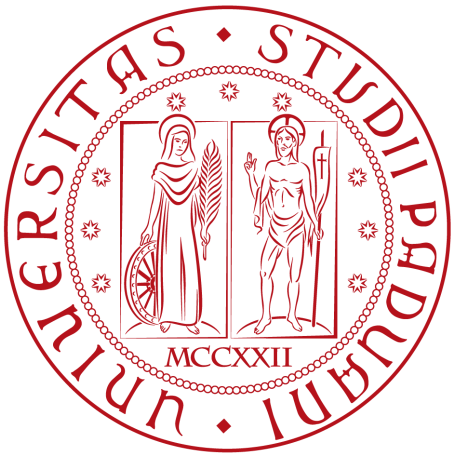


Sede Amministrativa: Università degli Studi di Padova
Dipartimento di Scienze Chimiche

SCUOLA DI DOTTORATO DI RICERCA IN SCIENZE MOLECOLARI
INDIRIZZO SCIENZE CHIMICHE
CICLO XXVIII

Structural and dynamic modeling of molecular systems at different length scales

Direttore della Scuola: ***Ch.mo Prof. Antonino Polimeno***
Coordinatore d'indirizzo: ***Ch.mo Prof. Antonino Polimeno***
Supervisore: ***Ch.mo Prof. Antonino Polimeno***

Dottorando: ***Mauro Torsello***

“Let us learn to dream, gentlemen,
then perhaps we shall find the truth”

—August Kekulé

Contents

Contents i

List of acronyms v

Sommario xi

Abstract xiii

1 Introduction 1

1.1 Multiscale modeling . . . 2

1.2 Investigated systems . . . 2

1.2.1 Dioxopeptides and thionated analogous . . . 3

1.2.2 Diphenyldiselenides and diphenylditellurides . . . 3

1.2.3 Connexin 26 hemichannel protein . . . 4

1.2.4 Reactions in microfluidic devices . . . 4

2 Software, hardware and methods 7

2.1 Employed software packages . . . 8

2.2 Computational facilities . . . 8

2.3 Quantum chemistry methods . . . 9

2.3.1 Hartree-Fock . . . 10

2.3.2 Density functional theory . . . 18

2.4 ONIOM multiscale method . . . 22

2.5 Molecular dynamics simulations . . . 23

2.6 Continuum mechanics simulations . . . 29

3 Molecular systems: a quantum chemistry approach 31

3.1 Energetics of dipeptides and thiodipeptides . . . 31

3.1.1 Experimental results . . . 33

3.1.2 Computational details . . . 34

3.1.3 Standard enthalpies . . . 35

3.1.4 Fragment analysis . . . 40

3.2 Flexibility and stability of secondary structure in oxo-/thio-dipeptides 42

4 Multi-layer quantum chemistry 49

4.1 Energetics and calcium binding of a portion of connexin 26 49

4.1.1 Fragments binding analysis 52

5 From quantum Hamiltonians to classic Hamiltonians 57

5.1 Molecular dynamics simulation of the connexin 26 57

5.1.1 Parametrization of the γ-carboxylated glutamate 59

5.1.2 Molecular dynamics simulations 60

5.1.3 Force field validation 61

5.1.4 Channel dimensions 61

5.1.5 N-terminal domain motion 62

5.2 General amber force field for diphenyldichalcogenides 63

5.2.1 Computational Details 67

5.2.2 Molecular structures 68

5.2.3 Quantum Potential energy surfaces 71

5.2.4 Molecular dynamics distributions 77

6 Continuum mechanics: microfluidic reactors 81

6.1 Fluid motion and transport equations 81

6.2 Numerical Methods 84

6.2.1 Navier-Stokes resolution 84

6.2.2 Advection-diffusion-reaction resolution 85

6.3 Genetic algorithm and optimization of the reaction process 99

6.3.1 Reaction 1 101

6.3.2 Reaction 2 103

7 Final remarks 105

A Electron repulsion integrals 107

B Energetics of oxo/thio-dipeptides: additional information 109

C Energetics and calcium binding of a portion of connexin 26: additional information 111

D General amber force field for diphenyldichalcogenides: additional information 113

E Advection-Diffusion-Reaction equation 119

F Stiff problems: Robertson reaction 121

List of publications i

Bibliography **iii**

Acknowledgements **xxix**

List of acronyms

AA Amino Acid. 40, 42, 43, 51, 52, 54

ACM Adiabatic Connection Method. 21

ADI Alternate Direction Implicit. 95–97

ADR Advection Diffusion Reaction. 4, 8, 82–84, 88, 95, 99, 101, 120

AMBER Assisted Model Building with Energy Refinement. 3, 8, 25, 26, 66–68, 73, 76

B3LYP Becke 3-parameter Lee Yang Parr. 10, 21, 22, 51, 67

BC Boundary Conditions. 82, 83

BDF Backward Differentiation Formula. 97, 98

BLYP Becke Lee Yang Parr. 21

BSSE Basis Set Superposition Error. 51, 53

CC Coupled Cluster. 8, 35, 36

CCFD Cell-Centered Finite Difference. 87, 88

CFD Computational Fluid Dynamics. 81, 99

CFL Courant-Friedrichs-Lewy. 90, 92, 95

CG Conjugate Gradient. 60, 61

CHARMM Chemistry At HARvard Molecular Mechanics. 25, 26

CI Configuration Interaction. 8

CSD Cambridge Structural Database. 68–70, 118

Cx Connexin. 49, 57–59

DFT Density Functional Theory. 7–10, 18, 20, 21, 28, 51, 52, 66, 107

DRK Dupuis Rys King. 107

ERI Electron Repulsion Integrals. 16, 17, 107

ESP ElectroStatic Potential. 60, 67, 72

FE Fully Extended. 44–46

FF Force Field. 25, 26, 59, 61, 69, 76

GA Genetic Algorithm. 5, 9, 99–104

GAFF General Amber Force Field. 3, 67, 68, 73, 75, 77, 78, 117

GGA Generalized Gradient Approximation. 21, 22

GPU Graphic Processing Unit. 7–9, 97

GPx Glutathione Peroxidase. 2, 64–66, 105

GROMACS GROningen MAchine for Chemical Simulations. 8, 61

GROMOS GROningen MOlecular Simulation. 25, 26

GTO Gaussian Type Orbital. 17, 107

HF Hartree Fock. 7–10, 12, 13, 15–17, 19–21, 28, 36, 39, 67, 107

HGP Head-Gordon Pople. 107

HPC High Performance Computing. 106

LDA Local Density Approximation. 20

LJ Lennard-Jones. 73, 75

LYP Lee Yang Parr. 21

MD Molecular Dynamics. 4, 7, 8, 26–29, 50, 51, 58, 66–68, 73, 77, 79

MGGA Meta Generalized Gradient Approximation. 21

MM Molecular Mechanics. 8, 22, 23, 25, 61, 66, 67, 77, 79

MP Møller Plesset. 8, 35, 36

MPI Message Passing Interface. 9, 99

NMM N-MethylMorpholine. 33

NMR Nuclear Magnetic Resonance. 33, 60

NS Navier Stokes. 4, 8, 81, 83, 84, 86, 95, 101

ODE Ordinary Differential Equation. 89, 97, 98, 121

ONIOM Our own N-layered Integrated molecular Orbital and molecular Mechanics. 23, 24, 49, 51–54, 57, 105

OS Obara Saika. 107

PBC Periodic Boundary Conditions. 28, 29

PBE Perdew Burke Ernzerhof. 22

PDE Partial Differential Equation. 81, 82, 84, 90, 95

PE Petroleum Ether. 33

PES Potential Energy Surface. 44, 60, 68, 71, 75, 76, 78

PH Pople Henre. 107

PM3 Parametric Model 3. 51, 54

PME Particle Mesh Ewald. 7, 8, 28, 29

PTM Post-Translational Modification. 58–60

QM Quantum Mechanics. 4, 8, 22, 23, 59, 61, 66, 67, 69, 72, 76–79

RESP Restrained ElectroStatic Potential. 60, 61, 67, 72, 73, 77

RHF Restricted Hartree Fock. 10, 16

RK Runge Kutta. 97, 98

RMSF Root Mean Square Fluctuation. 61

SCF Self Consistent Field. 9, 14

SIMPLE Semi-Implicit Method for Pressure Linked Equation. 84, 85

SPC Simple Point-Charge. 10

STO Slater Type Orbital. 16

TDMA Tridiagonal Matrix Algorithm. 95, 97

TGA ThermoGravimetric Analysis. 32, 34, 41–43

THF TetraHydroFuran. 33

TIP3P Transferable Intermolecular Potential 3-Point. 10, 60, 68, 77, 78, 117

UFF Universal Force Field. 73

UHF Unrestricted Hartree Fock. 15, 16

VMD Visual Molecular Dynamics. 8, 61

WT Wild Type. 61–63

Sommario

La continua crescita della potenza di calcolo, in termini di risorse hardware e software, ha reso l'approccio computazionale (*in-silico*) ai complessi problemi scientifici, uno strumento molto conveniente che permette di ottenere informazioni utili al fine di affiancare, interpretare ed, in alcuni casi, addirittura riprodurre i dati sperimentali a partire da principi primi. I metodi sono stati resi più veloci ed efficienti negli ultimi vent'anni, grazie anche allo sviluppo di algoritmi sempre più efficienti, in grado di sfruttare al meglio la potenza computazionale racchiusa nelle nuove soluzioni hardware (ad esempio architetture parallele basate sulle GPU), e di fornire pacchetti software di semplice utilizzo per molteplici applicazioni. Al giorno d'oggi l'approccio computazionale è impiegato in numerose aree scientifiche, che spaziano tra le più disparate discipline applicate come medicina, ingegneria, chimica, fisica, scienze dei materiali e molte altre.

In particolare in questo lavoro di tesi, alcuni degli approcci della chimica computazionale quali meccanica quantistica, dinamica molecolare classica e metodi ibridi, sono applicati allo studio di biomolecole e macromolecole, al fine di investigare differenti aspetti come struttura, dinamica, energetica e in particolare la *flessibilità*. In aggiunta ai metodi su menzionati è stato anche esplorato un approccio fluido-dinamico al fine di descrivere e simulare sistemi microfluidici, focalizzando l'attenzione sulla reattività dei sistemi presi in esame. Tutti questi approcci sono dipendenti dall'estensione del sistema e, poiché hanno un differente costo computazionale, la loro applicazione dovrebbe essere limitata ad una ragionevole dimensione dei sistemi studiati. Le profonde differenze in termini di costo/accuratezza sono discusse, fornendo un collegamento tra le *scale spaziali* delle diverse metodologie, al fine di esplicare come le informazioni ottenute a scale spaziali inferiori possano essere considerate come punto di partenza accurato per effettuare simulazioni a scale spaziali maggiori, in un approccio che è oggi comunemente noto come modellazione multiscala. La connessione tra i metodi ad alta accuratezza/alto costo e quelli a bassa accuratezza/basso costo è commentata, illustrando così come un approccio multiscala possa permettere, in casi specifici, di incrementare al contempo l'accuratezza del dato calcolato e la dimensione del sistema simulato.

Abstract

The continuous growth of computing power, both in terms of hardware and software resources, has made the computational (*in-silico*) approach to complex scientific problems a very profitable tool, which provides useful information to support, interpret or in some cases even reproduce the experimental datum from first principles. Methods have become cheaper and faster in the last two decades, thanks also to the development of more efficient algorithms, able to extract in full the computational power contained in novel hardware solutions (e.g. parallel computing and GPUs-based hardware), and to provide relatively easy-to-use software packages for diverse applications. Nowadays the computational approach is employed in several scientific areas, covering many different applied disciplines such as medicine, engineering, chemistry, physics, materials science and many others.

In particular in this thesis work, some of the main approaches of computational chemistry, namely quantum mechanics, classical molecular dynamics and hybrid methods, are applied to the study of biomolecules and macromolecules, in order to investigate different aspects like structure, dynamics, energetics and in particular *flexibility*. In addition to the aforementioned methods we also explore a fluido-dynamic approach to describe and simulate microfluidic systems, focusing the attention on the reactivity of the systems studied. All these approaches are size-dependent and because they have different computational costs, their application should be limited to a reasonable size of the studied system. The profound difference in terms of cost/accuracy are discussed, providing a link between the different methodologies *scales*, in order to exemplify how information gathered at smaller length scale can be considered as an accurate starting point to perform simulations at larger spatial scales, in what is nowadays know popularly as multiscale modeling. The connection between the high accuracy/high cost and low accuracy/low cost methods is commented upon, to illustrate how a multiscale modeling approach can allow, in specific cases, to augment at the same time the accuracy of the data calculated and the size of the system simulated.

Chapter 1

Introduction

Theories and experiments are tools that allow to inspect and describe important phenomena and shed light on the unknown aspects of the nature. Like a microscope of given magnifying power focuses best on objects near a given length-scale, so specific theories and experiments can concentrate on aspects of matter of a well given size, which allows to simplify our interpretation of some natural phenomena, but limits our capability of getting an overall picture of them. For example if we drop a solution containing nanosystems (e.g. macromolecules or nanoparticles) in water, we are dealing with a very complex system which contains many objects of different sizes. If the drop is colored by the presence of nanoparticles, we can observe the system from a "macroscopic" point of view, using our eyes that permit to see blur and color shades of the drop dissolving, proportionally to the concentration of nanoparticles. Instead, if we want to observe the "nanoscopic" nature of our system, using for instance transmission electron microscopy, we can see the nanosystems as beads dissolved in a continuous medium; finally if we suppose to be able to observe the "atomistic-molecular" nature of the drop, we can try to glimpse the detailed nature of this system, where water molecules vibrate, proteins are folded filaments whose shape change and nanoparticle are no longer spheres but complex molecular constructs with irregular surfaces, covered by fluctuating ligands. A similar train of thoughts can be followed using time instead of space as discriminating parameter. Usually processes that happen at smaller spatial scales are faster than those that take place at larger spatial scales.

This thesis work is dedicated to explore studied different chemical systems which cover a broad range of sizes, going from small peptides (characterizing structural and energetical properties) to proteins (studying the related structures, dynamics and binding energies to substrates) up to fluids and reacting solutions containing chemical substances (optimizing the product yield with respect to geometrical parameters). For each system a suitable computational method is chosen, depending on the spatial/temporal scales and also on the interesting aspects we want to investigate.

1.1 Multiscale modeling

Multiscale modeling is a computational approach that uses different models simultaneously, in order to solve complex problems that cover several scales in both space and time. This kind of modeling plays an important role in modern computational chemistry and it is usually employed when macroscale models are not accurate enough (because of their 'simplicity') and/or when microscale models are not efficients (because of their 'complexity'). In 2013 Martin Karplus, Michael Levitt and Arieh Warshel received the Nobel prize for the development of multiscale models, using classical and quantum mechanics theories for the description of complex chemical systems (e.g. proteins) and reactions.

Let us consider, as an example, a river flow. This in theory this can be described using different models of increasing complexity and decreasing spatial scale, e.g.

- water flow using Navier-Stokes equations (continuum mechanics)
- interaction between particles of fluid using Newton's equation (molecular dynamics)
- atomic and electronic dynamics using Schrödinger's equation (quantum mechanics)

but in practice only the first one is applicable to study the macroscopic flow. Due to their complexity, the computational operations necessary for the other two approaches, make possible (at least with present available hardware) to study at most a hundred thousand ($\approx 10^{-18}$ grams of water) using standard molecular dynamics and up to a hundred molecules using quantum mechanics, at least in the realistic time framework of a Ph.D. course (three-five years). It is clear that except for the first model, the other two approaches are not suitable to simulate directly the river flow. However molecular dynamics and quantum mechanics can, in principle, be useful in the problem at hand. Continuum-based approaches may be ameliorated if macroscopic parameters can be determined accurately (e.g. viscosity). Multiscale modeling, in this simple interpreation, is therefore simply equivalent to the process of extracting information from the most complex/accurate description that are used in the least expensive/accurate approach.[1].

The connection between different methodologies is not straightforward and the application of multiscale modeling is not an easy task. This thesis will be limited therefore to the exploration of a few specific examples. For reasons of clarity, each method employed in this thesis will be discussed highlighting the connections with other (Fig. 1.1) approaches.

1.2 Investigated systems

The chemical systems herein studied include biomolecules like peptides and their thioamidated analogs, biomolecules like Glutathion peroxidase (GPx) mimics, but also bigger systems like

[1] Other interpretation of the concept of 'multiscale approach' can be adopted in different contexts; for example in problems where a microscale model is not required everywhere but only to handle local defects or singularities, using elsewhere a macroscale model.

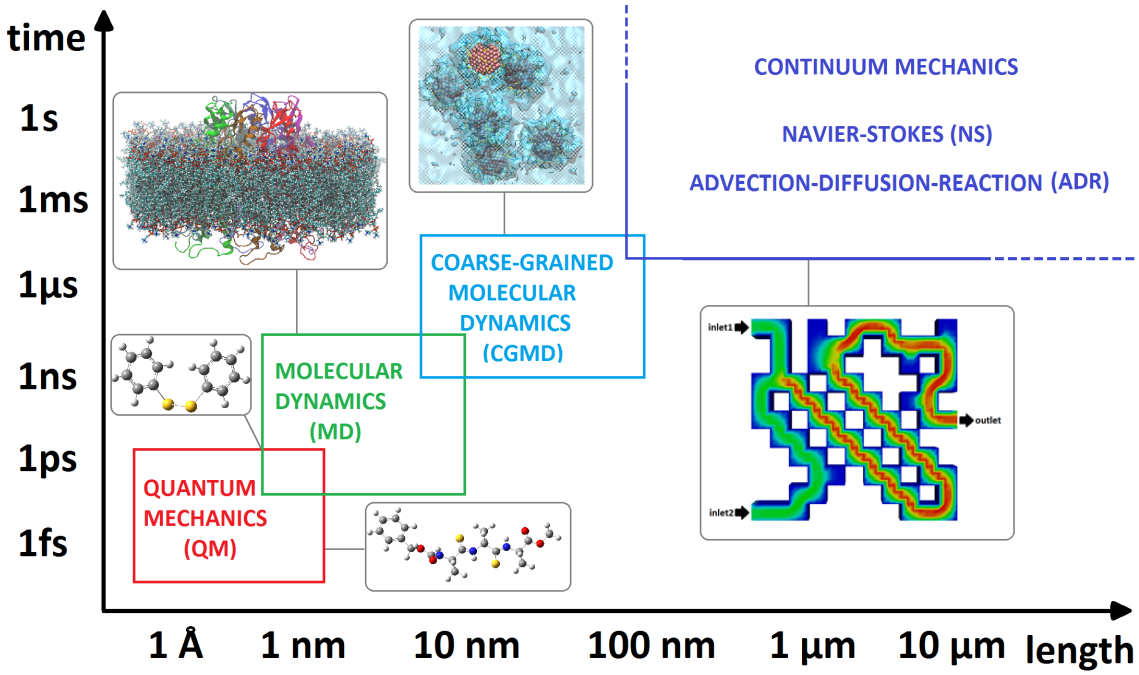


Figure 1.1: Modeling of chemical systems at different length scales.

macromolecules, in particular connexin 26, up to reactive macroscopic fluids. Focus will be on the the dynamics (flexibility), the structure and energetics of these systems.

1.2.1 Dioxopeptides and thionated analogous

Oxo-dipeptides and thio-dipeptides are built via condensation between couples of amino acids and amino thioacids, the latter with the carbonyl oxygen replaced by an sp^2 sulfur. We explored via *in silico* methods all the possible combinations and built 800 dipeptides, whose structures were fully optimized. Maps of condensation energies are presented to highlight optimal partners leading to stable dipeptides and critical situations for which lower stability or instability is predicted in terms of Gibbs reaction free energies. To validate the feasibility of the computational investigation, we synthesized and compared the stabilities of two thionated dimers, namely –Gly[C(CSNH)]Gly– and –Phe[C(CSNH)]Phe–, characterized by diverging physico-chemical properties. To the best of our knowledge, this is the first systematic analysis reported for dipeptides built from natural amino acids as well as for their corresponding thio-analogs.

1.2.2 Diphenyldiselenides and diphenylditellurides

The General AMBER Force Field (GAFF) has been extended to describe a series of selenium and tellurium diphenyldichalcogenides. These compounds display peroxidase activity, i.e. can reduce hydrogen peroxide and harmful organic hydroperoxides to water/alcohols and as such are very promising anti-oxidant drugs. Due to the lack of experimental data, the novel GAFF parameters are validated against quantum mechanically computed observables, i.e. molecular dipole

moments and through MD simulations in different solvents. The whole computational protocol is described in detail and integrated with in-house code to allow easy derivation of the force field parameters for analogous compounds as well as for Se/Te organocompounds in general.

1.2.3 Connexin 26 hemichannel protein

Connexin hemichannels are a class of membrane proteins important for auditory function, intercellular signalling and skin biology encoded by 21 genes in the human genome that, arrange together by non-covalent interactions; these quaternary aggregate forms are called connexons or hemichannels. The pathway generated by this membrane protein allows to connect cells' cytoplasm with the extracellular environment and are regulated by several gating mechanisms, some of which depend on the extracellular Ca^{2+} concentration ($[Ca^{2+}]_e$). It is known that hemichannel activity is inhibited at normal (i.e. about $\sim$1-2 mM) $[Ca^{2+}]_e$, instead lowering $[Ca^{2+}]_e$ to micromolar levels promotes hemichannel opening. Atomic force microscopy imaging shows significant and reversible changes of pore diameter at the extracellular part of Cx26 hemichannels exposed to different $[Ca^{2+}]_e$, however, the underlying molecular mechanisms are not fully elucidated. Analysis of the crystal structure of connexin 26 (Cx26) , together with MD simulations, suggests that several negatively charged amino acids create a favorable environment for Ca^{2+} binding within the extracellular vestibule of the Cx26 hemichannel. In particular a highly conserved glutammic acid, found in position 47 in most connexins, is thought to undergo post translational gamma carboxylation (γGlu47), and is thus likely to play an important role in Ca^{2+} coordination. γGlu47 may also form salt bridges with two conserved arginines (Arg75 and Arg184 in Cx26), which are considered important in stabilizing the structure of the extracellular region. Using a combination of QM and MD methods, we analyzed the interaction between γGlu47, Arg75 and Arg184 in a Cx26 hemichannel model in the absence and in the presence of Ca^{2+} for the wilde type and γ-carboxylatex Cx26.

1.2.4 Reactions in microfluidic devices

We simulated the transport and reaction phenomena that occur in microfluidic devices characterized by flat and regular geometries. In order to describe the solvent motion and each involved chemical substance, we performed numerical simulations using continuum models. The solvent was treated with Navier-Stokes NS equations for incompressible fluids, while the chemical substances were described using scalar Advection-Diffusion-Reaction ADR equations. The NS problem was solved using the OpenFOAM® software, a finite volume method library for computational continuum mechanics, instead the ADR equations were treated decoupled from the former and solved using an in-house developed C++ software. The complexity of the ADR problem was generously reduced adopting an operator splitting approach that permitted to decompose it in 3 simpler sub-problems, each handled using a specific method. These numerical simulations have been used within the context of a heuristic optimization approach; we used

a parallel implementation of the canonical genetic algorithm (GA), in order to modify the microchannels paths and ameliorate the outflow of a desired product. In particular we used a so called micro-GA, because the number of elements which constituted the population was low (8 individuals). As a preliminary test we studied the simple kinetic scheme $A + B \rightarrow C$, in order to test the whole methodology and to characterize the rate of convergence of the GA. This trial also permitted to find the optimal values for the genetic operator parameters (mutation probability, initial generation) that we employed, afterwards, in a more complex reaction, involving several chemical species.

Chapter 2

Software, hardware and methods

It is important for a computational chemist to know which algorithm is going to be used for a computation, how it scales increasing the number of computing cores, how much memory it requires and so on. Hardware architecture should be known, to test the implemented code and achieve the peak performance, i.e. build a scale-up curve in order to use efficiently the hardware at our disposal avoiding to include too many cores in the computation beyond a certain number. Using too many cores may be not so convenient because the scale-up curve at some point become almostly flat or even decreases, so adding computational resources (cores and/or *graphic processing units*, GPUs) will result in a waste of processing power [1] that may be used for other projects or by other users. It is very important to know and manage correctly the hardware and the software in order to obtain results efficiently and within the shortest amount of time.

A fundamental item is the adopted computational method. Going from the common classical to accurate quantum mechanical (e.g. post-HF) methods the scale cost raises up by several powers, as reported in Table 2.1. Nowadays, much effort has been put in improving the efficiency of most methods, often using straightforward ideas: for instance in molecular dynamics codes neglecting particle-particle interactions for atoms further than a certain cut-off, decided a priori or estimated during the calculations; or the efficient treatment of long range interactions, which allows to reduce the complexity of molecular dynamics (MD) simulations from $\mathcal{O}(N^2)$ to $\mathcal{O}(N \log N)$, using particle mesh Ewald (PME), withn N the number of particles. Analogously, in the standard Hartree-Fock (HF) method the computational cost can be lowered from $\mathcal{O}(N^4)$ to $\mathcal{O}(N^{2.7})$ because the basis functions decay at least exponentially, so most of the shell pairs to be calculated for large molecules are negligible; these shell pairs may be identified on the base of upper bounds calculated using the Cauchy–Schwarz inequality or other similar criteria. On the contrary DFT method scales formally as $\mathcal{O}(N^3)$ but in practice it is often added to existing HF code, so this reduction is not achieved in practice. In order to obtain very accurate results (bonds and molecular properties) computational chemists rarely use methods that go beyond the complexity of the coupled cluster method with single, double and perturbatively estimated triple

[1] The performance reduction is usually due to the overheads, i.e. the communication between each core slows down the rest of the calculus.

excitations (CCSD(T)). We could therefore say this hybrid coupled cluster/perturbation theory approach has become the "gold standard" because it is applicable to reasonably large molecules, for a reasonable computational cost.

Complexity	Method
$\mathcal{O}(N^2)$	MM MD
$\mathcal{O}(N^3)$	DFT
$\mathcal{O}(N^4)$	HF
$\mathcal{O}(N^5)$	MP2
$\mathcal{O}(N^6)$	MP3 MP4(SDQ) CISD CCSD
$\mathcal{O}(N^7)$	MP4 CCSD(T)
$\mathcal{O}(N^8)$	MP5 CISDT CCSDT
$\mathcal{O}(N^9)$	MP6
$\mathcal{O}(N^{10})$	MP7 CISDTQ CCSDTQ

Table 2.1: Complexity of some classical and quantum methods. N is the size of the system, i.e. number of atoms and basis functions for classical and quantum methods respectively.

2.1 Employed software packages

All calculations presented in this basis were obtained using pre-existing software, in-house developed C/C++ programs and scripts. In particular for the QM calculations we used Gaussian 09 Rev. B01 [1] together with C/C++ programs developed by ourselves in order to automate the launch of multiple jobs using differently substituted molecules and for post-production analysis. The MD simulations were performed using the GPU version of PME molecular dynamics (pmemd.cuda and pmemd.cuda.MPI) executables of the AMBER 14 package and the post-production analysis was done using AmberTools 14, GROMACS 4.6.5 and Visual Molecular Dynamics (VMD). The microfluidics simulations were done using OpenFOAM® 2.1.0 for NS resolution, in conjuction with the in-house developed solver MicReaX, to propagate the ADR equation and obtain the evolution of the concentrations in the time.

2.2 Computational facilities

Except post-production analysis, all the calculations were run on cluster systems in order to use parallelization with a convenient reduction of the required time. In particular all the quantum chemistry calculations, were run on the cluster Gollum, jointly managed by the Department of Physics and Astronomy and the Department of Chemical Sciences of the University of Padova, using at maximum 8 core per job, because increasing further this number lead to a much lower

scale-up efficiency. Some of these calculations were also executed at CINECA clusters EURORA and GALILEO, with the same setup. The molecular dynamics simulations were run on the cluster Curie hosted by the Department of Chemical Sciences employing at maximum 2 MPI processes and 2 Nvidia Tesla K20Xm GPUs, linked by a direct peer-to-peer communication with a resulting higher performance. The system 5.1, constituted by $\sim$200K atoms, led to the production of $\sim$16 ns/day. Finally the microfluidics simulations were run on the Avogadro cluster at the Department of Chemical Sciences using 24 core in order to run the GA.

2.3 Quantum chemistry methods

Quantum chemistry is a wide branch of chemistry that applies quantum mechanics to the study of molecules, in order to characterize their energetic, structural, electronic and spectroscopic properties. Supporting theories of this field were fully developed many years ago, starting in 1920 thanks to important scientists awarded with Nobel prize. The crisis of classical physics was due experiments showing that the light behaved as if it was made of particles and that the electrons behaved as waves, both highlighting the wave-particle duality that classical physics was unable to explain[2]. A first attempt to develop a quantum theory of matter was due to Niels Bohr who explained the spectral emission lines of hydrogen. A profound revision of previous physical theories was done by Erwin Schrödinger who formulated the equation describing a quantum particle, and Werner Karl Heisenberg who introduced the matrix mechanics formulation of quantum mechanics and developed the uncertainty principle. Subsequently Paul A. M. Dirac introduced the linear operators, the braket notation, and unified Heisenberg's matrix mechanics and Schrödinger's wave mechanics into a single mathematical formalism; he also predicted the existence of antimatter and proposed a relativistic equation of motion [2]. The work of Douglas Hartree and Vladimir Aleksandrovich Fock (end of 1920s), which developed the self consistent field method (SCF), was the first significant attempt to make the application of the theory possible to calculate properties of atoms or molecules. Many years later, in 1951, Clemens C. J. Roothaan and George G. Hall developed independently a new matrix form of the Hartree-Fock (HF) equations that, in contrast with the previous integro-differential form, were easily solvable with a computer. A few years later John C. Pople, R. K. Nesbet and Gaston Berthier extended the previously developed HF method to open shell systems, introducing the unrestricted HF method. In 1965, Walter Kohn and Lu Jeu Sham developed a parallel theory that is one of the most used nowadays to study molecular systems, known as *density functional theory*, DFT, in which the central role is played by the electron density and its functionals. Many other methods have been developed like semi-empirical approaches (Hückel, extended Hückel,

[2]The photoelectric effect experiment and theory were developed by Hertz and Albert Einstein respectively, the latter received the Nobel prize in 1921 for this achievement. The electron diffraction experiment have been done in 1921 by Davisson and Germer, while the theory was developed by De Broglie that received the Nobel prize in 1929

Austin model 1, Parameterized Model number 3, Intermediate Neglect of Differential Overlap, Modified Intermediate Neglect of Differential Overlap and many others) that include empirical parameters to reduce computational effort and the so-called post-HF methods (Configuration interaction, coupled cluster, Møller–Plesset perturbation theory and others) which ameliorate the HF method improving the description of electron correlation. Thanks to the growth in computer performance and development of efficient algorithms, modern DFT calculations may be done on hundreds of atoms and post HF computations may be applied to several tens of atoms (e.g. glycylglicine) obtaining observable very close to experimental values (see pag. 36). If the Moore's law[3] will continue to apply and algorithms to improve, it is highly probable that direct simulation via quantum methods of molecular and supramolecular will become a common item in the arsenal investigative tools of any practitioner in chemistry.

