot rising from 14 ns to 20 ns of SuMD simulation, the agonist reaches a deeper position inside the orthosteric binding site and explores different conformations (included a temporary anti-syn transition about the glycoside linkage), until engages Phe168 (ECL2) side chain in a π-π stacking interaction and Asn253 (TM6) side chain in hydrogen bond interactions through its exocyclic amine and the N7 position of the purine scaffold (point c, Figures 11A and Figures 11B). This complex orientation (associated with the minimum RMSD value in Figure 11E, with respect to the NECA crystallographic conformation) is followed by an alternative stabilized conformation (point d, Figure 11A and Figure 11B) which involves also hydrophobic interactions with Leu249 (TM6), Leu85 (TM3) and Val84 (TM3).

During the remaining SuMD simulation time, the protein-ligand complex geometry remain almost unaltered, with the exception of a reorientation of the N-ethylcarboxamidoribose moiety, pointing toward TM4, and the loss of the aromatic π-π interaction due to a conformational change occurring to Phe168 (EL2) side chain. In Figure 10S are reported other ligand-protein energy interaction analysis.

At the minimum RMSD value, NECA pyrimidine scaffold coincides with the crystallographic orientation, while the ribose moiety is oriented in an alternative conformation (Figure 11F).

Replicas Analysis (Figure 11S) highlights also a different NECA recognition pathway, which involves residues located at the ECL2 and characterized by comparable energetic stabilizations.

Conclusion

In the present work, we have demonstrated the general applicability of SuMD simulations using different types of proteins, including both globular and membrane proteins. Moreover, we have presented the SuMD-Analyzer tool that helps, also a non-expert user, in the analysis of the SuMD trajectories. Even if various other MD methods have also been used to characterize binding pathways, SuMD has the great advantage of being able to explore the ligand-

protein approaching path in nanosecond simulation time scale. Furthermore, SuMD simulations enable the investigation of ligand-protein binding events independently from the starting position, chemical structure of the ligand, and also from its target binding affinity. As described for each key study, SuMD simulations are able to characterize multiple ligand-protein binding pathways identifying a variety of metastable intermediate states (meta-binding sites). These informations may be an interesting starting point for further argumentations regarding the pharmacological consequences of that specific ligand-protein recognition process. Moreover, it is worthy to underline that, contrary to expectations, not all SuMD trajectories converge to the structure of the complex obtained crystallographically. Indeed, there are several plausible reasons that may be argued to describe this particular unexpected aspect: *a)* the crystallographic pose of the ligand is not the only minimum of the potential energy surface described by the force field during the SuMD simulations; *b)* the crystallographic conformation of the protein in its bound state is remarkably different respect its apo-form. This could be interpreted as the sign of an important induce-fit process during the ligand recognition; and *c)* the boundary conditions that led to the formation of the crystallographic ligand-protein complex (solvent and co-solvent, pH, ionic strength, or temperature just as a few examples) are not well described during the SuMD simulations. This must always be kept in mind when any conjecture is made starting from the analysis of SuMD trajectories. Currently, a major effort is underway to estimate, from SuMD simulations, binding kinetics properties (in particular on-rate values) in approximate agreement with experimental measurements.

Hopefully, the future of drug design will involve detailed characterization of not only the bound state but also the whole ligand–protein network of recognition pathways, including all metastable intermediate states (meta-binding sites). With such a complete understanding we hope expand our perspectives in several scientific areas from molecular pharmacology to drug discovery.

Abbreviations

3D	three-dimensional
CK2	caseine kinase 2
CPU	central processing unit
GPCRs	G protein-coupled receptors
GPU	graphics processor unit
GSTP1-1	P1-1 isoform of glutathione S-transferase
hA2AAR	human A2A adenosine receptor
HSA	human serum albumin
IE	interaction energy
K_d	equilibrium dissociation constant
K_{off}	dissociation rate constants
K_{on}	association rate constants
LeuT	leucine transporter
MD	molecular dynamics
NECA	5'-N-ethylcarboxamidoadenosine
PDB	protein data bank
POPC	1-palmitoyl-2-oleoyl-snglycero-3-phosphocholine
PRDX5	eroxiredoxin 5
RMSD	root-mean-square deviation
SASP	sulphasalazine
SuMD	supervised molecular dynamics

Bibliography

1. Protein-Ligand Interactions: From Molecular Recognition to Drug Design; Böhm, H.-J., Schneider, G., Eds.; Methods and Principles in Medicinal Chemistry; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, FRG, 2003.

2. Pan, A. C.; Borhani, D. W.; Dror, R. O.; Shaw, D. E. Molecular Determinants of Drug– receptor Binding Kinetics. Drug Discov. Today 2013, 18 (13–14), 667–673.

3. Moro, S.; Hoffmann, C.; Jacobson, K. A. Role of the Extracellular Loops of G Protein- Coupled Receptors in Ligand Recognition: A Molecular Modeling Study of the Human P2Y1 Receptor. Biochemistry (Mosc.) 1999, 38 (12), 3498–3507.

4. Dror, R. O.; Jensen, M. Ø.; Borhani, D. W.; Shaw, D. E. Exploring Atomic Resolution Physiology on a Femtosecond to Millisecond Timescale Using Molecular Dynamics Simulations. J. Gen. Physiol. 2010, 135 (6), 555–562.

5. Buch, I.; Giorgino, T.; De Fabritiis, G. Complete Reconstruction of an Enzyme-Inhibitor Binding Process by Molecular Dynamics Simulations. Proc. Natl. Acad. Sci. U. S. A. 2011, 108 (25), 10184–10189.

6. Johnston, J. M.; Filizola, M. Beyond Standard Molecular Dynamics: Investigating the Molecular Mechanisms of G Protein-Coupled Receptors with Enhanced Molecular Dynamics Methods. In G Protein-Coupled Receptors - Modeling and Simulation 796; 95–125, Springer Netherland, 2014.

7. 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. 2014, 54 (2), 372–376.

8. Ciancetta, A.; Sabbadin, D.; Federico, S.; Spalluto, P.; Moro, S. Advances in Computational Techniques to Study GPCR-Ligand Recognition. In Trends in Pharmacol Sci (2015) in press DOI: 10.1016/j.tips.2015.08.006.

9. Sabbadin, D.; Ciancetta, A.; Deganutti, G.; Cuzzolin, A.; Moro, S. Exploring the Recognition Pathway at the Human A2A Adenosine Receptor of the Endogenous Agonist Adenosine Using Supervised Molecular Dynamics Simulations. MedChemComm 2015, 6 (6), 1081–1085.

10. 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. 2015.

11. Paoletta, S.; Sabbadin, D.; von Kügelgen, I.; Hinz, S.; Katritch, V.; Hoffmann, K.; Abdelrahman, A.; Straßburger, J.; Baqi, Y.; Zhao, Q.; Stevens, R. C.; Moro, S.; Müller, C. E.; Jacobson, K. A. Modeling Ligand Recognition at the P2Y12 Receptor in Light of X- Ray Structural Information. J. Comput. Aided Mol. Des. 2015, 29 (8), 737–756.

12. Harvey, M. J.; Giupponi, G.; Fabritiis, G. D. ACEMD: Accelerating Biomolecular Dynamics in the Microsecond Time Scale. J. Chem. Theory Comput. 2009, 5 (6), 1632– 1639.

13. Case, D.; Babin, V.; Berryman, J.; Betz, R.; Cai, Q.; Cerutti, D.; Cheatham Iii, T.; Darden, T.; Duke, R.; Gohlke, H.; others. Amber14, Version AMBER14; Http://ambermd.org/ (accessed October 2015); University of California, San Francisco, 2014.

14. 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 (9), 1157–1174.

15. MacKerell, A. D.; Banavali, N.; Foloppe, N. Development and Current Status of the CHARMM Force Field for Nucleic Acids. Biopolymers 2000, 56 (4), 257–265.

16. Vanommeslaeghe, K.; Raman, E. P.; MacKerell, A. D. Automation of the CHARMM General Force Field (CGenFF) II: Assignment of Bonded Parameters and Partial Atomic Charges. J. Chem. Inf. Model. 2012, 52 (12), 3155–3168.

17. Vanommeslaeghe, K.; MacKerell, A. D. Automation of the CHARMM General Force Field (CGenFF) I: Bond Perception and Atom Typing. J. Chem. Inf. Model. 2012, 52 (12), 3144–3154.

18. 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 (1), 235–242.

19. CCG Inc. Molecular Operating Environment (MOE), 2014.09; Http://www.chemcomp.com (accessed October 2015).

20. Labute, P. Protonate3D: Assignment of Ionization States and Hydrogen Coordinates to Macromolecular Structures. Proteins 2009, 75 (1), 187–

205. Stewart, J. J. P. MOPAC2012, Version 2012; http://OpenMOPAC.net (accessed October 2015); Stewart Computational Chemistry: Colorado Springs, CO, USA, 2012.

21. Stewart, J. J. P. Optimization of Parameters for Semiempirical Methods V: Modification of NDDO Approximations and Application to 70 Elements. J. Mol. Model. 2007, 13 (12), 1173–1213.

22. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian 09, Revision B.01; Gaussian, Inc.: Wallingford, CT, 2009. (24).

23. MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz- Kuczera, J.; Yin, D.; Karplus, M. All-Atom Empirical Potential for Molecular

Modeling and Dynamics Studies of Proteins. J. Phys. Chem. B 1998, 102 (18), 3586–3616.

24. Vanommeslaeghe, K.; MacKerell, A. D. Automation of the CHARMM General Force Field (CGenFF) I: Bond Perception and Atom Typing. J. Chem. Inf. Model. 2012, 52 (12), 3144–3154.

25. Head-Gordon, M.; Pople, J. A.; Frisch, M. J. MP2 Energy Evaluation by Direct Methods. Chem. Phys. Lett. 1988, 153 (6), 503–506.

26. Mayne, C. G.; Saam, J.; Schulten, K.; Tajkhorshid, E.; Gumbart, J. C. Rapid Parameterization of Small Molecules Using the Force Field Toolkit. J. Comput. Chem.2013, 34 (32), 2757–2770.

27. Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. J. Mol. Graph. 1996, 14 (1), 33–38, 27–28.

28. Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. Proteins 2006, 65 (3), 712–725.

29. 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. 1983, 79 (2), 926.

30. Berendsen, H. J. C.; Postma, J. P. M.; Gunsteren, W. F. van; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. J. Chem. Phys. 1984, 81 (8), 3684–3690.

31. Loncharich, R. J.; Brooks, B. R.; Pastor, R. W. Langevin Dynamics of Peptides: The Frictional Dependence of Isomerization Rates of N-Acetylalanyl-N'-Methylamide. Biopolymers 1992, 32 (5), 523–535.

32. 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 (5), 501–508.

33. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. J. Chem. Phys. 1995, 103 (19), 8577–8593.

34. Lomize, M. A.; Lomize, A. L.; Pogozheva, I. D.; Mosberg, H. I. OPM: Orientations of Proteins in Membranes Database. Bioinforma. Oxf. Engl. 2006, 22 (5), 623–625.

35. Grubmüller, H.; Groll, V. Solvate, Version 1.0.1; Http://www.mpibpc.mpg.de/grubmueller/solvate (accessed October 2015); 1996. Williams, T.; Kelley, C. Gnuplot 4.5: An Interactive Plotting Program, Version 4.5; Http://gnuplot.info (accessed October 2015).

36. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera--a Visualization System for Exploratory Research and Analysis. J. Comput. Chem. 2004, 25 (13), 1605–1612.

37. Seeber, M.; Felline, A.; Raimondi, F.; Muff, S.; Friedman, R.; Rao, F.; Caflisch, A.; Fanelli, F. Wordom: A User-Friendly Program for the Analysis of Molecular Structures, Trajectories, and Free Energy Surfaces. J. Comput. Chem. 2011, 32 (6), 1183–1194.

38. Cozza, G.; Bortolato, A.; Moro, S. How Druggable Is Protein Kinase CK2? Med. Res.Rev. 2010, 30 (3), 419–462.

39. 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 (11), 2920–2923.

40. Wilce, M. C.; Parker, M. W. Structure and Function of Glutathione S-Transferases. Biochim. Biophys. Acta 1994, 1205 (1), 1–18.

41. Laborde, E. Glutathione Transferases as Mediators of Signaling Pathways Involved in Cell Proliferation and Cell Death. Cell Death Differ. 2010, 17 (9), 1373–1380.

42. Oakley, A. J.; Bello, M. Lo; Nuccetelli, M.; Mazzetti, A. P.; Parker, M. W. The Ligandin (non-Substrate) Binding Site of Human Pi Class Glutathione Transferase Is Located in the Electrophile Binding Site (H-Site). J. Mol. Biol. 1999, 291 (4), 913–926.

43. Aguirre, C.; Brink, T. ten; Guichou, J.-F.; Cala, O.; Krimm, I. Comparing Binding Modes of Analogous Fragments Using NMR in Fragment-Based Drug Design: Application to PRDX5. PLOS ONE 2014, 9 (7), e102300.

44. Declercq, J.-P.; Evrard, C.; Clippe, A.; Stricht, D. V.; Bernard, A.; Knoops, B. Crystal Structure of Human Peroxiredoxin 5, a Novel Type of Mammalian Peroxiredoxin at 1.5 Å resolution1. J. Mol. Biol. 2001, 311 (4), 751–759.

45. Shichita, T.; Hasegawa, E.; Kimura, A.; Morita, R.; Sakaguchi, R.; Takada, I.; Sekiya, T.; Ooboshi, H.; Kitazono, T.; Yanagawa, T.; Ishii, T.; Takahashi, H.; Mori, S.; Nishibori, M.; Kuroda, K.; Akira, S.; Miyake, K.; Yoshimura, A. Peroxiredoxin Family Proteins Are Key Initiators of Post-Ischemic Inflammation in the Brain. Nat. Med. 2012, 18 (6), 911–917.

46. Sjöholm, I.; Ekman, B.; Kober, A.; Ljungstedt-Påhlman, I.; Seiving, B.; Sjödin, T. Binding of Drugs to Human Serum albumin:XI. The Specificity of Three Binding Sites as Studied with Albumin Immobilized in Microparticles. Mol. Pharmacol. 1979, 16 (3), 767– 777.

47. Lejon, S.; Cramer, J. F.; Nordberg, P. Structural Basis for the Binding of Naproxen to Human Serum Albumin in the Presence of Fatty Acids and the GA Module. Acta Crystallograph. Sect. F Struct. Biol. Cryst. Commun. 2008, 64 (2), 64–69.

48. Kanner, B. I.; Zomot, E. Sodium-Coupled Neurotransmitter Transporters. Chem. Rev. 2008, 108 (5), 1654–1668.

49. Zhou, Z.; Zhen, J.; Karpowich, N. K.; Law, C. J.; Reith, M. E. A.; Wang, D.-N. Antidepressant Specificity of Serotonin Transporter Suggested by Three LeuT-SSRI Structures. Nat. Struct. Mol. Biol. 2009, 16 (6), 652–657.

50. Jacobson, K. A.; Gao, Z.-G. Adenosine Receptors as Therapeutic Targets. Nat. Rev. Drug Discov. 2006, 5 (3), 247–264.

51. Cooke, R. M.; Brown, A. J. H.; Marshall, F. H.; Mason, J. S. Structures of G Protein- Coupled Receptors Reveal New Opportunities for Drug Discovery. Drug Discov. Today 2015.

52. Sabbadin, D.; Ciancetta, A.; Deganutti, G.; Cuzzolin, A.; Moro, S. Exploring the Recognition Pathway at the Human A Adenosine Receptor of the Endogenous Agonist Adenosine Using Supervised Molecular Dynamics Simulations. Med Chem Commun 6,1081-1085, 2015.