2.3.1 Hartree-Fock

This method was developed at the end of 1920s by the physicists Douglas Hartree and Vladimir Aleksandrovich Fock and allows to obtain the wave function and energy of a quantum molecular system in a stationary state. This method is used often as a starting point for much more complex and accurate calculations in the so called post HF methods, which allow to obtain solutions very close to the true non-relativistic wavefunction, or for the semi-empirical functionals (as PBE0, B3LYP etc.) employed to also include the exchange energy in the DFT calculations. Despite the errors resulting in computed observables, this is still a standard for parameterizing partial charges of molecules that are to be solvated in water using molecular dynamics simulations. This is due to the fact that the error between charges in gas phase and in solution is compensated using the HF/6-31G(d) level of theory, obtaining an overestimation of the molecules polarity. For example the dipole moments of TIP3P and SPC water models is enhanced with respect to the water molecule in gas-phase[3], which is desiderable in order to study molecules in condensed phase and in general an increase of 10-20% is achived using this methodology. We also developed a stand-alone free 64-bit executable of the HF method, useful to compute observables for small molecules (dipole moments, energy, electronic levels), that may be found at this link [4]. The HF method has been employed in this thesis for the modeling of the systems reported in chapter 5. This method was initially designed for closed-shell systems, i.e. with all electrons spin-paired and is known as restricted Hartree-Fock (RHF). The total Hamiltonian operator of the system is given by the sum of the nuclear and electronic Hamiltonians

$$\hat{H} = \hat{H}_{el} + \hat{H}_{nuc} = \hat{T}_e + \hat{V}_{Ne} + \hat{V}_{ee} + V_{NN} \tag{2.1}$$

and if the nuclei positions are almost fixed because they are much heavier than the electrons, or as happens in quantum chemistry, are parameters that may be adjusted iteratively to obtain the

[3] Moore's law states that processor speeds, or overall processing power for computers will double every 18 months.

lowest energy (i.e. optimized structure) of the molecule, the (2.1) may be semplified applying the Born-Oppenheimer approximation. The Hamiltonian is split in the nuclear and electronic parts which are expressed in atomic units as

$$H_{nuc} = \sum_{A=1}^{M} \sum_{B>A}^{M} \frac{Z_A Z_B}{R_{AB}} \tag{2.2}$$

$$\hat{H}_{el} = -\frac{1}{2} \sum_{i=1}^{N} \nabla_i^2 - \sum_{i=1}^{N} \sum_{A=1}^{M} \frac{Z_A}{r_{iA}} + \sum_{i=1}^{N} \sum_{j>i}^{N} \frac{1}{rij} \tag{2.3}$$

where the (2.2) is the classical electrostatic potential of the nuclei that may be easily calculated and the (2.3) is given by the kinetic and potential energy terms of the electrons that require a more sophisticated treatment. We seek for a solution to the stationary Schrödinger equation for the electrons

$$\hat{H}_{el}\, \Psi_{el} = E_{el}\, \Psi_{el} \tag{2.4}$$

that has analytical solution for very simple cases but for practical purposes have to be solved iteratively using the variational principle. Given the spin orbital χ that describes the spatial distribution and the spin of one electron[4], the many-electron wavefunction was early expressed by Hartree as products of spin orbitals but an improvement was done thanks to John C. Slater that proposed to express the wavefunction as a determinant

$$|\Psi_{el}\rangle = \Psi_{el}(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_1(\mathbf{x}_1) & \chi_2(\mathbf{x}_1) & \cdots & \chi_N(\mathbf{x}_1) \\ \chi_1(\mathbf{x}_2) & \chi_2(\mathbf{x}_2) & \cdots & \chi_N(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_1(\mathbf{x}_N) & \chi_2(\mathbf{x}_N) & \cdots & \chi_N(\mathbf{x}_N) \end{vmatrix} \tag{2.5}$$

where **x** are the cartesian plus the spin coordinates. If we interchange the coordinates of two electrons (i.e. $\mathbf{x}_i = \mathbf{x}_j$) we swap two columns and the determinant change sign; instead if two electrons occupy the same orbital (i.e. $\chi_i = \chi_j$) than two columns are identical and the determinant goes to zero. For these reasons this mathematical object allows to respect the antisymmetric behaviour of fermions and the Pauli exclusion principle. The electronic Hamiltonian may be further simplified introducing the one-electron and two-electron operators expressed as

$$\hat{h}_i = -\frac{1}{2} \nabla_i^2 - \frac{Z_A}{r_{iA}} \tag{2.6}$$

$$\hat{g}_{ij} = \frac{1}{rij} \tag{2.7}$$

[4]The spin orbitals are simply given by the spatial orbital multiplied by the spin function $\chi_i(\mathbf{x}_i) = \phi_i(\mathbf{r}_i)\,|\alpha\rangle$, $\chi_i(\mathbf{x}_i) = \phi_i(\mathbf{r}_i)\,|\beta\rangle$ where **r** are the cartesian coordinates and **x** are the cartesian plus the spin coordinates.

so $\hat{H}_{el}$ may be expressed as

$$\hat{H}_{el} = \sum_{i}^{N} \hat{h}_i + \sum_{i}^{N} \sum_{j>i}^{N} \hat{g}_{ij} \tag{2.8}$$

The expectation value of the electronic energy is given by

$$E_0 = \langle \Psi_{el} | \hat{H}_{el} | \Psi_{el} \rangle = \sum_{i}^{N} \langle \Psi_{el} | \hat{h}_i | \Psi_{el} \rangle + \sum_{i}^{N} \sum_{j>i}^{N} \langle \Psi_{el} | \hat{g}_{ij} | \Psi_{el} \rangle \tag{2.9}$$

and if we express the integrals over the spin-orbitals in a compact form

$$\langle i | \, \hat{h}_i \, | j \rangle = \int \chi_i^*(\mathbf{x}_1) \, \hat{h}_i \, \chi_j(\mathbf{x}_1) \, d\mathbf{x}_1 \tag{2.10}$$

$$\langle ij | \, \hat{h}_i \, | kl \rangle = \int \chi_i^*(\mathbf{x}_1) \, \chi_j^*(\mathbf{x}_2) \, \frac{1}{r_{ij}} \, \chi_k(\mathbf{x}_1) \, \chi_l(\mathbf{x}_2) \, d\mathbf{x}_1 \, d\mathbf{x}_2 \tag{2.11}$$

we can cancel out normalization constant of Ψ_{el} and the (2.9) is expressed as

$$E_0 = \langle \Psi_{el} | \hat{H}_{el} | \Psi_{el} \rangle = \sum_{i}^{N} \langle i | \, \hat{h}_i \, | j \rangle + \frac{1}{2} \sum_{ij}^{N} \left(\underbrace{\langle ij | \, \frac{1}{r_{ij}} \, | ij \rangle}_{J_{ij}} - \underbrace{\langle ij | \, \frac{1}{r_{ji}} \, | ji \rangle}_{K_{ij}} \right) \tag{2.12}$$

where the disequality in the second summation disappears because $J_{ii} = K_{ii}$ and those terms cancel out. J_{ij} and K_{ij} are the Coulomb and exchange integrals respectively. The Coulomb integral is easily comparable to the classical analog, instead the exchange integral is a stabilizing term which has not an immediate classical counterpart to be comparable to and is also negligible at long distances. We need now to minimize the HF energy by modifying the orthonormal orbitals so that any change in the expectation value of the electronic energy will be zero.

$$\chi_i = \chi_i + \delta_{\chi_i} \Rightarrow \delta \, \langle \Psi_{el} | \, \hat{H}_{el} \, | \Psi_{el} \rangle = 0 \tag{2.13}$$

This may be addressed using the Lagrange multipliers method

$$\delta F := \delta \left[\langle \Psi_{el} | \hat{H}_{el} | \Psi_{el} \rangle - \sum_{ij} \epsilon_{ij} (\langle i | j \rangle - \delta_{ij}) \right] = 0 \tag{2.14}$$

where ϵ_{ij} are the Lagrange multipliers, $\langle i | j \rangle$ is the overlap between spinorbitals i and j and δ_{ij} is the Kronecker symbol ($\delta_{ij} = 1$ if $i = j$ else $\delta_{ij} = 0$), after some algebraic manipulation $\frac{\delta F}{\delta \chi_k^*}$ is expressed as

$$\left[\hat{h}_1 + \sum_{i}^{N} \hat{J}_i - \hat{K}_i \right] \chi_k(\mathbf{x}_1) = \sum_{i}^{N} \epsilon_{ki} \, \chi_i(\mathbf{x}_1) \tag{2.15}$$

if we define the Fock operator as $\hat{h}_1 + \sum_i^N (\hat{J}_i - \hat{K}_i) =: \hat{f}$ and if we concentrate on the solutions that satisfy $\epsilon_{ki} = \delta_{ki}\, \epsilon_k$ we obtain the HF equation

$$\hat{f}\, \chi_k = \epsilon_k\, \chi_k \tag{2.16}$$

To solve this complex integro-differential equation we introduce now the basis functions which is a set of functions linearly combined to generate the molecular orbitals. For practical purposes these functions are centered on the atoms but in theory they can be placed anywhere. The basis set also defines the space in which the problem is solved, so the complexity and accuracy of the calculus depend on it. The basis functions are given as a linear combination of contracted primitives muliplied by the expansion coefficients $C_{\mu i}$ that we will optimize in order to obtain the minimum energy of the molecule

$$\chi_i = \sum_{\mu=1}^{N_{BF}} C_{\mu i} \tilde{\chi}_\mu \tag{2.17}$$

where N_{BF} is the number of basis function employed. The higher N_{BF} the more accurate and complex the calculus will be. Thanks to the findings of Clemens C. J. Roothan we can convert the HF equation into a linear algebra problem

$$\sum_\nu F_{\mu\nu} C_{\nu i} = \epsilon_i \sum_\nu S_{\mu\nu} C_{\nu i} \tag{2.18}$$

where the summation goes over all the basis functions and

$$S_{\mu\nu} = \int \tilde{\chi}_\mu(\mathbf{x}_1)\, \tilde{\chi}_\nu(\mathbf{x}_1)\, d\mathbf{x}_1 \tag{2.19}$$

$$F_{\mu\nu} = \int \tilde{\chi}_\mu(\mathbf{x}_1)\, F(\mathbf{x}_1)\, \tilde{\chi}_\nu(\mathbf{x}_1)\, d\mathbf{x}_1 \tag{2.20}$$

are the densiy and Fock matrices, $C_{\nu i}$ is the coefficient matrix and the eigenvalues ϵ_i may be expressed as a diagonal matrix, obtaining the matrix form of the HF equation expressed as

$$\mathbf{F}\,\mathbf{C} = \mathbf{S}\,\mathbf{C}\,\boldsymbol{\epsilon} \tag{2.21}$$

To solve the (2.21) we use the Löwdin canonical orthogonalization, i.e. we rewrite the equation in order to obtain an eigenvalue problem

$$\mathbf{S}^{-1/2}\,\mathbf{F}\,\mathbf{C} = \mathbf{S}^{1/2}\,\mathbf{C}\,\boldsymbol{\epsilon} \tag{2.22}$$

$$\mathbf{S}^{-1/2}\,\mathbf{F}\,\mathbf{S}^{-1/2}\,\mathbf{S}^{1/2}\,\mathbf{C} = \mathbf{S}^{1/2}\,\mathbf{C}\,\boldsymbol{\epsilon} \tag{2.23}$$

so if we define the transformed Fock matrix as $\mathbf{F}' = \mathbf{S}^{-1/2}\,\mathbf{F}\,\mathbf{S}^{-1/2}$ and the transformed coefficient matrix as $\mathbf{C}' = \mathbf{S}^{1/2}\,\mathbf{C}$ the (2.21) may be expressed as an eigenvalue equation

$$\mathbf{F}' \, \mathbf{C}' = \mathbf{C}' \, \boldsymbol{\epsilon} \tag{2.24}$$

In practice given the orthogonalization matrix $\mathbf{S}^{-1/2}$ we transform the Fock matrix to $\mathbf{F}'$, that may be diagonalized to obtain the eigenvector $\mathbf{C}'$ and eigenvalues $\boldsymbol{\epsilon}$, after that we use again the orthogonalization matrix to get back $\mathbf{C} = \mathbf{S}^{-1/2} \, \mathbf{C}'$. so provided an initial guess for $\mathbf{F}'$ (using for example the core Hamiltonian and setting $\mathbf{D} = 0$) we can solve this equation, calculate new coefficients **C** and resolve iteratively until, given a tolerance, the change in **D** between two subsequent iteration cycles is very small. This is why the solution to Hartree-Fock-Roothan equation is also called SCF procedure. If the number of electrons in the molecule is even, the Fock matrix is constructed using the following formula

$$F_{\mu\nu} = h_{\mu\nu} + \sum_{i}^{N/2} [2(ij|kl) - (il|kj)] \tag{2.25}$$

and introducing the basis set

$$\begin{aligned} F_{\mu\nu} = h_{\mu\nu} + \sum_{\mu\nu} \left[2 \sum_{i=1}^{N/2} C_{\mu i} C_{\nu i} \right] \left[(\mu \, \nu|\sigma \, \lambda) - \frac{1}{2} (\mu \, \lambda|\sigma \, \nu) \right] = \\ h_{\mu\nu} + \sum_{\mu\nu} D_{\mu\nu} \left[(\mu \, \nu|\sigma \, \lambda) - \frac{1}{2} (\mu \, \lambda|\sigma \, \nu) \right] \end{aligned} \tag{2.26}$$

So the SCF algorithm (without geometry optimization) may be summarized as follows where

Algorithm 1 SCF algorithm

1: **procedure** CALCULATE INTEGRALS
2: Calculate the overlap matrix **S**
3: Calculate the orthogonalization matrix $\mathbf{S}^{-1/2}$
4: Calculate the core Hamiltonian $h_{\mu\nu}$
5: Calculate the electron repulsion integrals $(a \, b|c \, d)$
6: **end procedure**
7: Generate an initial guess $D_{\mu\nu}$
8: **while** $max|D^{n}_{\mu\nu} - D^{n-1}_{\mu\nu}| > \epsilon$ **do**
9: Calculate the Fock matrix $F_{\mu\nu} \leftarrow h_{\mu\nu} + \sum_{\mu\nu} D_{\mu\nu} \left[(\mu \, \nu|\sigma \, \lambda) - \frac{1}{2} (\mu \, \lambda|\sigma \, \nu) \right]$
10: Orthogonalize the Fock matrix $\mathbf{F}' \leftarrow \mathbf{S}^{-1/2} \, \mathbf{F} \, \mathbf{S}^{-1/2}$
11: Diagonalize $\mathbf{F}'$ to obtain eigenvectors $\mathbf{C}'$ and eigenenergies $\boldsymbol{\epsilon}$
12: Transform back $\mathbf{C} \leftarrow \mathbf{S}^{-1/2} \, \mathbf{C}'$
13: Form a new $D_{\mu\nu} \leftarrow 2 \sum_{i=1}^{N/2} C_{\mu i} C_{\nu i}$
14: **end while**

convergence criterion on the density matrix variation has been used, altough a more strict one may be used like the commutator $[\mathbf{F}, \mathbf{D}]$, e.g. $max|\mathbf{FDS} - \mathbf{SDF}| > \epsilon$ that is equal to 0 when the wavefunction has converged. The total electronic Hartree-Fock energy of the system is not

given by the sum of the eigenenergies ϵ, due to the double counting of the electron-electron interaction and if we use (2.12), (2.15) and (2.26) we see that the energy is given by

$$E_0 = \sum_k \epsilon_k - \frac{1}{2}\sum_{\mu\nu} D_{\mu\nu}\left[(\mu\,\nu|\sigma\,\lambda) - \frac{1}{2}(\mu\,\lambda|\sigma\,\nu)\right] = \frac{1}{2}\,\mathrm{tr}\,\mathbf{D}(\mathbf{h}+\mathbf{F}) \tag{2.27}$$

where it should be noticed that this is not the total energy of the system, which also include V_{NN}. Summarizing the Hartree–Fock method take into account five simplifications in order to get rid of the problem:

- relativistic effects are completely neglected, the momentum operator is assumed to be completely non-relativistic;
- the previously described Born–Oppenheimer approximation is always included;
- the variational solution is assumed to be a linear combination of a finite number of orthogonal basis functions which are assumed to be approximately complete;
- each energy eigenfunction is assumed to be describable by a single Slater determinant, an antisymmetrized product of one-electron wave functions;
- the mean field approximation is implied. Effects arising from deviations from this assumption, known as electron correlation, are completely neglected for the electrons of opposite spin, but are taken into account for electrons of parallel spin.

the so-called post-Hartree–Fock methods foresee a relaxation of the last two approximations.

Unrestricted Hartree-Fock

The HF method previously described cannot be applied to radicals or systems with different number of α and β electrons (i.e. open-shell systems), because the molecular orbitals are used twice for each spin. In the 1954 J. A. Pople and R. K. Nesbet developed a new method that allowed to treat α and β electrons separately, this method is known as unrestricted Hartree-Fock (UHF) [5]. In practice different molecular orbitals are used for both spins, solving two coupled Hartree-Fock-Roothan equations given as

$$\mathbf{F}^\alpha(\mathbf{D}^\alpha,\mathbf{D}^\beta)\,\mathbf{C}^\alpha = \mathbf{S}\,\mathbf{C}^\alpha\,\boldsymbol{\epsilon}^\alpha \tag{2.28}$$

$$\mathbf{F}^\beta(\mathbf{D}^\alpha,\mathbf{D}^\beta)\,\mathbf{C}^\beta = \mathbf{S}\,\mathbf{C}^\beta\,\boldsymbol{\epsilon}^\beta \tag{2.29}$$

where there are two Fock, density and energy matrices and the coupling is given in the Fock matrices expressed as follows

$$F^\alpha_{\mu\nu} = h_{\mu\nu} + \sum_{\mu\nu}\left[(D^\alpha_{\mu\nu}+D^\beta_{\mu\nu})(\mu\,\nu|\sigma\,\lambda) - D^\alpha_{\mu\nu}(\mu\,\lambda|\sigma\,\nu)\right] \tag{2.30}$$

$$F^{\beta}_{\mu\nu} = h_{\mu\nu} + \sum_{\mu\nu} \left[(D^{\alpha}_{\mu\nu} + D^{\beta}_{\mu\nu})(\mu\,\nu|\sigma\,\lambda) - D^{\beta}_{\mu\nu}(\mu\,\lambda|\sigma\,\nu) \right] \tag{2.31}$$

which depend from both $\mathbf{D}^{\alpha}$ and $\mathbf{D}^{\beta}$. In this case the electronic UHF energy is given by

$$E_0 = \frac{1}{2}\,\mathrm{tr}\,\mathbf{D}^{\alpha}(\mathbf{h} + \mathbf{F}^{\alpha}) + \frac{1}{2}\,\mathrm{tr}\,\mathbf{D}^{\beta}(\mathbf{h} + \mathbf{F}^{\beta}) \tag{2.32}$$

Differently from the RHF, the drawback of the UHF is that the two-electron wave function is not an eigenfunction of the $\hat{S}^2$ spin operator, because the ground state is always contaminated by higher spin states, unless an equal number of α and β electrons are described. To get rid of this problem a different type of method (even if less employed), called restricted-open Hartree-Fock, may be used or the UHF wave function may be corrected with several methods, in order to reduce the spin contamination effect.

Electron repulsion integrals

In order to describe molecules it is common to use a basis set constituted by a finite set of atomic orbitals linearly combined and centered on each atom. John C. Slater first introduced during the 1930, the so called Slater type orbitals (STO), a peculiar function that behave well at short and long distances that in cartesian coordinates is expressed as

$$\phi^{STO} = N x^a y^b z^c e^{-\zeta r} \tag{2.33}$$

where N is a normalizing constant, r is the electron-nucleus distance and ζ is a constant related to the effective charge of the nucleus. Differently from the hydrogen-like functions STOs have not radial nodes but they

- have an exponential decay which accurately describe the long-range overlap between atoms
- have a cusp on the nucleus as do the hydrogen-like atom functions

Although STOs describe very well the electron, the integrals involved in the computation of observables may only be calculated numerically and this is the main reason why this type of function is used in few software package[5]. S. Francis Boys in 1950 found out that it was convenient to approximate STOs as linear combination of Gaussian functions. Gaussians have an important property summarized in the "Gaussian product theorem", i.e. the product of two Gaussians is given by another Gaussian centered on a point along the axis connecting them, so that all the integrals may be computed analitically with a much lower computational effort. A normalized

[5] Amsterdam density functional software can do HF calculations using STOs. We should mention that the complexity of calculating electron repulsion integrals (ERI) has nothing to do with the time due to slow convergence of the basis set which may be more efficient using STOs especially for d-metal complexes [6]

cartesian GTO is expressed as

$$\phi^{GTO} = N x^a y^b z^c e^{-\zeta r^2} \tag{2.34}$$

where r is now squared, which also make worse the shape at short and long distances, so that several GTOs are necessary to obtain a good shape of the atomic orbitals; this collection of functions is known as contracted GTO, expressed as a linear combination of simpler functions called primitive GTOs

$$\phi^{CGTO} = \sum_{i=1}^{K} c_i \phi^{GTO} \tag{2.35}$$

where c_i are the contractions coefficients (fixed parameters usually obtained by least-squares fitting) and K is the contraction degree which reflects the accuracy/complexity of the calculus.

The computation of the repulsion energy and also their derivatives with respect to nuclear motion, requires the computation of matrix elements of the form

$$(a\ b|c\ d) = \int\int \phi_a(\mathbf{r}_1)\ \phi_b(\mathbf{r}_1) \frac{1}{r_{12}} \phi_c(\mathbf{r}_2)\ \phi_d(\mathbf{r}_2)\ d\mathbf{r}_1\ d\mathbf{r}_2 \tag{2.36}$$

and their n-th derivatives with respect to the displacement of the basis functions. There is a permutational symmetry for the two electron integrals expressed as

$$(a\,b|c\,d) = (b\,a|c\,d) = (a\,b|d\,c) = (d\,c|b\,a) = (c\,d|a\,b) = (d\,c|a\,b) = (c\,d|b\,a) = (b\,a|d\,c) \tag{2.37}$$

and since there are N^{BF} basis functions in the basis set, there are $N^{BF}(N^{BF}+1)/2$ distinct basis function pairs and similarly there are

$$N_{total} = \frac{1}{2}\left[\frac{N^{BF}(N^{BF}+1)}{2}\right]\left[\frac{N^{BF}(N^{BF}+1)}{2}+1\right] \tag{2.38}$$

distinct integrals. So the simmetry of the integrals may be exploited to reduce the computation by a factor 8. The contracted integrals $(a\ b|c\ d)$ are expressed as a sum of their component primitive integrals $[ab|cd]$

$$(a\,b|c\,d) = \sum_{i=1}^{K_A}\sum_{j=1}^{K_B}\sum_{k=1}^{K_C}\sum_{l=1}^{K_D} c_{ai}\ c_{bj}\ c_{ck}\ c_{dl}[a_i\ b_j|c_k\ d_l] \tag{2.39}$$

where c are the contraction coefficients. It is clear that the computation of each electron repulsion integral (ERI) requires $\mathcal{O}(K_A K_B K_C K_D)$ operations so this is the main heavy task that need to be executed to do HF calculations and this was also the main research topic in order to improve the speed-up. Different algorithms for the computation of ERI are described in appendix A.

2.3.2 Density functional theory

Density functional theory is widely used in chemistry, physics and material science to study the electronic structure of a many-body systems, using functions of another function (i.e. functionals) of the electron density. In this thesis the DFT method have been abundantly used to optimize the geometry of all the molecules. This approach have been very popular to model solid-state systems because large system sizes required its simplicity, but was not considered accurate enough to be also applied in quantum chemistry until the 1990s, when the approximate functionals employed were deeply refined. Although many improvements have been done some issues still remain, as a poor description of the dispersion interactions (many functionals have been developed to get rid of this inaccuracies, e.g. the Grimme's functionals [7–9]) and other drawbacks. This theory is based on the model of Llewellyn Thomas and Enrico Fermi (1927) describing electrons around a nucleus as a system of fermions interacting with a potential acting on each single fermion, that contains the interaction of all the others fermions (mean field approximation). Pierre C. Hohenberg and Walter Kohn laid the foundation of this model thanks to their theorems. The Hohenberg-Kohn theorems state that

1. the external potential $\hat{V}_{Ne}(\mathbf{r})$ and consequently the total energy, is a unique functional of the total density $\rho(\mathbf{r})$. This may be easily demonstrated by reduction to the absurd. Let's consider two external potentials $\hat{V}_{Ne}(\mathbf{r})$ and $\hat{V}'_{Ne}(\mathbf{r})$ with ground states $|\Psi_0\rangle$ and $|\Psi'_0\rangle$ that are functionals of the same density $\rho_0(\mathbf{r})$. The expectation values of the energies are given as

$$E_0 = \langle\Psi_0|\hat{H}|\Psi_0\rangle = \langle\Psi_0|\hat{F} + \hat{V}_{ext}|\Psi_0\rangle \tag{2.40}$$

$$E'_0 = \langle\Psi'_0|\hat{H}'|\Psi'_0\rangle = \langle\Psi'_0|\hat{F} + \hat{V}'_{ext}|\Psi'_0\rangle \tag{2.41}$$

   using as trial wave function $\langle\Psi'_0|$ for $\hat{H}$ the following inequality holds

$$E_0 < \langle\Psi'_0|\hat{H}|\Psi'_0\rangle = \langle\Psi'_0|\hat{H}'|\Psi'_0\rangle + \langle\Psi'_0|\hat{H} - \hat{H}'|\Psi'_0\rangle = E'_0 + \int [\hat{V}_{ext}(\mathbf{r}) - \hat{V}'_{ext}(\mathbf{r})]\, \rho(\mathbf{r})\, d\mathbf{r} \tag{2.42}$$

   and similarly taking $\langle\Psi_0|$ as trial wavefunction for $\hat{H}'$ we obtain

$$E'_0 < \langle\Psi_0|\hat{H}'|\Psi_0\rangle = \langle\Psi_0|\hat{H}|\Psi_0\rangle - \langle\Psi_0|\hat{H} - \hat{H}'|\Psi_0\rangle = E_0 - \int [\hat{V}_{ext}(\mathbf{r}) - \hat{V}'_{ext}(\mathbf{r})]\, \rho(\mathbf{r})\, d\mathbf{r} \tag{2.43}$$

   and adding together (2.42) and (2.43) we obtain

$$E_0 + E'_0 < E_0 + E'_0 \tag{2.44}$$

   which is absurd, so the ground-state density determines the external potential.

2. the ground-state energy can be obtained variationally, i.e. $\rho(\mathbf{r})$ that minimizes the total energy is the exact ground-state density; this may be easily demonstrated using variational

theorem. We know from the previous theorem that $\hat{V}_{ext}(\mathbf{r})$ and the ground state $|\Psi_0\rangle$ are determined by $\rho_0(\mathbf{r})$. If we try this ground state for the Hamiltonian with another external potential $\hat{V}$ we have

$$\langle\Psi_0|\hat{H}|\Psi_0\rangle = \langle\Psi_0|\hat{F}|\Psi_0\rangle + \langle\Psi_0|\hat{V}|\Psi_0\rangle = F[n] + \int \hat{V}\,\rho(\mathbf{r})\,d\mathbf{r} = E_V[n] \geq E_0 \quad (2.45)$$

Similarly as in the HF method, we use the Born-Oppenheimer approximation writing the Schrödinger equation as

$$\hat{H}_{el}\,\Psi_{el} = \left[\hat{T}_e + \hat{V}_{Ne} + \hat{V}_{ee}\right]\Psi_{el} = \left[-\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2 - \sum_{i=1}^{N}\sum_{A=1}^{M}\frac{Z_A}{r_{iA}} + \sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{rij}\right]\Psi_{el} = E\,\Psi_{el} \quad (2.46)$$

Instead of using the wavefunction the key variable is the electron density given as

$$\rho(\mathbf{r}) = N\int\cdots\int|\Psi_{el}(\mathbf{x}_1,\mathbf{x}_2,...,\mathbf{x}_N,)|^2 ds_1 d\mathbf{x}_1\,d\mathbf{x}_2\,...\,d\mathbf{x}_N \quad (2.47)$$

The total energy is expressed given by

$$E[\rho] = T[\rho] + E_{Ne}[\rho] + E_{ee}[\rho] \quad (2.48)$$

where the interaction with the external potential is given by $E_{Ne}[\rho] = \int \hat{V}_{Ne}(\mathbf{r})\,\rho(\mathbf{r})\,d\mathbf{r}$ but T and E_{ee} are unknowns. Walter Kohn and Lu J. Sham proposed a set of equations of non-interacting particles very similar to the HF equations using the Lagrange multipliers method. The total energy is given as

$$E[\rho] = T_S[\rho] + \int \hat{V}_{Ne}(\mathbf{r})\,\rho(\mathbf{r})\,d\mathbf{r} + J[\rho] + E_{XC}[\rho] \quad (2.49)$$

where $T_S[\rho] = -\frac{1}{2}\sum_i^N \langle\phi_i|\nabla^2|\phi_i\rangle$ is the kinetic energy of the N non-interacting electrons, $J[\rho] = \frac{1}{2}\int\frac{\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}{r_{12}}\,d\mathbf{r}_1\,d\mathbf{r}_2$ is the Hartree energy, i.e. the classical Coulomb interaction, similarly to J in (2.12) but expressed in terms of the density. The exchange-correlation term is given by

$$E_{XC} = (T[\rho] - T_S[\rho]) + (V_{ee}[\rho] - J[\rho]) \quad (2.50)$$

and represents the sum of the error made using a non-interacting kinetic energy T_S and a classic electron-electron interaction J. In order to minimize (2.49) under the constraint $\langle\phi_i|\phi_j\rangle = \delta_{ij}$ we obtain the Kohn-Sham equations

$$\left[-\frac{1}{2}\nabla^2 + V_{Ne}(\mathbf{r}) + \int\frac{\rho(\mathbf{r}')}{r_{rr'}}\,d\mathbf{r}' + V_{XC}(\mathbf{r})\right]\phi_i(\mathbf{r}) = \epsilon_i\,\phi(\mathbf{r}) \quad (2.51)$$

where $V_{XC}(\mathbf{r}) = \frac{\delta E_{XC}[\rho]}{\delta \rho}$ is the functional derivative of the exchange-correlation energy with respect to the density. If we define the Kohn-Sham operator $\hat{h}_{KS} = -\frac{1}{2}\nabla^2 + V_{Ne}(\mathbf{r}) + \int \frac{\rho(\mathbf{r}')}{r_{rr'}}\, d\mathbf{r}' + V_{XC}(\mathbf{r})$, we obtain an eigenvalue equation similar to (2.16)

$$\hat{h}_{KS}\, \phi_i(\mathbf{r}) = \epsilon_i\, \phi_i(\mathbf{r}) \tag{2.52}$$

where $\phi_i(\mathbf{r})$ are the Kohn-Sham orbitals and ϵ_i are the eigenenergies. For what concern $\phi(\mathbf{r})$ they have no physical meaning strictly speaking, altough some effort to determine a theorem similar to Koopmans', that is valid for HF orbitals, has been made.

Exchange-correlation functionals

The Kohn-Sham equations (2.52) are exacts because no approximations have been taken into account but as we already noticed, the unknown and most desired ingredient of the DFT, is the exchange-correlation functional. It is expressed in the (2.50), as a correction term added to the Hamiltonian in order to include the difference between the classical and quantum electron-electron repulsion and the kinetic energy of the simplified non-interacting system with respect to the real system. An approximate form of this operator have to be introduced, depending somehow on the electron density, e.g. using functional forms which directly depend on the density, on its gradient or Laplacian. The exchange-correlation energy is expressed as

$$E_{XC}[\rho(\mathbf{r})] = \int \rho(\mathbf{r})\, \epsilon_{XC}[\rho(\mathbf{r})]\, d\mathbf{r} \tag{2.53}$$

where we introduced ϵ_{XC} that is the exchange and correlation energy per particle. One of the first simplest forms of E_{XC} introduced was the local density approximation (LDA) which approximates the exchange-correlation energy at each position in space as the energy of a uniform electron gas[6] with the same density in that point [10]. On the contrary on what Kohn and Sham expected, this functional provided a good description for chemical bonds. The exchange and correlation energy per particle may be further split in the exchange and correlation contributions

$$\epsilon_{XC}[\rho(\mathbf{r})] = \epsilon_X[\rho(\mathbf{r})] + \epsilon_C[\rho(\mathbf{r})] \tag{2.54}$$

where ϵ_X for the uniform electron gas was exactly derived in the 1920s by F. Bloch and P. A. M. Dirac namely

$$\epsilon_X^{LDA}[\rho(\mathbf{r})] = -\frac{3}{4}\left[\frac{3\rho(\mathbf{r})}{\pi}\right]^{\frac{1}{3}} \tag{2.55}$$

instead for ϵ_C^{LDA} no analytical expression was found, even for this semplified uniform electron gas but in 1980, D. M. Ceperley and B. J. Alder found the correlation energy with a high accuracy for some systems using quantum Monte-Carlo method, determining the total energy and sub-

[6] The uniform electron gas is also known as Jellium.

tracting the analytical expression for ϵ_X. The assumption of the locality for the electronic density does not take into account the spatial variation of ρ which is far from being constant for the real systems, thus the previous functional may be improved considering also local variations of the density, i.e. its gradient (this operator will be discussed later, see the (6.2))); these functionals are known as generalized gradient approximations (GGA)

$$\epsilon_{XC}^{GGA}[\rho(\mathbf{r})] = \epsilon_{XC}^{LDA}[\rho(\mathbf{r})] + \Delta\epsilon_{XC}\left[\frac{|\nabla\rho(\mathbf{r})|}{\rho^{\frac{4}{3}}(\mathbf{r})}\right] \tag{2.56}$$

which were introduced by Becke [11] and exhibits correct asymptotic behaviour at long range for the energy density. Since the work of Becke many other exchange functionals have been developed, including or not empirical parameters. For what concern the correlation instead, several other functionals have been developed, like the well known LYP [12], thus the complete exchange-correlation functional BLYP was obtained using Becke's exchange and Lee, Yang, Parr correlation GGA functional. A next improvement was made introducing a dependence on the Laplacian (operator expressed as in (6.3)), these family of functionals are identified as meta-GGA (MGGA) but instead of using the density, e.g. $\nabla\rho(\mathbf{r})$ that leads to numerical instabilities, the kinetic energy of the Kohn-Sham orbitals was used, expressed as

$$z(\mathbf{r}) = \frac{1}{2}\sum_{i}^{occupied} |\nabla\phi(\mathbf{r})|^2 \tag{2.57}$$

The last step to get closer to the true functional, was the inclusion of an ingredient coming from the HF method, i.e. the exchange computed using the Kohn-Sham orbitals, derived from the antisymmetry of the wave function. This functionals are named "hybrid " or adiabatic connection method (ACM) functionals. The sum of the exchange and correlation functionals is given by

$$E_{XC} = \underbrace{E_X^{HF}}_{exchange} + \underbrace{z(E_{XC}^{DFT} - E_X^{HF})}_{correlation} \tag{2.58}$$

where we do not know z together with the energy of the exchange-correlation potential applied to the fully interacting wave function, i.e. E_{XC}^{DFT}. The first may be considered as an empirical parameter to be derermined, while the latter may be computed by chosing a DFT functional. If we define $a = 1 - z$, we may rearrange the (2.58) as

$$E_{XC} = (1-a)E_{XC}^{DFT} + aE_X^{HF} \tag{2.59}$$

but we should ask if the inclusion of additional parameters may improve the functional as Becke did. The work of Stevens et al. together with the previous developed LYP (which calculated completely the correlation energy), lead to the birth of an overused functional: B3LYP [13]. This hybrid functional includes also some of the improvements aforementioned, i.e. dependence on

the electron density, on its gradient and and was expressed as

$$E_{XC}^{B3LYP} = E_X^{LDA} + E_C^{LDA} + a(E_X^{HF} - E_X^{LDA}) + b\Delta E_X^B + c(E_C^{GGA} - E_C^{LDA}) \tag{2.60}$$

where the three parameters $a = 0.20$, $b = 0.72$, $c = 0.81$ have already been optimized by Becke using 56 atomization energies, 42 ionization potentials, 8 proton affinities, and 10 total atomic energies of first- and second-row systems [14]. B3LYP performance is quite good and have been used within the context of this thesis work for the optimizations of the systems reported in chapter 4 and 5. We used also another hybrid functional, characterized by a good accuracy, for the systems reported in chapter 3, i.e. the PBE1PBE [15, 16] also called PBE0, because it combines the PBE GGA functional with a predefined amount of exact exchange given as

$$E_{XC}^{PBE0} = \frac{1}{4}E_X^{HF} + \frac{3}{4}E_X^{PBE} + E_C^{PBE} \tag{2.61}$$

2.4 ONIOM multiscale method

One of the main practical problems in computational sciences, even more so computational chemistry, is the time required by the available computer to do its job. It is important to know how much time calculations require in order to plan our work and to reduce worthless delays; it is also necessary to find a good balance between the complexity of the calculations and the accuracy of the results. As already mentioned, the computational cost of the employed method dictate if we have to wait days or months before obtaining the results and each method has a very specific scale complexity (Tab. 2.1).

A practical solution is to partition the system in different regions, each treated using a different level of theory. In this way the reaction center of a protein for example, may be treated using a high accuracy quantum chemistry method, instead the surrounding residues may be included using a lower accuracy method, in order to take also into account their structural and electrostatic contribution. Using a meticulous separation of the system in different regions, very accurate results may be obtained in a fraction of the time required by conventional methods. This kind of methods were introduced by Warshell and Levitt [17], Singh and Kollman [18] and Karplus [19] using a combination of QM/MM higher and lower accuracy regions where the model system is described using an accurate and expensive QM method (that permit to describe bond breaking/formation and electronic states) and the surrounding using a much more efficient MM method. In this framework the total energy of the system is given by

$$E_{\text{QM/MM}} = E_{\text{QM}} + E_{\text{MM}} + E_{\text{QM-MM}} \tag{2.62}$$

where E_{QM} and E_{MM} are the Hamiltonians ralated only to the QM and MM parts respectively, while $E_{\text{QM-MM}}$ is an additive term that has to be evaluated to include the interaction Hamiltonian

between the two parts. Differently from the QM/MM method, any kind of methods and number of regions may be used in principle. A class of these approaches, which we adopt throughout this work, is called ONIOM (*Our own N-layered Integrated molecular Orbital and molecular Mechanics*) or ONIOMn methods (where the n indicates the number of layers/methods that are employed), developed by Morokuma et al. [20]. As opposed to the additive QM/MM method, the ONIOM is an "extrapolative" or "subtractive" method, because for example in a ONIOM2(QM:MM) method the QM energy of the "real" system (i.e. the whole system including all the atoms) is calculated as

$$E_{\text{ONIOM2(QM:MM)}} = E_{\text{QM,model}} + E_{\text{MM,real}} - E_{\text{MM,model}} \tag{2.63}$$

where $E_{\text{QM},model}$ is the QM Hamiltonian referred only to the "model" system, i.e. the higher accuracy layer, $E_{\text{MM},real}$ is the MM Hamiltonian referred to the "real" system, i.e. the whole system including all the atoms and the last term, $E_{\text{MM},model}$, remove the already counted MM contribution included in the previous term. In practice, as it is clear for its name, the system is partitioned in an onion-like structure, where each layer includes all the smaller layers, so that for example, the $E_{\text{MM},real}$ Hamiltonian is referred to the whole system, not to a portion of it like in the previously discussed QM/MM method. Following the same line of reasoning, the energy for a 3-layers ONIOM3(high:medium:low) model system composed by a "model", "intermediate" and "real" systems as represented in Fig. 2.1, may be expressed as

$$E_{\text{ONIOM3(high:medium:low)}} = E_{\text{high,model}} + E_{\text{medium,intermediate}} + E_{\text{low,real}} - E_{\text{medium,model}} - E_{\text{low,intermediate}} \tag{2.64}$$

Using this approach, except the largest layer which requires just one calculation at the lowest accuracy level of theory, for all the other layers two calculations of increasing accuracy are needed to extrapolate the energy of the real system, i.e.

$$E_{\text{ONIOM}n} = \sum_{i=1}^{n} E_{\text{level}(i),\text{model}(n+1-i)} - \sum_{i=2}^{n} E_{\text{level}(i),\text{model}(n+2-i)} \tag{2.65}$$

A schematic representation of the n-layer ONIOM method is depicted in Fig. 2.2. In section 4.1 the use of the ONIOM2(QM:QM) methods and the energetical characterization of the calcium binding to a portion of the connexin 26 channel protein are discussed.

2.5 Molecular dynamics simulations

This methodology allows to study the dynamics of systems composed by hundreds thousand of particles (for example the system reported in section 5.1 is composed by 200K atoms), using Newton's equation of motion. The trajectories of atoms are determined solving numerically the equations of motion given the forces and the potential energy. Interactions between

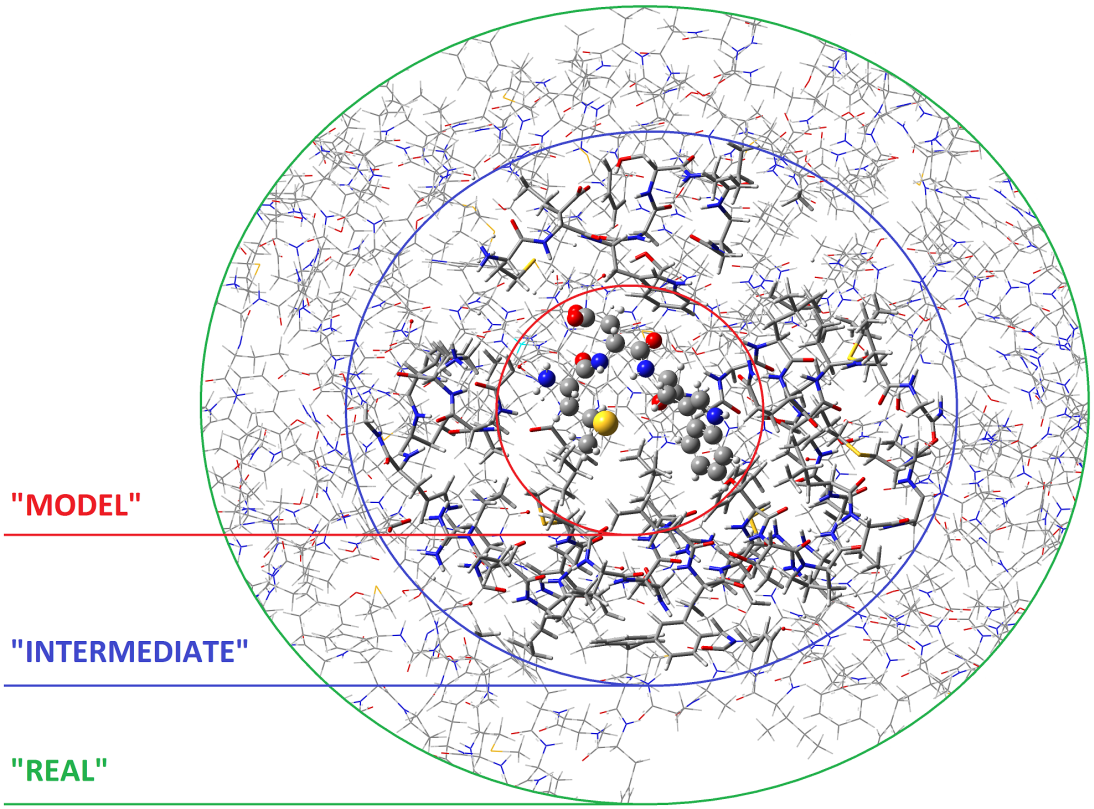


Figure 2.1: Representation of the different layers that constitute a ONIOM3 model system.

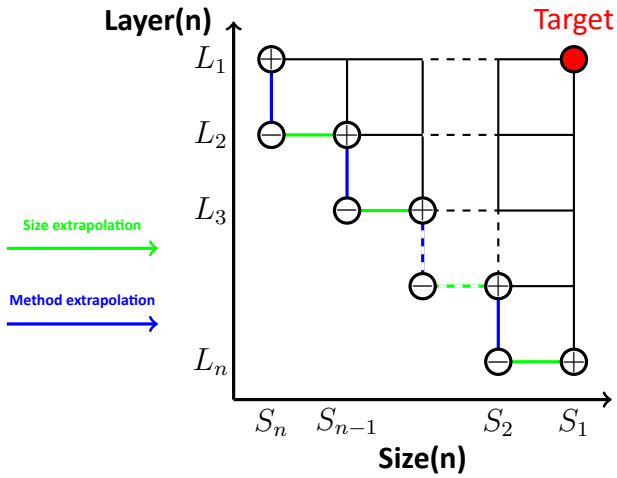


Figure 2.2: Representation of the ONIOM partition for an n-layer system.

non bonded atoms are based on standard potentials, electrostatic (Coulomb) and dispersion-like (e.g. Lennard-Jones potential); bonded-atoms interactions are treated as springs, using harmonic potentials with equilibrium distances and angles. The ensemble of parameters (charges, Lennard-Jones parameters, force constants, equilibrium distances and angles) are grouped into a functional form called *force-field*. Notice that most MD approaches cannot therefore describe bond formation/disruption[7]. A *Molecular Mechanics* study usually defines the energy of the system while a Molecular Dynamics uses this energy to calculate forces and determine how atoms (molecules) move during time.

Force Field Hamiltonians

The MM potential energy functions are defined as the sum of non covalent and covalent interactions, where the non covalent are given by electrostatic and van der Waals, while the covalent part is given by the sum of bonds, angles and dihedrals. For example some well known MM potentials are defined as follows

$$\begin{aligned} V_{\text{AMBER/GROMOS}} = & \sum_{i<j}^{N_{atoms}} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} + 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \\ & + \sum_{i}^{N_{bonds}} b_i(r_i - r_{i.eq})^2 + \sum_{i}^{N_{angles}} a_i(\theta_i - \theta_{i,eq})^2 \\ & + \sum_{i}^{N_{dihedrals}} \sum_{n}^{N_{i,max}} \frac{V_{i,n}}{2}[1 + cos(n\phi_i - \gamma_{i,n})] \end{aligned} \tag{2.66}$$

$$\begin{aligned} V_{\text{CHARMM}} = & \sum_{i<j}^{N_{atoms}} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} + 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \\ & + \sum_{i}^{N_{bonds}} k_b(r_i - r_{i.eq})^2 + \sum_{i}^{N_{angles}} k_\theta(\theta_i - \theta_{i,eq})^2 \\ & + \sum_{Urey-Bradley} k_u(u - u_0)^2 + \sum_{impropers} k(\omega - \omega_0)^2 \\ & + \sum_{\phi,\psi} V_{CMAP} + \sum_{i}^{N_{dihedrals}} \sum_{n}^{N_{i,max}} k_\phi[1 + cos(n\phi_i - \gamma_{i,n})] \end{aligned} \tag{2.67}$$

where the AMBER/GROMOS force field are expressed similarly, while the CHARMM model is slightly different. The AMBER Hamiltonian have been used for the system studied in chapter 5 and for the FF parametrization performed in section 5.2. Beyond the mathematical expression it is also important to identify the precise version of the FF because subsequent improvements in the parametrization bring changes in the values of the parameters, in order to get better de-

[7] Although force flids able to describe at least partially bond formation are starting to appear: e.g. ReaxFF attempts to simulate bond formation/breaking [21].

scription of simulated systems[8]. Once all parmeters, physical constants, and energy expression have been clearly defined, the MD simulation have to deal with all the motion variables, i.e. positions $\mathbf{r}_i$, velocities $\mathbf{v}_i$ and forces $\mathbf{f}_i$ related to each atom particle. In particular, the first are obtained from crystallographic data for proteins or randomly generated for solvents[9] and the latter are calculated from the gradient of the potential energy previously defined. The velocities are randomly generated on the base of the atoms velocity in gas-phase, in particular the Maxwell-Boltzmann equation is usually employed, given as

$$f(v_i) = \sqrt{\frac{m_i}{2\pi kT}}\,\exp\left\{-\frac{m_i v_i^2}{2kT}\right\} \tag{2.68}$$

where k is the Boltzmann's constant and the velocities v_i are generated for $i = 1, 2, ..., 3N$ given the absolute temperature T.

Integration Schemes

As previously mentioned the MD is based on Newton $\mathbf{f} = m\,\mathbf{a}$ equation, thus quantum effects are not treated. So we discretize the position and velocity by the time step Δt using the Taylor expansion viz.

$$\mathbf{r}_i(t+\Delta t) = \mathbf{r}_i(t) + \Delta t\,\dot{\mathbf{r}}_i(t) + \frac{(\Delta t)^2}{2}\,\ddot{\mathbf{r}}_i(t) + \frac{(\Delta t)^3}{3!}\,\dddot{\mathbf{r}}_i(t) + \mathcal{O}[(\Delta t)^4] \tag{2.69}$$

$$\mathbf{v}_i(t+\Delta t) = \mathbf{v}_i(t) + \Delta t\,\dot{\mathbf{v}}_i(t) + \frac{(\Delta t)^2}{2}\,\ddot{\mathbf{v}}_i(t) + \frac{(\Delta t)^3}{3!}\,\dddot{\mathbf{v}}_i(t) + \mathcal{O}[(\Delta t)^4] \tag{2.70}$$

where Newton's dot notation indicate the time derivatives and if substitute the expressions for the velocity and the force, namely $\mathbf{v}_i = d\mathbf{r}_i/dt$ and $\mathbf{f}_i = -\nabla_i V = d\mathbf{v}_i/dt$ in the (2.69) and (2.70) we obtain

$$\mathbf{r}_i(t+\Delta t) = \mathbf{r}_i(t) + \Delta t\,\mathbf{v}_i(t) + \frac{(\Delta t)^2}{2m_i}\,\mathbf{f}_i(t) + \frac{(\Delta t)^3}{3!}\,\dddot{\mathbf{r}}_i(t) + \mathcal{O}[(\Delta t)^4] \tag{2.71}$$

$$\mathbf{v}_i(t+\Delta t) = \mathbf{v}_i(t) + \frac{\Delta t}{m_i}\,\mathbf{f}_i(t) + \frac{(\Delta t)^2}{2}\,\ddot{\mathbf{v}}_i(t) + \frac{(\Delta t)^3}{3!}\,\dddot{\mathbf{v}}_i(t) + \mathcal{O}[(\Delta t)^4] \tag{2.72}$$

Two of the most used algorithms to integrate the motion equations are obtained rearranging the last two expressions, i.e. the **leap-frog Verlet**, given by the following scheme

$$\mathbf{r}_i(t+\Delta t) = \mathbf{r}_i(t) + \Delta t\,\mathbf{v}_i\left(t + \frac{\Delta t}{2}\right) \tag{2.73}$$

[8] It is important to notice that many versions and improvements of each FF expression aforementioned have been developed, so the acronyms AMBER, GROMOS and CHARMM do not define univocally which exact parameters are used.

[9] In particular solvent boxes are equilibrated first, and subsequently are included in the simulations as repeated replicas until the empty space is filled. Possible clashes with previous molecules are avoided removing the new solvent molecules closer than a given distance.

$$\mathbf{v}_i\left(t+\frac{\Delta t}{2}\right)=\mathbf{v}_i\left(t-\frac{\Delta t}{2}\right)+\frac{\Delta t}{m_i}\,\mathbf{f}_i(t) \tag{2.74}$$

named this way because the calculation of position and velocity are not calculated at the same time but jump between steps dephased by $\Delta t/2$. The calculation of the velocity at the integer time step is possible though, by doing an interpolation

$$\mathbf{v}_i(t)=\frac{\mathbf{v}_i(t+\frac{\Delta t}{2})+\mathbf{v}_i(t-\frac{\Delta t}{2})}{2} \tag{2.75}$$

The other common used method is the **velocity Verlet** defined by the scheme

$$\mathbf{r}_i(t+\Delta t)=\mathbf{r}_i(t)+\Delta t\,\mathbf{v}_i(t)+\frac{(\Delta t)^2}{2m_i}\,\mathbf{f}_i(t)+\mathcal{O}[(\Delta t)^3] \tag{2.76}$$

$$\mathbf{v}_i(t+\Delta t)=\mathbf{v}_i(t)+\frac{\Delta t}{2m_i}\,[\mathbf{f}_i(t)+\mathbf{f}_i(t+\Delta t)]+\mathcal{O}[(\Delta t)^3] \tag{2.77}$$

where now velocities and positions are calculated at the same time. Both leap-frog and velocity are second order methods because the error is $\mathcal{O}[(\Delta t)^2]$, moreover these are explicit methods, thus a stability condition must be fulfilled in order to prevent numerical instabilities (for an in-depth overview of implicit/explicit schemes see pag. 89). This is an important limit of this kind of integration, indeed some effort has been done in finding implicit integrators [22] but due probably to their high cost, explicit schemes are still the commonly used nowadays. The algorithm 2 depicts the instructions required in order to produce the trajectories of the atoms.

Algorithm 2 Molecular Dynamics

```
Set initial positions r_i and velocities v_i
while Termination not reached t < t_end do
    procedure PROPAGATE MOTION
        Compute force over each atom f_i = ∇_i V
        Apply thermostat and volume change
        Update r_i and v_i
        Move time forward t = t + Δt
    end procedure
end while
```

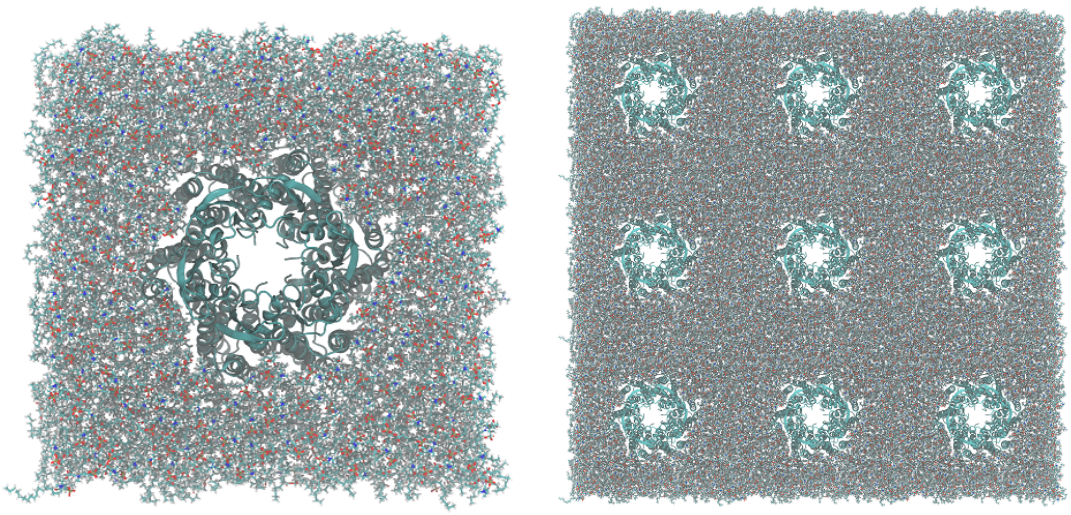
The MD limit is mainly due to time rather than the space, because bigger systems may be simulated increasing the number of working processors, instead the time has a limit due to the numerical instability of the integration scheme employed, which can not be overcome increasing the number of processors. In particular the time step should be small compared to the period of the fastest motion simulated, usually H-involving stretchings, which occur within 10 fs and the time step should be about 0.5 fs. Given that bond stretchings are rather foundamental for the description of macromolecules, they are preferably kept forzen using contraints; the limit of the time step may be partially overcome in this way and, up to 2 fs may be used as integration time step.

Electrostatic interactions and Periodical boundary conditions

Molecular dynamics method is cheaper than the quantum chemistry methods and it scales as $\mathcal{O}(N^2)$ or as $\mathcal{O}(N \log N)$ using particle mesh ewald (PME), instead of the $\mathcal{O}(N^4)$ operations required for HF and DFT with hybrid functionals. Molecular dynamics (MD) is widely used to study principally proteins dynamics in water, domain movements, folding, docking with ligands, mutations but also to model nucleic acids and many others biological macromolecules. More recently a particular type of MD methods were used to study inorganic systems like surfaces, nanoparticles and graphene sheets [23–25]; these methods are known as coarse-grained MD, which allow to do simulations for a much longer time and size of the systems.

To study a protein in water for example we start using a crystallographic structure obtained from the protein data bank [26], we define the box dimensions in order to contain the protein whithin and also leave 10-20 Å of water buffer between the proteic countour and the box side [10]; after that we add the solvent (water usually) using a pre-equilibrated box solvent that is replicated inside the box protein, excluding the region of space that contains the protein atoms. After that some ions are added to eventually neutralize the net charge of the macromolecule and the potential energy of the whole system is minimized moving the atoms in order to remove the bad contacts between the amino acids and the added water molecules. Once this is done a preparative step of the MD need to be addressed in order to add the thermostat and barostat because we want to simulate the macromolecule at physiological conditions. According to the kinetic theory of gases the initial velocities of atoms are generated randomly using a Maxwell-Boltzmann distribution at a starting low temperature. Before using the NPT ensemble (i.e. constant number of particles, pressure and temperature), we need to use the NVT ensemble (constant volume instead of the pressure) to slowly heat the system, in order to prevent spikes in the temperature that may lead to bad phenomena as for example inversion of chiral centers. Once the production temperature is reached the barostat is added and a pre-equilibration dynamics is performed before the real production dynamics. To simulate the finite system in bulk liquid conditions the periodic boundary conditions (PBC) are applied, i.e. the system box is virtually surrounded in each direction by copies of itself so that atoms interacts with atoms in another PBC image and there are not confinement effects. The non-bonded interactions are calculated if the atom pairs are closer than a spherical cut-off that prevents each atom to interact with itself and also permit to reduce the computation due to particle-particle interactions but a suitable box must be used in order to prevent that the protein interact with itself leading to an incorrect description. This cut-off method is inaccurate for the calculation of electrostatic interactions of charged groups and so this kind of semplification is not applicable and lead to the development of the particle mesh ewald (PME), developed by Darden et al. [27], which is a method that split this problem in two parts: short and long range regions. The dependence on the distance may be decomposed in two addends as

[10] As we will see this is necessary in order to prevent that the solute interacts with itself.



(a) Originary system (b) Replicated system along x and y

Figure 2.3: Replication of the system using PBC (replication along z is not shown for semplicity).

$$\frac{1}{r} = \underbrace{\frac{f(r)}{r}}_{\text{short range}} + \underbrace{\frac{1-f(r)}{r}}_{\text{long range}} \tag{2.78}$$

and the potential energy is given by

$$V = \underbrace{V_{dir}}_{\text{direct space}} + \underbrace{V_{rec}}_{\text{Fourier space}} \tag{2.79}$$

where the V_{dir} term is calculated in the direct space, while the second term is calculated using the Fourier space and a discretization procedure. Thanks to this separation the cut-off and its advantages may still be used while keeping a high accuracy. Despite using the PME, the electrostatic interactions are still the most computationally expensive part in each modern MD simulation ($\approx 85\%$ of the total CPU time for the system in section 5.1, using Amber 14, was spent to compute electrostatic interactions).

2.6 Continuum mechanics simulations

Continuum mechanics is a branch of mechanics that describe the behaviour of materials on a macroscopic scale, treated as continuous mass instead of discrete particles, as if they fill all the space region they occupy. This kind of description is very accurate at length-scales above the micrometer and is used to simulate the dynamics of fluids, allowing all sort of applications from weather predictions to the simulation of cardiovascular systems [28], going through structural and dynamics of buildings, bridges et cetera. Direct chemical applications of fluidodynamics coupled with reaction-advection-diffusion are common in chemical engineering. Here we shall limit

ourselves to present an application of these methods for the modeling of transport and reaction processes in order to simulate microfluidic devices. The fluid is treated as a continuum media as well as for the chemical species dissolved. The methodologies will be discussed extensively in chapter 6.

Chapter 3

Molecular systems: a quantum chemistry approach

Thioamides are analogous to amides but their different chemical properties make these molecules intriguing and interesting for many aspects. In this chapter we will discuss the energetics and structural properties of oxopeptides and the changes that occur after thioamidation of the molecules. All the observables will be determined employing quantum chemistry methods, which provides very accurate results and are suitable for these systems constituted by a reasonable number (from 20 up to 136) of atoms.

3.1 Energetics of dipeptides and thiodipeptides

The search for non-invasive, spectroscopic probes to shed light on the mechanism of natural complex systems, such as enzymes or other functional proteins, has brought much attention to the thioamide [–C(=S)–NH–] group. It is a highly attractive probe, being one of the closest mimics of an amide. In addition, it can be inserted into a peptide backbone, rather than into flexible amino acid side-chains, thus reporting precisely any conformational change occurring in a peptide or protein. Due to its interesting physical and chemical properties,[29–31] a thioamide can be investigated using various spectroscopic techniques.[31–35] It is a useful probe in circular dichroism and infrared spectroscopy studies, its absorption being located in spectral regions different from those of the amide.[31–34] It efficiently quenches the fluorescence of suitable chromophores in a distance-dependent manner.[35] This property allows one to follow the conformational changes occurring in a protein while performing its biological functions.[36] The thioamide can also undergo cis/trans isomerism in response to light stimuli.[31, 37] The insertion of a thioamide into a bioactive compound may affect both activity and conformation. While the resistance to enzymatic degradation is usually increased, thus improving the bioavailability of a drug,[38–40] no clear cut information is available on the bias a thioamide imparts to a polypeptide 3D-structure. In general, thioamides accommodate easily into common secondary

structures (β-turn and α-helix).[29] However, old[41, 42] and recent [43] experimental studies suggest that thioamides may promote the uncommon fully-extended conformation (2.0_5-helix). The growing interest in thioamides makes it necessary to understand in depth the conformational changes and the energies involved in the substitution of a peptide bond with a thioamide group. Numerous studies have focused on the former aspects.[44–49] Among these, in a computational study focused on the conformational effect of the S-for-O substitution on dialanines and diglycines[50] it has been described how the conformations change to accommodate the longer hydrogen bonds of thiodipeptides. In addition, since the thiopeptide group is a poorer hydrogen bond acceptor and a better hydrogen bond donor than the normal peptide group, thio-substitution at the amino terminal influences the conformations in an opposite way with respect to thio-substitution at the carboxyl terminal. In contrast, studies on the thermochemical properties of dipeptides are less common. Surprisingly enough, experimental condensation enthalpies for thio-dipeptides are not known. For oxo-dipeptides, energy values have been published for some amino acids, but scarce information is available.[51] The main reason for this gap seems to be the presence of the highly polar amino and carboxylic groups. By decreasing the volatility and favoring thermal instability they make the energy measurement fairly difficult. In this

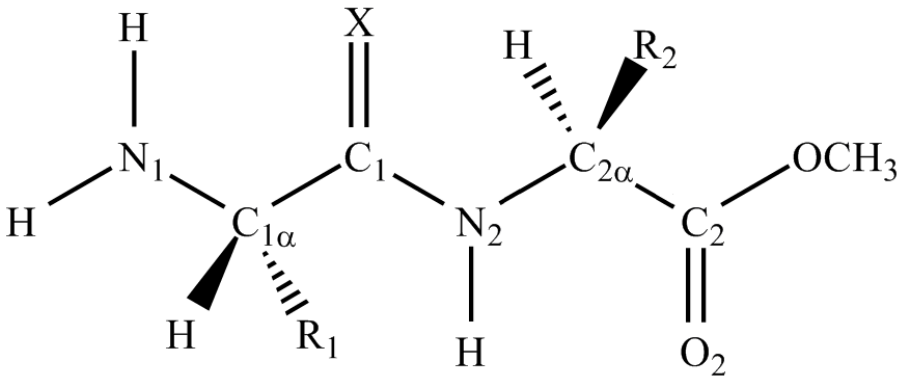


Figure 3.1: X = O_1, S; R_1 (R_2) corresponds to all different amino acid side chains.

contribution, we present a systematic exploration of the energetics of oxo- and thio-dipeptide formation (Fig. 3.1) entirely based on *in silico* methodologies. Our values refer to the gas phase reaction, but complementary information has been obtained by introducing solvent within a continuum approach, as well. The effect of increasing dielectric constant is also discussed. The aim of our systematic analysis is to provide the research community with energy maps obtained by rigorous calculations, thus filling the gap due to the abovedescribed experimental difficulties. The accuracy of the results has been validated against the few experimental enthalpies reported in the literature and accurate calculations. In addition, our overview of the energies involved in the formation of thioamide bonds allows considerations on their thermal stability. Two selected dipeptides, the simplest –Gly-Gly– and the bulkier and aromatic –Phe-Phe–, and their corresponding thionated analogs were synthesized and fully characterized. Their thermal stabilities were studied by thermogravimetric analysis (TGA) and compared. Although the TGA

did not afford a direct value for the bond energies, our measurements highlighted the remarkable differences in the energetics of oxo- and thioamides, thus indirectly confirming the results obtained *in silico*.

3.1.1 Experimental results

Peptide synthesis Boc-L-Phe-L-Phe-OMe and Boc-Gly-Gly-OMe were obtained in solution following standard procedures. The thioamide group was inserted by treating the corresponding carbonyl compound with Lawesson's O/S exchange reagent.[39, 52–54]

Boc-L-Phe-L-Phe-OMe. To a solution of Boc-L-Phe-OH (0.67 g, 2.5 mmol) in anhydrous CH_2Cl_2 at 0 °C, N-ethyl-N'-(3-dimethylaminopropyl)carbodiimide (0.49 g, 2.5 mmol) and 1-hydroxybenzotriazole (0.34 g, 2.5 mmol) were added. After about 5 minutes, HCl·H-L-Phe-OMe (0.5 g, 2.3 mmol) and N-methylmorpholine (NMM, 0.5 ml, 4.6 mmol) were added. The mixture was stirred at room temperature overnight. The solvent was then removed and the residue dissolved in EtOAc. The solution was washed with 5% $KHSO_4$, H_2O, 5% $NaHCO_3$ and H_2O, dried over Na_2SO_4, and evaporated to dryness. The product was recrystallized from EtOAc/petroleum ether (PE). Yield 90%. ^{1}H NMR (400 MHz, $CDCl_3$): δ 7.28–7.18 (m, 8H); 6.99–6.97 (m, 2H); 6.28–6.27 (d, 1H, NH); 4.94 (d, 1H, NH); 4.79–4.77 (q, 1H, α-CH); 4.33-4.32 (m, 1H, α-CH); 3.67 (s, 3H, OMe); 3.06–2.99 (m, 4H, β-CH); 1.40 (s, 9H, Boc). ESI-TOF MS: $[M + H]^+$ calculated for $C_{24}H_{31}N_2O_5^+$ = 427.2155, found 427.2279. HCl·H-L-Phe-L-Phe-OMe. Boc-L-Phe-L-Phe-OMe (0.2 g, 0.47 mmol) was dissolved in a 3 M solution of HCl in MeOH. The reaction mixture was stirred for two hours. The solvent was removed by evaporation and the product recrystallized from Et_2O. Yield: 95%. ESI-TOFMS: $[M + H]^+$ calculated for $C_{19}H_{23}N_2O_3^+$ = 327.1630, found 327.1746.

Boc-L-Phe[Ψ(CSNH)]-L-Phe-OMe. To a solution of Boc-L-Phe-L-Phe-OMe (0.2 g, 0.47 mmol) in anhydrous tetrahydrofuran (THF) kept under N_2 Lawesson's reagent (0.114 g, 0.28 mmol) was added. The reaction mixture was stirred for 1d, then Lawesson's reagent (0.114 g, 0.28 mmol) was added again. The reaction mixture was stirred for another day, then THF was evaporated under N_2 and the crude product purified by flash chromatography (eluent: EtOAc/PE 1 : 5). Yield: 55%. ESITOF MS: $[M + H]^+$ calculated for $C_{24}H_{31}N_2O_4S^+$ = 443.1926, found 443.1949.

HCl·H-L-Phe[Ψ(CSNH)]-L-Phe-OMe. Yield: 85%. ESI-TOF MS: $[M + H]^+$ calculated for $C_{19}H_{23}N_2O_2S^+$ = 343.1402, found 343.1455.

Boc-Gly-Gly-OMe. Yield 91%. ^{1}H NMR (400 MHz, CDCl3): δ 6.77 (t, 1H, NH); 5.29 (t, 1H, NH); 4.06–4.04 (d, 2H, α-CH), 3.84 (m, 2H, α-CH), 3.74 (s, 3H, OMe); 1.44 (s, 9H, Boc). ESI-TOF MS: $[M + H]^+$ calculated for $C_{10}H_{19}N_2O_5{}^+$ = 247.1216, found 247.1340.

H-Gly-Gly-OMe. Yield 93%. ESI-TOF MS: $[M + H]^+$ calculated for $C_5H_{11}N_2O_3^+$ = 147.0691, found 147.0904.

Boc-Gly[Ψ(CSNH)]Gly-OMe. Yield: 65%. ESI-TOF MS: $[M + H]^+$ calculated for $C_{10}H_{19}N_2O_4S^+$ = 263.0987, found 263.1059.

HCl·H-Gly[Ψ(CSNH)]Gly-OMe. Yield: 85%. 1HNMR(400MHz, DMSO-d6): d 8.5 (bs, 3H);

4.40–4.39 (d, 2H); 3.87 (d, 2H); 3.66 (s, 3H, OMe). ESI-TOF MS: $[M + H]^+$ calculated for $C_5H_{11}N_2O_2S^+$ = 163.0463, found 163.0534.

TGA analysis

The thermal stability of the dipeptides was evaluated by TGA, performed using a SDT2960 system from TA Instruments. Samples of about 5 mg were put in alumina pans. The scans were recorded at a heating rate of 10 °C min^{-1} and in the temperature range 30–600 °C, under an air atmosphere. Indium and zinc of high purity were used for calibrating the temperature scale. Standard masses were used to calibrate the TGA mass loss. Three replicates were performed for each sample.

3.1.2 Computational details

All the calculations were performed using Gaussian 09 rev B.01.[1] Conventional labels are used for standard amino acids (AA); a superscript prefix t is added to denote the corresponding thio amino acids (^{t}AA). The dipeptides derived from all the possible combinations of two amino acids (using 20 N-terminus as well as 20 C-terminus structures) were built, i.e. 400 oxo-dipeptides and 400 thio-dipeptides. All the carboxylic groups are methylated in the C-terminus part of all amino acids as well as in the side chain groups of aspartic and glutamic acids. The N-terminus partis in all cases the unmodified amino group (Fig. 3.1).

The initial structures of all dipeptides were prepared using a home-developed tool that links together all C-terminus amino acids to the N-terminus amino acids, starting from separate files containing internal coordinates of each amino acid. Full geometry optimization of all 800 structures was carried out, followed by frequency calculation to confirm the nature of the minima and to obtain thermochemical data. The same calculations were performed for the 40 isolated amino acids and for the water molecule. The functional PBE0 was chosen[16, 55–57] with 6-31G(d,p) basis for all the atoms. Geometry optimizations and frequency calculations were carried out at the same level of theory (PBE0/6-31G(d,p)). The level of theory was validated by comparing optimized structures of $^{(t)}$Gly–Gly with crystallographic data of analogous dipeptides (Table 3.1), which were shown to be in very nice agreement. It is worth noting that the deviation for the C1–S bond (0.03 Å) is substantially larger than those of the other bonds (<0.01 Å). Subsequent single point energy calculations were carried out for all the species at PBE0/6-311G(d,p) level of theory; the zero point and thermochemical corrections (PBE0/6-31G(d,p)), calculated at 298 K and 1 atm, were added to these results to calculate E + ZPC, enthalpies H and Gibbs free energies G, respectively. Solvent effects were accounted for only in single point energy calculations using the Polarizable Continuum Model (PCM).[58] A standard cavity was used, and the dielectric constants of diethylether, acetonitrile and water employed were 4.24, 35.688 and 78.3553, respectively. The inclusion of dispersion and its effects has been tested by running single point calculations at PBE0-D3/6-311G(d,p) and PBE0-D3(BJ)/6-311G(d,p).[59, 60] The limits of the com-

putational method were explored using MP2, MP4 and CCSD, CCSD(T) methods in combination with the cc-pVTZ basis set. We employed these highly accurate but computationally expensive methods only on the Gly–Gly and ^{t}Gly–Gly dipeptides. Energy decomposition analysis[61–63] was carried out on selected oxo/thiodipeptides by homolytically breaking the oxo/thiopeptide bond and generating two radical fragments. According to the scheme employed here, the total binding energy (ΔE) is made up of two major components, ΔE_{prep} and ΔE_{int}:

$$\Delta E = \Delta E_{\text{prep}} + \Delta E_{\text{int}} \tag{3.1}$$

ΔE_{prep} is the amount of energy required to deform the individual (isolated) fragments from their equilibrium structure to the geometry that they acquire in the dipeptide. The interaction energy, ΔE_{int}, corresponds to the actual energy change when these geometrically deformed fragments are combined to form the overall dipeptide.