53. Lebon, G.; Warne, T.; Edwards, P. C.; Bennett, K.; Langmead, C. J.; Leslie, A. G. W.; Tate, C. G. Agonist-Bound Adenosine A2A Receptor Structures Reveal Common Features of GPCR Activation. Nature 2011, 474 (7352), 521–525.

54. Barelier, S.; Linard, D.; Pons, J.; Clippe, A.; Knoops, B.; Lancelin, J.-M.; Krimm, I. Discovery of Fragment Molecules That Bind the Human Peroxiredoxin 5 Active Site. PLoS ONE 2010, 5 (3), e9744.

55. Krishnamurthy, H.; Gouaux, E. X-Ray Structures of LeuT in Substrate-Free Outward- Open and Apo Inward-Open States. Nature 2012, 481 (7382), 469–474.

56. Kober, A.; Sjöholm, I. The Binding Sites on Human Serum Albumin for Some Nonsteroidal Antiinflammatory Drugs. Mol. Pharmacol. 1980, 18 (3), 421–426.

3.2 Exploring the recognition pathway at the human A_{2A} adenosine receptor of the endogenous agonist adenosine using supervised molecular dynamics simulations.

Davide Sabbadin, Antonella Ciancetta, Giuseppe Deganutti, Alberto Cuzzolin and Stefano Moro*

Abstract

Adenosine is a naturally occurring purine nucleoside that exerts a variety of important biological functions through the activation of four G protein-coupled receptor (GPCR) isoforms, namely the A_1, A_2, A_{2B} and A_3 adenosine receptors (ARs). Recently, the X-ray structure of adenosine-bound hA_{2A} AR has been solved, thus providing precious structural details on receptor recognition and activation mechanisms. To date, however, little is still known about the possible recognition pathway the endogenous agonist might go through while approaching the hA_{2A} AR from the extracellular environment. In the present work, we report the adenosine-hA_{2A} AR recognition pathway through the analysis of a series of Supervised Molecular Dynamics (SuMD) trajectories. Interestingly, a possible energetically stable meta-binding site has been detected and characterized.

Introduction

Adenosine is a naturally occurring purine nucleoside that forms primarily from the metabolism of adenosine triphosphate (ATP), both intracellularly and extracellularly[1]. Consequently, the extracellular levels of adenosine are regulated by its synthesis, metabolism, release and uptake[1,2]. Adenosine exerts pleiotropic functions throughout the body. In the central nervous system (CNS), the nucleoside plays important functions, such as modulation of neurotransmitter release, synaptic plasticity and neuroprotection in ischemic, hypoxic and oxidative stress events[1,3,4]. In addition, adenosine plays different roles in a large variety of tissues. In the cardiovascular system, adenosine produces either

Published - Sabbadin D, Ciancetta A, Deganutti G, Cuzzolin A, Moro S. *MedChemComm* 6, 1081-1085 (2015)

vasoconstriction or vasodilation of veins and arteries. Moreover, adenosine regulates T cell proliferation and cytokine production, inhibits lipolysis and stimulates bronchoconstriction[1,3,4].

Adenosine mediates its biological effects by recognizing four G protein-coupled receptor (GPCR) isoforms, namely the A_1, A_{2A}, A_{2B} and A_3 adenosine receptors (ARs). Each subtype has a unique pharmacological profile, tissue distribution and effector coupling[1,4]. Considering receptor sequence similarity, among the human ARs (hARs), the most similar are the A_1 and A_3 ARs (49% similarity), and the A_{2A} and A_{2B} ARs (59% similarity). Conversely, the A_1, A_{2A} and A_3 ARs possess relatively high affinity for adenosine whereas the A_{2B} AR shows relatively lower affinity for adenosine, as summarized in Table 1.

Recently, the crystallographic structure of adenosine- bound hA_{2A} AR has been solved (PDB code: 2YDO)[5]. Although this structural data is extremely precious to interpret both receptor recognition and activation mechanisms of the endogenous agonist, little is still known about the possible recognition pathway between adenosine coming from the extracellular environment and the hA_{2A} AR embedded in the cytoplasmic membrane.

Table 2 - Adenosine affinities at the four receptor subtypes

	hA_1 , K_i (nM)	**hA_{2A}, K_i (nM)**	**hA_{2B}[a]**	**hA_3**
Adnosine	***ca.* 100**	**310**	**15,000**	**290**

[a] Data from functional studies. [b] ref. 4

In this context, Supervised Molecular Dynamics (SuMD) has been recently presented as an alternative computational method that allows the exploration of ligand–receptor recognition pathway investigations on a nanosecond (ns) time scale[6]. In addition to speeding up the acquisition of the ligand–receptor recognition trajectory, this approach facilitates the identification and the structural characterization of multiple binding events (such as meta-binding, allosteric, and orthosteric sites) by taking advantage of the all-atom MD

simulation accuracy of a GPCR–ligand com- plex embedded into an explicit lipid–water environment[6].

In the present study, in order to better understand how adenosine approaches the orthosteric binding site of the hA_{2A} AR, its recognition pathway has been described through the analysis of a series of SuMD trajectories. Interestingly, a possible energetically stable meta-binding site has been detected and characterized. The meta-binding site concept was introduced several years ago to describe those binding events that chronologically anticipate the orthosteric binding event[7].

Results and discussion

As anticipated, recently the crystallographic structure of adenosine-bound hA_{2A} AR has been solved. The attempt to apply MD methodology to address the problem of ligand dissociation from its receptor is subjected to some limitations. First of all, ligand dissociation dynamics is usually a slow event in comparison to the timescales accessible to current simulation techniques and computer resources. This does not mean necessarily that the actual event of ligand dissociation takes so long, but it is clear that conformational sampling cannot be done effectively in a conventional MD simulation. On the other hand, the recognition process between a ligand and its receptor is a very rare event to describe at the molecular level and, even with the recent GPU-based computing resources, it is necessary to carry out classical molecular dynamics (MD) experiments on a long microsecond time scale[6]. For this reason, in order to better under- stand how adenosine approaches the orthosteric binding site of the hA_{2A} AR, its recognition pathway was explored using a SuMD study (Video 1).

In particular, following the ligand recognition pathway emerged by the analysis of SuMD trajectories (Fig. 1 and Video 1), the third extracellular loop (EL3) of hA_{2A} AR plays an essential role in directing the agonist toward the orthosteric binding site. In particular, His264, Ala265, Pro266 (EL3) and Leu267 (7.32) (Fig. 1, panel A) establish favourable hydrophobic contacts with the adenine core of adenosine. Such interactions orient the ribose ring towards the entrance of the

orthosteric binding site. The hydroxyl group in the C3' position of the ribose ring is engaged in a direct hydrogen bond interaction with Glu169 (EL2). Not surprisingly, the described extracellular site corresponds to the previously reported meta-binding site located in EL3[6,7], which enables high-potency hA_{2A} AR antagonists, such as ZM 241385, 6-IJ2,6- dimethylpyridin-4-yl)-5-phenyl-1,2,4-triazin-3-amine (T4G), and 4-IJ3-amino-5-phenyl-1,2,4-triazin-6-yl)-2-chlorophenol (T4E), to reach the orthosteric binding cleft from the extracellular vestibule. As already described, once the antagonists reach the orthosteric binding site, they adopt binding conformations that match the geometric positions observed in the corresponding X-ray structures[6].

By approaching the orthosteric binding site, adenosine explores receptor-bound states that only partially overlap – RMSD < 3.5 Å – (Fig. 1, panel B–C) with the crystallographic bound conformation. In such conformational states, the ribose moiety explores the bottom part of the binding pocket ("ribose-down" conformation) and is in close contact with Thr88 (3.36). Glu169 (EL2) and Asn53 (6.55) are involved in key polar interactions with the endo and exocyclic nitrogen atoms of the aromatic core. Hydrophobic interactions are established with Met174 (5.35), Met177 (5.38), Ala59 (2.57), Ala63 (2.61), Val84 (3.32) and Ile160 (El2). In particular, Phe168 (EL2) is involved in π-stacking interaction with the adenine core.

Notably, the role of several key residues (such as Thr88 (3.36), Phe168 (EL2) and Met177 (5.38)) herein highlighted is consistent with the available mutagenesis data for agonist binding, which have been recently analysed and clarified by means of MD/FEP calculations[8].

As reported in Fig. 1, panel D–F, once inside the orthosteric site, adenosine dynamically flips between two different binding modes: the one above reported – the so-called "ribose-down" conformation – and the "ribose-up" conformation (Fig. 1, panel D) where the ribose moiety is directed towards the extracellular space. The hydroxyl group, attached at the C2' position of the ribose ring, establishes a hydrogen- bond interaction with Glu169 (EL2) and the exocyclic nitro- gen atom of the adenine ring interacts with the Ser67 (2.65) side

chain. The agonist aromatic ring is involved in a π-stacking interaction with Phe168 (EL2). Val84 (3.32), Ala63 (2.61) and Met174 (5.35) are responsible of the majority of non-polar ligand–receptor contacts.

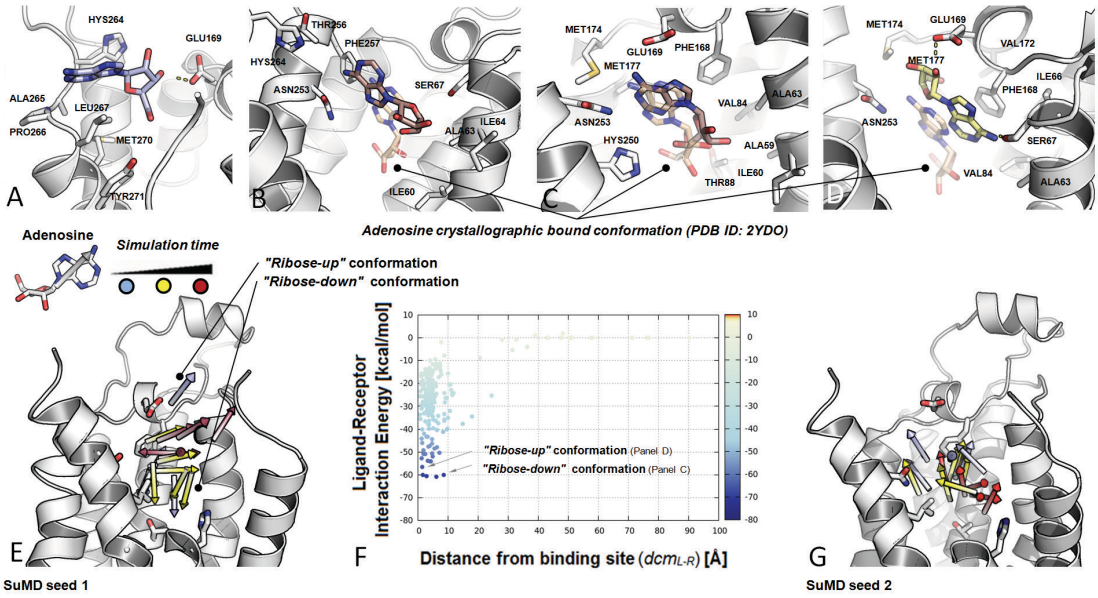

Figure 1 (Panel A to D) - Overview of multiple adenosine binding conformation inside the hA_{2A} AR binding pocket generated from SuMD simulation trajectories in comparison with X-ray crystal structure, PDB ID: 2YDO (wheat sticks). Stick colouring scheme is based on simulation progression (time). Hydrogen atoms are not displayed, whereas hydrogen bond interactions are highlighted as yellow dashed lines. (Panel E and G) Overview of multiple discrete binding states that occur during ligand–receptor recognition. Arrow colouring scheme is based on simulation progression (time). Receptor ribbon representation is viewed from the membrane side facing transmembrane domain 6 (TM6) and transmembrane domain 7 (TM7). (Panel F) Ligand−receptor interaction energy landscape for the nonbiased adenosine-hA_{2A} AR recognition process. The most energetically stable binding conformations of adenosine inside the hA_{2A} AR binding pocket are highlighted by arrow. Interaction energy values are expressed in kcal mol^{-1}.

Therefore, although the described ligand–receptor contacts provide sufficient energetic protein–ligand complex stabilization to reach the global protein–ligand interaction energy minimum (Fig. 1, panel F), the recognition of

the agonist is not accompanied by subsequent stabilization of the ligand conformation within the orthosteric site, as adenosine dynamically flips between the "ribose down" and "ribose up" binding modes (Fig. 1, panel E). Therefore, as also elucidated by a clustering analysis of the space explored by adenosine during the binding pathway, the agonist recognition process does not show the same behaviour of potent hA_{2A} AR antagonists. The adenosine binding profile, instead, is more similar to the one observed for a weak binder such as caffeine (Fig. 1, panel F)[6]. Moreover, this peculiar conformational landscape along with the emerged major interaction sites, which anticipate the orthosteric binding site, is independent from ligand placement and orientation at the beginning of the SuMD simulation (Fig. 1, panel G).

Experimental

General

The numbering of the amino acids follows the arbitrary scheme by Ballesteros and Weinstein: each amino acid identifier starts with a helix number, followed by the position relative to a reference residue among the most conserved amino acids in that helix, to which the number 50 is arbitrarily assigned[9].

Trajectory analysis and figure and video generation have been performed using several functionalities implemented by Visual Molecular Dynamics[10], WORDOM[11], the PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC (http://www.pymol.org/) and the Gnuplot graphic utility (http://www.gnuplot.info/). Ligand-hA_{2A} AR interaction energies were calculated by extrapolating the non-bonded energy interaction term of CHARMM27 Force Field[12] using NAMD[13].

Computational facilities

All computations were performed on a hybrid CPU/GPU cluster. Molecular dynamics simulation has been performed with a 2 NVIDIA GTX 680 and 3 NVIDIA GTX 780 GPU cluster engineered by Acellera[14].

Human A_{2A} adenosine receptor–ligand complex preparation

The selected agonist-bound crystal structures (PDB IDs: 2YDO[5]) and the FASTA sequence of the hA_{2A} AR (Uniprot ID: P29274) were retrieved from the RCSB PDB database[15] (http://www.rcsb.org) and the UniProtKB/Swiss-Prot[16,17], respectively. The co-crystallized ligand structure was extracted from the orthosteric binding site and randomly placed in the space above the receptor, at least 40 Å away from protein atoms. Ionization states and hydrogen positions were assigned by using the MOE-sdwash utility (pH 7.0). The FASTA sequence was aligned, using BLAST (Blosum 62 matrix)[18], with the template sequence. Backbone and conserved residue coordinates were copied from the template structure, whereas newly modelled regions and non-conserved residue side chains were modelled and energetically optimized by using CHARMM 27 force field[12] until a r.m.s. of conjugate gradient <0.05 kcal mol^{-1} $Å^{-1}$ was reached. Missing loop domains were constructed by the loop search method implemented in the Molecular Operating Environment (MOE, version 2012.10) program[19] on the basis of the structure of compatible fragments found in the Protein Data Bank. N-terminal and C-terminal were deleted if their lengths exceeded those found in the crystallographic template. The "Protonate-3D" tool[20] was used to appropriately assign ionization states and hydrogen positions to the build models. Then, the structures were subjected to energy minimization with CHARMM 27 force field[12] until the r.m.s. of conjugate gradient was <0.05 kcal mol^{-1} $Å^{-1}$. Protein stereochemistry evaluation was then performed by employing several tools (Ramachandran and χ plots measure j/ψ and $\chi 1/\chi 2$ angles, clash contact reports) implemented in the MOE suite[19].