	Gly-Gly[a]	^{t}Gly-Gly[b]
C_1–X[c]	1.22 (1.211)	1.66 (1.630)
$C_{1\alpha}$–C_1	1.52 (1.524)	1.52 (1.523)
$C_{2\alpha}$–C_2	1.51 (1.508)	1.51 (1.508)
C_1–N_2	1.35 (1.352)	1.33 (1.326)
$C_{2\alpha}$–N_2	1.43 (1.434)	1.43 (1.431)
C_2–O_2	1.21 (1.195)	1.21 (1.209)

Table 3.1: [a] Crystallographic data are given in parentheses.[64] [b] Crystallographic data are given in parentheses.[65] [c] X = O_1, S.

3.1.3 Standard enthalpies

The oxo-dipeptides and thio-dipeptides sketched in Fig. 3.1 have been built as described in the computational methods and full geometry optimizations followed by frequency calculations at PBE0/6-31G(d,p) were systematically performed. The individual amino acids, thio-amino acids and the water molecule were also optimized at the same level of theory with the aim of calculating the reaction energy of this condensation:

$$^{(t)}AA + AA \rightarrow {}^{(t)}AA\text{–}AA + H_2O \tag{3.2}$$

where $^{(t)}$AA indicates one of the twenty amino (thio-amino) acids. The reaction energy is defined as:

$$\Delta_r E = E(^{(t)}AA\text{–}AA) + E(H_2O) - E(^{(t)}AA) + E(AA) \tag{3.3}$$

where E represents the electronic energy.[1] By comparing the complete map of $\Delta_r E$ calculated at PBE0/6-31G(d,p) and PBE0/6- 311G(d,p) levels (Fig. 3.2A and 3.2B and Fig. B.1 in the appendix B), no important difference is revealed (less than 0.3 Kcal/mol) when using the triple quality basis set, and this further supports our choice of using the double quality basis set to run the full geometry optimizations and frequency calculations. All the thermal corrections have been applied to the single point energies obtained using the triple quality basis set. To evaluate the accuracy of the calculated energetics maps, we have run highly accurate *ab initio* calculations on the dimers Gly-Gly and ^{t}Gly-Gly. The results are shown in Table 3.2. Taking as reference the $\Delta_r H$ calculated at CCSD(T)/cc-pVTZ, one can notice that HF and MP2 systematically underestimate and overestimate the energy value, respectively. For Gly–Gly, PBE0/6-311G(d,p) results fall approximately 2 Kcal/mol below the CCSD(T) enthalpy, which is in excellent agreement with the experimental value. Also for ^{t}Gly–Gly the agreement with the CCSD(T) value is very good, the discrepancy being only 1 Kcal/mol. Thus, 1-2 Kcal/mol can be considered the uncertainty of our calculatedmaps, a value which is comparable to the chemical accuracy. The inclusion of dispersion effects (PBE0-D3/6-311G(d,p) and PBE0-D3(BJ)/6-311G(d,p), 3.2) generally lowers $\Delta_r E$ and particularly in the case of D3(BJ) correction on ^{t}Gly-Gly leads to excellent agreement with the CCSD(T) value . The important difference emerging from the maps of Fig. 3.2 is the excess of blue-cyan in the case of thio-dipeptides contrasting the orange-green aspect in the case of oxo-dipeptides. This reflects neatly negative $\Delta_r E$ in the former and slightly negative or positive $\Delta_r E$ in the latter case. In Fig. 3.2C and 3.2D the maps of $\Delta_r H$, i.e. including the enthalpic correction, are shown. If compared to the maps of Fig. 3.2A and 3.2B, the effect of the corrections is a fading of the map; the same effect is evident in Fig. B.2 and B.3 (appendix B) where maps of $\Delta_r E$ + ZPC, i.e. including the zero point correction, and of $\Delta_r G$, i.e. including Gibbs energy correction, are shown. Condensation enthalpies of oxopeptides are approximately in the range from -12 to 3 Kcal/mol while enthalpies of thiopeptides are in the range from -18 to about -1.5 Kcal/mol, ($\sim$4.5–6 Kcal/mol lower); importantly, energy patterns are very similar.

[1] The same equation is used to compute reaction enthalpies and Gibbs reaction free energies using thermal corrections at 1 atm and 298 K for H and G to replace E, respectively.

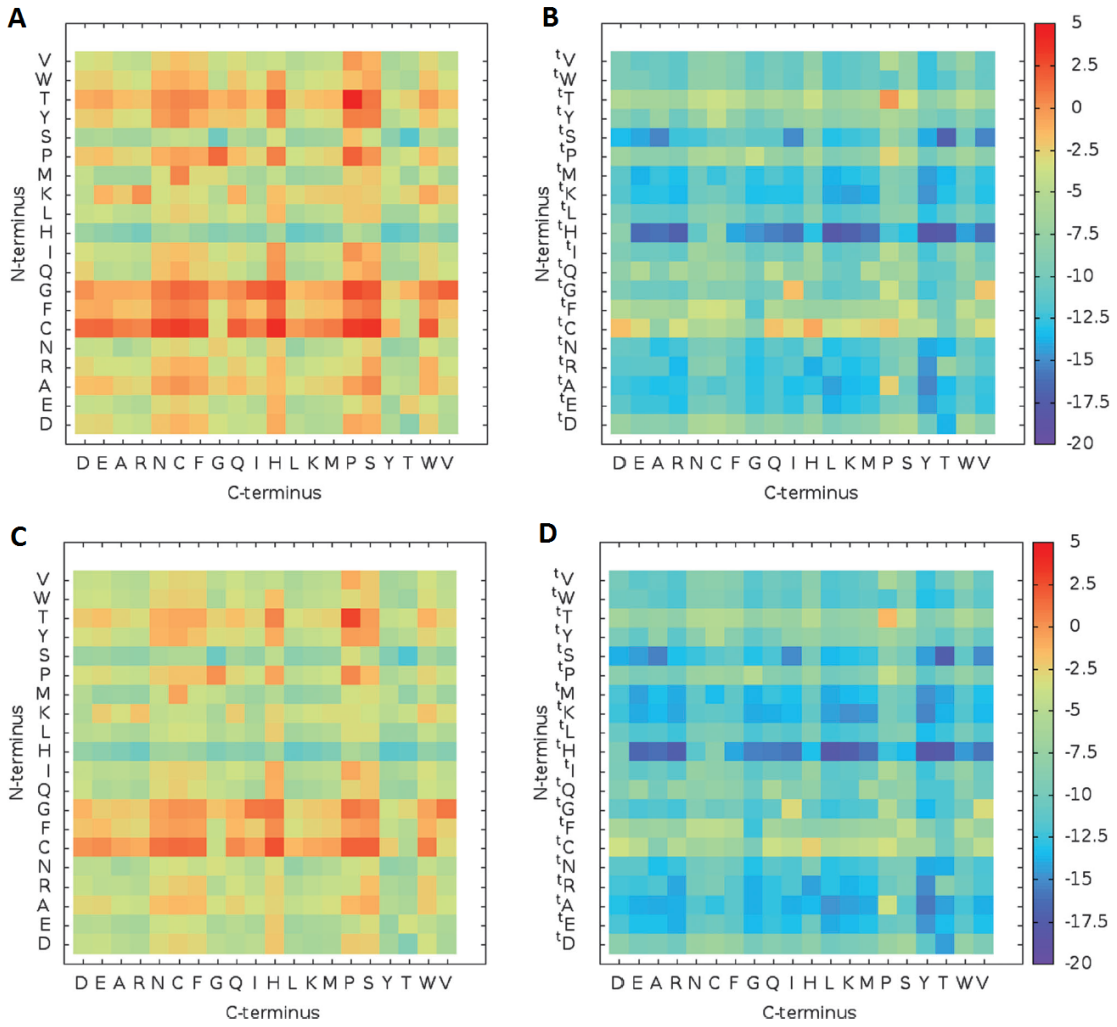


Figure 3.2: Maps of $\Delta_r E$ (kcal/mol) calculated at the PBE0/6-31G(d,p) level of theory (a and b) and $\Delta_r H$ (kcal/mol) calculated at the PBE0/6-311G(d,p) level of theory (c and d) for oxo- and thio-dipeptides (left and right, respectively).

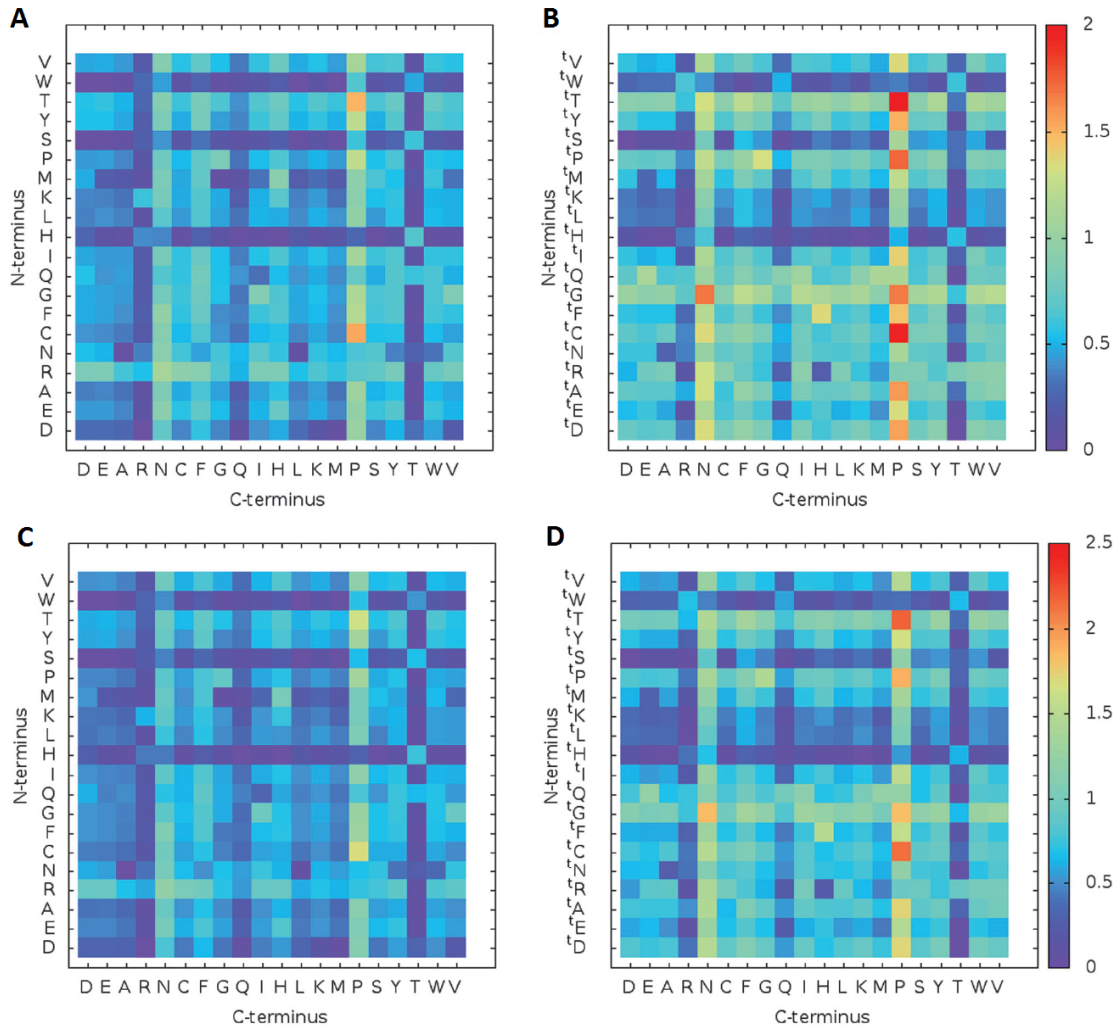


Figure 3.3: Maps $\Delta\Delta_r E$ calculated at PCM-PBE0/6-311G(d,p) for oxo- and thio-dipeptides (left and right, respectively); values in a solvent are in acetonitrile vs. diethylether (a and b) and in water vs. diethylether (c and d).

Method	Gly-Gly	^{t}Gly-Gly
HF	-1.16	-9.91
PBE0/6-311G(d,p)	-1.92	-12.15
PBE0-D3/6-311G(d,p)	-1.80	-12.68
PBE0-D3(BJ)/6-311G(d,p)	-2.14	-13.13
MP2	-5.35	-15.09
MP4SDQ	-3.66	-12.55
CCSD	-3.78	-12.77
CCSD(T)	-4.19	-13.13
Experimental	-4.42$\pm$2.47	—

Table 3.2: *Ab initio* calculated condensation enthalpies (Kcal/mol) of Gly–Gly and ^{t}Gly–Gly; the basis set is cc-pVTZ for all atoms. For comparison PBE0/6-311G(d,p) and available experimental values are shown.

Among the thio-dipeptides $\Delta_r H$ never have positive values and the smallest negative values are those of ^{t}Thr–Pro and ^{t}Cys–His, i.e. -1.5 and -2.5 Kcal/mol respectively; in contrast ^{t}His–Tyr and ^{t}His-Leu have the most negative $\Delta_r H$, i.e. -18.1 and -17.9 Kcal/mol respectively. Among oxo-dipeptides, Thr–Pro and Cys–His have the most positive $\Delta_r H$, i.e. 2.9 and 2.4 Kcal/mol respectively, while the most negative $\Delta_r H$ are calculated for Ser–Thr and His–Tyr, i.e. $^{-1}$2.0 and $^{-1}$1.3 Kcal/mol respectively. The effect of the environment on the energetics of formation of all these dipeptides was taken into account while performing single point calculations employing a continuum solvation model, i.e. at the PCM-PBE0/6-311G(d,p) level. Three solvents were considered, whose dielectric constant span a large range, i.e. diethylether, (ϵ = 4.24), acetonitrile (ϵ = 35.688) and water (ϵ = 78.3553).[58] In order to emphasize the solvent effect, we calculated the module of the difference between energies of diethylether–acetonitrile and water–diethylether (Fig. 3.3).[2] From the patterns, it can be noticed that trends for oxoand thiopeptides are very similar but the difference between diethylether and acetonitrile is one order of magnitude higher with respect to the difference between acetonitrile and water.

		Calculated			
	Experimental	PBE0/6-311G(d,p)	PBE0-D3/6-311G(d,p)	PBE0-D3(BJ)/6-311G(d,p)	CCSD(T)/cc-pVTZ
Gly-Gly	-4.42$\pm$2.47	-1.92	-1.80	-2.14	-4.19
DL-Ala-Gly	-9.80$\pm$7.39	-3.49	-3.72	-4.07	—
DL-Ala-DL-Val	-12.95$\pm$4.6	-3.57	-4.76	-5.06	—

Table 3.3: Experimental standard enthalpies (Kcal/mol) for the condensation reaction in the gas phase[51, 66–70, 238] and calculated values from this work.

[2] We neglected possible variations in the molecules due to geometry relaxation or proton transfer that may occur in water.

From Fig. 3.3 it emerges that for oxopeptides (thiopeptides) relevant differences due to the change of solvent are found in particular for $^{(t)}$Thr–Pro and $^{(t)}$Cys–Pro and in general the variation is higher for thio-peptides, especially for those including Thr and Gly (N-terminus) and Asn and Pro (C-terminus). As discussed in the Introduction, experimental enthalpies of formation of both oxodipeptides and thiodipeptides are scarce in the literature. The few data available for the former systems are reported in Table 3.3, together with our corresponding computed values. This comparison involves only few dipeptides, i.e. Gly–Gly, Ala–Gly and Ala–Val, mainly because there is a lack of experimental thermodynamic values for the solid–vapor phase transition for amino acids and peptides.[51] In addition, with the exception of Gly–Gly, the data refer to diastereomer mixtures (Table 3.3). The experimental condensation enthalpies in the gas phase were estimated using the 3.3 replacing E with the standard formation enthalpy $\Delta_f H^{\ominus}(g)$ in the gas phase, and propagating reported errors. When $\Delta_f H^{\ominus}(g)$ values were not available, enthalpies were computed using the following expression

$$\Delta_f H^{\ominus}(g) = \Delta_f H^{\ominus}(s) + \Delta_{sub} H^{\ominus} \tag{3.4}$$

where $\Delta_f H^{\ominus}(s)$ is the standard solid phase formation enthalpy and $\Delta_{sub} H^{\ominus}$ is the standard sublimation enthalpy; for the chiral amino acids we chose to use the data of the L enantiomer. The calculated values shown in Table 3.3 are in nice agreement with experimental results, also taking into account the large error affecting them. Finally, in order to get some insight into the effect of thionation, we carried out an energy decomposition analysis on selected oxo- and thiodipeptides; they were chosen among the couples spanning a range of significant, intermediate or negligible $\Delta\Delta_r E$ (Table 3.4). We considered each dipeptide as composed of two radical fragments obtained by breaking homolytically the C(X)–NH bond (X = O, S). We started our analysis from the model systems N-methyl-acetamide (MAc) and N-methyl-thioacetamide (^{t}MAc), whose fragments have very small conformational freedom if compared to the AA radicals forming the dipeptides. The presence of sulfur rather than oxygen implies a more negative ΔE_{int} and larger ΔE_{prep}, resulting in a more negative ΔE (Table 3.4).

3.1.4 Fragment analysis

In order to compare analogous oxo- and thio-dipeptides it is useful to consider the $\Delta\Delta E$ values, reported in the last three columns of Table 3.4. For the couple Phe–Phe/$^{(t)}$Phe–Phe, $\Delta\Delta E_{\text{int/prep}}$ values, and consequently $\Delta\Delta E$ values, are very similar to those computed for the 'rigid' model couple MAc/$^{(t)}$MAc; in fact Phe–Phe and $^{(t)}$Phe–Phe nicely overlap (Fig. 3.4). Larger structural differences between analogous oxo- and thio-dipeptides, due to conformational rearrangements upon dipeptide formation, are associated with smaller $\Delta\Delta E_{\text{prep}}$. In all cases however the trend is maintained, i.e. thiodipeptides have more negative ΔE_{int} and larger ΔE_{prep}, resulting in more negative ΔE than the corresponding oxodipeptides. $\Delta\Delta E_{\text{int}}$ values span a range of 1.3 Kcal/mol and $\Delta\Delta E_{\text{prep}}$ values a larger range of 4.5 Kcal/mol.

	ΔE_{int}	ΔE_{prep}	ΔE	$\Delta\Delta E_{int}$	$\Delta\Delta E_{prep}$	$\Delta\Delta E$
CysVal	-112.9	11.5	-101.4	-4.8	3.0	-1.8
^{t}CysVal	-117.7	14.5	-103.2			
PhePhe	-110.6	9.8	-100.8	-6.1	5.5	-0.6
^{t}PhePhe	-116.7	15.3	-101.4			
LysArg	-109.1	10.1	-99.0	-5.4	1.0	-4.4
^{t}LysArg	-114.5	11.1	-103.4			
MAc	-106.3	5.4	-100.9	-7.8	5.2	-2.6
^{t}MAc	-114.1	10.6	-103.5			

Table 3.4: Energy decomposition analysis of selected oxo- (thio)dipeptides (PBE0/6-311G(d,p)); all values are in Kcal/mol.

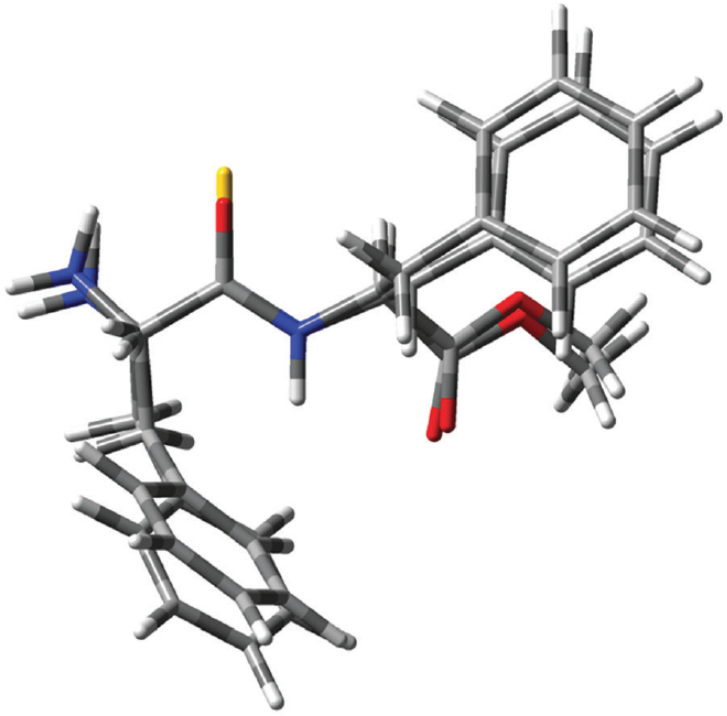


Figure 3.4: Superimposed views of PhePhe and tPhePhe structures optimized at PBE0/6-31G(d,p) level of theory.

The two hydrochloride, thionated, dipeptide esters, HCl·H-Gly[Ψ(CSNH)]Gly-OMe and HCl·H-Phe[Ψ(CSNH)]Phe-OMe, and their parent oxo-dipeptides underwent TGA. Both in the case of Gly and of the bulky and hydrophobic Phe amino acid, the oxo-dipeptides display remarkably diverging behaviors from those of their thionated analogs. As an example, the TGA patterns of the Phe dipeptides are reported in Fig. 3.5. The weight losses at about 200 1C (calculated 18.9% for the oxo- and 18.1% for the thionated dipeptide) correspond to the release of HCl and CH3OH upon-cyclization of the dipeptides to diketopiperazines (chemical structures in Fig. 4). Indeed, the formation of cyclic dipeptides is a well-known behavior of dipeptide methyl esters.[71] However,

when the temperature is raised above 200 °C, the oxo-diketopiperazine is almost completely degraded at 340 °C, whereas the monothionated diketopiperazine requires a temperature of about 500 °C. A similar behavior is observed also in the case of the Gly dipeptides. The higher thermal stability of the thionated compounds nicely fits with the results of our computational analysis: indeed the presence of the thioamide enhances the stability of the dipeptide, although the initial cyclization process prevents the TGA of the original compounds. Conclusions We have presented maps of computed energies *in vacuo* as well as in solvent for the formation of 800 oxo-dipeptides and thiodipeptides built via condensation between couples of amino acids and amino thioacids, the latter having the carbonyl oxygen replaced by an sp^2 sulfur. These maps highlight optimal partners leading to stable dipeptides and critical situations for which a lower stability or instability is predicted in terms of Gibbs reaction free energies. Energy decomposition analysis carried out on selected oxo- and thio-dipeptides reveals that the replacement of O with S is accompanied by a larger increase of the interaction than the deformation contribution, which combines to give an overall more negative ΔE. The synthesis of four oxo- and thio-dipeptides and their TGA are also presented. While the condensation reaction *in silico* is designed starting from two oxo-amino acids/one oxo- and one thio-amino acid and gives the oxo-dipeptide/thio-dipeptide products, respectively, the synthetic route is significantly different and in particular thio-dipeptides are prepared from oxo-dipeptides treated with Lawesson's O/S exchange reagent. To the best of our knowledge no experimental condensation enthalpies of thiodipeptides have been reported, and, surprisingly, scarce data are available for oxodipeptides, mainly due to the presence of the highly polar amino and carboxylic groups which decrease the volatility and favor thermal instability. TG analysis, which could provide information in this sense, shows that cyclization of oxo- and thio-dipeptides initially occurs, and thus reveals, upon further heating, only the higher thermalstability of thionated species when compared to oxo-species. Nevertheless, this finding is in agreement with our computational results, which support increased stability in the presence of a thioamide group, and, validated against accurate *ab initio* calculations and the few experimental data available, offers an exhaustive overview of the energetics of these dipeptides.

3.2 Flexibility and stability of secondary structure in oxo-/thio-dipeptides

After the previous work on oxo-/thio-peptides based essentially on an energetical characterization of all possible couples of AA we decided to explore the conformation of these molecules and the stability of the secondary structure [1]. This is still an ongoing work, thus some preliminary results will be shown in this section.

The rotation about the Φ/Ψ angles (Fig. 3.6) was investigated using three different molecules, i.e. BOC-Gly_2-OMe, the same molecule with thionation of amide group X_2 atom and the

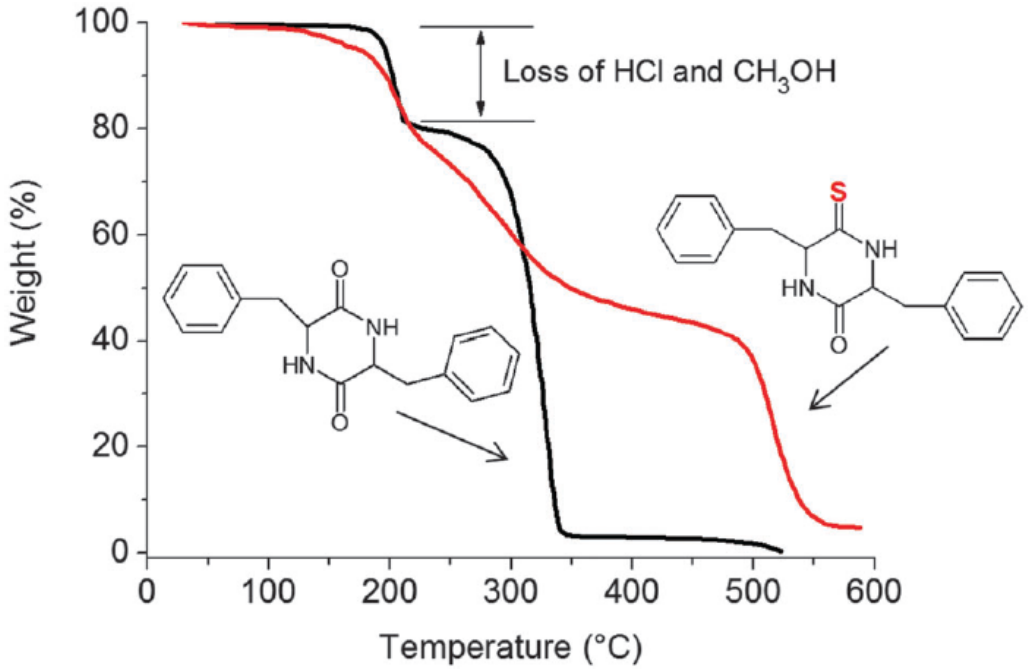


Figure 3.5: TGA patterns of the PhePhe (black) and ^{t}PhePhe (red) dipeptides. The molecular structures of the cyclic dipeptides formed upon heating are also shown.

same molecule with both X_1 and X_2 replaced by sulfur atom. The substitution of urethane X_1 atom, was done to mimic the presence of a previous thioamine group and its influence on the following Cα Φ/Ψ angles along the peptide chain; this approach was also employed in order to keep the same number of atoms. All the calculations have been performed using Gaussian 09 rev B.01 [1]. Since these calculations are very time consuming, we used the PBE0/6-31G(d,p) level of theory. Each rotation along the dihedral was performed 35 times by 10°, starting from the optimized structure and obtaining a grid composed by 1296 total points for each graph. The energy values related to unconverged calculations have been determined using bilinear interpolation when 4 adjacent points were available, otherwise using linear interpolation between 2 points instead. Afterwards we studied energetic differences and conformation for 40 oligopeptides, including glycine and alanine, with different chain length going from BOC-AA_2-OMe to BOC-AA_6-OMe; the calculations were done using 20 oxo- and 20 full-thioamidated molecules with two different conformations. Each molecule was prepared in right handed α-helices (α_R) and β-sheets (β) conformation by imposing as starting angles Φ=-60°,Ψ=-50° and Φ=-80°,Ψ=+150° respectively. Afterwards these geometries were optimized at the PBE0/6-311G(d,p) level. In order to treat carefully the dihedral angles, which represented a critical degree of freedom leading to different structures as well as to mixtures of α_R and β conformations, we included few constraints in the optimization. In particular, for each molecule we performed $2N + 1$ subsequent optimizations, where N stands for the number of constituent AA, by fixing all the Φ/Ψ couples and removing the constraints at each step, until a full optimization was done. By doing so the first optimization was done by fixing all the Φ/Ψ angles, the second relaxing also Φ_1, the third relaxing Φ_1/Ψ_1, the

fourth relaxing $\Phi_1/\Psi_1/\Phi_2$ and so on for the remaining dihedrals, until all the constraints were removed. Using this approach, we were also able to optimize the oligopeptide while keeping its initial conformation, crucial for the structural and energetic properties of these systems and to avoid mixed conformations within the same molecule. For the thioamidated-analogs the same optimization approach has been adopted but no modification of the urethane has been done; only the [Ψ(CSNH)] group have been introduced. In order to be sure that optimized geometries were not transition states, all the final structures were checked using frequency calculation.

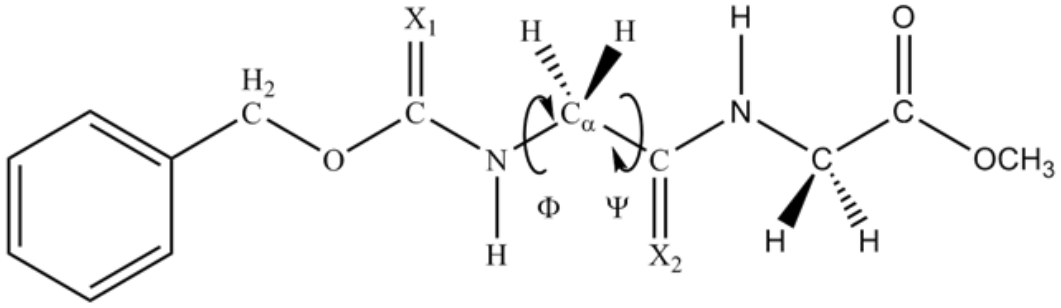


Figure 3.6: X_1 (X_2) = O, S.

The PES reported in Fig. 3.7 are symmetric because differently from the other amino acids, glycine has not a bulky side chain, moreover C_α has not this symmetry break. The rotation of Φ/Ψ in one direction or in the opposite for this amino acid, gives the same interactions and all the considerations referred to right handed α-helix herein studied are valid for the left-handed form. The PES is strictly related to Ramachandran plots, i.e. the lowest energy zones are the most frequently explored by glycine in proteins backbone. In particular the Φ=0, Ψ=$\pm$180° and Φ=Ψ=0° regions, are dominated by the O–O and O–H interactions respectively, between two subsequent amides. The thionation of the amide group (Fig. 3.7b) and of both the amide and urethane groups (Fig. 3.7c) raise up the maximum energy by about 6 Kcal/mol for each additional sulfur substitution, expecially in the zone Φ=0, Ψ=$\pm$180 because the atom dimension and steric hindrance increases going from oxygen to sulfur. The same happens in the region Φ=Ψ=0°, after the thionation of the urethane oxygen, that interacts with the hydrogen of the subsequent amide. The lowest energy region free from these bad contacts, is ruled by the electronic effects of the backbone, and is where α_R, β and fully extended (FE) conformations are located. Thionation of the amide produce smaller changes in this region but not less important, governing the interconversion between the aforementioned conformations. For the oxo-dipeptide there is a small barrier going from the α_R to the β regions, identified by the level set equal to 2 Kcal/mol, however both these zones are similarly stable (Fig. 3.7a). In the case of the thio-dipeptide this small barrier is absent, i.e. only two regions are found, identified by level set equal to 2 Kcal/mol and the highest stabilization is in the region of α-helices (Fig. 3.7b). Finally for the compound with amide/urethane thionated groups, similar results were found, i.e. the barrier is absent but a stability region is also found for Φ=Ψ=$\pm$180°, corresponding to the FE conformation. The absence of this barrier may be important for the α_R/β interconversion dynamics that these mo-

lecules can undergo in solution and it appears that thiopeptides are likely more stable in the FE form (Φ=Ψ=$\pm$180°) with respect to oxo-peptides.

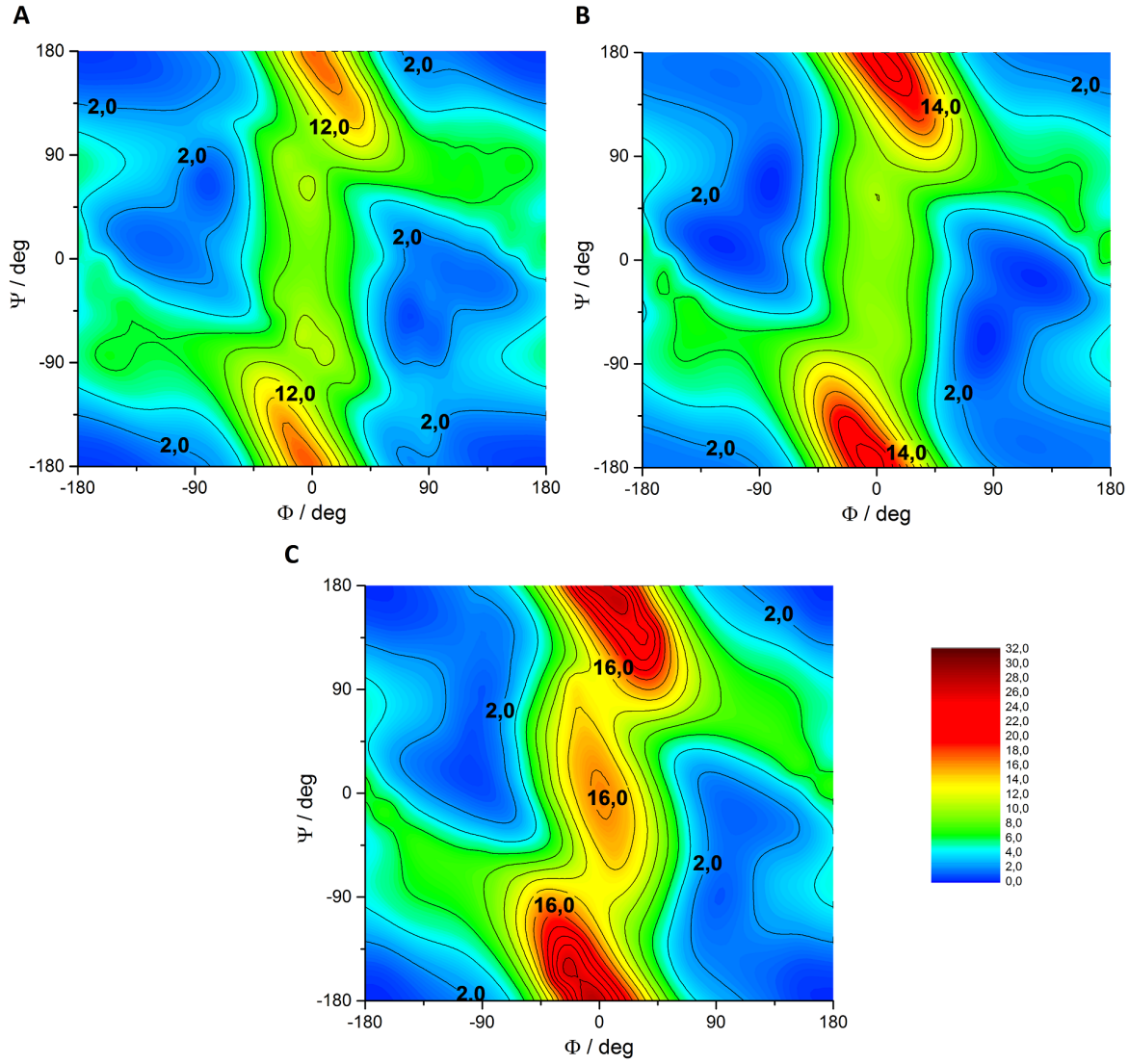


Figure 3.7: Potential energy surface expressed in Kcal/mol for the rotation of Φ/Ψ angles for oxo (A), thio (B) and thiothio (C) derivatives of BOC-Gly2-OMe.

After the first investigation, we calculated the destabilization energy of oxo and thionated oligomers constituted by 2 up to 6 glycine with respect to the most stable conformation (Fig. 3.8). All the molecules optimized starting from the α_R conformation kept the same secondary structure, with angles close to Φ=-60°,Ψ=-50° values. Instead, as expected, all the molecules with initial β-sheet conformation, after full optimization reached the FE form (both oxo- and thio-derivatives). For both oxo- and thio-oligopeptides ΔE for molecules in α_R conformation, going from 2 to 6 glycines lowers by approximately 2 Kcal/mol, instead the opposite bevahiour is found for molecules in the FE conformation. It should be noticed that the N–H–O enthalpy in vapor phase is of about 1.9 Kcal/mol, so quite similar to the energies involved in this analysis. It is interesting to notice that at the intersection point, i.e. for the tetrapeptides, thio-derivative

have a different stabilization. In particular, FE conformation is more stable than α_R, instead for the oxo-derivatives the opposite trend holds. This may be explained considering that the two major contributions to internal energy are given by Φ/Ψ destabilization energy and by the stabilization contribution due to the formation of intermolecular N–H–O hydrogen bonds. For the oxo-compounds, the first energy quantity is lower and exceeded by the stronger hydrogen bond, while for the thio-compounds the torsional strain energy is higher than the gain due to the weaker N–H–S hydrogen bond. The fact that all molecules in β-sheet conformation reached a FE secondary structure suggest that β-sheet conformation is preferred when aggregation between different peptides or within the same chain of much longer peptides occur. In this latter case the strain energy spent to assume angles the Φ=-80°,Ψ=+150° is largely compensated by the formation of hydrogen bonds. For the oligopeptides with length employed in this study, we are far from obtaining β-sheet structure with a single peptide chain for this reason.

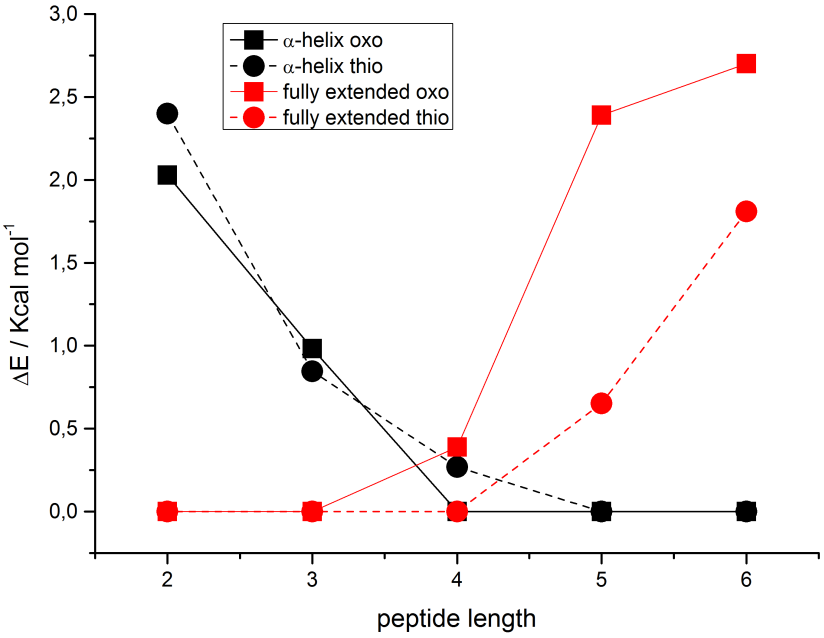


Figure 3.8: ΔE of oligoglycine calculated with respect to the most stable compound with the same chain length.

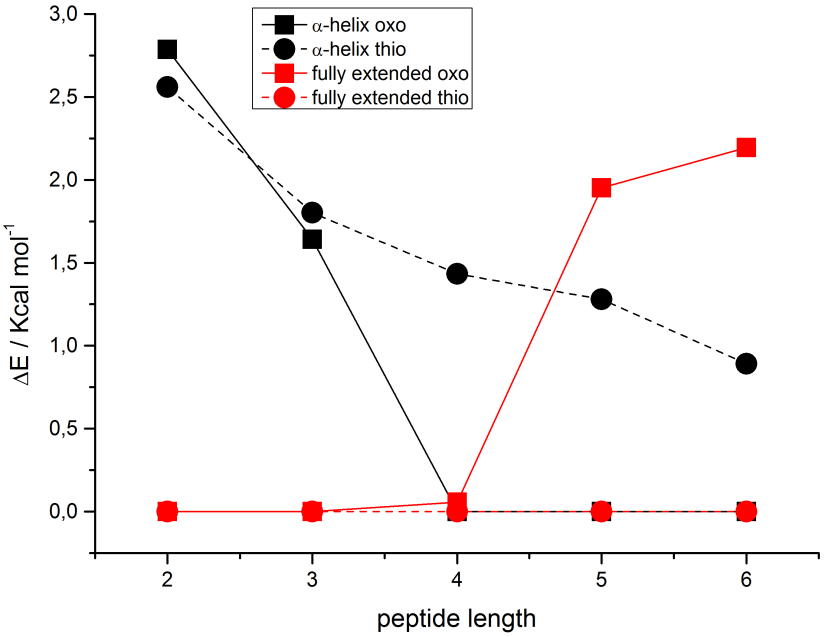


Figure 3.9: ΔE of oligoalanine calculated with respect to the most stable compound with the same chain length.

Chapter 4

Multi-layer quantum chemistry

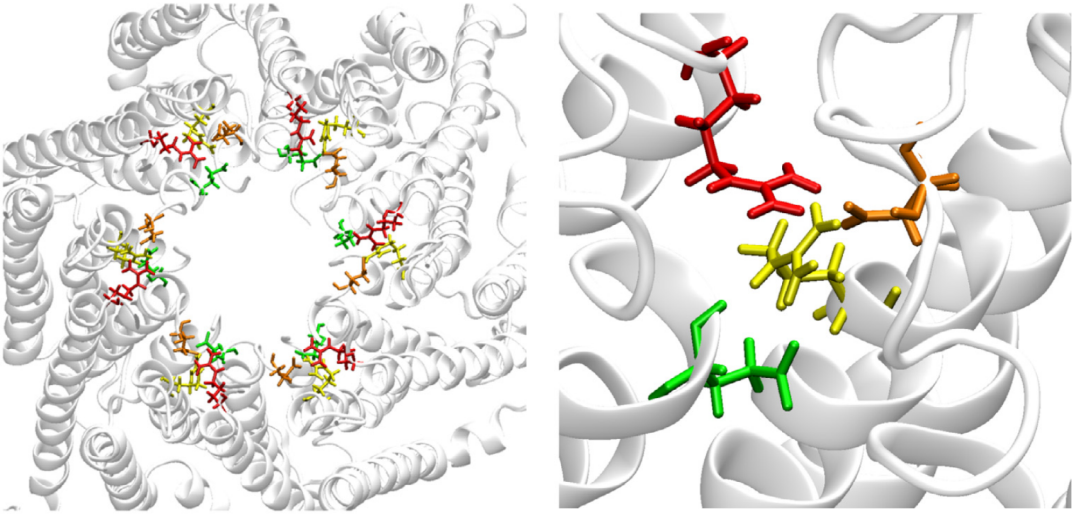
In order to study reaction sites or portions of proteins, the quantum chemistry approach become too expensive, thus a lighter method should be used. For this purpose it is convenient to partition the system by treating the atoms which belong to the important region with the highest accuracy. The other atoms which belong to the surrounding region, that does not take part directly to the reaction but are still important for the structural motif of the system, may be included using a less expensive/accurate method. This subdivision in principle, may be executed several times, altough 3 layers are sufficient for this purpose. In the following chapter we show a preliminary characterization of a portion of the connexin 26 protein, and the calcium binding study using the ONIOM multilayer method.

4.1 Energetics and calcium binding of a portion of connexin 26

Connexins, indicated with the abbreviation Cx followed by their molecular weight in kilodalton (e.g. Cx26 for the 26 kDa protomer), are tetraspan transmembrane proteins encoded by at least 21 genes in the human and 20 genes in the mouse genome, 19 of which can be grouped as sequence-orthologous pairs [72]. Connexin proteins oligomerize in the ER/Golgi or trans-Golgi network forming hexameric assemblies known as hemichannels or connexons, which are delivered to the plasma membrane by vesicular carriers travelling along microtubules [72–75]. Plasma membrane hemichannels form aqueous conduits with a pore diameter of $\approx$1.5 Å and open in response to various types of stimuli, including mechanical, shear, ionic and ischemic stress [74, 75]. Open hemichannels provide a pathway for the release from cells of ATP, glutamate, NAD^+ and prostaglandin E2, which act as paracrine messengers [74, 75]. Intercellular (gap junction) channels are formed by the head-to-head docking of two connexons in adjacent cells and mediate the exchange of a variety of cytoplasmic molecules (virtually all soluble second messengers, amino acids, nucleotides, ions, glucose and its metabolites) [73]. Connexin mutations that alter either hemichannel or gap junction function have been implicated in a variety of human diseases [76]. It is well established that hemichannel activity is inhibited at normal

($\approx$1 mM) $[Ca^{2+}]_e$ and this may serve as a protecting mechanism against the harmful effects of leaky hemichannels [74, 75, 77–79]. Lowering $[Ca^{2+}]_e$ to micromolar levels fosters hemichannel opening and promotes signaling [80]. Atomic force microscopy imaging shows significant and reversible changes of pore diameter at the extracellular mouth of Cx26 hemichannels exposed to different $[Ca^{2+}]_e$ [81]. Also the voltage gating of connexin hemichannels is affected by $[Ca^{2+}]_e$. Lowering $[Ca^{2+}]_e$ to zero causes a marked increase in the amplitude of hemichannel currents, shifts their activation to more negative potentials and alters the kinetics of activation and deactivation, whereas increasing $[Ca^{2+}]_e$ has opposite effects (reviewed in Ref. [82]). Molecular dynamics (MD) simulations [83] show that multiple Ca^{2+} ions may linger within the negatively charged extracellular mouth of Cx26 hemichannels at a membrane potential of -80 mV, and low-affinity binding of several Ca^{2+} ions near the point of narrowest pore constriction can occlude the pore. Upon depolarization to 0 mV the interactions weaken and Ca^{2+} ions shift towards the extracellular space [83]. Asp46 and Glu47, two highly conserved negatively charged amino acids facing the pore in the extracellular mouth, are strong candidates for Ca^{2+} binding [84]. Glu47 is also a candidate for gamma carboxylation [85], a post-translational modification that would increase its capability of coordinating Ca^{2+}. According to the crystal structure of Cx26 [86], Glu47 is found in the middle of the parahelix formed by residues 42–50 at the border between transmembrane helix 1 (TM1) and the first extracellular loop (E1). MD simulations of the Cx26 hemichannel suggest that Glu47, Arg75 and Arg184 are components of an extensive intra/ intersubunit electrostatic network that includes salt-bridge and hydrogen-bond formation and stabilizes the parahelix and TM1/E1 bend angle (4.1) [87]. Arg75 and Arg184 are conserved among different connexin isoforms and the deafness-associated R75Q and R75W mutations of Cx26 produce proteins that prevent the formation of functional gap junction channels [88–90]. The absence of gap-junctional communication caused by R75W expression is due to defective gapjunction formation by functional hemichannels [91]. In Cx32, mutations of Arg75 (R75Q, R75P, R75W) associated with X-linked Charcot Marie Tooth disease (CMT1X) result in trafficking problems, and mutant proteins fail to reach the plasma membrane [92]. A pathogenic role has been also attributed to mutations of Arg184 in Cx26 (R184P [93], R184Q [94], R184W [95]), as well as to mutation of the homologous Arg183 in Cx32 (R183C, R183H, R183S [96]). Here, we used quantum chemistry computations to investigate the effects exerted by coordination of Ca^{2+} to γGlu47. We examined the equilibrium configuration of a local cluster of amino acids surrounding γGlu47, including Arg75 and Arg184, and singled out a local rearrangement of arginines that can be linked to one of the hemichannel gating mechanisms.

All geometries (γGlu47, dyads and triads with and without Ca^{2+}) were fully optimized *in vacuo* at the B3LYP/6-31G(d,p) level of theory [11, 12] as implemented in the software package Gaussian [1]. The initial structures were derived from an equilibrated MD configuration of the Cx26 connexin hemichannel [83, 97] and Glu residues were gamma carboxylated before geometry optimization; a Ca^{2+} was subsequently added and the structure re-optimized. Single point energies and binding energies were computed using a larger basis set, i.e. B3LYP/6-311++G(d,p).



(a) Extracellular region. Protein backbone is shown in ribbon representation, while residues mentioned in the text are shown in licorice representation: color legend: γGlu42 (green), γGlu47(orange), Arg75 (yellow), Arg184 (red).

(b) Close up view of the residues interacting with γGlu47. Part of two connexins protomers are shown in ribbons. The configuration is taken from an equilibrium MD trajectory, in which in positions 42 and 47 there were two standard (non gamma carboxylated) Glu.

Figure 4.1: Connexin 26 backbone and γGlu42 (green), γGlu47(orange), Arg75 (yellow), Arg184 (red) residues.

Binding energies ΔE were computed by partitioning the molecular system into two suitable fragments (fragments are labeled as contiguous sequences of amino acids (AA) without hyphenation to distinguish them from dyads and triads and are not in bold); the sum of the energies of each fragment, calculated separately without relaxing their geometries, has been subtracted to the energy of the total system; according to this definition a negative ΔE corresponds to a favorable binding. Basis set superposition error (BSSE) [98], although negligible when neutral fragments are involved, was evaluated and taken into account in the results. The amino terminus of the isolated γGlu was saturated with an acetyl group replacing H, while the carboxylic group was saturated with a $NHCH_3$ group replacing OH. In all other larger molecular systems, the amino and carboxyl termini were described simply as $–NHCH_3$ and $–COOCH_3$, respectively, to reduce the computational effort. Unless otherwise stated, γGlu residues were always considered dianions. A two-layer QM/QM approach was set up for full geometry optimization of a larger AA cluster formed by 15 residues (γGlu42–Val43–Trp44–Gly45–Asp46–γGlu47–Gln48–Ala49–Phe51–Tyr65–Ser72–Arg75–Val182–Arg184–Lys188), using the ONIOM scheme implemented in Gaussian [99]. The inner layer (or high layer) was accurately described using the DFT method; B3LYP functional and 6-31G(d,p) basis set were employed. This layer includes part of the side chain of γGlu47 ($(COO)_2CHCH_2^{2-}$) and part of the side chain of arginines ($(NH_2)_2C(NH)CH_2^{+}$), and Ca^{2+} ion when present. The parametric model 3 (PM3) Hamiltonian [100] was used for the outer layer (or low layer), which is formed by the residues located within 4 Å from γGlu47 of the inner layer (γGlu42–Val43–Trp44–Gly45–Asp46–Gln48–Ala49–Phe51–Tyr65–Ser72–Val182–Lys188). This computational approach has been satisfactorily used before by some of us for structural

investigation of different protein systems [101, 102]. Single point and binding energies were calculated also on the triads γGlu47–Arg75–Arg184 (with and without Ca^{2+}), extracted from ONIOM optimized structures at the B3LYP/6-311++G(d,p) level of theory.

4.1.1 Fragments binding analysis

Our working hypothesis is that the peculiar and specific interaction of Ca^{2+} with a gamma carboxylated glutamate, in particular γGlu47, may impart distortions in the structure of Cx26 by modifying the interaction between γGlu47 and two proximal arginines (Arg75 and Arg184) in the AA cluster γGlu47-Arg75-Arg184. To investigate the binding mode of Ca^{2+}, a bottom-up approach was undertaken, based on state-of-the-art quantum chemistry methods, i.e. hybrid DFT calculations. The first step was the full (unconstrained) geometry optimization of a non-coded γGlu. A monoanionic structure, characterized by a neutral (protonated) and an anion carboxylate adjacent group, was fully optimized at the B3LYP/6-31G(d,p) level of theory. The calculation converged to the same molecular geometry (Fig. C.1A in appendix C), irrespective of the choice of the carboxylate group which can be initially protonated. In the T-shaped structure the carboxylate groups are arranged in a five-membered ring fashion; the angle O–H–O is close to 160- and the O–H distances are 1.04 and 1.46 Å, respectively. The optimization of the dianionic form, with both negatively charged adjacent carboxylates, converged, but involved a proton transfer from the N terminus to one carboxylate group (Fig. C.1B in appendix C). In this case the T-shaped form is preserved but the five membered ring is disrupted. Thus the dianion form with both negative charges on the carboxylate groups is unlikely astandalone species, but in the protein environment where protonated arginines are present in close proximity, it is reasonable totreat γGlu residues as dianions, which are also better candidates for binding calcium. In the optimized γGlu-Ca^{2+} geometry (Fig. C.1C in appendix C), the T-shaped structure is retained and the ion is chelated by three O atoms of the two negatively charged carboxylates; the average Ca–O distance is 2.3 Å; the fourth O atom is hydrogen bonded to the close amide group and cannot be involved in calcium binding for steric reasons. The ΔE *in vacuo* after Ca^{2+} coordination is -512.9 Kcal/mol, thus the binding is largely favored (Table 4.1).

Compound	Fragments	Binding energy ΔE (Kcal/mol)
γGlu-Ca^{2+} (0)	γGlu (-2)	-512.9
	Ca^{2+} (+2)	
γGlu47-Arg75 (-1)	γGlu47 (-1)	-14.7
	Arg75 (0)	
γGlu47-Arg75-Ca^{2+} (+1)	γGlu47Arg75 (-1)	-379.9
	Ca^{2+} (+2)	
γGlu47-Arg75-Arg184 (0)	γGlu47Arg75 (0)	-18.3
	Arg184 (0)	
	γGlu47Arg184 (-1)	-111.8
	Arg75 (+1)	
γGlu47-Arg75-Arg184-Ca^{2+} (+2)	γGlu47Arg75Arg184 (0)	-271.3
	Ca^{2+} (+2)	
	γGlu47Arg184Ca^{2+} (+1)	17.3
	Arg75 (+1)	
	γGlu47Arg75Ca^{2+} (+1)	-4.6
	Arg184 (+1)	
γGlu47-Arg75-Arg184 (0)^a	γGlu47Arg75 (-1)	-130.4
	Arg184 (+1)	
	γGlu47Arg184 (-1)	-88.6
	Arg75 (+1)	
γGlu47-Arg75-Arg184-Ca^{2+} (+2)^a	γGlu47Arg75Arg184 (0)	-229.3
	Ca^{2+} (+2)	
	γGlu47Arg184Ca^{2+} (+1)	+33.2
	Arg75 (+1)	
	γGlu47Arg75Ca^{2+} (+1)	+20.6
	Arg184 (+1)	

Table 4.1: Binding energies ΔE between defined fragments of the studied model compounds; total charges are given in parenthesis. All energies include BSSE correction. ^{a} Optimized geometries extracted from ONIOM model.

It is worth mentioning that this large value is mainly due to electrostatic contributions and is of the same order of magnitude of binding energies reported for analogous systems [103]. Subsequently, two dyads of amino acids were considered, i.e. γGlu47-Arg75 and γGlu42-Arg75 (Fig. 4.2). γGlu42 was examined because Glu42 is close to Glu47, it is another candidate for gammacarboxylation and it too, in principle, could bind Ca^{2+}, although it is not conserved in all connexin isoforms. Both Glu residues face the amino groups of Arg75 and in both dyads, a proton transfer occurs during geometry optimization from Arg75 to one of the carboxylate groups

of glutamate, which becomes a monoanion (Fig. 4.2). In the presence of a calcium ion, the two γGlu residues behave very differently (Fig. 4.2B,4.2D). γGlu42 uses its acetyl tail together with its carboxylate groups to bind Ca^{2+} (Fig. 4.2D), however this imposes a severe distortion to its structure, which is highly unlikely to occur in the protein. By contrast, in the dyad γGlu47-Arg75 with Ca^{2+} (Fig. 4.1(b)), the ion is chelated by one COO^{-} group of γGlu47 and two N atoms of Arg75, both having their electron pair available after a proton transfer from one of them to the second carboxylate group of glutamate. The binding energy between γGlu47 and Arg75 is -14.7 Kcal/mol, a value which accounts mainly for salt bridge formation since the total charges of the fragments are -1 and 0, respectively. As expected, the binding energy of Ca^{2+} to the γGlu47Arg75 fragment is much larger, i.e. -379.9 Kcal/mol, due to the dominant electrostatic contribution (Table 4.1). We finally examined the AA triad γGlu47-Arg75-Arg184 (Fig. 4.3A,4.3B; recall that Arg184, which can interact with γGlu47, belongs to an adjacent connexin). The salt bridge between γGlu47 and Arg75 is maintained in the isolated triad; the second COO^{-} is protonated after a proton transfer from Arg184 (Fig. 4.3A). Note thatthe backbone of Arg184 is significantly displaced from its initial position, however this rearrangement is unlikely to occur in the protein due to the interactions with the surrounding amino acids. In the presence of Ca^{2+} (Fig. 4.3B) the salt bridge is preserved, although slightly distorted. As in the dyad γGlu47-Arg75, Ca^{2+} interacts with three O atoms of γGlu47 and two N atoms of Arg75 (Fig. 4.3B). On the basis of the binding energies computed using two different fragmentations, i.e. γGlu47Arg75 and Arg184 and γGlu47Arg184 and Arg75, we conclude that the interaction between γGlu47 with Arg75 (-111.8 Kcal/mol) is stronger than the interaction with Arg184 (-18.3 Kcal/mol). However, the in presence of Ca^{2+}, the ion is chelated by γGlu47, the binding of the latter to Arg184 becomes less favorable, and the binding of γGlu47 to Arg75 is neatly unfavorable (-4.6 and +17.3 Kcal/mol,respectively). This strongly suggests that Ca^{2+} can alter the local structure, because its binding to γGlu47 prevents any stabilizing interaction between γGlu47 and the close arginines. To further validate these results, obtained for the model triad *in vacuo*, and to take into account properly the protein environment, we constructed a two-layer ONIOM model in which the core formed by portions of γGlu47-Arg75-Arg184 (with and without Ca^{2+}) is accurately described at B3LYP/6-31G(d,p), and the surrounding amino acids within a sphere centered on γGlu47 with a radius of 4 Å are treated using the PM3 Hamiltonian (Fig. 4.3C,4.3D). The geometry of the two-layer ONIOM model was fully optimized (without any constraint). Note that only portions of the three amino acids (and Ca^{2+}, when present) were included in the inner layer, as explained in Section 2, and shown in ball and stick representation in Fig. 4.3C, 4.3D. In the triad γGlu47-Arg75- Arg184 extracted from the ONIOM model, the salt bridge between γGlu47 and Arg75 is maintained and no proton transfer from Arg184 to γGlu47 takes place. Indeed, analysis of the binding energies shows the ΔE is much larger, due a stronger electrostatic contribution for this system than for the model triad discussed above (-130.4 vs -18.3 kcal/mol; see Table 4.1). In the presence of Ca^{2+} (Fig. 4.3D), the salt bridge is destroyed and Ca^{2+} binds two oxygen atoms, belonging to each COO^{-} group of γGlu47, as in the isolated triads (Fig. 4.3B). Importantly, this bind-

ing mode is identical to that observed in the isolated γGlu47-Ca^{2+} complex. As in the isolated triad, energy decomposition analysis reveals that the interaction between γGlu47 and the close arginines is disrupted (Table 4.1). This result supports the idea that a local bond rearrangement, i.e. binding of Ca^{2+} to γGlu47 and the breaking of salt bridges between γGlu47 and the neighboring arginines impacts on the energetics and induces a structure rearrangement in a critical region of connexin hemichannels. To conclude, our results suggest that the post-translationally gamma carboxylated Glu47 is a strong candidate to coordinate Ca^{2+} in the extracellular vestibule of a Cx26 hemichannel. A consequence of this putative coordination action is a rearrangement of the side chains of two highly conserved arginines (Arg75 and Arg184), that are required for the correct functioning of connexin hemichannels and gap junction channels, accompanied by large energetics variations. The salt bridges between Glu47 and Arg184 are considered important to maintain the quaternary structure of the connexin hemichannel [104, 105], and Glu47 is part of the parahelix that is thought to undergo a structural rearrangement to close the hemichannel in response to membrane hyperpolarization and extracellular calcium [106]. More extensive computations and simulations are currently underway to assess whether this distortion is sufficient to initiate the experimentally observed gating processes.

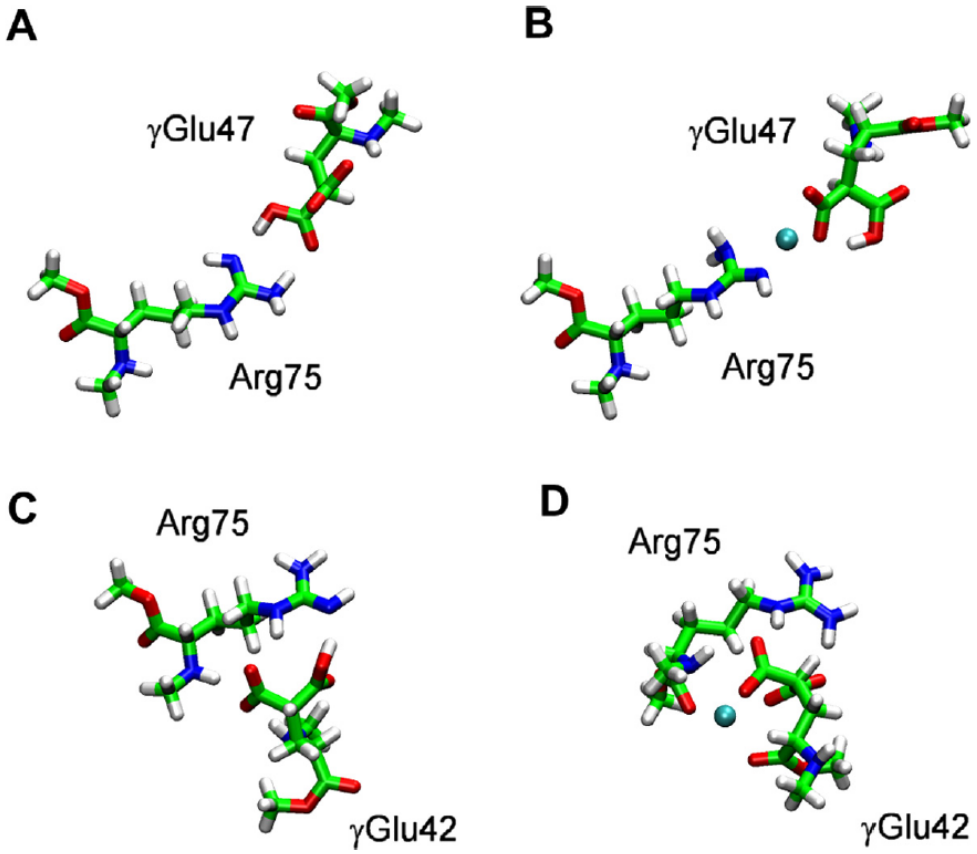


Figure 4.2: Connexin 26 backbone and γGlu42 (green), γGlu47(orange), Arg75 (yellow), Arg184 (red) residues.

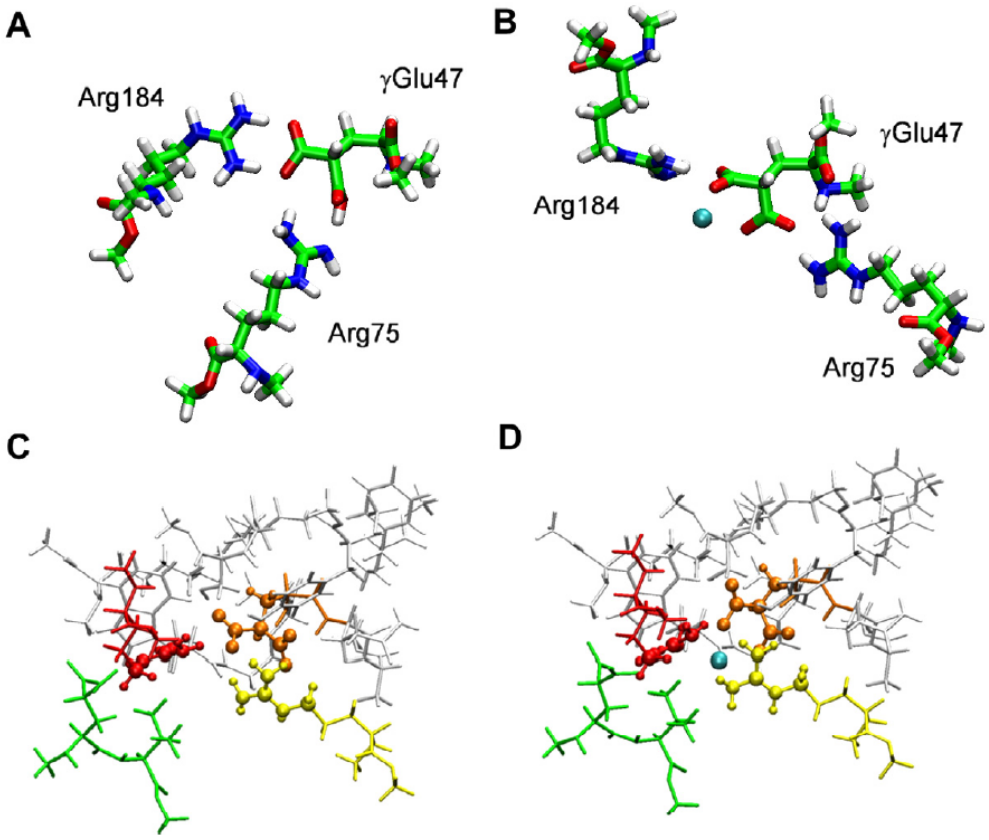


Figure 4.3: Connexin 26 backbone and γGlu42 (green), γGlu47(orange), Arg75 (yellow), Arg184 (red) residues.

Chapter 5

From quantum Hamiltonians to classic Hamiltonians

In the previous chapters we introduced the quantum chemistry for the study of small molecules and the ONIOM multiscale method which is useful to obtain static properties of large molecules. If we are interested in the dynamic properties of bio-/macro-molecules, we need to employ another approach described in section 2.5, i.e. the molecular dynamics. This method allows to treat atoms as point particle, which obey the Newton's law, thus a possible quantum nature is not considered within this type of description. This drawback is of course necessary if we want to model large systems, because the quantum mechanics implicate an enormous complexity, e.g. instead of a point we have a function describing the position, which is dramatic complication. It is still possible to use the quantum mechanics information somehow, in fact the force field employed for the simulation of the dynamics, may be parameterized using accurate *ab-initio* calculations, providing the link between quantum and classical mechanics.