Receptor membrane embedding and system preparation

Receptors were embedded in a 1-palmitoyl-2-oleoyl-sn-glycero-3 phosphocholine (POPC) lipid bilayer (85 × 85 Å wide) and placed into the membrane according to the suggested orientation reported in the "Orientations of Proteins in Membranes (OPM)" database[21] for the hA_{2A} AR in a complex with the antagonist T4G (PDB ID: 2YDV[7]). Overlapping lipids (within 0.6 Å) were removed upon insertion of the protein. The prepared systems were solvated with TIP3P water[22] using the program Solvate 1.0[23] and neutralized with Na^+/Cl^- counterions to a final concentration of 0.154 M. The total number of atoms per system was approximately 75,000. Membrane MD simulations were carried out on a GPU cluster with the ACEMD program using the CHARMM27 Force Field[18] and periodic boundary conditions. Initial parameters for the ligands were derived from the CHARMM General Force Field for organic molecules[24,25]. The system was equilibrated using a stepwise procedure. In the first stage, to reduce steric clashes due to the manual setting up of the membrane–receptor system, a 500 step conjugate-gradient minimization was performed. Then, to allow lipids to reach equilibrium and water molecules to diffuse into the protein cavity and to avoid ligand–receptor interaction in the equilibration phase, protein and ligand atoms were restrained for the first 8 ns by a force constant of 1 kcal mol^{-1} $Å^{-2}$. Then side chains were set free to move, while gradually reducing the force constant to 0.1 kcal mol^{-1} $Å^{-2}$ to the ligand and alpha carbon atoms up to 9 ns. Temperature was maintained at 298 K using a Langevin thermostat with a low damping constant of 1 ps^{-1}, and the pressure was maintained at 1 atm using a Berendsen barostat. Bond lengths involving hydrogen atoms were constrained using the M-SHAKE algorithm[26] with an integration time step of 2 fs. Harmonical constraints were then removed and Supervised MD was conducted in a NVT ensemble. Long-range Coulomb interactions were handled using the particle mesh Ewald summation method (PME)[27] with grid size rounded to the approximate integer value of cell wall dimensions. A non-bonded cutoff distance of 9 Å with a switching distance of 7.5 Å was used. In order to assess the biophysical validity of the built systems,

the average area per lipid headgroup (APL) and bilayer thickness measurements for each built system were measured using Grid-MAT-MD[28]. The corresponding calculated averaged area per lipid headgroup of the extracellular and intracellular leaflet during the production phase for all simulations was in agreement with the experimental values measured for 1-palmitoyl-2- oleoyl-sn-glycero-3-phosphocholine (POPC) lipid bilayers[29].

Conclusions

In the present work, we have carried out SuMD experiments to elucidate the recognition pathway of the naturally occurring purine nucleoside adenosine by the hA_{2A} AR. The analysis of the SuMD trajectories revealed that residues located in the third extracellular loop play an essential role in orienting the ribose ring of agonist toward the entrance of the orthosteric site, thus representing a possible energetically stable meta-binding site.

Our analysis has also revealed that, once the orthosteric site is reached, adenosine experiences a dynamic flip between two different binding modes: the "ribose-down" and the “ribose-up” conformation, with the ribose moiety pointing towards the intracellular and extracellular space, respectively. Consequently, the adenosine binding profile resulting from our analysis resembles that of a weak binder rather than the one previously observed for potent hA_{2A} AR antagonists.

Further work is underway in our lab to better elucidate the role of the meta-binding site that has been detected and characterized in this study. In particular, SuMD simulations with adenosine-hA_{2A} AR 2:1 stoichiometry are currently under evaluation. Moreover, we are carrying out a comprehensive SuMD exploration of the recognition pathway of adenosine against all other adenosine receptor subtypes to clarify the experimental selectivity profile provided by the natural agonist.

Bibliography

1. Jacobson, K. A. & Gao, Z.-G. Adenosine receptors as therapeutic targets. *Nat. Rev. Drug Discov.* **5,** 247–264 (2006).

2. Latini, S. & Pedata, F. Adenosine in the central nervous system: release mechanisms and extracellular concentrations. *J. Neurochem.* **79,** 463–484 (2001).

3. Sebastião, A. M. & Ribeiro, J. A. Fine-tuning neuromodulation by adenosine. *Trends Pharmacol. Sci.* **21,** 341–346 (2000).

4. Fredholm, B. B., IJzerman, A. P., Jacobson, K. A., Linden, J. & Müller, C. E. International Union of Basic and Clinical Pharmacology. LXXXI. Nomenclature and classification of adenosine receptors--an update. *Pharmacol. Rev.* **63,** 1–34 (2011).

5. Lebon, G. *et al.* Agonist-bound adenosine A2A receptor structures reveal common features of GPCR activation. *Nature* **474,** 521–525 (2011).

6. 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).

7. Moro, S., Hoffmann, C. & Jacobson, K. A. Role of the extracellular loops of G protein-coupled receptors in ligand recognition: a molecular modeling study of the human P2Y1 receptor. *Biochemistry (Mosc.)* **1,** 3498–3507.

8. Keränen, H., Gutiérrez-de-Terán, H. & Åqvist, J. Structural and Energetic Effects of A2A Adenosine Receptor Mutations on Agonist and Antagonist Binding. *PLoS ONE* **9,** e108492 (2014).

9. Juan A. Ballesteros, H. W. Integrated Methods for the Construction of Three-Dimensional Models and Computational Probing of Structure-Function Relations in G-Protein-Coupled Receptors. *Methods Neurosci.* **25,** 366–428 (1995).

10. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **14,** 33–38, 27–28 (1996).

11. Seeber, M. *et al.* Wordom: A user-friendly program for the analysis of molecular structures, trajectories, and free energy surfaces. *J. Comput. Chem.* **32,** 1183–1194 (2011).

12. MacKerell, A. D., Banavali, N. & Foloppe, N. Development and current status of the CHARMM force field for nucleic acids. *Biopolymers* **56,** 257–265 (2000).

13. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26,** 1781–1802 (2005).

14. Acellera. *Acellera* at <https://www.acellera.com/>

15. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28,** 235–242 (2000).

16. UniProt Consortium. The Universal Protein Resource (UniProt) in 2010. *Nucleic Acids Res.* **38,** D142–148 (2010).

17. Jain, E. *et al.* Infrastructure for the life sciences: design and implementation of the UniProt website. *BMC Bioinformatics* **10,** 136 (2009).

18. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25,** 3389–3402 (1997).

19. Chemical Computing Group. at <http://www.chemcomp.com/>

20. Labute, P. Protonate3D: Assignment of ionization states and hydrogen coordinates to macromolecular structures. *Proteins Struct. Funct. Bioinforma.* **75,** 187–205 (2009).

21. Lomize, M. A., Lomize, A. L., Pogozheva, I. D. & Mosberg, H. I. OPM: orientations of proteins in membranes database. *Bioinforma. Oxf. Engl.* **22,** 623–625 (2006).

22. 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).

23. Grubmuller, H. & Groll, V. Solvate. (1996). at <http://www.mpibpc.mpg.de/grubmueller/solvate>

24. Vanommeslaeghe, K. & MacKerell, A. D. Automation of the CHARMM General Force Field (CGenFF) I: Bond Perception and Atom Typing. *J. Chem. Inf. Model.* **52,** 3144–3154 (2012).

25. Vanommeslaeghe, K., Raman, E. P. & MacKerell, A. D. Automation of the CHARMM General Force Field (CGenFF) II: assignment of bonded parameters and partial atomic charges. *J. Chem. Inf. Model.* **52,** 3155–3168 (2012).

26. 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).

27. Essmann, U. *et al.* A smooth particle mesh Ewald method. *J. Chem. Phys.* **103,** 8577–8593 (1995).

28. Allen, W. J., Lemkul, J. A. & Bevan, D. R. GridMAT-MD: a grid-based membrane analysis tool for use with molecular dynamics. *J. Comput. Chem.* **30,** 1952–1958 (2009).

29. Kucerka, N., Tristram-Nagle, S. & Nagle, J. F. Structure of fully hydrated fluid phase lipid bilayers with monounsaturated chains. *J. Membr. Biol.* **208,** 193–202 (2005).

3.3 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.

Giuseppe Deganutti, Alberto Cuzzolin, Antonella Ciancetta, Stefano Moro*

Abstract

The search for G protein-coupled receptors (GPCRs) allosteric modulators represents an active research field in medicinal chemistry. Allosteric modulators usually exert their activity only in the presence of the orthosteric ligand by binding to protein sites topographically different from the orthosteric cleft. They therefore offer potentially therapeutic advantages by selectively influencing tissue responses only when the endogenous agonist is present. The prediction of putative allosteric site location, however, is a challenging task. In facts, they are usually located in regions showing more structural variation among the family members. In the present work, we applied the recently developed Supervised Molecular Dynamics (SuMD) methodology to interpret at the molecular level the positive allosteric modulation mediated by LUF6000 toward the human adenosine A_3 receptor (hA_3 AR). Our data suggest at least two possible mechanisms to explain the experimental data available. This study represent, to the best of our knowledge, the first case reported of an allosteric recognition mechanism depicted by means of molecular dynamics simulations.

Introduction

Besides the orthosteric site, which conventionally recognizes endogenous ligands, most G protein-coupled receptors (GPCRs) possess topographically distinct allosteric sites that can be recognized by small molecules and accessory cellular proteins. Pharmacologically speaking, an allosteric modulator does not have any activity by itself, thus needing the orthosteric binder to exhibit its action. Although the modulatory character of allosteric binders is not always

Published - Deganutti G, Cuzzolin A, Ciancetta A, Moro S. *Biorg Med Chem*. 23, 4065-4071 (2015)

clear-cut, true allosteric modulators increase or decrease the action of an agonist or an antagonist recognising the allosteric site(s) on the receptor. In facts, ligand binding to allosteric sites promotes a conformational reorganization in the GPCR that can alter orthosteric ligand affinity and/or efficacy. Although an allosteric modulator may not possess efficacy by itself, it can provide a powerful therapeutic advantage over orthosteric ligands, as they selectively influence tissue responses only when the endogenous agonist is present. Consequently, allosteric modulation of GPCRs has stimulated an intensive identification campaign for new classes of hit-candidates different from conventional agonists and antagonists. This has been the subject of several recent reviews[1–3].

However, natural allosteric sites are very difficult to identify because they are usually located far from the orthosteric sites. Moreover, allosteric sites resides in regions of the receptor that show more structural variation among family members and, consequently, this implies a general lack of success in predicting the locations of potential binding regions. Albeit the crystallographic structure of the M2 receptor simultaneously bound to the orthosteric agonist iperoxo and the positive allosteric modulator LY2119620 has been recently reported[4], little is known about the possible allosteric control regarding the activation mechanism of other GPCRs.

Within this framework, we have recently reported on an alternative computational method – the Supervised Molecular Dynamics (SuMD) – that allows to investigate the ligand– receptor recognition pathway in a nanosecond (ns) time scale[5]. In addition to speeding up the acquisition of the ligand–receptor recognition trajectory, this approach facilitates the identification and the structural characterization of multiple binding events (such as meta-binding, allosteric, and orthosteric sites) by taking advantage of the all-atom MD simulations accuracy of GPCR– ligand complexes embedded into explicit lipid–water environment[5].

Interestingly, adenosine receptors (ARs) were among the first GPCRs discovered to be allosterically regulated and, in particular, allosteric enhancers for A1 and A3 ARs have been widely investigated[1,2,6]. Among the most interesting

allosteric enhancers for the A3 AR, *N*-(3,4-dichlorophenyl)-2-cyclohexyl- 1*H*-imidazo[4,5-*c*]quinolin-4-amine (LUF6000, see Fig. 1) has been deeply characterized[7,8]. LUF6000 potentiates the maximum efficacy of the agonist Cl-IB-MECA by 45–50%, enhances agonist efficacy in functional assays and decreases the agonist dissociation rate without influencing agonist potency. Moreover, LUF6000 has been reported to act as allosteric enhancer of the maximal effect exerted by structurally diverse agonists at the A_3 AR, being more effective for low-efficacy than for high-efficacy agonists.

Very recently, in vivo studies have reported the ability of LUF6000 to act as allosteric modulator of rat and mice A_3 ARs by allowing the endogenous ligand adenosine to bind to the receptor with higher affinity[9].

With the aim to interpret at the molecular level the positive allosterism mediated by LUF6000 toward the human A_3 AR (hA_3 AR), possible recognition pathways have been explored by performing SuMD simulations in the absence and in presence of the natural agonist adenosine (Fig. 1). Interestingly, our results suggest two possible mechanisms by which LUF6000 might exert its positive allosteric modulator effects: according to the outcomes of our simulations, the ligand might either induce a loop rearrangement that stabilizes agonist placement into the orthosteric site, or form a ternary complex with the agonist bound receptor state, thus acting as orthosteric pocket cap.

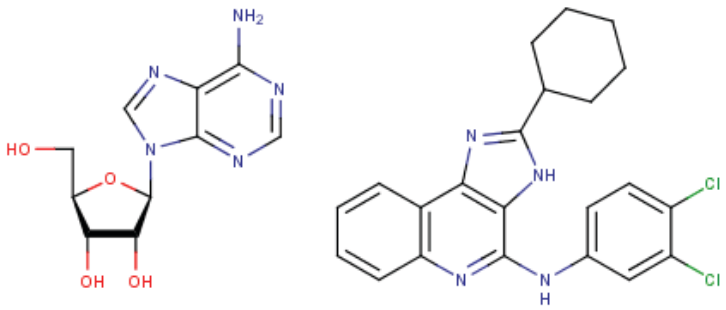

Figure 1- Structures of the endogenous hA_3 AR agonist adenosine (left) and the positive allosteric modulator LUF6000 (right).

SuMD Simulations

LUF6000-hA_3 AR recognition mechanism.

The imidazoquinolinamine allosteric modulator LUF6000 enhances agonist efficacy in functional assays and decreases agonist dissociation rate without influencing agonist potency[7,8]. Besides, LUF6000 presents a weak antagonist activity (ca. 45% inhibition at 10 μM)[7]. To explore LUF6000 attitude to recognize the orthosteric binding site of the hA_3 AR, we analysed its recognition pathway by performing SuMD experiments.

During the SuMD simulations, LUF6000 reached the orthosteric binding site in less than 20 ns. The corresponding energy landscape (Fig. 2A) highlights two major interaction sites (**a** and **b**). Prior reaching the orthosteric site, LUF6000 interacts with residues located in a region between the second and third extracellular loops (EL2 and EL3, respectively). Met151 (EL3), Thr154 (EL2), Met174 (5.35) side chains and the aliphatic portion of Arg173 (EL2) establish hydrophobic contacts with the imidazoloquinoline core of LUF6000, whereas the ligand exocyclic nitrogen atom is involved in a hydrogen bond interaction with the backbone of Lys152 (EL2) (**a** in Fig. 2A, Fig. 2B, Video S1). While reaching the orthosteric site, the Val169 (EL2) side chain facilitates ligand reorientation providing favourable hydrophobic contacts. The most stable conformation observed (**b** Fig. 2A, Fig. 2C, Video S1) is characterized by hydrophobic contacts with Phe168 (EL2), Met174 (5.35), Leu246 (6.51), Leu264 (7.35), Leu268 (7.39) and Trp243 (6.48), whereas Asn250 (6.55) is engaged in a hydrogen bond with the exocyclic nitrogen of the ligand. At the maximum energetic stabilization, the formed complex is characterized by energetic values of about -50 kcal/mol, a value previously observed for weak ARs binders. This is consistent with the LUF6000 antagonist activity observed through functional assays at the hA_3 AR[8].