5.1 Molecular dynamics simulation of the connexin 26

As already outlined in chap 4, connexins (Cx) are a class of membrane proteins important for auditory function, that arrange together by non-covalent interactions. These quaternary aggregate forms are called connexons or hemichannels. The pathway generated by this membrane protein allows to connect cells' cytoplasm with the extracellular environment. Moreover head-to-head assembly of two connexons form intercellular (gap junction) channels, that connect the plasma membranes of two adjacent cells and allow the passage of nucleotides, ions, small peptides and metabolites [107]. In particular the Cx26 (26 KDa protomer), codified by the gene GJB2, is found in the supporting cells of the organ of Corti, located in the mammalian cochlea, where at least four isoforms are expressed [85]. Mutations in the GJB2 gene that modify the channel function of the connexin, cause inherited non-sindromic deafness, leading to skin and sensorineural deseases. In addition many sites of the protein subject to post-translational modification

(PTM), have been identified expressing Cx26 in HeLa cells, and analyzed with mass spectrometry techniques [85]. The found PTMs include hydroxylation, methylation, phosphorylation, acetylation and γ-carboxyglutamation, the last deeply investigated in this work. The publication of the 3.5 Å human Cx26 crystal data [86] has been a long awaited result that shed light, elucidating the previous lower resolution structures [108–110] and provided an essential starting point for deepened future investigations. Each of the six connexins composing the hemichannel, has a particular tertiary structure that crosses the membrane four times. The main topology of a single connexin may be grouped in the transmembrane helices (TM1-4), the cytoplasmic N/C-terminal (NT, CT) domains, the cytoplasmic loop (CL) that links TM2-TM3, an extracellular loop (E1) between TM1-TM2 and a second extracellular loop (E2) between TM3-TM4 (Figure 5.1). The TM1-4 helices span the following residue range: 21-40, 76-98, 132-154 and 193-215; the cytoplasmic NT domains, composed by residues 1-20, are located in the pore channel, while the cytosolic CT domain, going from residue 216 to 226, is located at the external portion of the connexon. The CL domain, which goes from residue 99 to 131, links TM2 and TM3 helices and, how we will see, is the most mobile portion of the whole system. The extracellular loop E1, goes from residue 41 to 75, while E2 has the following residue range 155-192 (Fig. 5.1). In the primary sequence there are 226 total amino acids, from which a portion of NT, CL and CT, i.e. residues 1, 110-124 and 216-226, were not located in the crystal data together with a some atoms of the side chain of residues 15, 17 and 19 [86]. Altough the absence of these amino acids may not influence the pore size directly (except for Met1 which falls at the center of the channel), they may assume an important structural role with consequences on the overall quaternary structure, thus they were included in all the simulations performed in this work[2]. A first attempt to describe structure-functions relation after PTMs was done (without explicitly including the charges of the modified residues) where after relaxing the protein structure with MD simulations, the pore diameter was larger with respect to the cryistallographic data, expecially near the N-terminal residues (i.e. Met1). Moreover acetylation of the Met1 residue did not show significant changes to the pore size and together with γ-carboxyglutamation switched the channel selectivity to cations [87]. Other MD simulations highlighted that the open channel has a barrier due to the presence of Lys41 residue, that, in absence of an electrochemical gradient, obstruct the ion diffusion through the pore [111]; another work, instead, focused on the permeation of the calcein molecule through the channel, observing that the electrostatic properties of the pore did not allow the penetration of the molecule corroborating two hypoteses: the 3.5 Å structure of the protein is not related to an open channel, because of the crystallization and dehydratation procedures; the PTMs in living cells transform profoundly the permeation properties thanks to the modification of the charge in the side chain of many residues [112]. In addition two mutants of the Cx26 were studied by the same authors [113, 114] and, in particular, in the case of M34T mutation, the subtitution of an apolar amino acid with a polar, did change the pore bottleneck and the NT domain was separated more easily from the inner wall of the channel, increasing the ions permeation barrier; instead the C169Y mutation located spatially near the

extracellular portion, disrupted the disulfide bond with Cys64, compromising the protein structure. So the Cx26C169Y instead of being a polymorphism as previously identified, it was found both experimentally and computationally to be a dysfunctional channel with conseguent hearing loss due to problems in the intercellular head-to-head docking of two connexons. Some of us also characterized a simplified model of this system, using QM/QM calculations on a specific portion of the Cx26 to depict the Ca^{2+} binding process to the Glu42 and Glu47 residues (chapter 4). We observed that calcium binding, after modification of the salt bridges with Arg75 and Arg184, may be responsible for the rearrangement of the quaternary structure of the system, with a possible following closure of the channel [115]. In this work we tried to simulate the hemichannel behaviour in the presence of PTMs and Ca^{2+}. In particular we focused only on the γ-carboxylation of Glu42 and Glu47, which we believe to be responsible for the modification of the hemichannel structure. To do so we derived new partial charges and missing FF parameters of the γ-carboxylated glutamate residue which we substituted in the protein sequence. We did not take into account also the possible γ-carboxylation of Glu114, because that residue falls in the highly moving external CL domain, that is far from the pore axis. Finally the system was analyzed from many different point of view, highlighting rigidity/flexibility and structure.

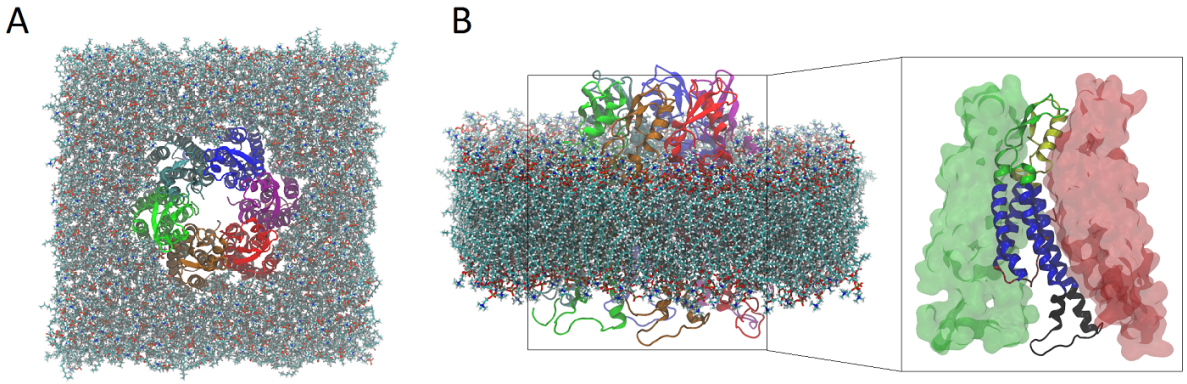


Figure 5.1: (A) top view of the 6 protomers in the bilayer (B) side view of the protein and zoom of the orange connexin unit. Colour code in the right panel according to [86]: red, NT; blue, TM1-TM4; green, E1; yellow, E2; black, CL and CT. For the sake of clarity water molecules and ions are not shown.

5.1.1 Parametrization of the γ-carboxylated glutamate

The parametrization for the γ-carboxylated glutamate was performed similarly to an our previous work [116] fitting all the partial charges for the residue and the missing C_β-C_γ-C angle bending and the C_β-C_γ-C-O torsion parameters. The peptide bonds of the residues were first capped using the -$COCH_3$ (ACE) and -$NHCH_3$ (NME) terminals, in agreement with the previous Amber force field parametrizations. The molecule was optimized first in 4 different conformations, 2 with Φ/Ψ angles in the region of right handed α-helix and 2 with Φ/Ψ angles in the

region of β-sheets with the B3LYP/6-311G(d,p) level of theory and, afterwards, electrostatic potential (ESP) was calculated at HF/6-31G(d) level. The data obtained were employed in the multiconfigurational restrained ESP (RESP) fitting, by fixing the charges of the capping groups to the values already present in the FF14SB [117] and fitting the partial charges on the remaining atoms [118]. The C_β-C_γ-C angle bending potential energy surface (PES) was calculated at the B3LYP/6-311G(d,p) level of theory varying the angle around its equilibrium value with 6 steps of 5 degrees, relaxing the structure and obtaining 7 points. Instead the dihedral angle PES was determined doing a relaxed scan at the same level of theory, starting from the optimized structure and rotating one of the two C_β-C_γ-C-O angles for 35 steps by 10°. The dihedral was described using up to 4 terms in the Fourier expansion. The angle bending and the dihedral parameters were fitted using `paramfit` [119].

5.1.2 Molecular dynamics simulations

The presence of the Ca^{2+} ions was studied by performing two different simulations for the wild type (Cx26wt) and the PTM connexin 26 (Cx26E42-47γc), in the absence and in the presence of 2 calcium ions, which correspond to a 2 mM concentration. We started from the crystallographic structure [86], from which residue sequences 1, 110-124, 218-226 and atoms from residues 15, 17, 19 were missing in each connexin. We first added missing residues, adjusting their postitions with conjugate gradient first (CG), then doing simulated annealing up to 1000K and, after cooling down to 300K, a second CG minimization was done. This procedure was repeated generating 10 different models from which we chose the best one on the basis of its Discrete Optimized Protein Energy score [120], implemented in the Modeller 9.15 software [121]. During these steps the atom positions of the early residues were kept fixed. We then added the missing atoms using the CHARMM-GUI PDB reader [122, 123]. The protein was inserted in a preequilibrated bilayer membrane that was composed by 400 1-palmitoyl,2-oleoyl-sn-glycero-3-phosphocholine (POPC) molecules, using the CHARMM-GUI membrane builder [122, 124–127]. The system has been solvated using 42291 TIP3P water molecules [128] and the biological ionic strength was reproduced including K^+ and Cl^- ions [129, 130] at a concentration of about 0.15 M, slightly unbalanced to neutralize the total charge. We studied a total number of six systems, simulating both the Cx26wt and the Cx26E42-47γc. Each of the two proteins was simulated using two different concentrations of Ca^{2+}, namely 0 and 2 mM, corresponding to 0 and 2 total Ca^{2+} ions. The total number of atoms which constituted the systems were approximately equal to 200K particles. In order to perform the molecular dynamics of the protein, we used the Amber 14 package [131] and fort the amino acids we used the recent FF14SB [117] parameters which provides more reliable proportions of helical with respect to extended structures and other improvements [117, 131] with respect to older force fields. The bilayer was modelled using the lipid14 force field that well describes observables as area per lipid, bilayer thickness, NMR order parameters, scattering data, lipid lateral diffusion and permit to avoid the use of artificial pa-

rameters as the surface tension or constant area term [131, 239]. We simulated first the Cx26wt without Ca^{2+}, doing first a partial minimization with steepest descent and CG for 10000 cycles by keeping fixed protein and bilayer atom positions and leaving water and ions to adjust freely, followed by 20000 cycles of full minimization without restraints on the atom positions. Afterwards two subsequent heatings were conducted with protein and bilayer atoms restrained; the first ramp to 100K in the NVT ensemble was reached in 5 ps using the Langevin thermostat and the second ramp to the final temperature of 300K in the NPT ensemble was reached in 100 ps, by setting the pressure to 1 atm, with the anisotropic Berendsen weak-coupling barostat. After heating the system to the production temperature, a long equilibration dynamics of 100 ns was conducted using the Langevin thermostat and the Monte-Carlo anisotropic barostat on the system, followed by 100 ns production dynamics with the same settings and without restraints on the atom positions. Long-range elctrostatic interactions were treated using the particle mesh Ewald method [27, 132–134] and the cut-off distance for the non-bonded interaction was set equal to 10.0 Å. The SHAKE constraint algorithm was applied to all hydrogen atoms, and 2 fs for the time-step were used. For the remaining three simulations, we repeated minimization, heating, equilibration and production dynamics steps as described before, starting from the last frame of the equilibration performed for Cx26wt-Ca0, and generating three replicas with different initial velocities. The analysis of the simulation was carried on the 100 ns production dynamics. We did check the root mean square deviation, obtained using VMD [135], while root mean square fluctuations (RMSF) and other information were obtained using GROMACS 4.6.5 [136].

5.1.3 Force field validation

In order to check the validity of the new developed parameters we compared the modulus of dipole moment obtained using the MM point charges and QM *ab-initio* density and dipole integrals. We chose this observable because it allows to take into account both structural and charge parameters. The dipole moment obtained using the MM optimized structure with new FF parameters and atom-centered RESP charges was equal to 18.61 Debye, while the dipole moment calculated at HF/6-31G(d) level using the most similar conformer, optimized using B3LYP/6-311G(d,p), was equal to 19.56 Debye; the difference between these two values was reasonably low and this suggest that the novel MM parameters reproduce well the QM values.

5.1.4 Channel dimensions

The pore size was measured using the HOLE algorithm [137] software. As pointed out by Oshima [138], in order to allows the permeation of hydrated ions, the channel diameter must be larger than 0.8 nm in diameter. It is well known that in the WT protein the closure mechanism is triggered by the extracellular concentration of calcium ions ($[Ca^{2+}]_e$). In particular when $[Ca^{2+}]_e \sim 2$ mM the closure of the hemichannel is promoted and this is also in agreement with

our results, where a narrowing of the pore to ∼0.7 nm was observed at ∼3.5 nm along the z-axis (Fig. 5.2A). In addition we unexpectedly found that in the γ-carboxylated protein, the closure mechanism is no longer activated by the presence of calcium ions, thus very small differences in the pore diameter have been found when $[Ca^{2+}]_e$ is equal to 0 and 2 mM (Fig. 5.2B).

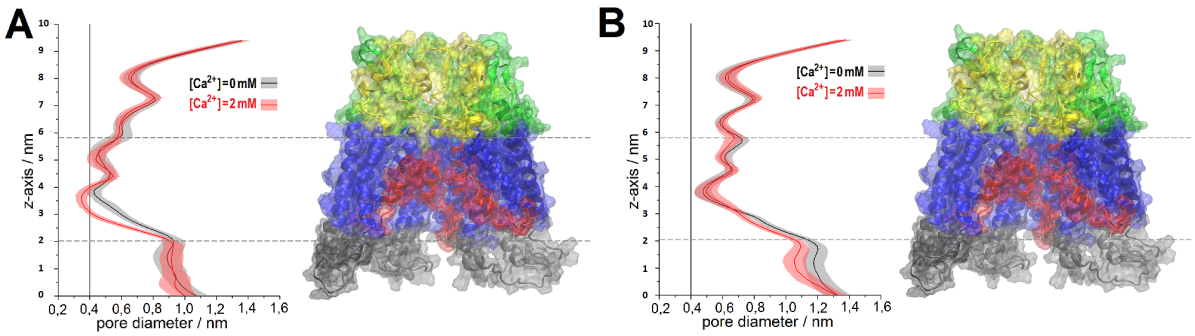


Figure 5.2: Pore diameter along the z-axis for the Cx26wt (A) and Cx26E42-47γc (B) at $[Ca^{2+}]$=0 mM (black) and $[Ca^{2+}]$=2 mM (red). Colored areas show the standard deviation over the 6 protomers.

Moreover the local flexibility of the WT hemichannel is reduced in presence of calcium ions, in particular for the E1/E2 and TM2/TM4 regions, as can be inspected from Fig. 5.3A; instead the addition of Ca^{2+} ions to the γ-carboxylated protein does not show significant changes in the local flexibility, thus the protein in this case is less sensible to the presence of the calcium ions (Fig. 5.3B).

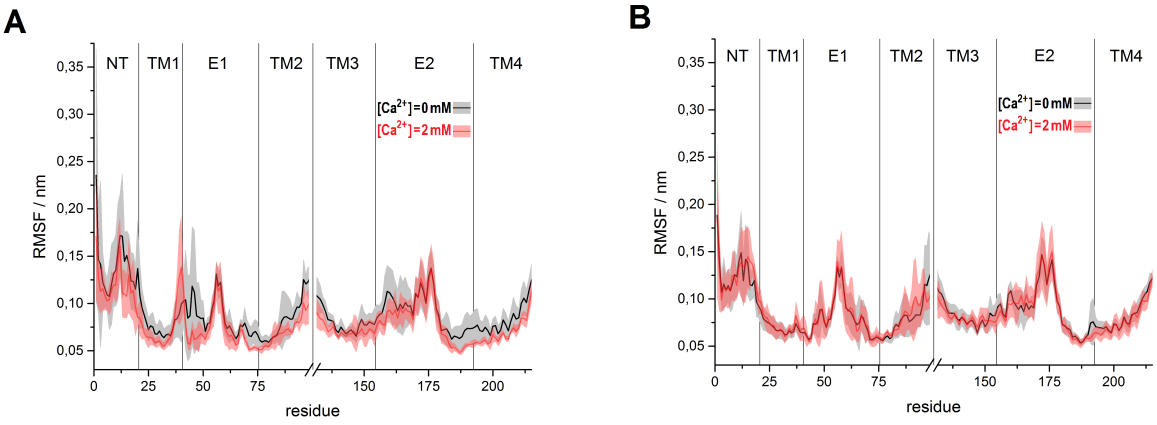


Figure 5.3: Root mean square fluctuations averaged over the 6 protomers of the Cx26wt (A) and Cx26E42-47γc (B) at $[Ca^{2+}]$=0 mM (black) and $[Ca^{2+}]$=2 mM (red). Colored areas show the standard deviation over the 6 protomers. CL and CT regions were not represented.

5.1.5 N-terminal domain motion

Several closing mechanisms have been proposed involving movements of small and large portions of the protein [138]. In particular the subunit rotation model foresees a high cooperation

between all the subunits that assume an identical conformational change at the same time, while the loop gating, the Vj-gating and the plug mechanisms require smaller rearrangements. The latter two involve the residues located in the NT portion of the hemichannel, constituted by a short α-helix (with about numero residues) and a small loop that comes before the TM1 region. The plug constituted by the structured α-helix may occlude the funnel in agreement with the cryo-electron microscopy (cryo-EM) data obtained for the Cx26M34A [110], where the NT residues are supposed to constitute an organized structure, due to the replacement of the Met34 with Ala34 which no longer interacts with Trp3 and produce a collapse in the NTs [86]. Our results highlight that the plug play an important role in the narrowing of the pore funnel *via* a Vj-gating like mechanism which require that the NT portions move towards the cytoplasmic side (Fig. 5.4). For the WT hemichannel the calcium produce an average shift of the Met1 Cα of $\sim$0.17 nm while in the γ-carboxylated Cx26 this difference is lower ($\sim$0.04 nm).

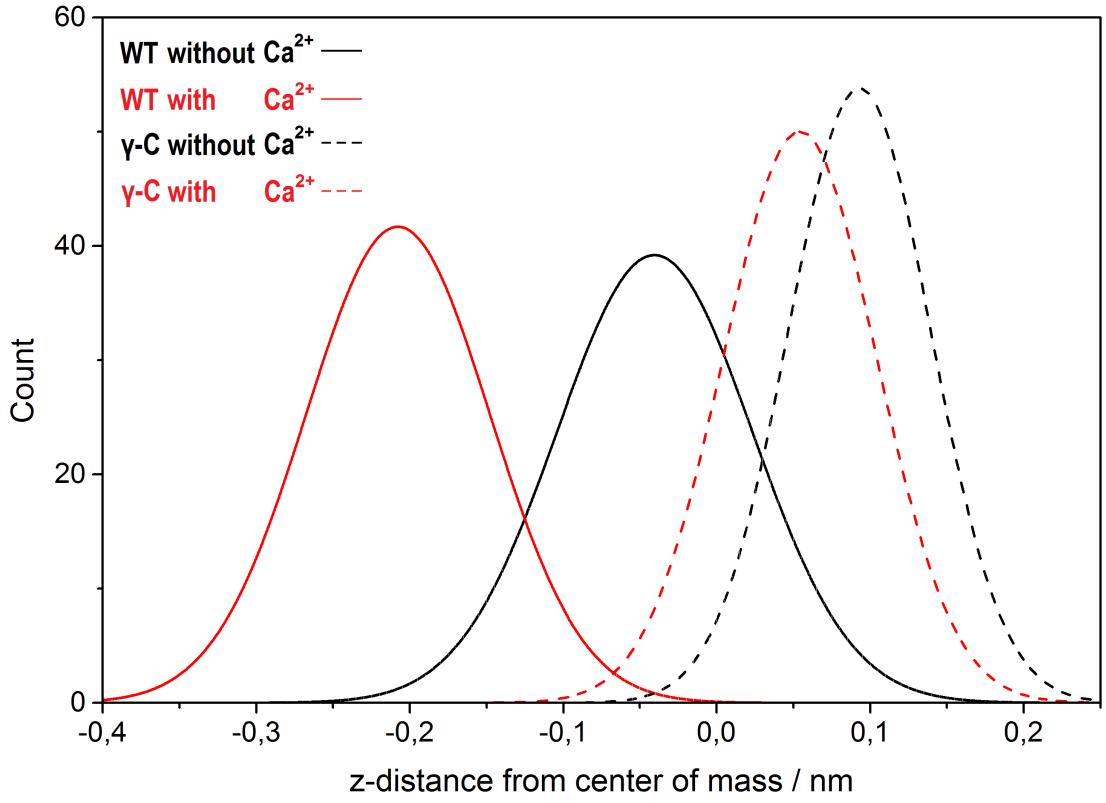


Figure 5.4: Average z position for Met1 Cα shifted by the center of mass calculated over all the atoms.

5.2 General amber force field for diphenyldichalcogenides

Organoselenium compounds are important catalysts in organic chemistry,[139, 140] with promising application in green chemistry,[141, 142] but are also intensively studied for their potential

use as therapeutic agents.[143, 144] Selenium is a bio–essential element and is present in 25 selenoproteins in the form of selenocysteine (Sec), the 21st aminoacid. The presence of Sec, typically ascribed to its superior nucleophilicity, superior leaving group and electron acceptor ability relative to Cys, is still under debate.[145] The specific role of selenium rather than sulfur is particularly intriguing in the family of glutathione peroxidase (GPx) enzymes. The biological function of these enzymes is crucial in signalling pathways, but is especially fundamental in the reduction of hydrogen peroxide and harmful organic hydroperoxides to water and alcohols, respectively.[146, 147] The overall reaction (Fig. 5.5) requires two equivalents of glutathione (GSH) as cofactor and involves three main steps, i.e.

1. oxidation of the selenol (E–SeH) to selenenic acid (E–SeOH), accompanied by the reduction of the hydroperoxide and the release of one water/alcohol molecule;
2. reaction of E–SeOH with a glutathione (GSH) to form a selenenylsulfide intermediate (E–SeSG) and release of a water molecule;
3. regeneration of the initial selenol E–SeH upon reaction with the second GSH accompanied by the formation of a disulfide product (GSSG).[148–152]

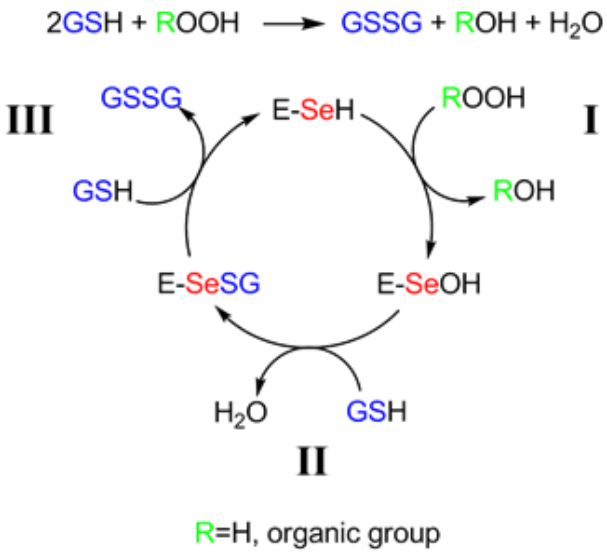


Figure 5.5: Mechanism of organic hydroperoxide reduction catalyzed by GPx (E–SeH).

It is well known that upon replacement of Sec with Cys, i.e. of selenium with sulfur, the enzymatic activity is reduced by several orders of magnitude, a phenomenon which might explain why in nature selenium has survived in the catalytic site of GPx, despite that it requires a complex and energetically very expensive co-translational insertion machinery for Sec.[153, 154] The biological importance of GPx has pushed the research in the last few decades to the design of organoselenium compounds that might be used in medicine as GPx mimics, maintaining in particular the peroxidase-like anti-oxidant activity, but also antitumor and antimicrobial properties.[155, 156] Among the synthetic small organoselenium compounds which mimic the

properties of GPx, ebselen (2-phenyl-1,2-benzisoselenazol-3(2H)-one) is well known also for its anti-inflammatory, anti-atherosclerotic and cytoprotective properties observed in both in vitro and *in vivo* models.[157] Ebselen was synthesized in 1924 [158] but its anti-oxidant properties were discovered only in 1984:[159, 160] since then it is undoubtedly the most studied GPx mimic.[161–164] The organoselenium synthetic GPx mimics reported in literature have been conveniently classified by Mugesh and Bhabak[155, 156] in three major categories:

1. cyclic selenenyl amides having a Se-N bond; ebselen belongs to this category;
2. diphenyldiselenides;
3. aromatic or aliphatic monoselenides.

Our interest is mainly on diphenyldiselenides, and more in general on diphenyldichalcogenides (Fig. 5.6), for the presence of a chalcogen-chalcogen bond, as well as for their peroxidase-like efficiency, the mechanistic details of which are not clear yet Fig. 5.6.[165–167]

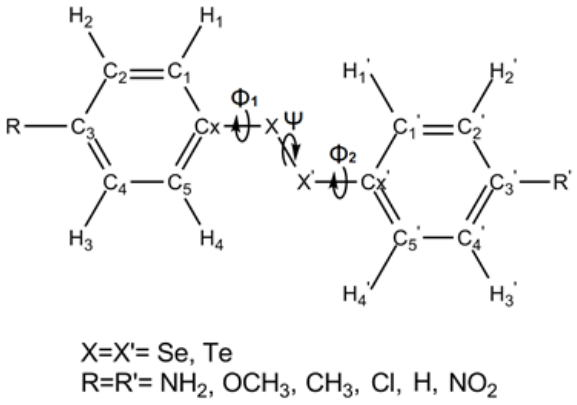


Figure 5.6: Diphenyldichalcogenides included in the present work.

Differently from selenium, tellurium has not any known biological function and is in general believed to be more toxic than selenium.[144] Also, Te compounds, whose pharmacological effects are less known, possess strong peroxidase activity, as first claimed by Andersson and co-workers.[168] Subsequently Engman [169] reported on diarylditellurides and very recently Rocha et al.[170, 171] presented several experimental studies on the anti-oxidant activity of diphenylditellurides. In addition, Te compounds have been investigated as chemopreventive agents and diphenylditelluride appears to have a very high capacity of inducing apoptotic cell death.[165, 171] Importantly, the toxicity of organotellurides is found to be comparable to organoselenides and this paves the route to their possible use as drugs, in particular as anti-oxidant GPx mimics, although careful control is warmly recommended.[165, 171] From a mechanistic point of view, small organoselenium compounds show some limitations and often their catalytic activity is relatively poor.[162] On the basis of the available experimental and computational studies this can be ascribed to

1. a competing undesired thiol exchange reaction which hampers the regeneration of the selenol,[172] due to the fact that nucleophilic attack at Se (reaction III, Fig. 5.6) is both thermodynamically and kinetically favored;

2. the nature of the thiol cosubstrate, which may have a dramatic effect on the catalytic activity of ebselen and its analogues;[173–176]

3. the non-specificity for the substrate which leads to undesirable reactions with Cys residues in different protein sites.

Mechanistic studies on organotellurium compounds and their GPx activity are still scarce; nevertheless the experimental evidence on diphenylditellurides reported very recently by Rocha et al. is undoubtedly attractive.[165, 171] Aiming at an accurate drug design through *in silico* methodologies adopting a bottom-up approach, a quantum chemical mechanistic investigation of the single molecule is a fundamental starting point to disclose in detail the intimate connection between the atomistic nature of the compound and its reactivity. For example, referring to the systems here discussed, accurate quantum mechanics studies have shed some light on the difference between sulfur and selenium in the reaction of hydroperoxide reduction by model thiols/selenols[177, 178] or in the reactions of thiolate nucleophilic attack to homo/hetero dichalcogenides containing sulfur as well as selenium.[179–181] Numerous DFT contributions can be found on the reactivity of ebselen.[182–184] Combined experimental and DFT studies allowed to identify suitable molecular features which enhance the peroxidase activity of amine based diphenyldiselenides.[185] At this point, the next step in drug design requires the docking of the molecule in a biological environment, but simulations of these drugs in contact with biomolecules or testing their reactivity once they are bioincorporated in a suitable protein scaffold is a challenging goal, which requires MD techniques as well as hybrid QM/MM approaches. The lack of parameters in the available force fields (e.g. force constants, equilibrium bond lengths and angles, torsions, charges) for diphenyldiselenides and diphenylditellurides, and more generally for Se and Te compounds, prompted us to derive them, setting up a computational protocol based on the most recent tools. The AMBER Hamiltonian has been chosen, due to its large success in the simulation of biomolecules, proteins and nucleic acids.[131] The AMBER potential has the form:

$$
\begin{aligned}
V_{\text{AMBER}} = & \sum_{i<j}^{N_{atoms}} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} + 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \\
& + \sum_{i}^{N_{bonds}} b_i(r_i - r_{i.eq})^2 + \sum_{i}^{N_{angles}} a_i(\theta_i - \theta_{i,eq})^2 \\
& + \sum_{i}^{N_{dihedrals}} \sum_{n}^{N_{i,max}} \frac{V_{i,n}}{2}[1 + cos(n\phi_i - \gamma_{i,n})]
\end{aligned}
\tag{5.1}
$$

where the energy is calculated as the sum of five contributions: pairwise electrostatic potential and Lennard-Jones potential for nonbonded forces (the sums of these two terms include only atoms belonging to different molecules or separated by at least three bonds and this is why the index is n'), harmonic potential for bonds (r) and angles (θ), and a truncated Fourier series for dihedrals (ϕ). The novel parameters are incorporated in the general AMBER force field (GAFF), which has been designed by Kollman et al.[186] to be compatible with existing AMBER force fields and has parameters for almost all the organic molecules consisting of C, N, O, H, S, P, halogens, so that it is suitable for pharmaceutical applications. In this paper we illustrate in detail the methodology to derive the GAFF parameters of twelve diphenyldiselenides/-ditellurides derivatives. The change in the extent of the rotational barriers of the phenyls, due to the different substituents, has been investigated using *ab initio* QM potentials. Moreover, MM and QM dipole moments of all the molecules have been compared and MD simulations of diphenyldiselenide and diphenylditelluride (i.e. the so-called parent compounds) have been done to validate the force field parameters and study the conformation of these molecules in solvent. Complete files of parameters to be used with AMBER software are provided for the reader who is interested in this important class of drug-like compounds. Also the entire procedure is described in detail, so that the reader can easily extend GAFF to other chalcogen based compounds, using a quantum chemistry software suite like Gaussian,[187] along with AMBER tools[131] like Antechamber and the recent Paramfit,[119] and the in-house code ("gout2mdcrd") provided by us for required file format conversions.

5.2.1 Computational Details

All quantum chemical calculations were performed with the Gaussian09 programs suite.[187] The molecular geometries of twelve dichalcogenides (see Table 5.1) were fully optimized using the B3LYP functional[12–14, 188] combined with the cc-pVTZ basis set for all atoms except Te, for which the cc-pVTZ-PP basis set, which includes also pseudo potentials, was used[189–191]; in the following we will refer to this level of theory as B3LYP/cc-pVTZ(-PP) basis set, meaning that pseudo potentials are only used for the tellurium center in tellurium derivatives. Vibrational analyses were subsequently performed to ensure that minima were properly identified. The atomic charges were obtained using the optimized structures and calculating the electrostatic potential (ESP) at HF level with 6-31G(d) basis set for all atoms except Se and Te, for which the cc-pVTZ(-PP) basis set was again used (level of theory: HF/6-31G(d),cc-PVTZ(-PP)). These ESP data were subsequently used in the multiconfigurational restrained ESP (RESP) fitting procedure for each molecule[118]. Constraints were imposed to obtain the same charge values on chemically equivalent atoms; this led to a reduction of the total number of different charges (approximately 1/3 with respect to the number of atoms for all the compounds). For bond stretchings and angle bendings we generated potential energy surfaces starting from the optimized equilibrium geometry, and linearly increasing the bond length or angle, while keeping all the other geom-

etry parameters frozen. The obtained data points were fitted using Gnuplot 5.0. The torsional energy profiles for the dihedrals Ψ(C_X-X-X'-C_X') and Φ_1 (C_1-C_X-X-X') (or Φ_2 (C_1'-C_X'-X'-X)) (Fig. 5.6) were calculated from relaxed scans starting from the global minimum at B3LYP/cc-pVTZ(-PP) in order to fit the force field dihedral functions to the relaxed quantum potential energy surface (PES). All the dihedral angles of the diphenyl compounds in the force field were described using the Fourier series truncated to the first four terms. This allowed to obtain a reasonably low value of the root mean square displacement (less than 1 Kcal/mol for all the molecules). Only the $V_{i,n}$ parameters were optimized, while $\gamma_{i,n}$ phases were set all equal to 180°. All the dihedral parameters were fitted simultaneously using the novel powerful AMBER tool Paramfit.[119] The remaining molecular parameters were set equal to the parameters included in GAFF. After obtaining all the parameters for the force field, MD simulations were carried out on the parent compounds, i.e. diphenyldiselenide and diphenylditelluride, in two different explicit solvents (chloroform[192] and TIP3P water[193]) using the AMBER 14 suite of programs.[131] In the MD simulation only a single diphenyldichalcogenide molecule was present in order to avoid aggregation effects. The system was prepared using a cubic box with a surrounding buffer of 12.0 Å. First, the potential energy was minimized keeping the diphenyldichalcogenide atoms fixed and relaxing the solvent; afterwards a full minimization was done. The temperature was increased using the Langevin thermostat up to 100 K for 50 ps first and, subsequently, the Berendsen barostat was added for 1 ns in order to reach the production temperature and pressure of 310 K and 1 bar, respectively. A pre-equilibration and production dynamics of 500 ps and 500 ns respectively were conducted switching to the Monte Carlo barostat. The cut-off distance for the non-bonded interactions was set to 10.0 Å, the shake constraints were applied to all hydrogen atoms and a time-step of 2 fs was used.

5.2.2 Molecular structures

Twelve diselenides and ditellurides have been chosen (Fig. 5.6): the so-called parent compounds, $(PhSe)_2$ and $(PhTe)_2$, in which no substituent is present on the phenyl rings and a series of para-substituted molecules with amino, methoxy, methyl, chlorine and nitro groups; these are denoted in the text (p-NH_2–PhSe/Te)$_2$, (p-OCH_3–PhSe/Te)$_2$, (p-CH_3–PhSe/Te)$_2$, (p-Cl–PhSe/Te)$_2$ and (p-NO_2–PhSe/Te)$_2$, respectively. The structures of most of these compounds are present in the Cambridge Structural Database (CSD);[194] their salient geometrical parameters are reported in Table 5.1. The average Te–Te and Te–C bond length is longer than the corresponding Se–Se and Se–C bond lengths by 0.4 and 0.2 Å, respectively. The parent compounds and their derivatives have two equivalent structures with values of the dihedral angle Ψ (Fig. 5.6) with opposite sign. Like in all dichalcogenides the absolute value of Ψ falls approximately in the range 75-110°.[195] In contrast, the values of the dihedrals Φ_1 and Φ_2 (Fig. 2) are less homogeneous, suggesting that the planes of the phenyl rings can be parallel to each other (absolute values of Φ close to 90°) or deviate from this orientation up to $\Phi_{1,2} \approx 0°, \pm 180°$, a conformation which is

not forbidden by sterical hindrance because of the characteristic value of Ψ, which prevents the coplanarity of the phenyl rings in all cases. We have named these extreme conformations closed and open, referring in figurative sense to the position of saloon doors, once the molecules are suitably oriented (Fig. 5.7). Due to symmetry, Φ_1 and Φ_2 have the same parametrization and will be also referred to simply as Φ, implying that we do not distinguish Φ_1 and Φ_2 (e.g. in the FF parameters, relaxed scans and distributions).