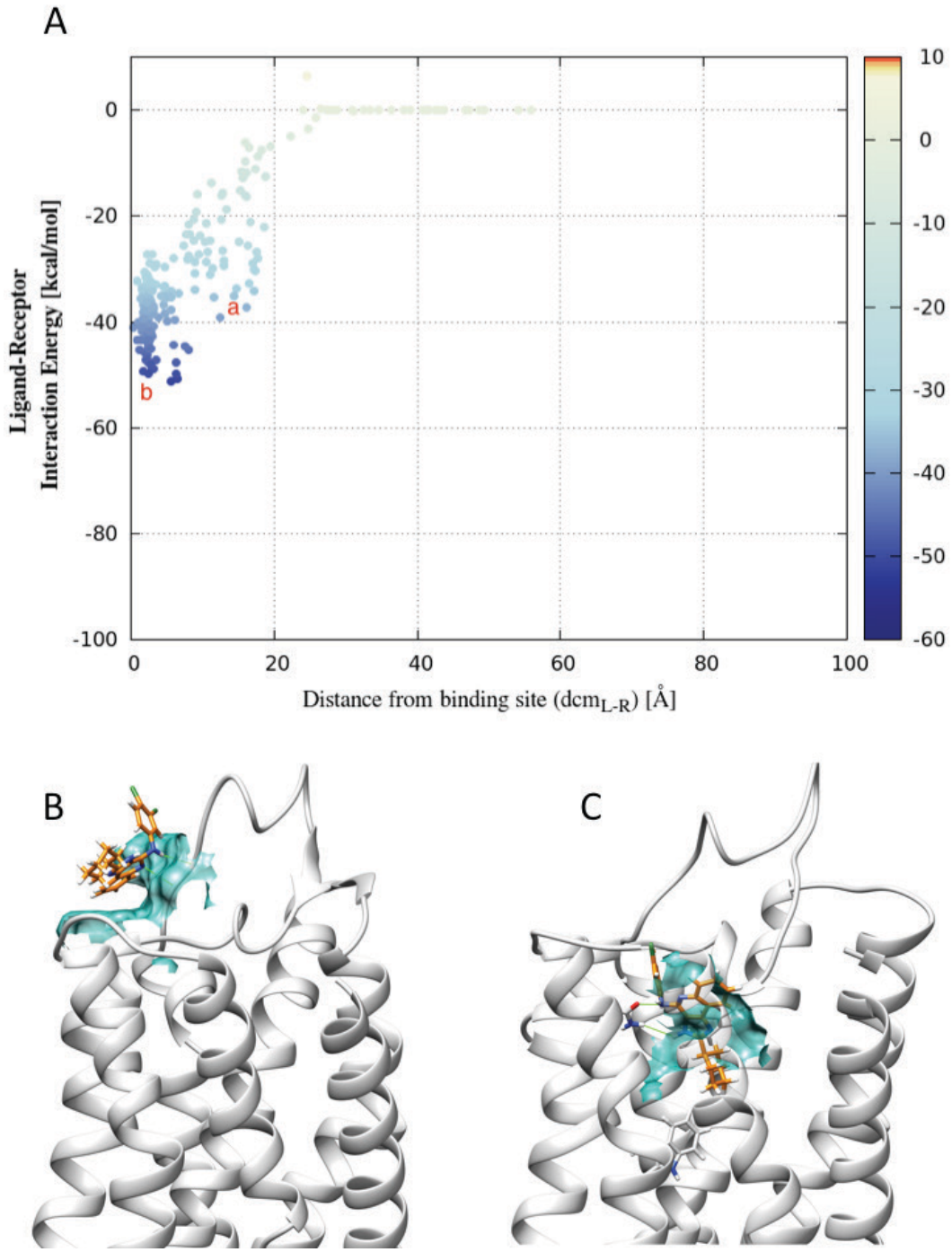

Figure 2. (A) Interaction Energy landscape for the recognition pattern of LUF6000 by the hA_3 AR.(B) LUF6000 binding mode in the meta-binding site.(C) LUF6000 binding mode in the orthosteric binding site. Ligand is displayed as orange stick, side chains of residues interacting through hydrogen bond or π-π stacking are depicted as grey stick, whereas side chains of residues interacting through hydrophobic contacts are rendered as coloured surfaces.

Adenosine-hA$_3$ AR recognition mechanism

Recently, the crystallographic structure of adenosine in complex with the hA$_{2A}$ AR has been solved[10]. This structural piece of information aid elucidating both the recognition and activation mechanisms of the ARs by their endogenous agonist adenosine. Although in principle the interaction pattern of the adenosine-hA$_{2A}$ AR complex can be transferred to the other ARs subtypes, in order to better understand how adenosine approaches the orthosteric binding site of the hA$_3$ AR, its recognition pathway as described by the obtained SuMD trajectories has been analyzed.

In our SuMD experiments, adenosine reached the orthosteric binding site in less than 20 ns. The corresponding energy landscape is reported in Fig. 3A. During the recognition pathway (Video S2), EL3 engages the ligand ribose moiety in favourable hydrogen bonds mainly through Val259 (EL3) and Gln261 (EL2) backbone atoms (**b** in Fig. 3A, Fig. 3B). This situation anticipates a change of adenosine orientation, triggered by hydrophobic contacts between the ligand purine core and Leu264 (7.35), Ile268 (7.39), Ile253 (6.58), and Ile249 (6.54) side chains. Once the orthosteric pocket is reached, adenosine interacts with the side chain of Trp185 (5.46), Leu246 (6.51) and conserved residues Asn250 (6.55), Phe168 (EL2), Trp243 (6.48), Ile268 (7.39) (Fig. 3C, Video S2). As already observed for adenosine recognition pathway by the hA$_{2A}$ AR[11], the agonist explores different conformational states once inside the pocket. In particular, adenosine experiences a dynamic flip between two different binding modes: the above described "ribose-down" and the "ribose-up" conformation, with the ribose moiety pointing towards the intracellular and extracellular space, respectively. The ribose down conformation (**b** in Fig. 3A, Fig. S1) is characterized by additional electrostatic interactions with Glu19 (1.39) and Ser73 (2.65) and represents the most energetically stable ligand-receptor complex observed in the analysed trajectories.

LUF6000-hA$_3$ AR (in complex with adenosine) recognition mechanism

With the aim to reproduce the experimental conditions that allow to measure LUF6000 PAM activity, SuMD simulations were performed considering the hA$_3$ AR receptor in complex with adenosine. LUF6000 was randomly placed 60 Å at least away from the barycentre of the orthosteric binding site. The starting adenosine-hA$_3$ AR complex was extracted from the previously described SuMD trajectory, and selected on the basis of its similarity with the X-Ray crystallographic conformation observed for the complex with the hA$_{2A}$ AR[10].

The LUF6000 recognition energy landscape is reported in Fig. 4A. The pathway described by the SuMD trajectories highlights three main situations: *i)* LUF6000 not interacting with the adenosine-hA$_3$ AR complex (point a in Fig.4A); *ii)* LUF6000 interacting with a meta-binding site (**b** in Fig. 4A); and *iii)* LUF6000 interacting with the orthosteric pocket (**c** in Fig. 4A).

Moreover, during the SuMD simulations, the interaction energy between adenosine and the hA$_3$ AR has been computed (Fig. 4B). The endogenous agonist reaches a stability maximum approximately after 10 ns of simulation (point B in Fig. 4B), which correspond to LUF6000-hA$_3$ AR meta binding site complex formation. The ligand approaches EL2 (Fig. 5A) by recruiting Tyr157 (EL2) and His158 (EL2) side chains, and establishing contacts with Arg173 (EL2), Met174 (5.35), Ile253 (6.58), Tyr254 (6.59) (Fig. 4B, Video S3). This loop rearrangement is accompanied by conformational changes in residues located farther in EL2, included Phe168 (EL2), that loses the capability of stabilizing adenosine through π-π stacking interactions. As a consequence, adenosine moves deeper in the orthosteric pocket (Fig. S2) and establishes favourable interactions with Leu246 (6.51), Trp243 (6.48), Ser247 (6.52), and Thr94 (3.36). The overall protein conformational and adenosine positional changes that occur after the interaction between LUF6000 and the hA$_3$ AR EL2 (*i.e.* system evolution from point a to point b) are reported in Fig. S3.

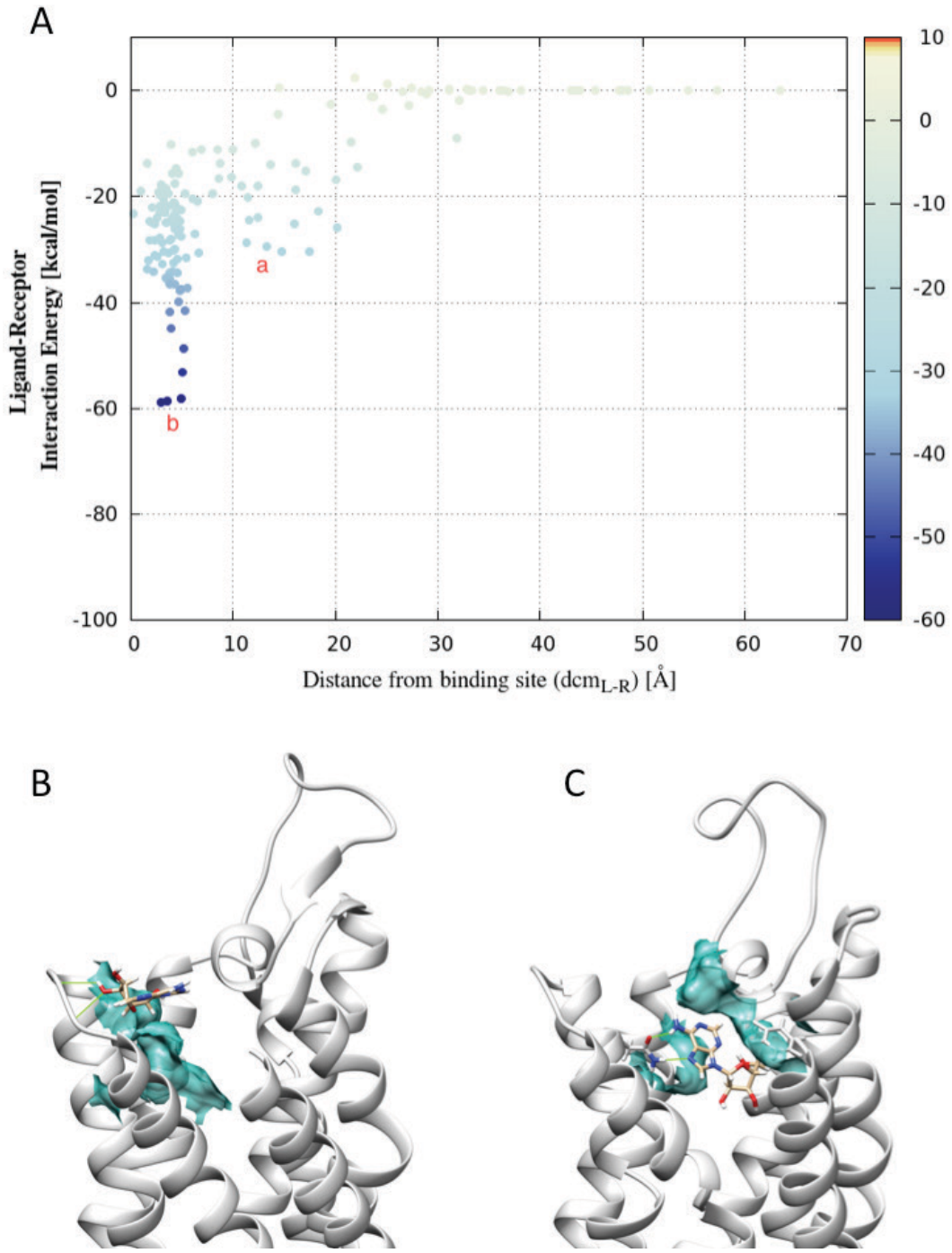

Figure 3. (A) Interaction Energy landscape for the recognition pattern of adenosine by the hA_3 AR.(B) Adenosine binding mode in the meta-binding site.(C) Adenosine binding mode in the orhtosteric binding site. Ligand is displayed as tan stick, side chains of residues interacting through hydrogen bond or π-π stacking are depicted as grey stick, whereas side chains of residues interacting through hydrophobic contacts are rendered as coloured surfaces

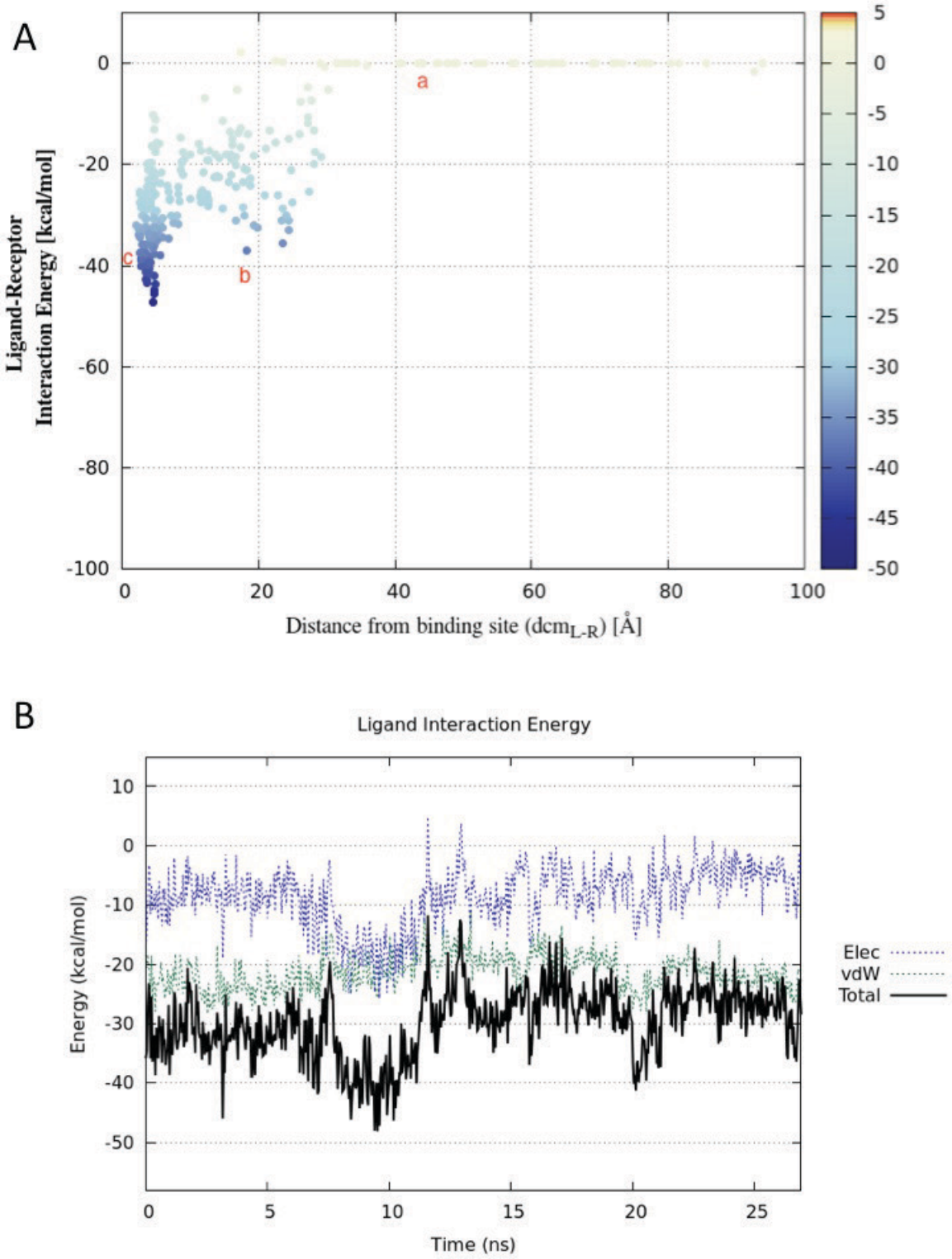

Figure 4. (A) Interaction Energy landscape for the recognition pattern of LUF6000 by the hA_3 AR-adenosine complex. (B) hA_3 AR-adenosine interaction energy.

Once the complex with the meta binding site is formed, approximately after 14 ns of simulations, LUF6000 moves to establish hydrophobic interactions with Tyr254 (6.59), Met174 (5.35), Val169 (EL2) and Ile253 (6.58), located at the top of orthosteric binding site. This allows the ligand to directly interact with adenosine. Simultaneously, the previously evidenced π-π stacking interaction between adenosine and Phe168 (EL2) is restored. In the tertiary complex just formed, the energy interaction between adenosine and the hA_3 AR is stabilized at values slightly lower than the starting complex, with the exception of a transitory stabilization after 21 ns (C in Fig. 4B, Fig. 5B). LUF6000 is therefore able to stabilize the interaction energy between adenosine and the hA_3 AR and to lock the agonist inside the orthosteric pocket for the remaining simulation time. A similar behaviour has been observed also for the tertiary complex formed with a LUF6000 close analogue LUF6069 (See Supplementary Information, Fig. S4-S5 and Video S4).

Conclusion

In the present work, we have utilized SuMD[5], a computational approach we have recently developed, with the aim to characterize and rationalize the activity of LUF6000, a hA_3 AR PAM, at a molecular level. We have analysed the ligand-receptor recognition pattern, both for LUF6000 and the endogenous agonist adenosine separately and also considering the recognition pathway of the PAM by the hA_3 AR in complex with adenosine. This represents, to date, the first case reported of an allosteric mechanism investigated by means of MD simulations.

Our results have highlighted that LUF6000 is able to establish favourable interactions with conserved residues located in the orthosteric binding site of the hA_3 AR, consistently with the experimentally observed weak inhibitor activity at this receptor subtype. The analysis of the interaction pathway of the endogenous agonist adenosine suggests a key role played by residues located in the EL2 in engaging the agonist and energetically promoting its approach to the orthosteric pocket.

The inspection of the interaction pathway obtained by simulating LUF6000 approaching the hA_3 AR in complex with the endogenous agonist adenosine suggests two possible mechanisms to explain the experimentally observed positive allosteric modulation[7,8]. According to our analysis, the ligand could: *i)* trigger conformational changes in the EL2 that would enable the agonist to form more energetically favourable interactions with residues located deeper in the orthosteric binding site; *ii)* establish a ternary complex with the agonist and the receptor, thus acting as orthosteric pocket cap.

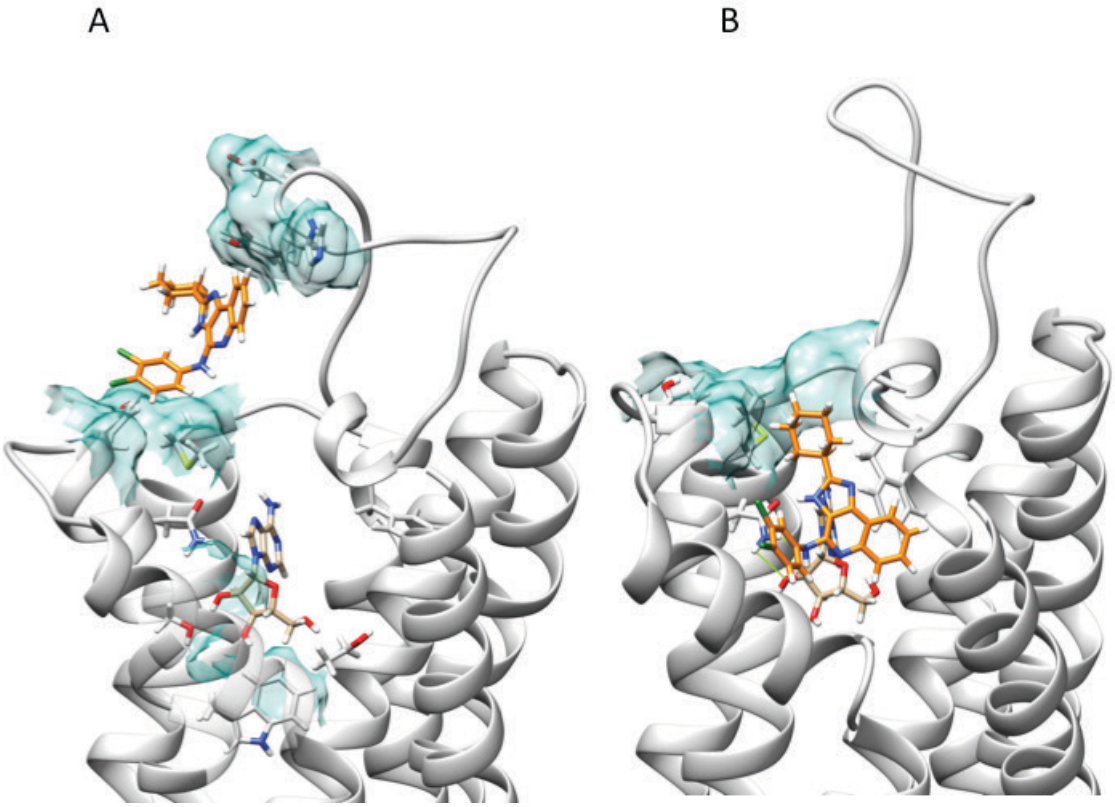

Figure 5. (A) LUF6000 binding mode in the hA_3 AR meta-binding site.(B) LUF6000 binding mode in the hA_3 AR orthosteric binding site occupied by adenosine. LUF6000 and adenosine are displayed as orange and tan stick, respectively. Side chains of residues interacting through hydrogen bond or π-π stacking are depicted as grey stick, whereas side chains of residues interacting through hydrophobic contacts are rendered as coloured surfaces.

The mutagenesis data available to date[12] apparently confute the first hypothesis, as it has been reported that the mutation of some residues located in the upper region of the receptor does not affect the allosteric activity of the imidazoquinoline compound DU124183 and the pyridinylisoquinoline compound VUF5455[10]. However, it is well accepted that a PAM activity is strictly depending on the structure of the agonist considered to perform the experiments.

Experimental Section

General

All computations were performed on a hybrid CPU/GPU cluster. Molecular dynamics simulations have been performed with GPU cluster equipped with 3 NVIDIA GTX 780 and 3 NVIDIA GTX 980.

Trajectory analysis, Figures and videos generation have been performed using several functionalities implemented by Visual Molecular Dynamics[13], WORDOM[14], the PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC (http://www.pymol.org/) and the Gnuplot graphic utility (http://www.gnuplot.info/). Ligand-hA_3 AR interaction energies were calculated extrapolating the non-bonded energy interaction term of CHARMM27 Force Field[15] using NAMD[16]. All molecular dynamics simulations have been carried out using ACEMD engine (http://www.acellera.com/).

The numbering of the amino acids follows the arbitrary scheme proposed by Ballesteros and Weinstein[17]: each amino acid identifier starts with the helix number (1-7), followed by a dot and the position relative to a reference residue among the most conserved amino acids in that helix, to which the number 50 is arbitrarily assigned.

Homology Model of hA_3 AR

As, to date, no crystallographic information about the hA_3 AR is available, we used a previously build homology model deposited in our web platform

dedicated to ARs, Adenosiland[18,19]. In particular, among all the currently available crystallographic structures of the hA_{2A} AR we selected the model built upon the complex with the endogenous agonist adenosine (PDB code: 2YDO, 3.00 Å resolution)[10].

Receptor membrane embedding and system preparation.

Receptors were embedded in a 1-palmitoyl-2-oleoyl-sn- glycero-3-phosphocholine (POPC) lipid bilayer (85x95 Å wide) and placed into the membrane according to the suggested orientation reported in the "Orientations of Proteins in Membranes (OPM)" database[20] for the hA_{2A} AR in complex with the endogenous agonist adenosine (PDB ID: 2YDO)[10]. Overlapping lipids (within 0.6 Å) were removed upon insertion of the protein. The prepared systems were solvated with TIP3P water[21] using the program Solvate 1.0[22] and neutralized by Na^+/Cl^- counter-ions to a final concentration of 0.154 M. The total number of atoms per system was approximately 110000. Membrane MD simulations were carried out on a GPU cluster with the ACEMD program using the CHARMM27 Force Field[15] and periodic boundaries conditions. Initial parameters for the ligands were derived from the CHARMM General Force Field for organic molecules. The system was equilibrated using a stepwise procedure. In the first stage, to reduce steric clashes due to the manual setting up of the membrane-receptor system, a 2500 steps conjugate-gradient minimization was performed. Then, to allow lipids to reach equilibrium, water molecules to diffuse into the protein cavity and to avoid ligand-receptor interaction in the equilibration phase, protein and ligand atoms were restrained for the first 8 ns by a force constant of 1 kcal/mol•Å^2. Then, force constant was gradually reduced to 0.1 kcal/mol•Å^2 for the next 9 ns. Temperature was maintained at 298 K using a Langevin thermostat with a low damping constant of 1 ps^{-1} and the pressure was maintained at 1 atm using a Berendensen barostat. Bond lengths involving hydrogen atoms were constrained using the M-SHAKE algorithm[23] with an integration timestep of 2 fs.

SuMD simulations

After the equilibration procedure, harmonical constraints were removed and SuMD simulations were conducted in a NVT ensemble. As previously described, the supervision of the trajectory is perpetuated until ligand-receptor distance is lower than 5 Å without introducing bias to the simulations. Long-range Coulomb interactions were handled using the particle mesh Ewald summation method (PME)[24] with grid size rounded to the approximate integer value of cell wall dimensions. A non-bonded cut-off distance of 9 Å with a switching distance of 7.5 Å was used. Ligand parametrization procedure and methodological insights on the quantitative estimate of the electrostatic and hydrophobic occurring ligand-protein interaction maps have been reported previously[5].

Abbreviations

ARs adenosine receptors

ATP adenosine triphosphate

hA3AR human A3 adenosine receptor

EL2 second extracellular loop

EL3 third extracellular loop

LUF6000 *N*-(3,4-dichlorophenyl)-2-cyclohexyl-1*H*-imidazo[4,5-*c*]quinolin-4-amine

LUF6096 *N*-{2-[(3,4-dichlorophenyl)amino]quinolin-4-yl}cyclohexanecarboxamide

GPCRs G protein-coupled receptors

GPU graphics processing unit

PAM positive allosteric modulator

RMSD root mean square deviation

SAR structure-affinity relationship;

TM transmembrane; ZM 241385, 4-[2-[7-amino-2-(2-furyl)-1,2,4-triazolo[1,5-*a*][1,3,5]triazin-5-yl-amino]ethylphenol.

Bibliography

1. Gao, Z.-G. & Jacobson, K. A. Keynote review: allosterism in membrane receptors. *Drug Discov. Today* **11,** 191–202 (2006).

2. Lewis, J. A., Lebois, E. P. & Lindsley, C. W. Allosteric modulation of kinases and GPCRs: design principles and structural diversity. *Curr. Opin. Chem. Biol.* **12,** 269–280 (2008).

3. Keov, P., Sexton, P. M. & Christopoulos, A. Allosteric modulation of G protein-coupled receptors: a pharmacological perspective. *Neuropharmacology* **60,** 24–35 (2011).

4. Kruse, A. C. *et al.* Activation and allosteric modulation of a muscarinic acetylcholine receptor. *Nature* **504,** 101–106 (2013).

5. 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).

6. Göblyös, A. & Ijzerman, A. P. Allosteric modulation of adenosine receptors. *Biochim. Biophys. Acta* **1808,** 1309–1318 (2011).

7. Kim, Y., de Castro, S., Gao, Z.-G., Ijzerman, A. P. & Jacobson, K. A. Novel 2- and 4-substituted 1H-imidazo[4,5-c]quinolin-4-amine derivatives as allosteric modulators of the A3 adenosine receptor. *J. Med. Chem.* **52,** 2098–2108 (2009).

8. Gao, Z.-G., Ye, K., Göblyös, A., Ijzerman, A. P. & Jacobson, K. A. Flexible modulation of agonist efficacy at the human A3 adenosine receptor by the imidazoquinoline allosteric enhancer LUF6000. *BMC Pharmacol.* **8,** 20 (2008).

9. Cohen, S. *et al.* A3 Adenosine Receptor Allosteric Modulator Induces an Anti-Inflammatory Effect: In Vivo Studies and Molecular Mechanism of Action. *Mediators Inflamm.* **2014,** (2014).

10. Lebon, G. *et al.* Agonist-bound adenosine A2A receptor structures reveal common features of GPCR activation. *Nature* **474,** 521–525 (2011).

11. Sabbadin, D., Ciancetta, A., Deganutti, G., Cuzzolin, A. & Moro, S. Exploring the recognition pathway at the human A2A adenosine receptor of the endogenous agonist adenosine using supervised molecular dynamics simulations. *MedChemComm* **6,** 1081–1085 (2015).

12. Gao, Z.-G. *et al.* Identification of essential residues involved in the allosteric modulation of the human A(3) adenosine receptor. *Mol. Pharmacol.* **63,** 1021–1031 (2003).

13. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **14,** 33–38, 27–28 (1996).

14. Seeber, M. *et al.* Wordom: A user-friendly program for the analysis of molecular structures, trajectories, and free energy surfaces. *J. Comput. Chem.* **32,** 1183–1194 (2011).

15. MacKerell, A. D., Banavali, N. & Foloppe, N. Development and current status of the CHARMM force field for nucleic acids. *Biopolymers* **56,** 257–265 (2000).

16. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26,** 1781–1802 (2005).

17. Juan A. Ballesteros, H. W. Integrated Methods for the Construction of Three-Dimensional Models and Computational Probing of Structure-Function Relations in G-Protein-Coupled Receptors. *Methods Neurosci.* **25,** 366–428 (1995).

18. Floris, M., Sabbadin, D., Medda, R., Bulfone, A. & Moro, S. Adenosiland: walking through adenosine receptors landscape. *Eur. J. Med. Chem.* **58,** 248–257 (2012).

19. Floris, M. *et al.* Implementing the 'Best Template Searching' tool into Adenosiland platform. *Silico Pharmacol.* **1,** 25 (2013).

20. Lomize, M. A., Lomize, A. L., Pogozheva, I. D. & Mosberg, H. I. OPM: orientations of proteins in membranes database. *Bioinforma. Oxf. Engl.* **22,** 623–625 (2006).

21. 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).

22. Grubmuller, H. & Groll, V. Solvate. (1996). at <http://www.mpibpc.mpg.de/grubmueller/solvate>

23. 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).

24. Essmann, U. *et al.* A smooth particle mesh Ewald method. *J. Chem. Phys.* **103,** 8577–8593 (1995).

3.4 New Trends in Inspecting GPCR-ligand Recognition Process: the Contribution of the Molecular Modeling Section (MMS) at the University of Padova.

Antonella Ciancetta, Alberto Cuzzolin, Giuseppe Deganutti, Mattia Sturlese, Veronica Salmaso, Andrea Cristiani, Davide Sabbadin and Stefano Moro*[a]

Abstract

In this review, we present a survey of the recent advances carried out by our research groups in the field of ligand-GPCRs recognition process simulations recently implemented at the Molecular Modeling Section (MMS) of the University of Padova. We briefly describe a platform of tools we have tuned to aid the identification of novel GPCRs binders and the better understanding of their binding mechanisms, based on two extensively used computational techniques such as molecular docking and MD simulations. The developed methodologies encompass: (i) the selection of suitable protocols for docking studies, (ii) the exploration of the dynamical evolution of ligand-protein interaction networks, (iii) the detailed investigation of the role of water molecules upon ligand binding, and (iv) a glance at the way the ligand might go through prior reaching the binding site.