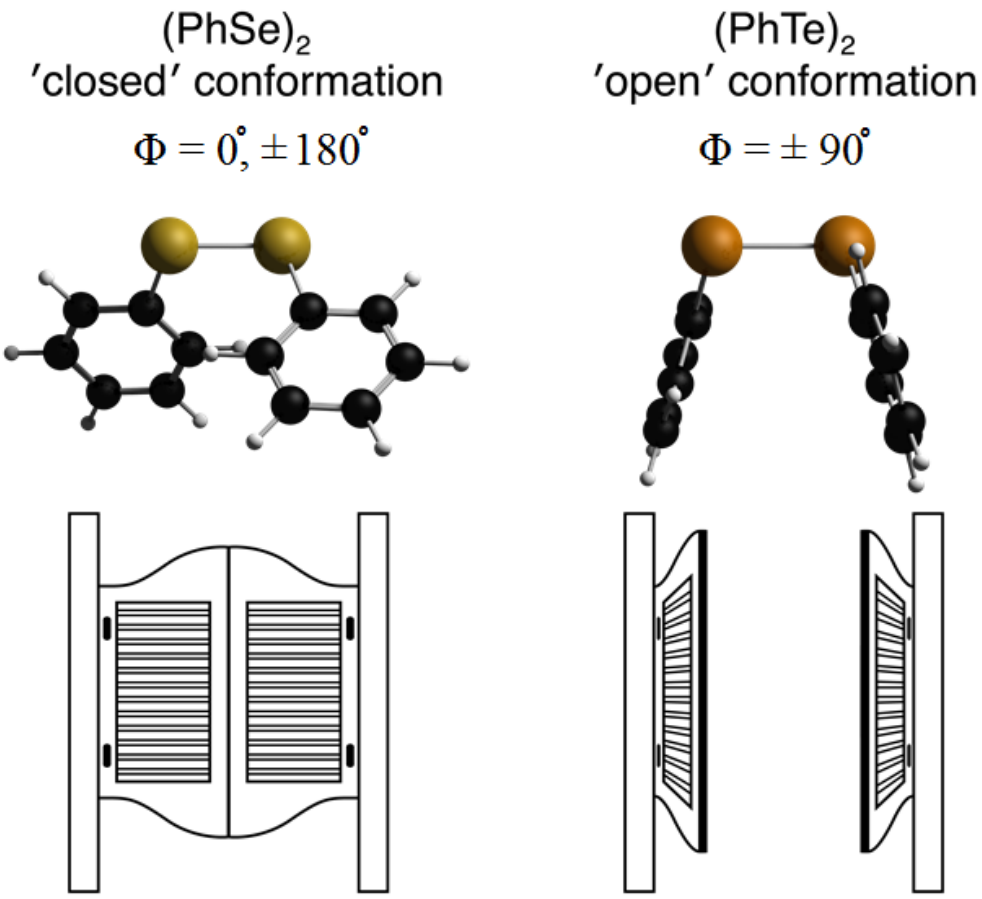


Figure 5.7: Closed and open conformation of the phenyl rings.

The geometries taken from the CSD have been fully optimized at B3LYP/cc-pVTZ(-PP) level of theory; the missing compounds, i.e. (p-NH_2-PhSe/Te)$_2$, (p-OCH_3-PhSe)$_2$, (p-NO_2-PhTe)$_2$, and (2-NaphSe)$_2$, were generated from the crystallographic structures of the parent compounds: the substituents were added manually; a full geometry optimization was then performed. The calculated structures are in very good agreement with the experimental ones and the relevant geometric features are maintained (Table 5.1). The largest differences appear when comparing the Φ values, but, as demonstrated in the next paragraphs, very little energy is required to rotate the phenyl rings and crystal packing as well as solvent effects might strongly influence this geometrical feature. We have immediately noticed that the crystallographic structures of (PhSe)$_2$ and (PhTe)$_2$ exhibit closed and open conformation, respectively. Thus, an investigation to assess the importance of this issue and validity of the QM method has been conducted through prelimi-

	CSD Identifier	X-X'	C_X-X C_X'-X'	C_X-X-X' C_X'-X'-X	C_X-X-X'-C_X' (Ψ)	C_1-C_X-X-X' C_1'-C_X'-X'-X $(\Phi_{1,2})$
(p-NH_2-PhSe)$_2$	—	2.39	1.93 1.93	103 104	-84	96 97
(p-NH_2-PhTe)$_2$	—	2.76	2.13 2.13	102 102	-85	92 93
(p-OCH_3-PhSe)$_2$	—	2.38	1.93 1.93	103 103	-86	97 97
(p-OCH_3-PhTe)$_2$	BAJHAV	2.75 (2.715)	2.14 (2.139) 2.14 (2.143)	102 (99.8) 102 (100.3)	-84 (-81)	93 (12) 95 (20)
(p-CH_3-PhSe)$_2$	QQQGBV01	2.38 (2.328)	1.93 (1.923) 1.93 (1.910)	103 (102.0) 103 (101.2)	84 (100)	83 (90) 83 (112)
(p-CH_3-PhTe)$_2$	DPTOTE	2.75 (2.696)	2.14 (2.125) 2.14 (2.131)	101 (101.5) 101 (100.0)	84 (86)	88 (37) 89 (37)
(p-Cl-PhSe)$_2$	CLPHSE	2.37 (2.332)	1.94 (1.929) 1.94 (1.911)	102 (98.6) 103 (98.9)	-83 (-74)	100 (119) 101 (74)
(p-Cl-PhTe)$_2$	CLPHTE01	2.74 (2.680)	2.14 (2.144) 2.14 (2.160)	101 (99.6) 101 (101.6)	-84 (-89)	95 (167) 96 (168)
(PhSe)$_2$	YUXPIR	2.37 (2.307)	1.94 (1.946) 1.94 (1.934)	102 (102.4) 103 (103.5)	84 (85)	80 (160) 80 (0)
(PhTe)$_2$	DPHDTE01	2.74 (2.707)	2.15 (2.131) 2.15 (2.115)	101 (100.7) 101 (97.7)	-85 (-90)	93 (95) 93 (90)
(p-NO_2-PhSe)$_2$	DUWKEL	2.33 (2.302)	1.94 (1.921) 1.94 (1.921)	105 (103.2) 105 (103.2)	-89 (-88)	6 (23) 5 (23)
(p-NO_2-PhTe)$_2$	—	2.71	2.15 2.15	103 103	-91	6 6

Table 5.1: Relevant interatomic distances (Å) and angles (°) of the fully optimized dichalcogenides at B3LYP/cc-pVTZ(-PP) level of theory. When available, parameters from the crystallographic structures taken from CSD are given in parentheses.

nary optimizations and Ψ, Φ relaxed scans for all the compounds at B3LYP/6-311G(d,p),SDD level (SDD is the Stuttgart-Dresden core potential basis function employed for the chalcogens). The resulting chalcogen-chalcogen bond length was not accurately described, differing up to 0.2 Å from the crystallographic data (Table D.7). In addition, we could find only minima with open conformation. Full geometry optimization of the parent compounds at MP2/6-311G(d,p),SDD level of theory was then attempted; both $(PhSe)_2$ and $(PhTe)_2$ converged with the conformation of the phenyl rings in agreement with the crystallographic data, i.e. closed and open respectively, but for $(PhTe)_2$, a value of -34° was computed for Ψ, which deviates too much from the experimental value (-90°). Finally we tried B3LYP/cc-pVTZ(-PP) which instead, gave good structures and represented a good compromise between accuracy and complexity. At this level of theory we succeeded in optimizing both open and closed conformations for the parent compounds, although the former always corresponds to the global minimum *in vacuo* (Tab. 5.1).

5.2.3 Quantum Potential energy surfaces

We generated different structures for the fit procedure doing relaxed scans about the Φ and Ψ dihedrals. It is interesting to notice that the rotational barriers for Φ show a dependence on the functional group in para position with the following trend NH_2 > CH_3O > CH_3 $\sim$ Cl > H > NO_2 (Fig. 5.8 and Tab. 5.2), which follows qualitatively the Hammett series (with the exception of Cl that should fall between H and NO_2) for substituents in para position. Indeed the more negative σpara is, the stronger electron donor the substituent is and this leads to a higher rotational barrier, e.g. in the case of the amino group, reaching at maximum about 2.6 Kcal/mol and 2.1 Kcal/mol for the Se and Te compounds respectively. Tellurium derivatives have a slightly lower rotational barrier. In the case of the para-nitro compounds there is also a shift of the PES along the abscissa by 90°, because, differently from the others, these molecules are more stable in the closed conformation. It is noteworthy that the motion of rotation of the phenyl rings about their axes in diphenyldichalcogenides was studied about 40 years ago by measuring the spin-lattice relaxation times (T1) of ^{13}C.[196] Using a rigid anisotropic model with Ψ set equal to 90°, the tumbling of the phenyl rings was investigated and found to be less and less hindered when going from S to Se and Te, a result which is in agreement with our findings on the energetics of these compounds. Instead, for the Ψ dihedrals, a preliminary inspection carried out at the B3LYP/SDD,6-311G(d,p) level of theory did not show any significant differences due to the substituent (Fig. D.1); thus the PES for rotation of the Ψ dihedral computed at B3LYP/cc-pVTZ(-PP) level of theory for the two parent compounds is employed for the substituted derivatives too. The height of the Ψ rotational barriers for the parent compounds is approximately one order of magnitude higher than for the Φ dihedral.

To obtain the missing parameters in GAFF, i.e. the charges and the Lennard-Jones constants, as well as the parameters required to describe the stretching, bending and torsional contributions to the potential involving at least one chalcogen atom, the following procedures were car-

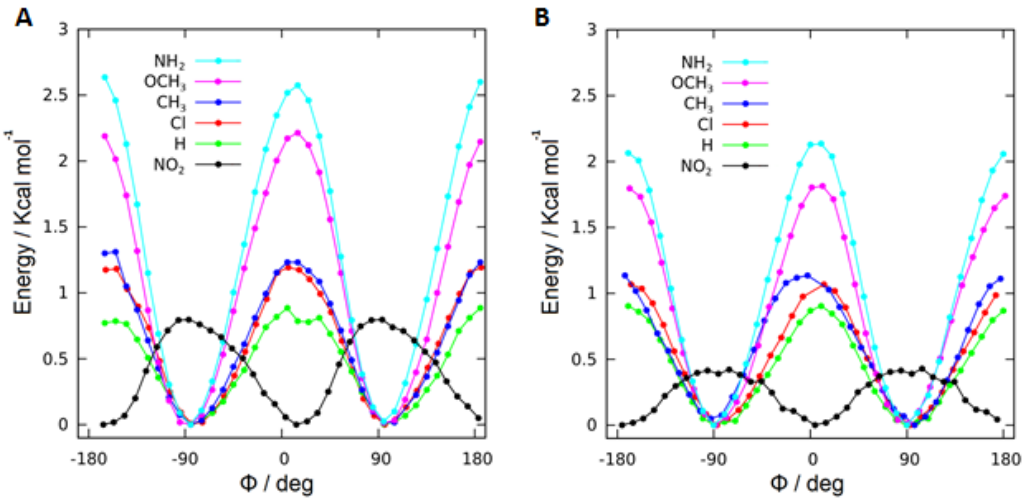


Figure 5.8: Energy profiles at B3LYP/cc-pVTZ(-PP) level of theory for the rotation of Φ of all Se (A) and Te (B) diphenyl compounds.

substituent	NH_2	CH_3O	CH_3	Cl	H	NO_2
barrier height	2.6	2.2	1.3	1.2	0.9	0.8
(Kcal/mol)	(2.1)	(1.8)	(1.1)	(1.1)	(0.9)	(0.4)
σ_{para}	-0.66	-0.27	-0.17	+0.23	0.00	+0.78

Table 5.2: Torsional barriers related to Φ and Hammet sigma constants for substituents in para position; values for Te compounds are given in brackets.

ried out.

Derivation of atom charges and estimate of Lennard-Jones parameters

The atom charges were first computed for all the optimized geometries of multiple conformers, at HF/6-31G(d),cc-pVTZ(-PP) level, fitting the ESP using the electrostatic charge computing method by Merz and Kollman.[118] This QM level of theory was chosen to be consistent with previous parametrization;[186] exceptionally, since the 6-31G(d) basis set is not available for Te, we have employed cc-pVTZ(-PP)[1] for Te and Se as well. First of all the number of layers and grid points for the electrostatic potential fitting has been investigated for Se and Te derivatives. Increasing by 10 the number of layers and grid points beyond the recommended values (10 and 17, respectively), did not produce relevant difference in the charges and dipole moments of the parent compounds, the latter changing by less than 1%. Multiorientation effects were taken into account to obtain more robust partial charges using two conformations (open and closed) for the parent as well as for all substituted compounds. Atom-centered partial charges were then derived according to the RESP methodology outlined by Kollman et al.[118] In Fig. D.3, the

[1]Prior the use of the cc-pVTZ(-PP) basis set for Se/Te for the determination of the RESP charges, we compared the charges of $(PhSe)_2$ calculated (i) at the usual HF/6-31G(d) level and (ii) at HF/6-31G(d),cc-pVTZ level, where the latter basis set has been used only for the Se atom. The relative differences of both the ESP charges and dipole moments were less than 3%.

electrostatic surface potential and the RESP charges of the parent compounds are shown; the RESP charge values for all compounds are listed in Table 5.3. Constraints were used during the fitting, so that chemically equivalent atoms have the same charge. The charges on the chalcogens become approximately less negative when going from the strongest electron donor to the strongest electron withdrawing group in the series.

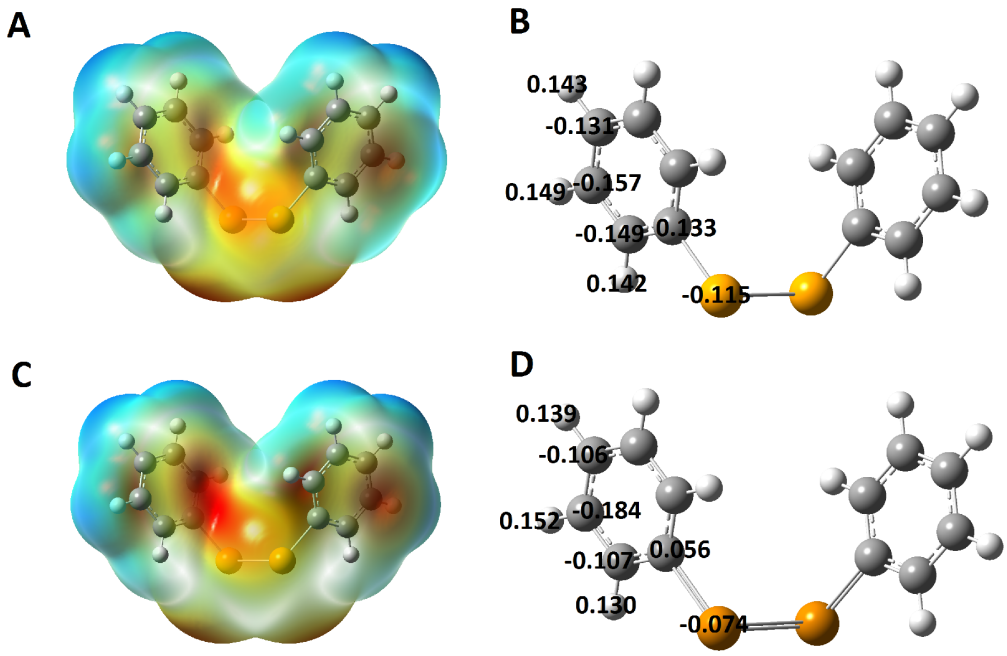


Figure 5.9: Parent compounds: electrostatic surface potential and RESP charges of (PhSe)2 (a and b) and (PhTe)2 (c and d); the global minimum geometry is used for the representation.

The ϵ parameters were taken equal to those included in the UFF,[197] instead for σ we found that σ_{UFF} values divided by 2 shows excellent linear correlation with the σ_{GAFF} of C, H, N, O, P, S (Fig. D.2, Table D.6); so the σ parameters for Se and Te were set equal to half the σ_{UFF} values.[2] For the other atoms we kept the GAFF parameters. The cross terms were obtained using the Lorentz/Berthelot mixing rules, i.e. $\sigma_{ij} = 0.5(\sigma_{ii} + \sigma_{jj})$ and $\epsilon_{ij} = \sqrt{(\epsilon_{ii} \cdot \epsilon_{jj})}$.

Parametrization of the harmonic contributions

As stated above, in agreement with the AMBER philosophy, the stretching and bending parameters have been determined only for the parent compounds and transferred to all twelve molecules reducing the number of the new atomtypes. The first harmonic term in the AMBER potential

[2]To the best of our knowledge, these parameters are not available in literature. In L. Zhao, A. G. Cox, J. A. Ruzicka, A. A. Bhat, W. Zhang, E. W. Taylor PNAS 2000 97 6356–6361 the same LJ parameters of Zn, developed by S. Amirjalayer, M. Tafipolsky and R. Schmid (Angew. Chem. Int. Ed. 2007, 46, 463–466), are employed for Se, that is σ=2.29 Å and ϵ=0.276 Kcal/mol, and no validation is provided. Our strategy of using these modified UFF parameters has been very recently employed by some of us for the parametrization of ff14SB of Sec and Tec. Among the preliminary results of extensive MD simulations, carried out to study the Se and Te based glutathione peroxidase, we observe the preservation of the catalytic pocket in which Sec/Tec are buried, and this enhances our confidence in our choice of the LJ parameters.

	X	C_X	C1(=C5)	C2(=C4)	C3	H1(=H4)	H2(=H3)		R(=R')	
							N	H		
(p-NH_2-PhX)$_2$	-0.140095	-0.119705	-0.142116	-0.356906	0.473450	0.160820	0.190714	-0.910606	0.376261	
	(-0.100312)	(0.035058)	(-0.090951)	(-0.386461)	(0.492143)	(0.147236)	(0.194704)	(-0.903764)	(0.373910)	
								O	C	H
(p-OCH_3-PhX)$_2$	-0.122505	0.036504	-0.072294	-0.399946	0.530932	0.152239	0.190539	-0.396038	-0.015674	0.075235
	(-0.084124)	(-0.025935)	(-0.050396)	(-0.408257)	(0.543914)	(0.147382)	(0.189615)	(-0.411340)	(0.004244)	(0.072184)
								C	H	
(p-CH_3-PhX)$_2$	-0.121651	-0.113947	-0.137652	-0.323703	0.329828	0.155804	0.178847	-0.426264	0.119183	
	(-0.079513)	(0.027097)	(-0.088486)	(-0.353883)	(0.359165)	(0.143699)	(0.181825)	(-0.437973)	(0.121638)	
								Cl		
(p-Cl-PhX)$_2$	-0.109071	0.225231	-0.277640	-0.005680	-0.069194	0.177065	0.132893	-0.100242		
	(-0.168276)	(0.488516)	(-0.598903)	(-0.363502)	(0.209714)	(0.408684)	(0.381204)	(-0.184921)		
(PhX)$_2$	-0.115244	0.133007	-0.148740	-0.156617	-0.130699	0.141678	0.148814	0.142666		
	(-0.073755)	(0.056209)	(-0.106937)	(-0.183825)	(-0.105684)	(0.130557)	(0.152151)	(0.139338)		
								N	O	
(p-NO_2-PhX)$_2$	-0.059245	0.082713	-0.115092	-0.220180	0.114185	0.151408	0.200601	0.701682	-0.436404	
	(-0.008853)	(-0.013892)	(-0.059956)	(-0.255448)	(0.149797)	(0.137494)	(0.203458)	(0.690059)	(-0.434104)	

Table 5.3: RESP charges; values for Te compounds are given in parentheses.

(Eq. 5.1) represents the stretching of the bonds. Constrained scans, i.e. keeping all other degrees of freedom frozen, were performed starting from the optimized geometry and elongating the C_X-X and the X-X' bonds in 20 steps of 0.02 Å increments. These curves were truncated at the point at which the energy had increased roughly 5 Kcal/mol. The selected points (11 to 12 for C_X-X bonds, and 13 to 17 for X-X') were fitted using a quadratic polynomial, yielding by construction a value of req corresponding to the actual optimized equilibrium geometry. The same procedure was adopted for the C_X-X-X' and C_1-CX-X angles to parametrize the second harmonic term representing the bending of the angles; the corresponding equilibrium value for θeq was used in this case. For the C_X-X-X' angles, datapoints within 5 Kcal/mol of the equilibrium value were selected (resulting in 9 to 13 points), whereas for the more rigid C_1-C_X-X angles the threshold was raised to 20 Kcal/mol in order to have sufficient datapoints (10 to 12). The parameters obtained for the parent compounds are listed in Table 5.4. When going from Se to Te the stretching force constants become smaller, due to the weakening of the chalcogen-chalcogen and carbon-chalcogen bonds. This trend is in agreement with the force constants of the analogous S-S (r_{eq}=2.050 Å; b=161.7 Kcal/mol Å^{-2}) and C(aromatic)-S (r_{eq}=1.770 Å; b=256.6 Kcal/mol Å^{-2}) bonds already present in GAFF, which have the largest force constants.

$(PhX)_2$									
Bonds		Angles		Dihedrals					
r_{eq}	b	θ_{eq}	a	IDIVF[a]	$V_1/2$	$V_2/2$	$V_3/2$	$V_4/2$	γ
x-x		x-x-ca		ca-x-x-ca					
2.330	109.0	104.53	67.13	1	-1.8498	-1.6208	-0.6648	0.0015	180
(2.743)	(71.6)	(100.92)	(40.47)	1	(-1.4681)	(-2.7095)	(-0.5895)	(0.0168)	180
x-ca		x-ca-ca		x-x-ca-ca					
1.948	143.4	124.14	86.16	1	2.5629	0.2734	0.8091	0.0679	180
(2.145)	(122.8)	(119.94)	(78.24)	1	(0.2825)	(0.0821)	(-0.0954)	(0.0650)	180
ca-ca		ca-ca-ca		x/ca-ca-ca-ca					
1.387	478.4	119.97	67.18	x/ca/ha-ca-ca-ha					
ca-ha		ca-ca-ha		4	–	14.5	–	–	180
1.087	344.3	120.01	48.46						

Table 5.4: Force-field parameters for the parent compounds; values for Te compounds are given in parentheses. GAFF atomtypes nomenclature is used: C_X=ca, H=ha. a This is the factor which divides the barriers.

Parametrization of the torsional contributions

After deriving the charges, the parameters of the harmonic contributions and the LJ constants for Se and Te, the determination of the parameters for the Ψ and Φ dihedrals was left. The torsional term of Eq. 5.1 must be obtained considering the difference between the quantum mechanics PES and the molecular mechanics PES computed using the complete potential without the torsional contribution. In fact, this last term is the correction to the molecular mechanics PES that will allow to reproduce the accurate quantum potential. Given the preliminary finding that the

Ψ rotation PES did not show significant differences (Figure D.1) when varying the substituent in para position, we calculated the potential only for the parent compounds. Conversely, Φ was computed for each molecule, that is all the bonded FF parameters obtained for diphenyldiselenide and diphenylditelluride have been transferred to the other molecules except for the Φ dihedral. In order to obtain the quantum PES, relaxed scans were run separately for rotation around the Ψ and Φ dihedrals starting from the global minima; the rotation was done increasing each dihedral separately with 35 steps of 10°. All the energies and structures were calculated at B3LYP/cc-pVTZ(-PP) level of theory. Afterwards, they were extracted from the Gaussian output file and converted to a suitable format (i.e. .dat and .mdcrd files) manageable with the Paramfit suite,[119] using an in-house C++ program (provided in the appendix D). For all the molecules the Fourier expansion was truncated to the fourth term, in order to get a reasonably low value of the least square sum (less than 1 Kcal/mol). The Vi,n parameters were optimized, while setting all the γ phases equal to 180°, because their optimization was not necessary; this strategy allowed to keep a low number of free parameters. The fitting of the quantum PES is very good; the conformational space sampling of the parent compounds based on the QM scans, which can be inspected in Fig. 5.10a and 5.10b, resulted is of high quality.

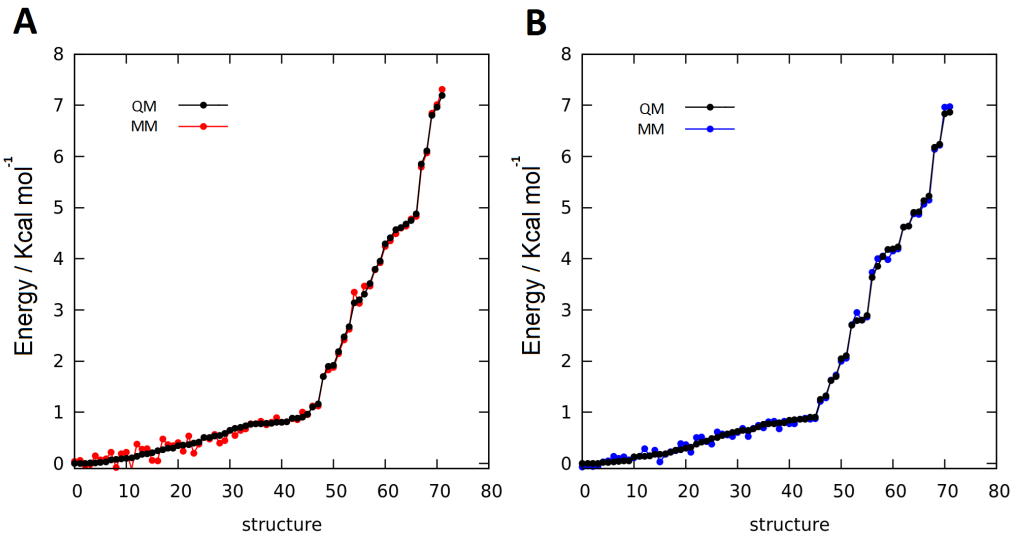


Figure 5.10: Conformational space sampling based on the QM scans of $(PhSe)_2$ (A) and $(PhTe)_2$ (B).

Based on AMBER philosophy, the parameters of Table 5.4 are valid for the class of compounds here described. In any case, due to the low barriers associated to the rotation of the phenyls about their axes, for an accurate description we recommend to use the parameters reported in Tables D.1-D.5, which have been derived taking into account the dependence of the energy profile upon Φ for all the different substituted dichalcogenides.

Validation of the new GAFF parameters

Unluckily there are no thermodynamic data (or other suitable experimental result) available for these molecules to validate the derived force field parameters, mainly because these compounds are used in in vitro/in vivo tests or catalytic processes. Thus, as first benchmark of our novel GAFF parameters, we calculated the molecular dipole moment from the optimized MM models. We compared the electrostatic dipole moment of all twelve diphenyl derivatives calculated using (A) the *ab initio* density and dipole integrals computed at HF/6-31G(d),cc-pVTZ(-PP) level with the B3LYP/cc-pVTZ(-PP) optimized structures to (B) the values obtained using RESP atom-centered point charges and the MM optimized molecular structure (minimization of a single molecule *in vacuo*). The results are shown in Table 5.5. Comparison between the dipole moments shows an average difference of about 0.3 Debye, with the highest discrepancy for (p-NH2-PhTe)2 equal to 0.66 Debye. Taking into account that the QM value is based on a single conformer, while MM value is calculated for the same conformer but the RESP charges were obtained by a multiconfigurational fit, we consider these discrepancies acceptable. The agreement improves even further when the MM value is obtained using the RESP charges fitted for only the closed conformation.

	(p-NH_2-PhX)$_2$	(p-OCH_3-PhX)$_2$	(p-CH_3-PhX)$_2$	(PhX)$_2$	(p-Cl-PhX)$_2$	(p-NO_2-PhX)$_2$
A	4.26	1.76	2.66	2.11	0.90	5.10
	(2.34)	(1.40)	(2.49)	(1.86)	(1.10)	(5.73)
B	4.55	1.35	3.11	2.54	0.63	5.02
	(3.00)	(1.51)	(2.82)	(2.29)	(0.84)	(5.70)

Table 5.5: Molecular dipole moments in Debye calculated with *ab initio* density and dipole integrals using the B3LYP/cc-pVTZ(-PP) equilibrium geometry (A) and from RESP charges using the MM equilibrium geometry (B). Values for the Te compounds are given in parentheses.

5.2.4 Molecular dynamics distributions

We furthermore tested our extension of GAFF by performing four MD simulations of the parent compounds in two different solvents, i.e. chloroform[192] and TIP3P water.[193] In each simulation a single molecule in a box was considered in order to avoid the formation of aggregates that may lead to unbalanced distributions. The MD simulation was carried out for 500 ns. The X-X bond length is very well maintained in the two different solvents with the average values falling at 2.33 Å and 2.74 Å for (PhSe)$_2$ and for (PhTe)$_2$, respectively. The same holds for the C$_X$-X bonds and C$_X$-X-X' and C$_X$'-X'-X angles, the average values of which are approximately equal to 1.95 Å, 105° for (PhSe)$_2$ and 2.14 Å, 102° for (PhTe)$_2$. From the analysis of the trajectories it emerges that Ψ values fall at $\pm$90° for both parent compounds in both solvents (Fig. 5.11 and Fig. D.2). The Φ values exhibit very similar distributions for both compounds in chloroform (Fig. 5.12), where the closed conformer is more easily found (peaks at 0°,$\pm$180°). Instead, using

TIP3P water, we find that $(PhSe)_2$ has the same behaviour, while $(PhTe)_2$ surprisingly assumes more frequently the open conformation (peaks at $\pm 90^\circ$). These results, in agreement with the crystallographic data (Table 5.1), were not found using the QM calculations *in vacuo* and highlight that the stabilization due to the medium is a very important ingredient to be taken into account when involved energies are in the order of few Kcal/mol.

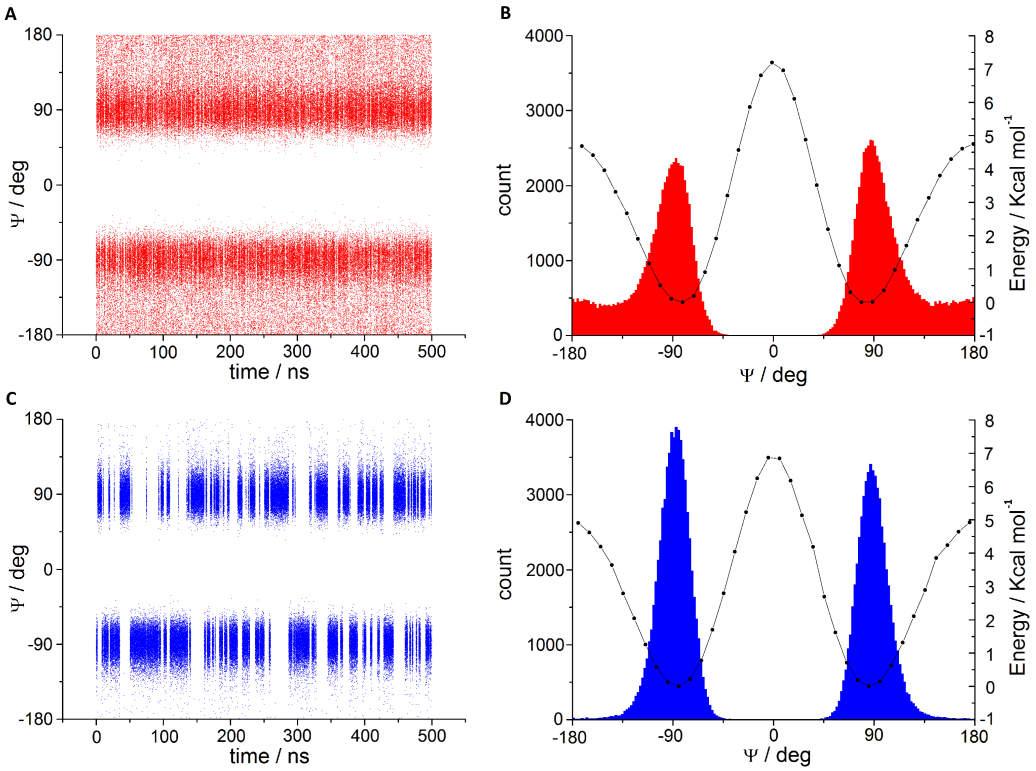


Figure 5.11: Trajectories (a and c) and distributions/QM PES (b and d) of Ψ in water for $(PhSe)_2$ (a and b) and $(PhTe)_2$ (c and d); the vertical axis on the left is used for the distributions and the vertical axis on the right is used for the energies.

We have extended the popular GAFF to a series of diphenyldiselenides/-ditellurides. Besides the well known use in organic catalysis of Se compounds, recently evaluated also in green reactions, this class of Se and Te compounds has significant anti-oxidant peroxidase activity, and some of them also have efficient anti-tumor and anti-microbial properties. The intrinsic symmetry, the presence of the chalcogen-chalcogen bond and the aromatic rings with different para-substituents make these compounds and their chemistry intriguing. After testing different levels of theory, i.e. B3LYP/6-311G(d,p),SDD, MP2/6-311G(d,p),SDD and B3LYP/cc-pVTZ(-PP), we chose B3LYP/cc-pVTZ(-PP) which offers good geometry accuracy at an affordable computational cost. Based on QM calculations we proved that

1. the barrier associated to the rotation of the phenyl rings about their axis is small

2. this barrier is affected by changes in the electron-donating/withdrawing character of the para substituents, the trend reflecting quite well the Hammett series.

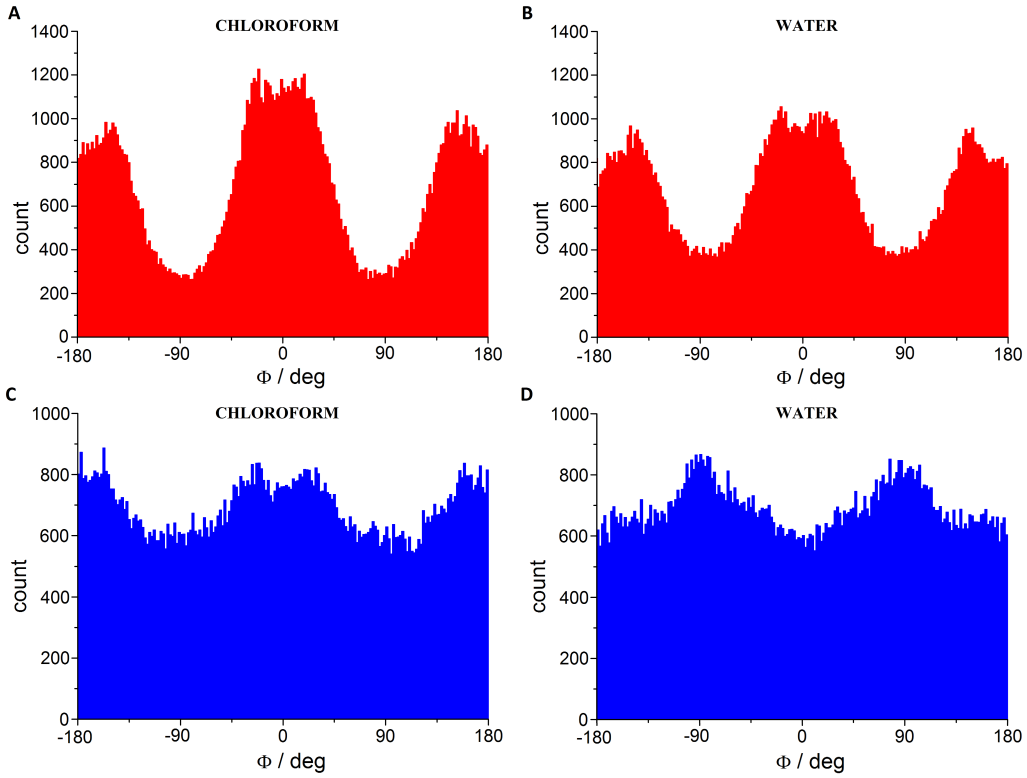


Figure 5.12: Distributions of Φ in chloroform and water for $(PhSe)_2$ (a, b) and $(PhTe)_2$ (c, d).