Introduction

Today, it is largely recognized that G protein-coupled receptors (GPCRs) represent the largest family of surface receptors with more than 800 members in humans.[1] They respond to different extracellular stimuli ranging from small molecules to lipids, peptides, proteins, and even light.[2]

The binding event triggers the activation of cytoplasmic heterotrimeric GTP binding proteins (G proteins) and mediates the signal transduction through the modulation of several downstream effectors. The participation of GPCRs in numerous physio-pathological processes entails a potential role for their modulation by agonist, antagonists and inverse agonists in the treatment of

Published - Ciancetta A, Cuzzolin A, Deganutti G, Sturlese M, Salmaso V, Cristiani A, Sabbadin D, Moro S., *Mol Inform*, 35, 440-448 (2016)

several diseases, including cardiovascular and mental disorders,[3] cancer,[4] and viral infections.[5] Nowadays, about more than 50% of the drugs in clinical use targets a GPCR.[6]

According to the GRAFS classification,[7] human GPCRs are commonly grouped into five main classes: Glutamate (Class C), Rhodopsin (Class A), Adhesion (Class B), Secretin (Class B), and Frizzled/Taste2 (Class F). From a structural point of view, all members share a common architecture represented by seven membrane-spanning helices connected by three intracellular and three extracellular loops with the N-Term domain exposed toward the extracellular side.

The insertion into the cell membrane along with receptors dynamism have hampered for long time the structural determination of GPCRs by X-ray crystallography. To overcome these limitations, several techniques have been developed: the use of fusion proteins such as T4 lysozyme or apocytochrome,[8,9] complexation with antibody fragments,[10] and the receptor thermostabilization through systematic scanning mutagenesis.[11] The advances in protein engineering and crystallography have represented a breakthrough for the research focused on GPCRs and yielded numerous X-ray structures.[12] The availability of ligand-bound three-dimensional structures provides invaluable insights to understand GPCRs function and pharmacology and enables the application of structure-based drug design approaches to aid the discovery of novel candidates with improved pharmacological profiles.[13] In particular, molecular dynamics (MD) simulations have become a helpful complement for the study of GPCRs biophysics and molecular pharmacology, by enriching our understanding of, among other aspects, ligand-receptor interaction and ligand-subtype selectivity.[14,15] In addition, the recent exploitation of the commodity, graphics processing units (GPUs), a technology firstly designed to improve video game performances, in the molecular modeling field represents an important step forward for the simulation of GPCRs in explicit lipid-water environments within a reasonable computation time.[16] In this paper, we briefly survey the recent advances carried out by our research groups in the field of ligand-GPCRs

recognition process simulations.[17] Following the description of the tools we have developed to aid the identification of novel binders of GPCRs binders and the better understanding of their binding mechanisms, we will discuss their use in a case study: the comparison between ZM 241385 (4-(2-(7-Amino-2-(furan-2-yl)-[1,2,4]triazolo[2,3-a][1,5-a][1,3,5]triazin-5-yl-amino)ethyl) phenol) and caffeine, a strong and weak human Adenosine 2A Receptor (hA2A AR) antagonists, respectively.

Methods

Docking Protocols Validation: the "Quality Descriptors"

The availability of ligand-bound crystal structures enables to perform docking simulations to rationalize structure-activity relationships of known binders or to conduct virtual screening campaigns to identify novel candidates. It is highly recommended to assess the performances of a docking protocol in reproducing the available experimental data prior to applying it. This procedure is best known as benchmark study. We have recently developed a pipeline that allows a fast graphical evaluation of different docking protocols, based on two newly defined quality descriptors: the "Protocol Score" and the "Interaction Energy Map" (IEM).[18,19]

The "Protocol Score" is a RMSD based descriptor that assigns a 0–3 score to each docking protocol according to the following criteria: (i) if the protocol returns either a RMSDave value lower than the crystal structure resolution (R) or generates at least 10 (out of 20) conformations having RMSD<R, a score 1 is assigned; (ii) if a protocol satisfies both the above mentioned requirements, a score 2 is assigned; (iii) if a protocol satisfies none of the above mentioned requirements, a score 0 is assigned. Moreover, a score 3 is conferred to the best protocols, i.e. those returning at the same time the lowest RMSDave value and the highest number of conformers with a RMSD<R. The scores are then converted in a color code and the data visualized as a colored map (Figure 1A): protocols corresponding to white and light green spots are not

suitable for the system under consideration, dark green spots highlight good protocols, whereas blue spots identify the best among the tested ones.

The IEMs are based on the analysis of ligand-protein interactions and are derived as follows. Firstly, per residue electrostatic and hydrophobic contributions to the interaction energy (denoted IEele and IEhyd, respectively) are computed for residues surrounding the binding site or known to play a role in the binding. The analysis is performed for both the crystallographic binding modes and the docking poses. These pieces of information are then graphically transferred into heat-like maps reporting the key residues involved in the binding with the considered ligands along with a color code reflecting the quantitative estimate of the occurring interactions (the more intense the color, the stronger the interaction). The comparison is therefore based on the quality of the interactions in terms of number of established interactions and their relative strength among the X-Ray binding mode and the generated docking poses (Figure 1B).

The main advantage of the proposed pipeline resides in the full automation of the benchmark procedure: the user is provided with pre-compiled input files for several docking programs, thus minimizing the required expertise to carry out the benchmark study. To this aim, the results are presented as easy to interpret colored maps enabling a fast graphical inspection of large amount of data. The results are analyzed on the basis of the above described quality descriptors.

Binding Modes Inspection: the Dynamic Scoring Function (DSF)

The docking approach suffers from several limitations.[20] Although is a valuable method to get insights on the final stage of ligand-protein recognition, it lacks the description of two fundamental aspects that might play a significant role in ligand binding: water molecules mediated interactions and protein flexibility. To complete the description provided by the docking method with such contributions, we have recently developed the "dynamic scoring function"(DSF), an approach that enables to follow the dynamical evolution of a

docking pose in a realistic environment, i.e.the solvated membrane embedded ligand-protein complex.[21] The DSF provides a dynamic estimate of both the ligand position and the strength of the interaction network while accounting for the interplay of water molecules and protein side-chains flexibility. The procedure envisages the dynamic selection of residues within a range of 4.5 Å from the ligand during the MD simulation, starting from a previously obtained docking pose. The DSF is the cumulative sum of electrostatic and hydrophobic contributions to ligand-protein interaction (DSF_{ele} and DSF_{hyd}, respectively) and is calculated at frames extracted every 100 ps as follows:

$$DSF_{ele} = \sum_{t=0}^{n} IE_{ele} \quad (1)$$

$$DSF_{hyd} = \sum_{t=0}^{n} IE_{hyd} \quad (2)$$

The DSF value corrected for the ligand fluctuation (RMSD) with respect to the starting position yields the weighted DSF (wDSF), a number that highlights differences between stable and unstable poses. The corresponding weighted electrostatic and hydrophobic DSFs (denoted as wDSFele and DSFhyd, respectively) are therefore obtained as reported below:

$$wDSF_{ele} = \frac{\sum_{t=0}^{n} IE_{ele}}{RMSD} \quad (3)$$

$$wDSF_{hyd} = \frac{\sum_{t=0}^{n} IE_{hyd}}{RMSD} \quad (4)$$

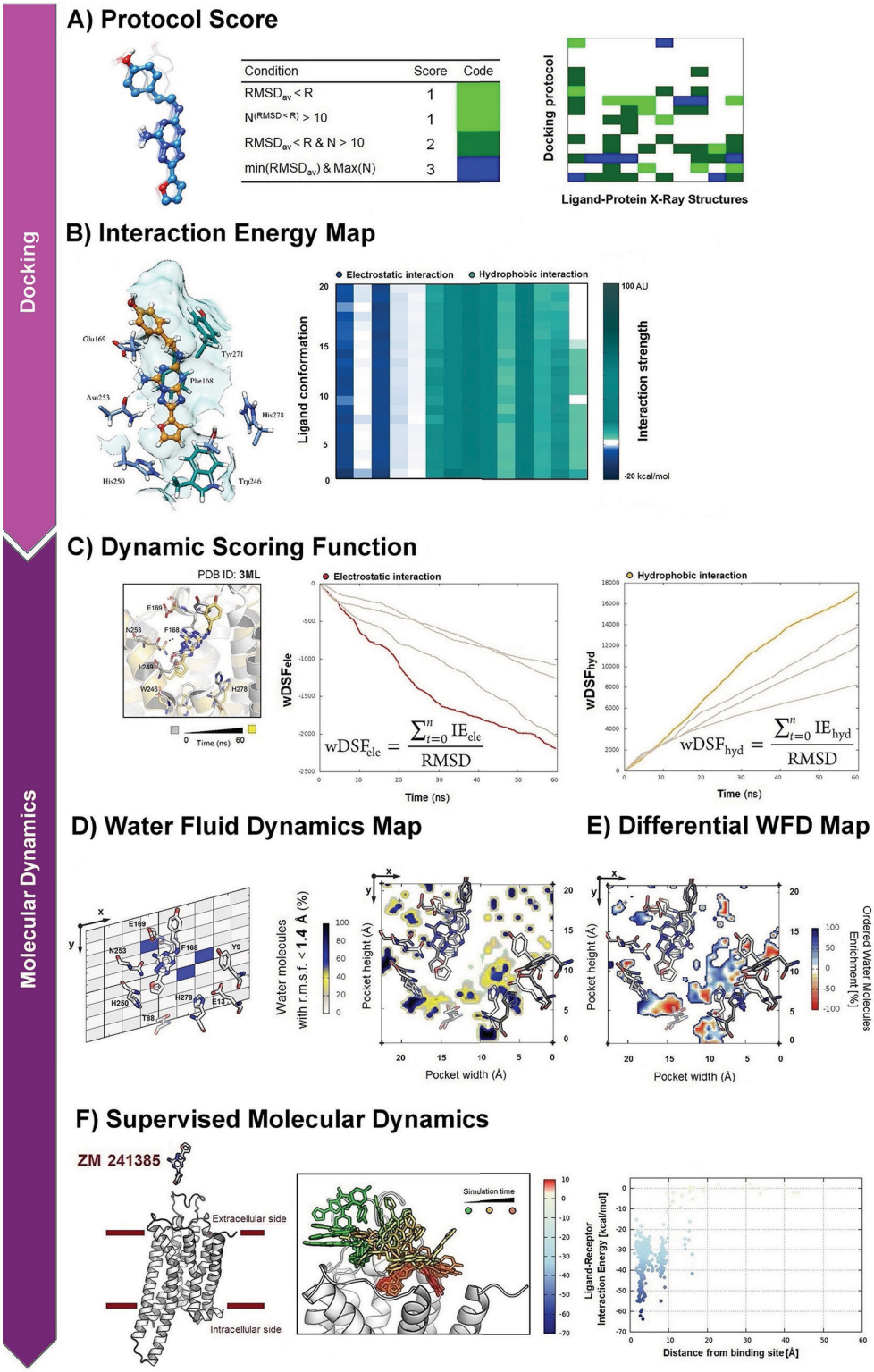

Figure 1. Schematic representation of the developed tools: A) Protocol Score; B) Interaction Energy Maps (IEMs); C) RMSD weighted Dynamic Scoring Function (wDSF); D) Water Fluid Dynamics (WFD) maps; E) Differential WFD maps; F) Ligand-receptor interaction energy landscape from supervised MD (SuMD) simulations. The figures were adapted from the original papers.[18,21,27,35]

The DSFs can be computed during the MD simulations or performed as a post-processing procedure, so that in principle any trajectory that has been previously produced can be re-analyzed with this approach. It can be regarded as an alternative to conventional scoring functions, as it is able to take into account both the complex flexibility in the membrane environment as well as water-driven interactions. The resulting graphs (Figure 1C) obtained by plotting the DSFs against the simulation time enable a graphical comparison of the relative stability of docking poses. This representation can help in detecting and validating the feasibility of alternative binding conformations proposed by the docking algorithm. We have recently exploited this feature to support an apparently less plausible binding mode of a series of 5-alkylaminopyrazolo[4,3-e]1,2,4-triazolo[1,5-c]pyrimidine at the hA3 AR.[22] Moreover, we tested the applicability of our approach by taking part in the community- wide 2013 GPCR Dock Assessment.[23] Among the proposed targets, we focused on the 5HT2B/ergotamine complex, whose X-Ray structure has been released after the predictions were submitted. Therefore we tested the applicability of our tool to homology models of a system on which our laboratory did not hold expertise. We submitted several alternative ligand-protein complexes suggested by the docking protocol to membrane MD simulations and selected the best final poses according to the outcomes of the DSFs analysis. Our predictions ranked 8th among 254, suggesting the portability of our approach to homology models as well as to other GPCRs.[23]

A Closer Look at Water Molecules: Water Fluid Dynamics (WFD) Maps

It is generally recognized that water molecules contribute to protein-ligand binding in at least two ways: they either stabilize the complex by forming hydrogen bond networks,[24] or are replaced by the ligand once the complex is formed.[25,26] It is therefore crucial in a drug design process to be able to distinguish between water molecules that mediate protein-ligand interactions and those that can be targeted for being displaced. To this aim, we have very recently tuned a tool that inspects the time-dependent variation of fluid dynamics

properties of water molecules as a consequence of the binding event by means of MD simulations.[27] Our approach detects structural water molecules inside the orthosteric binding site of the receptor and collects these pieces of information in a bi-dimensional graph, that we called water fluid dynamics (WFD) map. Unlike other existing MD based methodologies,[28,29] our approach is aimed at localizing protein "hot-spots" – i.e. regions where water molecules playing a key role in ligand binding mostly reside – rather than estimating their binding affinity. The WFD maps have been therefore mainly conceived as qualitative tool to drive ligand design to avoid substituents disrupting key water molecules' networks.

The WFD maps are derived as follows: residues within a range of 5 Å from the ligand are selected and a box surrounding the binding site is created and split into a threedimensional grid. During the MD simulations the diffusion of water molecules in each grid cell is followed. The data are acquired by saving the MD trajectories at regular intervals (every 10 ps) and by projecting the averaged position of water molecules showing a RMSF value below 1.4 Å into a bi-dimensional grid. The overlap of these grids yields a map (Figure 1D) with cell colored according to the residence time of water molecules on a 0–100% scale. White zones (0%) are occupied by water molecules with a residence time equivalent to bulk, whereas blue regions (100%) are occupied by trapped water molecules showing the maximum residence time of the considered trajectory. The maps allow a fast graphical identification of water distribution inside the orthosteric binding pocket. Moreover, "differential" WFD maps representing by a color code the enrichment or displacement of water molecules as a consequence of ligand binding (Figure 1E) are derived by comparing the WFD maps of the receptor in the apo and bound states.

Exploring the Ligand-Receptor Recognition Process: the Supervised Molecular Dynamics (SuMD) Approach

One of the most challenging tasks for ligand-GPCRs modeling is the prediction of the recognition pathway, an event which knowledge would ease the

development of drug candidates with better pharmacodynamic profiles. Unfortunately, the recognition of a ligand by a receptor is a process hard to simulate as it requires classical MD experiments in a long microsecond time scale.[14,30,31]. To overcome this technical limitation, enhanced sampling methods that facilitate the crossing of energy barriers through the introduction of biased potentials have been developed.[23,33]. Another approach,[34] induces ligand unbinding by applying external forces to the system, thus requiring knowledge of the ligand-receptor complex final state. Within this framework, we have recently proposed an alternative strategy – the "supervised molecular dynamics" (SuMD)[35] – that enables to follow the ligand-GPCR approaching path by considerably reducing the simulation time scale and without introducing bias. SuMD performs standard simulations in which the distance between the center of masses of the ligand atoms and the receptor binding site is monitored by a tabu-like algorithm. If the location of the binding site is unknown, several simulations are run by setting the centers of previously detected cavities. An arbitrary number of distance points is collected "on the flight" and fitted into a linear function f(x)=mx. The tabu-like algorithm is applied to increase the probability to produce ligand-receptor binding events as follows: If the slope (m) is negative, the ligand-receptor distance is likely to be shortened and a classic MD simulation is restarted from the last set of coordinates. Otherwise, the simulation is restored from the original set of coordinates and random velocities are reassigned to each atom. The supervision is repeated until the ligand-receptor distance is less than 5 Å.

The results of a SuMD simulation are displayed in a graph reporting the interaction energy toward the distance between the ligand and the binding site (Figure 1F). This approach can be exploited to analyze binding events to both orthosteric and allosteric sites and to assist the design of site-directed mutagenesis experiments in order to infer the role of specific residues on the molecular recognition process.

Application to Drug Design

To explain the applicability of the described tools, we discuss here as case study the comparison between ZM 241385 and caffeine, a strong and a weak hA2A AR binder with pKD values 9.18.0.23 and 5.31.0.44, respectively.[36] Among the hA2A AR available crystal structures we have selected the two co-crystallized with the ligands of interest identified by the following PDB IDs: 3EML and 3RFM.[36,37]. The starting point of the study is the evaluation of the reproducibility of the X-Ray binding modes through docking calculations. To accomplish this task we compare the IEMs computed for the best performing docking protocols. We then proceed by evaluating the dynamic evolution of alternative binding modes proposed by the docking algorithm, thus imaging the common case where X-Ray structures are not available for comparison. As anticipated, MD simulations allow taking into account the flexibility of the receptor and the role of water molecules in the binding. A more careful inspection of water dynamics is then performed by deriving differential WFD maps from the computed trajectories of a selected docking pose for each structure. Finally, we move outside the receptor and try to reproduce the binding pathways from the extracellular side through SuMD experiments.

Assessing the Reproducibility of a Binding Mode: IEMs Comparison

We start our case study by assessing the performance of a previously selected docking algorithm through IEMs inspection.[18] Figure 2 displays the comparison between the computed IEMs for the two considered structures. As shown, the binding mode of ZM 241385 (3EML, Figure 2A) encompasses a tight interaction network that is correctly reproduced by the majority of the 20 generated poses. On the other hand, the caffeine binding mode (3RFM, Figure2B) is more challenging to be reproduced and implies a lower number of less intense interactions with the binding site residues. The comparison of IEMs therefore helps evaluating in a fast graphical fashion both the docking protocol performances and the reproducibility of X-Ray observed binding modes. Interaction patterns without interruptions are clue of binding modes easy to

reproduce and indicate good protocol performances, whereas discontinuous patterns suggest binding modes challenging to be predicted and unsatisfactory protocol performances.

Following the Dynamics of Ligand-receptor Interactions:wDSFs Profiles

When the X-ray structures of the ligand-protein complex of interest are not available, usually a modeler is asked to select among several feasible binding modes suggested by the docking protocol slightly differing for the assigned scores. How to recognize the solution best approaching the "real" binding mode? Figure 3 displays the exercise we have conducted to address this issue: in order to identify as many different as possible binding modes, we forced the docking protocol to return ten poses that differed in terms of RMSD for at least 1.75 Å.[27] Nevertheless, the protocol assigned to the generated conformations scores differing at most for ten units. We subjected each docking pose to MD simulation and evaluated the wDSFs. Figure 3A–B displays the results for the two considered structures: the different values of both the cumulative electrostatic and the hydrophobic contributions reflect the different affinities of the two binders. In particular, ZM 241385 exhibits higher absolute values for both contribution types consistently with its higher affinity for the receptor. Moreover, for both structures, the wDSFs trends enable to graphically recognize the pose that best reproduces the X-ray observed binding mode, i.e. the one showing the slope with the highest absolute value.

What About the Role of Water Molecules? WFD Maps Inspection

A detailed inspection of the WFD maps of the two considered compounds further contributes to explain their different binding affinities to the hA2A AR. The WFD map for the ZM 241385 complex (Figure 4A) highlights the presence of water molecules bridging the aromatic scaffold to key residues in the binding site, namely Tyr9, Glu13, His278, Asn253, and Glu169.[27] The interactions with some of those residues were already detected from the docking pose, whereas other ones arose from MD simulations. The differential WFD

map (Figure 4B) highlights that the ligand displaces water molecules close to Thr88 while binding. The WFD map corresponding to the caffeine complex (Figure 4C), instead, shows a high propensity of bulk water molecules to solvate the fragment-like compound.[27] This is a direct consequence of the lack of strong interactions with the residues of the binding site detected during the MD simulation and aid explaining the lower affinity of the compound.

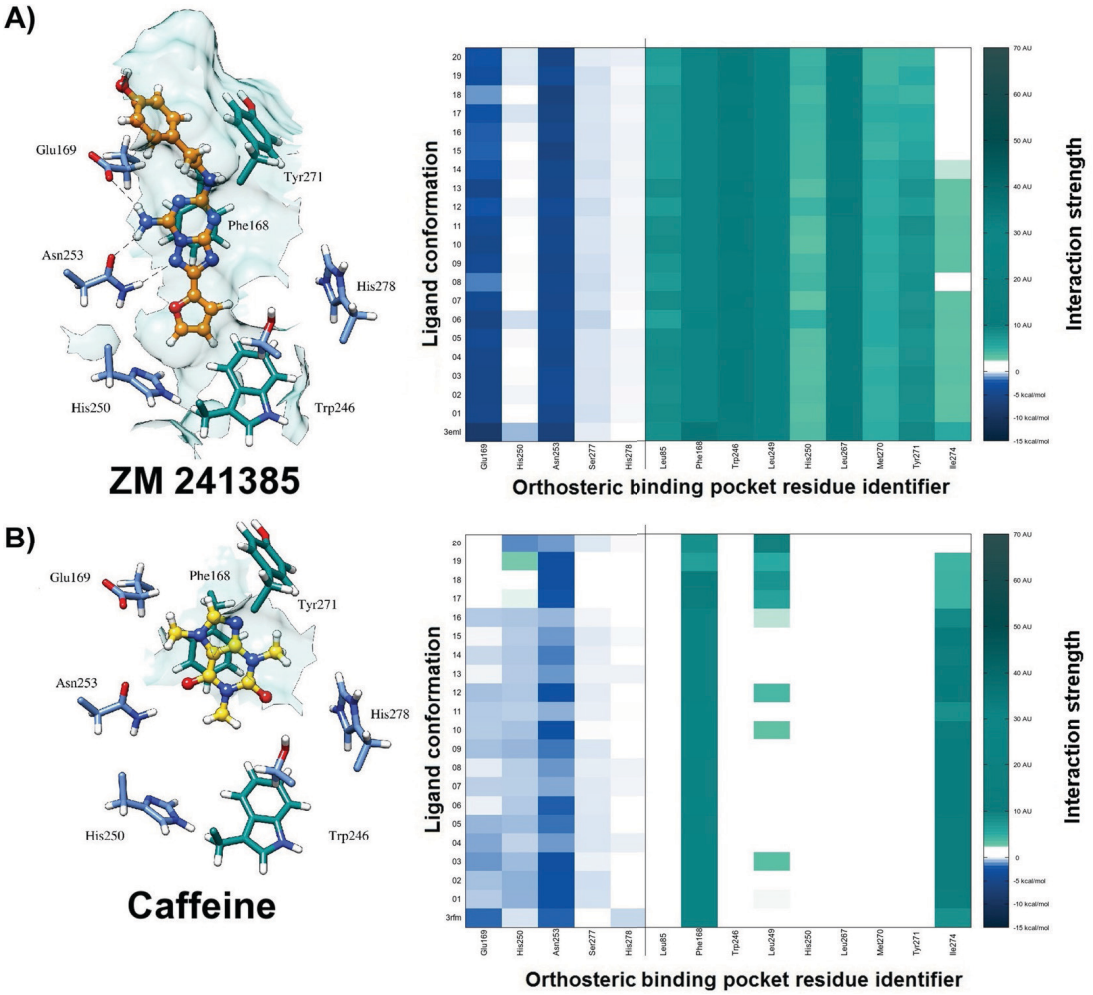

Figure 2. Comparison of IEMs for two hA2A AR ligands: ZM 241385 (A) and caffeine (B). While ZM 241385 establishes a strong interaction network conserved among the 20 generated poses, caffeine finds lower number and less intense interactions. IEele values: kcal $Å^{-1}$ mol^{-1}, IEhyd values: arbitrary units. The figures were adapted from the original papers.[18]

On the Extracellular Side of hA2A AR: the SuMD Approach

On its way to the orthosteric binding site, the ligand might interact with the so-called meta-binding sites,[38] which in some cases, may coincide with possible allosteric sites. The SuMD path we have computed for ZM 241385 highlights two major interaction sites: the second and third extracellular loop (EL2 and EL3, respectively, Figure 5A).[35] As depicted in the diagram in Figure 5B, although a higher interaction (less favorable) energy is associated to these metabinding sites, they seem to play a role in tuning the correct orientation of the ligand scaffold while approaching the orthosteric site. The EL3 also takes part in the caffeine recognition pathway (Figure 5C),[35] which, however, lacks strong interactions with the orthosteric site (Figure 5D). The SuMD simulations thus recognize the critical role of the hA2A AR extracellular loops in the ligand recognition process, role that has been postulated in the past by using site-directed mutagenesis.[39,40] We have recently applied the SuMD approach to interpret the binding of two challenging ligands: (i) the natural agonist and a (ii) imidazoquinolinamine derivative acting as positive modulator (LUF6000). The binding of the natural agonist adenosine at the hA2 AR revealed a possible energetically stable meta-binding site.[41] The SuMD simulations suggested at least two possible mechanisms to explain the available experimental data for the positive allosteric modulation mediated by LUF6000 toward the hA3 AR.[42]

Summary and Outlook

Through this paper, we have surveyed the recent advances carried out by our research groups in modeling the ligand- GPCRs recognition process. The crystallographic revolution of the last decade, on one side, and the advent of graphics processing units (GPUs) in the molecular modeling field, on the other side, allowed us to tune several tools to assist the drug design procedure.

The proposed approaches enrich the pool of molecular modeling techniques currently available to disclose the factors influencing the ligand-GPCRs recognition process and exploit two computational methodologies extensively used by modelers such as molecular docking and membrane MD

simulations. The majority of the methods herein presented are conceived as post-processing procedures, so that in principle any docking output or MD trajectory previously obtained can be rapidly re-analyzed using these tools. Moreover, the full automation of the procedures as well as the presentation of the results as easy to interpret colored maps are aimed at broadening their applicability within the scientific community encouraging non-expert users to approach them. A different philosophy is instead at the basis of the SuMD approach, which introduces a supervision of the MD trajectory through a tabu-like algorithm to speed up the computation time required to inspect the ligand-GPCRs recognition event. The comparison between ZM 241385 and caffeine, a strong and a weak hA2A AR antagonists, has been presented as case study to explain the usefulness and potentiality of our approaches.

As a future perspective we foresee to extend and improve the applicability of these computational tools to address other fascinating open questions in GPCRs field. We would like to summarize some of the hottest topics in the area: a) clarify at the molecular level the orthosteric and the allosteric control mediated by different binders on GPCR functionality; b) elucidate the implication of phosphorylation and glycosylation in both ligand binding and receptor activation; c) understanding the physio-pathological meaning of monomer-oligomer (homo and/or hetero) receptor equilibrium; d) identification of novel second messengers involved in G protein-alternative signaling path- ways; e) explore the possibility to perform high-throughput SuMD (HTSuMD) simulations for virtual screening applications as well as for real-time interpretations of mutagenesis data.

Concluding, we hope that these computational approaches carefully integrated with all other experimental GPCRs competencies will broaden our perspectives in several scientific areas from molecular pharmacology to drug discovery.

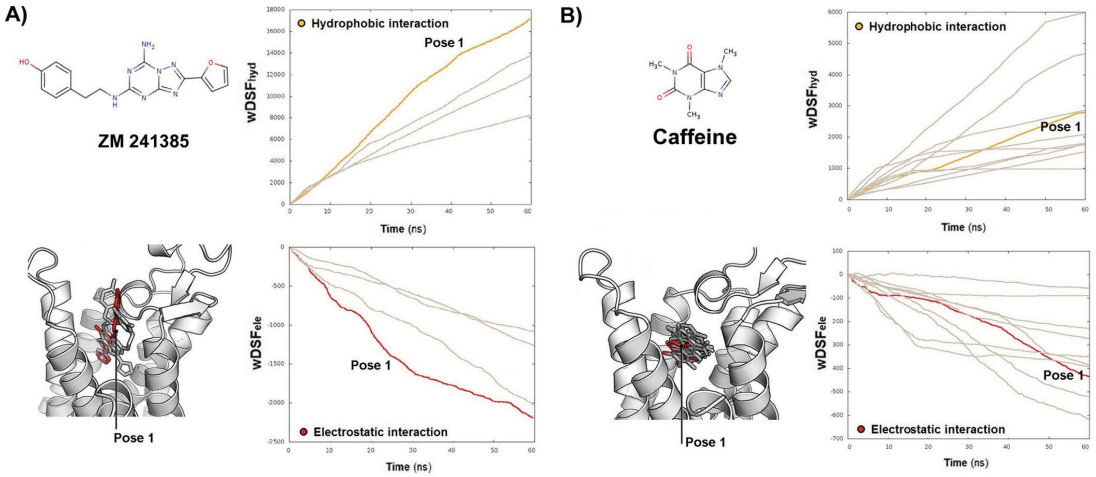

Figure 3. wDSFs comparison: A) ZM 241385 $wDSF_{hyd}$ (top) and $wDSF_{ele}$ (down); B) caffeine $wDSF_{hyd}$ (top) and $wDSF_{ele}$ (down). IEele values: kcal $Å^{-1}$ mol^{-1}. IEhyd values: arbitrary units. For both ligands the boundle of poses subjected to MD are rendered coloring pose number 1 in red. The same color scheme is used in the plots to identify pose 1 among the others. The figures were adapted from the original paper.[21]

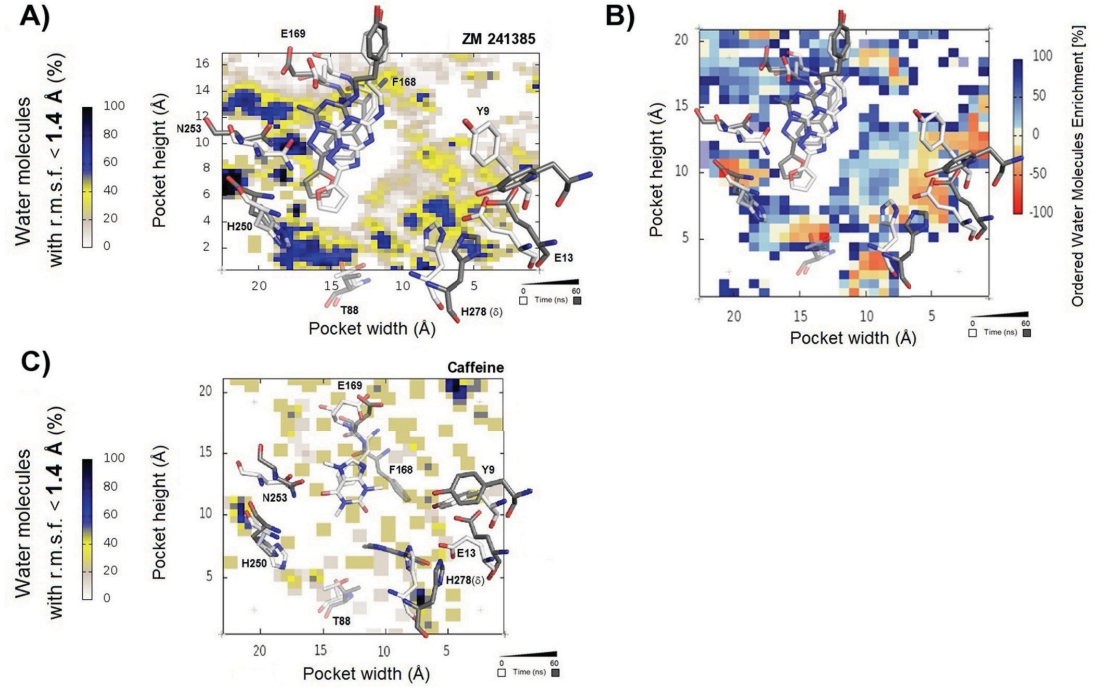

Figure 4. WFD maps comparison: A) position of water molecules experimentally determined for ZM 241385 complex structure; B) differential WFD for ZM 241385 in comparison to the apo-state of hA_{2A} AR; C) differential WFD for caffeine in comparison to the apo-state of hA_{2A} AR. Receptors are viewed from the membrane side facing TM6 and TM7. Side chains of key residues are displayed as gray sticks. Hydrogen atoms are not displayed. The figures were adapted from the original paper.[27]

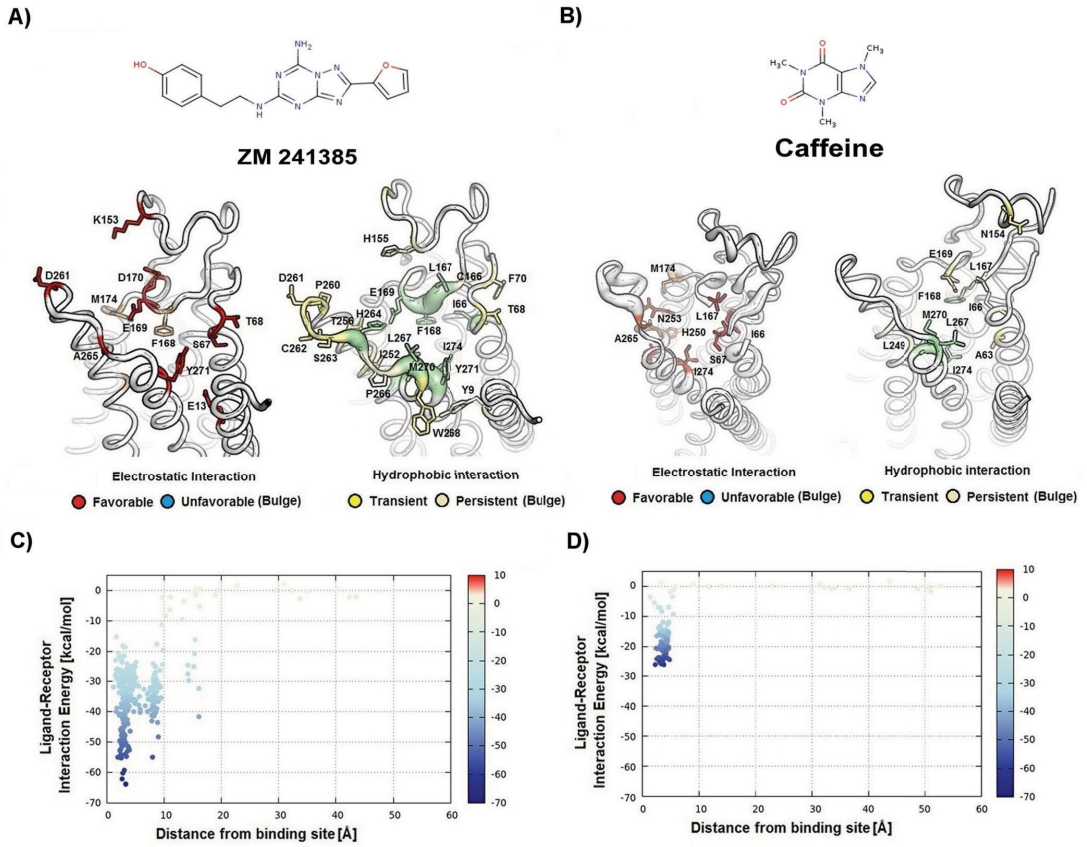

Figure 5. SuMD experiments on hA$_{2A}$ AR. Top: electrostatic and hydrophobic contributions to the interaction energy of each receptor residue involved in the binding with A) ZM 241385 and B) caffeine. Down: SuMD ligand-receptor interaction energy landscape for C) ZM 241385 and D) caffeine. Interaction Energy values: kcalmol^{-1}. The figures were adapted from the original paper.[35]

References

[1] K. L. Pierce, R. T. Premont, R. J. Lefkowitz, Nat. Rev. Mol. Cell Biol. 2002, 3, 639–650.

[2] K. Kristiansen, Pharmacol. Ther. 2004, 103, 21–80.

[3] J. L. Moreno, T. Holloway, J. GonzÇlez-Maeso, Prog. Mol. Biol. Transl. Sci. 2013, 117, 187–205.

[4] M. O'Hayre, M. S. Degese, J. S. Gutkind, Curr. Opin. Cell Biol. 2014, 27, 126–135.

[5] A. Sodhi, S. Montaner, J. S. Gutkind, Nat. Rev. Mol. Cell Biol. 2004, 5, 998–1012.

[6] K. Lundstrom, Curr. Protein Pept. Sci. 2006, 7, 465–470.

[7] R. Fredriksson, M. C. Lagerstrçm, L.-G. Lundin, H. B. Schiçth, Mol. Pharmacol. 2003, 63, 1256–1272.

[8] D. M. Rosenbaum, V. Cherezov, M. A. Hanson, S. G. F. Rasmussen, F. S. Thian, T. S. Kobilka, H.-J. Choi, X.-J. Yao, W. I. Weis, R. C. Stevens, et al., Science 2007, 318, 1266–1273.

[9] A. A. Thompson, W. Liu, E. Chun, V. Katritch, H. Wu, E. Vardy, X.-P. Huang, C. Trapella, R. Guerrini, G. Calo, et al., Nature 2012, 485, 395–399.

[10] T. Hino, T. Arakawa, H. Iwanari, T. Yurugi-Kobayashi, C. Ikeda- Suno, Y. Nakada-Nakura, O. Kusano-Arai, S. Weyand, T. Shimamura, N. Nomura, et al., Nature 2012, 482, 237–240.

[11] T. Warne, M. J. Serrano-Vega, J. G. Baker, R. Moukhametzianov, P. C. Edwards, R. Henderson, A. G. W. Leslie, C. G. Tate, G. F. X. Schertler, Nature 2008, 454, 486–491.

[12] R. C. Stevens, V. Cherezov, V. Katritch, R. Abagyan, P. Kuhn, H. Rosen, K. Wüthrich, Nat. Rev. Drug Discov. 2013, 12, 25–34.

[13] K. A. Jacobson, S. Costanzi, Mol. Pharmacol. 2012, 82, 361–371.

[14] R. O. Dror, A. C. Pan, D. H. Arlow, D. W. Borhani, P. Maragakis, Y. Shan, H. Xu, D. E. Shaw, Proc. Natl. Acad. Sci. U.S.A 2011, 108, 13118–13123.

[15] B. Selvam, J. Wereszczynski, I. G. Tikhonova, Chem. Biol. Drug Des. 2012, 80, 215–226.

[16] I. Buch, M. J. Harvey, T. Giorgino, D. P. Anderson, G. De Fabritiis, J. Chem. Inf. Model. 2010, 50, 397–403.

[17] A. Ciancetta, D. Sabbadin, S. Federico, P. Spalluto, S. Moro, in Trends Pharmacol Sci 2015 Press DOI 101016jtips201508006, n.d..

[18] A. Ciancetta, A. Cuzzolin, S. Moro, J. Chem. Inf. Model. 2014, 54, 2243–2254.

[19] A. Cuzzolin, M. Sturlese, I. Malvacio, A. Ciancetta, S. Moro, Molecules 2015, 20, 9977–9993.

[20] G. L. Warren, C. W. Andrews, A.-M. Capelli, B. Clarke, J. La- Londe, M. H. Lambert, M. Lindvall, N. Nevins, S. F. Semus, S. Senger, et al., J. Med. Chem. 2006, 49, 5912–5931.

[21] D. Sabbadin, A. Ciancetta, S. Moro, J. Chem. Inf. Model. 2014, 54, 169–183.

[22] S. Federico, A. Ciancetta, D. Sabbadin, S. Paoletta, G. Pastorin, B. Cacciari, K. N. Klotz, S. Moro, G. Spalluto, J. Med. Chem. 2012, 55, 9654–9668.

[23] I. Kufareva, V. Katritch, R. C. Stevens, R. Abagyan, Struct. Lond. Engl. 1993 2014, 22, 1120–1139.

[24] Y. Lu, R. Wang, C.-Y. Yang, S. Wang, J. Chem. Inf. Model. 2007, 47, 668 675.

[25] P. W. Snyder, J. Mecinovic, D. T. Moustakas, S. W. Thomas, M. Harder, E. T. Mack, M. R. Lockett, A. H|roux, W. Sherman, G. M. Whitesides, Proc. Natl. Acad. Sci. U.S.A 2011, 108, 17889 – 17894.

[26] G. V. De Lucca, P. K. Jadhav, R. E. Waltermire, B. J. Aungst, S. Erickson-Viitanen, P. Y. Lam, Pharm. Biotechnol. 1998, 11, 257– 284.

[27] D. Sabbadin, A. Ciancetta, S. Moro, J. Chem. Inf. Model. 2014, 54, 2846–2855.

[28] T. Young, R. Abel, B. Kim, B. J. Berne, R. A. Friesner, Proc. Natl. Acad. Sci. U.S.A 2007, 104, 808–813.

[29] J. Michel, J. Tirado-Rives, W. L. Jorgensen, J. Phys. Chem. B 2009, 113, 13337–13346.

[30] I. Buch, T. Giorgino, G. De Fabritiis, Proc. Natl. Acad. Sci. U.S.A 2011, 108, 10184–10189.

[31] R. O. Dror, H. F. Green, C. Valant, D. W. Borhani, J. R. Valcourt, A. C. Pan, D. H. Arlow, M. Canals, J. R. Lane, R. Rahmani, et al., Nature 2013, 503, 295–299.

[32] A. Laio, M. Parrinello, Proc. Natl. Acad. Sci. U.S.A 2002, 99, 12562–12566.

[33] D. Hamelberg, J. Mongan, J. A. McCammon, J. Chem. Phys. 2004, 120, 11919–11929.

[34] S. Park, K. Schulten, J. Chem. Phys. 2004, 120, 5946–5961.

[35] D. Sabbadin, S. Moro, J. Chem. Inf. Model. 2014, 54, 372–376.

[36] A. S. Dore, N. Robertson, J. C. Errey, I. Ng, K. Hollenstein, B. Tehan, E. Hurrell, K. Bennett, M. Congreve, F. Magnani, et al., Structure 2011, 19, 1283–1293.

[37] V.-P. Jaakola, M. T. Griffith, M. A. Hanson, V. Cherezov, E. Y. T. Chien, J. R. Lane, A. P. Ijzerman, R. C. Stevens, Science 2008, 322, 1211–1217.

[38] S. Moro, C. Hoffmann, K. A. Jacobson, Biochemistry (Mosc.) 1999, 38, 3498–3507.

[39] J. Kim, Q. Jiang, M. Glashofer, S. Yehle, J. Wess, K. A. Jacobson, Mol. Pharmacol. 1996, 49, 683–691.

[40] J. Kim, J. Wess, A. M. van Rhee, T. Schçneberg, K. A. Jacobson, J. Biol. Chem. 1995, 270, 13987–13997.

[41] D. Sabbadin, A. Ciancetta, G. Deganutti, A. Cuzzolin, S. Moro, MedChemComm 2015, 6, 1081–1085.

[42] G. Deganutti, A. Cuzzolin, A. Ciancetta, S. Moro, Bioorg. Med. Chem. 2015, 23, 4065–4071.

Published - Sabbadin D, Ciancetta A, Deganutti G, Cuzzolin A, Moro S. *MedChemComm* 6, 1081-1085 (2015)

4 CONCLUSION AND FURTURE PERSPECTIVES

The power of molecular dynamics simulation to tackle complicate SBDD aspects linked to the intrinsic dynamicity of macromolecular targets, ligands and environment at molecular level has been demonstrated in this Ph.D. thesis.

The SuMD technique showed a broad applicability being able to reproduce final bound configurations in good agreement with the experimental X-ray crystallographic complexes for both cytosolic and membrane proteins. First innovative aspect of this method is the time needed for simulate the stochastic event represented by the ligand - protein binding: it falls in the nanosecond timescale, while it usually needs microseconds of non-supervised molecular dynamics simulation. Moreover, SuMD has the value to allow simulate multiple binding events, therefore to sample different putative recognition pathway and multiple metastable binding sites that can act as modulators of the overall kinetics of the binding, whose quantitative estimation should be the next issue addressed by means of the method.

The importance of the metastable sites, encountered on its way to the orthosteric site of the A_{2A} AR, was highlighted for the endogenous effector adenosine. More precisely, various protein residues locate on the EL3 and EL2 participate to the ligand transition from the bulk solvent to the final bound state (in accordance with mutagenesis studies). An interesting topic outlined by the simulations is the adenosine difficulty to gain a stable orthosteric conformation, preferring to experience an ensemble of different orientations. Further investigations on this peculiar behaviour are ongoing, considering a recently disclosed X-ray crystal structure of the A_{2A} AR bound to an engineered G protein[1]: the aim is to determine if an almost completely active receptor conformation is able to better stabilize the agonist.

SuMD outcomes for the PAM LUF6000 on A_3 AR, to the better of our knowledge, for the first time addresses the allosterically-enhanced agonist affinity with the formation of a ternary intermolecular complex, that involves protein, agonist and allosteric modulator. This mechanism, if further confirmed, should

[1] Carpenter B, Nehmé R, Warne T, Leslie AG, Tate CG. Structure of the adenosine A2A receptor bound to an engineered G protein. *Nature*, **536**, 104–107**.**

be considered as an alternative way to mediate allosterism, beside the well established model that consider protein conformational changes occurring upon PAM binding.

Moving to the last paper presented, the molecular dynamics' utility in an integrated drug design strategy was described. Indeed, it is a useful tool for the quality evaluation of molecular docking results as well as a reliable method for discriminate strong binders from weak binders, as demonstrated by the example of caffeine and ZM241385 on the A_{2A} AR. In our opinion, high throughput SuMD simulations should become an important part of SBDD programs.

Finally, a SuMD-like approach combined to metadynamics is currently under development[2], with the aim of sampling unbinding events and to obtain useful insights also from ligand- receptor dissociation mechanisms.

[2] Giuseppe Deganutti, Andrei Zhukov, Stephanie Federico, Giampiero Spalluto, Stefano Moro, Jonathan S. Mason and Andrea Bortolato. Identification of Putative Adenosine A_{2A} Ligand Binding Kinetics Bottlenecks Explain Their Transition State Thermodynamics Properties. *Manuscript in preparation.*