Due to the lack of experimental data we used QM dipole moments as a benchmark: molecular dipole moments calculated at MM level of theory with the novel parameters well reproduce the QM values obtained using the HF/6-31G(d),cc-pVTZ(-PP) density and dipole integrals evaluated on the B3LYP/cc-pVTZ(-PP) optimized structures. In addition, MD simulations in chloroform and water show that the main structural features of the parent Se and Te compounds are conserved in presence of explicit solvent. From the trajectories frequent jumps between the two more stable conformers (Ψ roughly $\pm 90^{\circ}$), are observed, as expected from the values of the lowest interconversion barrier (about 5 Kcal/mol). This is drawn by the two peak distribution of Ψ, barely different for $(PhSe)_2$ and $(PhTe)_2$. In contrast, the distributions of Φ are less sharp and the closed conformer is more frequently found in chloroform. ImportantImportantly, upon switching the solvent to water, the $(PhTe)_2$ compound is found more frequently in the open conformation, while for $(PhSe)_2$ the distribution remains similar to that found in chloroform. This behavior, which incidentally reproduces the crystallographic finding, may be explained considering that the local solvent-solute interactions play an important role. Therefore, upon increasing the solvent polarity, this effect is emphasized, leading to different Φ distributions because the partial charges are neatly different for the two compounds. Ongoing work, i.e. bioincorporation of these drugs in a protein scaffold to design *de novo* semi-natural enzymes, will provide further insight into the structure and reactivity of these dichalcogenides, in order to rationalize and support the goldmine of available experimental data related to their biomimetic anti-oxidant activity.

Chapter 6

Continuum mechanics: microfluidic reactors

In this chapter physicochemical systems are described using continuum methods. The object of this study is the description of microreactors, useful tools widely employed for very different purposes in modern chemical applications. These objects are also know as “lab on a chip” with dimensions of the order of the millimeter and with section of the channels of about hundreds of micrometers. Thanks to their small dimensions and to the high surface/volume ratio, these devices exhibit many useful characteristics such as efficient thermal and mass exchange, low volumes of solvents employed and the fluid dynamics is dominated by dissipative forces instead of viscous forces, resulting in a laminar regime[198, 199]. In this context we study these devices adopting continuum methods, in order to model the transport and reaction of chemicals. We also attempt a general procedure for improving the performance of such devices adopting a very robust optimization technique, which relies on evolutionary algorithms, allowing to modify geometrical parameters to optimize some well defined observable (e.g reaction yield, outflows).

6.1 Fluid motion and transport equations

The Navier-Stokes (NS) equations are partial differential equations (PDE) used in computational fluid dynamics (CFD) to describe the motion of viscous fluids modeled as continuum media and were first developed by the French engineer Claude-Louis Navier in 1822 and re-derived later in 1845 by Irish mathematician-physicist George Gabriel Stokes. These equations are commonly solved numerically[1] and are widely used to study air, oceans, blood stream, design the shape of cars or aircrafts and for many other applications.

For an incompressible fluid and neglecting the gravity force, the NS equations may be expressed as follows

$$\rho\left(\frac{\partial}{\partial t}\mathbf{v} + \mathbf{v}\cdot\nabla\mathbf{v}\right) = \mu\nabla^2\mathbf{v} - \nabla p, \quad \nabla\cdot\mathbf{v} = 0 \tag{6.1}$$

[1]Their existence and smoothness has not been demonstrated yet and is part of one of the seven millennium prize problems. Mukhtarbay Otelbaev, originary from Kazakhstan, tried to solve this problem [200] but he is still working to fix his proof.

where $\mathbf{v} = \mathbf{v}(\mathbf{x}, t)$ is the velocity vectorial field, $p = p(\mathbf{x}, t)$ is the pressure scalar field, ρ is the density of the fluid, μ is the dynamic viscosity coefficient, ∇ is the gradient in cartesian coordinates given as

$$\nabla = \left(\frac{\partial}{\partial x}\hat{\mathbf{u}}_x, \frac{\partial}{\partial y}\hat{\mathbf{u}}_y, \frac{\partial}{\partial z}\hat{\mathbf{u}}_z \right)^T \tag{6.2}$$

and ∇^2 is the laplace operator, given as the divergence of the gradient viz.

$$\nabla^2 = \nabla \cdot \nabla = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \tag{6.3}$$

Usually together with the Navier-Stokes equations, in order to describe the behaviour of a complex reactor in the presence of chemical processes, the Advection-Diffusion-Reaction (ADR) are employed. They are PDE which describe chemical substances transported by the solvent motion, following the streamlines, coupled with diffusion and chemical transformation (reactions). Assuming constant (translation) diffusion coefficients and a standard description based on elementary chemical kinetics of the reaction mechanism, with given elementary reaction and relative chemical kinetics constants, the ADR equations for N chemicals can be expressed as

$$\frac{\partial}{\partial t}c_i + \mathbf{v} \cdot \nabla c_i = D_i \nabla^2 c_i + \mathbf{R}(c_i), \quad i = 1, 2, ..., N \tag{6.4}$$

where $c_i = c_i(\mathbf{x}, t)$ and D_i are the concentration scalar field and the translational diffusion coefficient of the i-th chemical substance and **R** represents the reaction operator (for a complete derivation of 6.4 see Appendix E). The reaction term is in general non linear[2], does not depend on space and time, and allows to couple together all the different reactants. It should be noticed that in the 6.1 and 6.4 the left terms represent the total derivative[3]. The (6.1)(6.4) are not sufficient for the description of the whole problem because the solution of a PDE problem is not unique in general, thus a set of additional constraints are required, which force the solution of the equations (or their derivatives) to be equal to some fixed values with a physical meaning, on the boundary regions. From now on we will refer to a generic scalar (or vectorial) field as u and to its domain and the boundary region as Ω and $\partial\Omega$ respectively. The main types of boundary conditions (BC) may be expressed in general as

$$u|_{\partial\Omega} = a \tag{6.5}$$

$$\nabla_n u|_{\partial\Omega} = b \tag{6.6}$$

$$(au + b\nabla_n u)|_{\partial\Omega} = g \tag{6.7}$$

[2] In some cases the reaction term may be linear, e.g. for unimolecular reaction $A \rightarrow P$

[3] The total derivative $d/dt = \partial/\partial t + v \cdot \nabla$ is also known as material derivative or lagrangian derivative because in continuum mechanics it describes the time evolution of a volume particle that follows the field lines.

where ∇_n is equal to the projection of the gradient to the normal outgoing from the domain **n**, i.e. $\nabla_n u = (\nabla u) \cdot \mathbf{n}$. The (6.5) is the simplest constraint known as Dirichlet BC, the (6.6) does not impose a constraint on the value of u but on its derivative and in particular in this case on the normal outgoing component of the gradient of u and is known as Neumann BC; the (6.7) is a mixing of the (6.5)(6.6) and is known as Robin BC.
In order to solve the NS problem the BC counditions were imposed as follows

- for the pressure p
  - null gradient everywhere, namely $\nabla_n p|_{\partial\Omega} = 0$, except at the outlet;
  - null value at the outlet, i.e. $p|_{\partial\Omega_{\text{outlet}}=0}$[4].
- for the velocity **v**
  - no-slip condition $\mathbf{v}|_{\partial\Omega} = 0$ is imposed everywhere exept at the inlets and outlets;
  - inflow condition at inlet $\mathbf{v}|_{\partial\Omega_{\text{inlet}}} = \mathbf{v}_{\text{in}}$
  - null gradient at outlet $\nabla_n \mathbf{v}|_{\partial\Omega_{\text{outlet}}} = 0$

these conditions state that close to the boundary the wall and the fluid move with the same velocity and the wall prevent fluid from leaving the domain[5]; moreover these BC require that we only know the fluid velocity at the inlet while the pressure may change freely during the resolution of the problem.
Similarily for the ADR equations another set of BC must be imposed for the concentration c_i or its flux j_i, in order to be able solving the problem

- outgoing flux null everywhere, i.e. $\mathbf{j}_n|_{\partial\Omega} = 0$
- inflow at inlets $\mathbf{j}_n|_{\partial\Omega_{\text{inlet}}} = \mathbf{j}_{\text{in}}$
- outflow condition at outlets, i.e. $\mathbf{j}_n|_{\partial\Omega_{\text{outlet}}} = \mathbf{j}_{\text{out}}$

these conditions state that the wall prevent the chemicals from leaving the domain and that we know the inflow of the reactants. In practice to impose the inflow condition it is sufficient to impose a Dirichlet BC on the concentration c_i because the inflow is given as $\mathbf{j}_{\text{in}} = c\mathbf{v}_{\text{in}}$. As we will see in chapter 6.2, physical and numerical BC may be slightly different. Finally the initial conditions must me imposed and in our case it is sufficient to set null velocity and concentrations in all the domain, i.e. $\mathbf{v}|_\Omega = 0$ and $c_i|_\Omega = 0$. So given all the initial and boundary conditions we were able to treat the whole problem, solving the NS equations first until steady state, in order to obtain **v** and p, afterwards the velocity **v** was used to solve the ADR equations and to obtain

[4] In principle the pressure at outlet should be set to the atmospheric pressure, but this is not necessary because within this context all the pressure field will be shifted by the atmospheric pressure, obtaining the same result up to an additive constant.

[5] There may be some cases in which this condition does not hold.

c_i until steady state was reached too. In this model we neglected the possible production of heat, due to the reactions that could modify the velocity field and thus required a simultaneous solution of both NS and ADR equations. This condition holds with diluted solutions and if the produced heat is easily dissipated through the system, both easily verified if we are modeling microfluidic devices.

6.2 Numerical Methods

The equations previously introduced cannot be solved analytically, except for a few cases, with simple geometries, thus for general and practical purpose we need a numerical scheme to manage the resolution of the problem. There are many different types of numerical schemes, each with its own strengths and weaknesses. First of all it is necessary to introduce the concept of discretization, so that instead of determining the solution in each point of space and time the PDE are reduced to a system of algebraic equations that may be easily solved with the computer. So with this reduction, instead of calculating the solution for infinite points that compose the domain, we just need to determine a given number of solutions to some given locations in both space and time. Various discretization schemes exist and the most famous are: finite differences, finite volumes, finite elements and spectral elements. Different methods can be combined. For instance, in this work the NS equations were solved via a finite element method, the ADR equations via a finite difference approach.

6.2.1 Navier-Stokes resolution

The NS equations were solved using an existing open source software known as OpenFOAM [201], which is a library package based on different 'solver' subpackages and allows to treat several type of physical problems, going from combustion, to compressible/incompressible fluids, particle tracking, electromagnetism, molecular dynamics and Monte Carlo. The method used for the resolution of the problem will be briefly discussed in the following sections without too many technicalities, to focus on the chemical aspects of the problem.
The solver chosen to treat the NS equations was SIMPLE (Semi-Implicit Method for Pressure Linked Equation), which allows to deal with stationary incompressible fluids. It is called semi-implicit because the discretized momentum and pressure correction equations are solved implicitly, instead the velocity correction is solved explicitly (explicit/implicit methods are discussed at page 89). The NS equations for an incompressible fluid are expressed as

$$\frac{\partial \mathbf{v}}{\partial t} + \nabla \cdot (\mathbf{vv}) = -\nabla p + \nabla \cdot (\nu \nabla \mathbf{v}), \quad \nabla \cdot (\rho \mathbf{v}) = 0 \tag{6.8}$$

given

$$H(\mathbf{v}) = -\sum_n a_n \mathbf{v}_n + \frac{\mathbf{v}^o}{\Delta t} \tag{6.9}$$

we can rewrite the momentum equation in a semidiscrete form

$$a_p \mathbf{v}_p = H(\mathbf{v}) - \nabla p \Rightarrow \mathbf{v}_p = \frac{H(\mathbf{v})}{a_p} - \frac{\nabla p}{a_p} \tag{6.10}$$

the continuity equation may be expressed in discretized form as

$$\nabla \cdot \mathbf{v} = \sum_f \mathbf{S} \cdot \mathbf{v}_f = 0 \tag{6.11}$$

where **S** is the face area vector, pointing outward and $\mathbf{v}_f$ is the velocity on the face that may be expressed interpolating the semi discrete form

$$\mathbf{v}_f = \left(\frac{H(\mathbf{v})}{a_p} \right)_f - \frac{\nabla f}{a_f} \tag{6.12}$$

by substituting (6.12) into (6.11) we obtain the pressure equation

$$\nabla \cdot \left(\frac{1}{a_p} \nabla p \right) = \nabla \cdot \left(\frac{H(\mathbf{v})}{a_p} \right) = \sum_f \mathbf{S} \left(\frac{H(\mathbf{v})}{a_p} \right)_f \tag{6.13}$$

The SIMPLE algorithm is described as follows

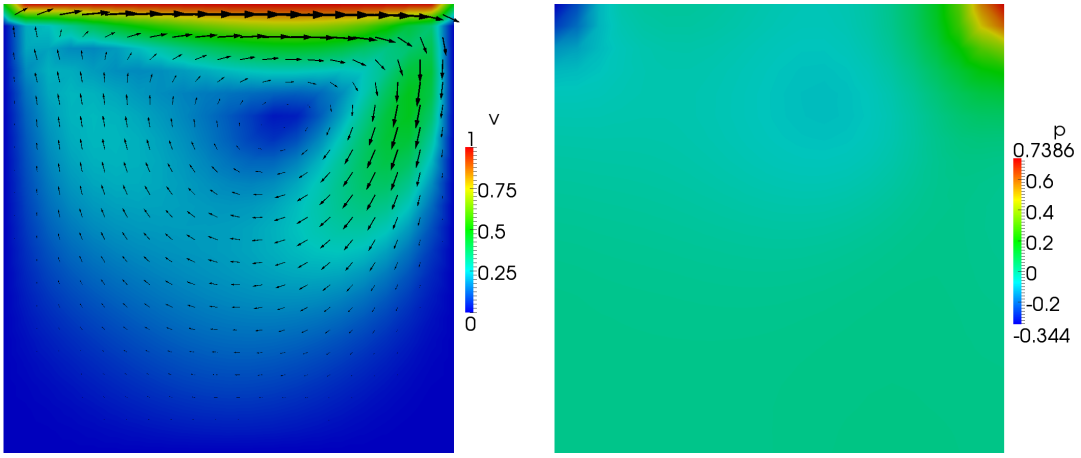
Algorithm 3 SIMPLE algorithm

1: Set the boundary conditions
2: solve (6.12) to compute the intermediate velocity field
3: calculate the uncorrected mass fluxes at cells faces
4: Solve the pressure equation to produce cell values of the pressure correction
5: update the pressure field applying under-relaxation $p^{\text{new}} = p^{\text{old}} + \alpha_{\text{UR}}(p^p - p^{\text{old}})$
6: update the boundary pressure conditions
7: Correct the mass fluxes at the cell faces
8: Correct the cell velocities on the basis of the new pressure field

In figure 6.1 are reported some examples of solutions of the Navier-Stokes equations for a bidimensional airfoil exposed to a fluid flow and the lid-driven cavity flow, where the motion inside the square is generated from the upper lid moving from left to right. Both the problems have been solved using the OpenFOAM [201] package.

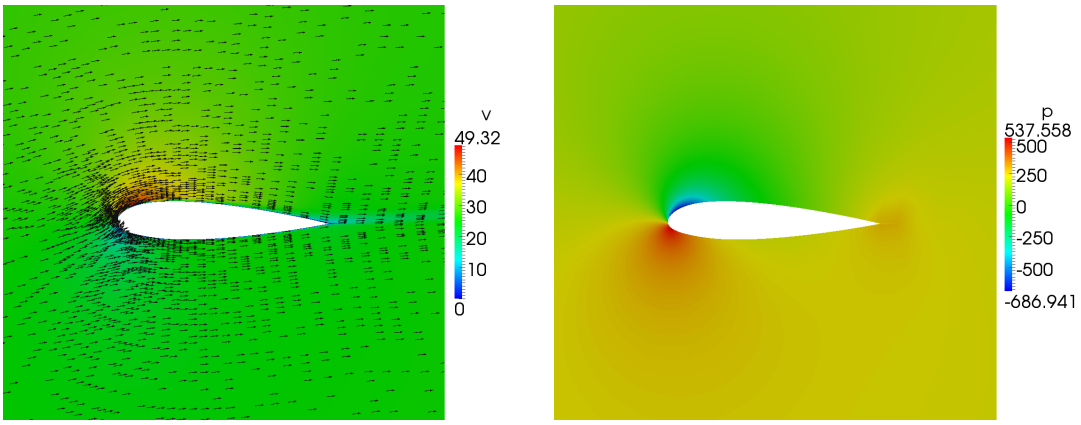
6.2.2 Advection-diffusion-reaction resolution

To solve the (6.4) we employ the finite difference method, based on the Taylor series expansion of the function, so that the differential operators is simplified by substituting derivatives with differential quotients. So given the previous introduced scalar (or vectorial) field u, we denote its discrete points in space and time as $u(x_i, y_j, z_k; t_n) =: u^n_{i,j,k}$. The partial derivative of u in



(a) velocity modulus scalar field and vectors

(b) pressure scalar field



(c) velocity modulus scalar field and vectors

(d) pressure scalar field

Figure 6.1: Lid driven cavity flow (a-b) and 2-dimensional airfoil (c-d) example problems. The solutions to the NS equations have been obtained using OpenFOAM [201].

time may be expressed as

$$\frac{\partial u}{\partial t} = \frac{u^{n+1} - u^n}{\Delta t} + \mathcal{O}(\Delta t) \tag{6.14}$$

where $\mathcal{O}$ is the so called big O term and gives an indication of the error as a function of Δt; if this term is sufficiently small we can approximate the (6.14) as

$$\frac{\partial u}{\partial t} \approx \frac{u^{n+1} - u^n}{\Delta t} \tag{6.15}$$

A similar expression is written for the partial derivative with respect to a given spatial coordinate. The second derivative with respect to x or any other spatial coordinate may be expressed as

$$\frac{\partial^2 u}{\partial x^2} = \frac{u_{i+1} - 2u_i + u_{i-1}}{(\Delta x)^2} + \mathcal{O}[(\Delta x)^2] \tag{6.16}$$

and if the error is sufficiently small as in the previous case

$$\frac{\partial^2 u}{\partial x^2} \approx \frac{u_{i+1} - 2u_i + u_{i-1}}{(\Delta x)^2} = \frac{\delta_x^2 u_i}{(\Delta x)^2} \tag{6.17}$$

where $\delta_x^2 u_i = u_{i+1} - 2u_i + u_{i-1}$. The same second derivatives are expressed along y, z axis, acting on the j, k indexes. From now, instead of using the general letter u, we will use c because we will deal with concentrations.

Mesh

The domain is subdivided using a Cartesian staggered grid given as

$$(x, y, z; t) \rightarrow (x_i, y_j, z_k; t_n) \tag{6.18}$$

$$x_i = \frac{\Delta x}{2} + i\Delta x, \;\; 1 \leq i \leq N_i$$

$$y_j = \frac{\Delta y}{2} + j\Delta y, \;\; 1 \leq j \leq N_j$$

$$z_k = \frac{\Delta z}{2} + k\Delta z, \;\; 1 \leq k \leq N_k$$

where $N_{i,j,k}$ are the maximum grid points along x, y, z; the time grid instead is given as $t_n = n\Delta t$, $n \geq 0$. We refer to the whole numeric scheme as cell-centered finite difference (CCFD) where the spatial discretization has been done using the grid points centered in the cells. In order to evaluate the advective fluxes, the velocities were linearly interpolated, doing an average, between 2 neighbor grid points, in order to obtain the net fluxes on the cell edges. At each timestep and position in the grid, the concentration and velocities are denoted respectively as $c(x_i, y_j, z_k; t_n) =: c^n_{i,j,k}$ and $v(x_i, y_j, z_k) =: v_{i,j,k}$ as in Fig. 6.2 (it should be noticed that the velocity does not depend on time).

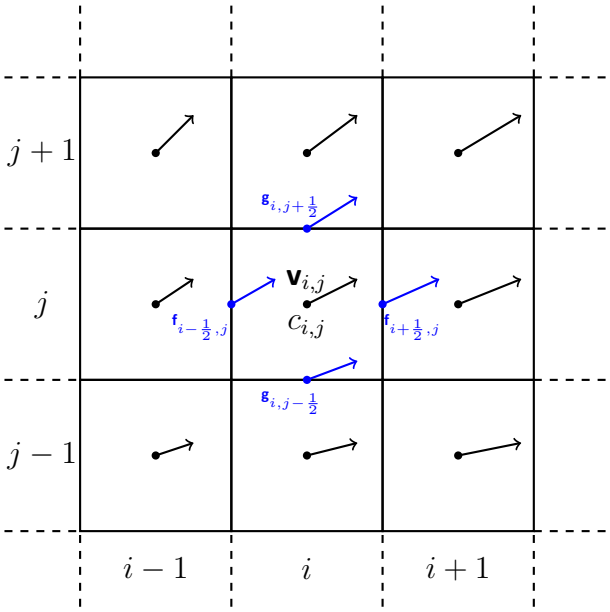


Figure 6.2: 2D representation of the mesh used for the CCFD calculations to solve the ADR equations.

Operator splitting

As already mentioned the whole problem is simplified by splitting it in two parts: the fluid motion and the chemical motion. In addition to this reasonable approximation we simplify further the ADR equations using the so called operator splitting method. Segregated solvers, usually employed in commercial software, may suffer from strongly coupled nonlinear problems (i.e. fast kinetics) and require a large amount of memory by the computing machine, because they often adopt direct matrix-solving algorithms. To overcome this kind of issues, the whole problem was decomposed in three subproblems of lower complexity, treating separately advection, diffusion and reaction processes and each associated to a corresponding operator which required few memory. Given the following ADR problem

$$\frac{\partial c}{\partial t} + \mathcal{L}_{\mathsf{ADR}} c = 0, \quad \mathcal{L} = \mathcal{L}_{\mathsf{A}} + \mathcal{L}_{\mathsf{D}} + \mathcal{L}_{\mathsf{R}} \tag{6.19}$$

We subdivide $\mathcal{L}_{\mathsf{ADR}}$ in three different operators that apply the advection, diffusion and reaction processes in a separated fashion. This approach permit to chose which process apply and eventually simulate limit cases with advection, diffusion or reaction only process or a combination of them. Each operator acts subsequently with a corresponding scheme that updates the concentration c from timestep n to timestep $n+1$ as

$$c^n \xrightarrow{\mathcal{U}_{\mathsf{A}}} c^{n+1/3} \xrightarrow{\mathcal{U}_{\mathsf{D}}} c^{n+2/3} \xrightarrow{\mathcal{U}_{\mathsf{R}}} c^{n+1} \tag{6.20}$$

$$c^{n+1/3} = \mathcal{U}_{\mathsf{A}}(c^n, \Delta t)$$

$$c^{n+2/3} = \mathcal{U}_{\mathsf{D}}(c^n, c^{n+2/3}, \Delta t)$$

$$c^{n+1} = \mathcal{U}_{\mathsf{R}}(c^{n+2/3}, c^{n+1}, \Delta t)$$

If the update scheme requires only the concentration at the previous step, it is **explicit**; otherwise if the scheme requires the concentration at the current time step it is **implicit**. If both concentrations at the previous and current time step are required, it is considered **semi-implicit**. As we will see the advection operator is explicit, the diffusion operator is semi-implicit and the reaction operator may be explicit or implicit depending on the solver used to treat the ODE problem. The splitting error is zero if all the operators commute and the final concentration does not depend on their order. The commutation conditions are the following [202]:

1. $\mathcal{U}_{\mathsf{A}}$ commutes with $\mathcal{U}_{\mathsf{D}}$ if both the velocity and the translational diffusion coefficient do not depend on the position; this is only true when the velocity field is constant in the space and the translational diffusion coefficient is a constant. For our purposes only the last condition is verified.

2. $\mathcal{U}_{\mathsf{A}}$ commutes with $\mathcal{U}_{\mathsf{R}}$ if the velocity field is divergence-free and $\mathcal{U}_{\mathsf{R}}$ does not depend on the position; this is only true with incompressible fluids because the divergence is null in each spatial point and in chemical reactions there is no dependence on the spatial coordinates.

3. $\mathcal{U}_{\mathsf{D}}$ commutes with $\mathcal{U}_{\mathsf{R}}$ if the reaction operator is linear in c and does not depend on the position; the reaction term is linear only when we consider a unimolecular reaction and it never depends on the spatial coordinates.

In our case the velocity field is divergence-free and the reaction is independent from the position, so only the second condition is verified, but this is not sufficient to guarantee commutation between any process.
The drawback of the operator splitting method is due to the splitting error. For the (6.20) update scheme the splitting error is $\mathcal{O}(\Delta t)$, so this is a first order scheme. By using the (6.20) we are losing precision because, as we will see, all the A, D, R operators are at least of second order. To avoid this loss we can use the improvement given by a more accurate scheme as the Strang splitting method [203], that allows to reduce the error to $\mathcal{O}[(\Delta t)^2]$ and recover the second order accuracy. To apply this method two operators must be applied twice with half time step, so it is convenient to chose the two with less computational effort; instead the remaining operator is applied only once. In the non-stiff case (for a short discussion on *stiff problems* see appendix F) we can use the explicit reaction operator, so that the advection operator is the most expensive. The five update of the concentration that we performed with the RDADR sequence was done in the following way

$$c^n \xrightarrow{\mathcal{U}_{\mathsf{R}}} c^{n+1/5} \xrightarrow{\mathcal{U}_{\mathsf{D}}} c^{n+2/5} \xrightarrow{\mathcal{U}_{\mathsf{A}}} c^{n+3/5} \xrightarrow{\mathcal{U}_{\mathsf{D}}} c^{n+4/5} \xrightarrow{\mathcal{U}_{\mathsf{R}}} c^{n+1} \tag{6.21}$$

$$c^{n+1/5} = \mathcal{U}_{\mathsf{R}}(c^n, \Delta t/2)$$

$$c^{n+2/5} = \mathcal{U}_{\mathsf{D}}(c^{n+1/5}, c^{n+2/5}, \Delta t/2)$$

$$c^{n+3/5} = \mathcal{U}_{\mathsf{A}}(c^{n+2/5}, \Delta t)$$

$$c^{n+4/5} = \mathcal{U}_{\mathsf{D}}(c^{n+3/5}, c^{n+4/5}, \Delta t/2)$$

$$c^{n+1} = \mathcal{U}_{\mathsf{R}}(c^{n+4/5}, \Delta t/2)$$

where the reaction updates are explicit. Instead for stiff problems we have to use the implicit reaction operator that is more computational expensive than the advection operator and the suitable update sequence will be ADRDA expressed as follows

$$c^n \xrightarrow{\mathcal{U}_{\mathsf{A}}} c^{n+1/5} \xrightarrow{\mathcal{U}_{\mathsf{D}}} c^{n+2/5} \xrightarrow{\mathcal{U}_{\mathsf{R}}} c^{n+3/5} \xrightarrow{\mathcal{U}_{\mathsf{D}}} c^{n+4/5} \xrightarrow{\mathcal{U}_{\mathsf{A}}} c^{n+1} \tag{6.22}$$

$$c^{n+1/5} = \mathcal{U}_{\mathsf{A}}(c^n, \Delta t/2)$$

$$c^{n+2/5} = \mathcal{U}_{\mathsf{D}}(c^{n+1/5}, c^{n+2/5}, \Delta t/2)$$

$$c^{n+3/5} = \mathcal{U}_{\mathsf{R}}(c^{n+2/5}, c^{n+3/5}, \Delta t)$$

$$c^{n+4/5} = \mathcal{U}_{\mathsf{D}}(c^{n+3/5}, c^{n+4/5}, \Delta t/2)$$

$$c^{n+1} = \mathcal{U}_{\mathsf{A}}(c^{n+4/5}, \Delta t/2)$$

Advection

Transport phenomena present spatial derivatives that require to solve numerically the related PDE. In order to do that we used the symmetrized dimensionally-split scheme proposed by Kuchařík et al. [204] combined with the superbee flux limiter, introduced by Roe [205], which is mandatory in order to prevent oscillations in the solution due to discontinuities of the concentration. This scheme is explicit and in order to prevent numerical instabilities, the Courant-Friedrichs-Lewy (CFL) condition must be satisfied. All the transport processes are described using the more general continuity equation given as

$$\frac{\partial c}{\partial t} + \nabla \cdot \mathbf{j}_{\mathsf{A}} = 0 \tag{6.23}$$

where $\mathbf{j}_{\mathsf{A}}$ is the advective flux; given the (6.14), If we rewrite the (6.23) for one dimension we obtain

$$\frac{c^{n+1}_{i,j,k} - c^{n}_{i,j,k}}{\Delta t} = -\frac{f(c^{n+1/2}_{i+1/2,j,k}) - f(c^{n+1/2}_{i-1/2,j,k})}{\Delta x} \tag{6.24}$$

so if we rearrange the (6.24) we obtain the three update scheme L^x, L^y and L^z (leap-frog corrector steps) defined using a conservative formulation, in order to produce advection along x, y and z respectively; they are very similar and act in the same way, except that the augmented or

diminished index are i, j and k respectively.

$$L^x c_{i,j,k}^{n+1} = c_{i,j,k}^n - \frac{\Delta t}{\Delta x}\left[f(c_{i+1/2,j,k}^{n+1/2}) - f(c_{i-1/2,j,k}^{n+1/2})\right] \tag{6.25}$$

$$L^y c_{i,j,k}^{n+1} = c_{i,j,k}^n - \frac{\Delta t}{\Delta y}\left[g(c_{i,j+1/2,k}^{n+1/2}) - g(c_{i,j-1/2,k}^{n+1/2})\right] \tag{6.26}$$

$$L^y c_{i,j,k}^{n+1} = c_{i,j,k}^n - \frac{\Delta t}{\Delta z}\left[h(c_{i,j,k+1/2}^{n+1/2}) - h(c_{i,j,k-1/2}^{n+1/2})\right] \tag{6.27}$$

where g and h are the fluxes along y and z. $L^{x,y,z}$ does not commute for different directions, so in order to symmetrize the application of the three operators, we do an average over all the 6 permutations using the following expression

$$c^{n+1} = \mathcal{U}_{\mathsf{A}}(c^n, \Delta t) = \frac{1}{6}\big(L^x L^y L^z + L^x L^z L^y + L^y L^x L^z + L^y L^z L^x + L^z L^y L^x + L^z L^x L^y\big)c^n \tag{6.28}$$

instead for the bidimensional case the symmetrization is simply expressed by

$$c^{n+1} = \mathcal{U}_{\mathsf{A}}(c^n, \Delta t) = \frac{1}{2}\big(L^x L^y + L^y L^x\big)c^n \tag{6.29}$$

f, g and h are the fluxes at the cell edges along x,y,z. The fluxes are calculated by applying the superbee flux-limiter that avoids oscillations due to possible shocks (i.e. discontinuities in the concentration). Low order schemes are diffusive, i.e. they smooth the concentration profile near discontinuities while higher order schemes are dispersive thus they produce spurious oscillations near discontinuities. The flux limiter permits to take advantage of the numerical diffusion produced by the low order schemes (Fig. 6.3A), avoids the oscillatory behavior of the higher order schemes (Fig. 6.3B) and, mixing these two low and high order schemes, produce nice results. indeed the numerical and analytical solution for the linear advection are barely distinguishable in Fig. 6.3C. The fluxes were be expressed as

$$f\big(c_{i\pm\frac{1}{2},j,k}^{n+\frac{1}{2}}\big) = \big(1 - \phi_{sb}(r_{i\pm\frac{1}{2}})\big)f^L(c_{i\pm\frac{1}{2},j,k}^n) + \phi_{sb}(r_{i\pm\frac{1}{2}})f^H(c_{i\pm\frac{1}{2},j,k}^n) \tag{6.30}$$

$$g\big(c_{i,j\pm\frac{1}{2},k}^{n+\frac{1}{2}}\big) = \big(1 - \phi_{sb}(r_{j\pm\frac{1}{2}})\big)g^L(c_{i,j\pm\frac{1}{2},k}^n) + \phi_{sb}(r_{j\pm\frac{1}{2}})g^H(c_{i,j\pm\frac{1}{2},k}^n) \tag{6.31}$$

$$h\big(c_{i,j,k\pm\frac{1}{2}}^{n+\frac{1}{2}}\big) = \big(1 - \phi_{sb}(r_{k\pm\frac{1}{2}})\big)h^L(c_{i,j,k\pm\frac{1}{2}}^n) + \phi_{sb}(r_{k\pm\frac{1}{2}})h^H(c_{i,j,k\pm\frac{1}{2}}^n) \tag{6.32}$$

where f^L and f^H are respectively the fluxes generated by the lower order upwind and the higher order Lax-Wendroff scheme, defined as follows [206]

$$f^L(c_{i+\frac{1}{2},j,k}^n) = v_{i+\frac{1}{2},j,k}^x \begin{cases} c_{i,j,k}^n, v_{i+\frac{1}{2},j,k}^x > 0 \\ c_{i+1,j,k}^n, v_{i+\frac{1}{2},j,k}^x < 0 \end{cases} \tag{6.33}$$

$$f^L(c^n_{i-\frac{1}{2},j,k}) = v^x_{i-\frac{1}{2},j,k} \begin{cases} c^n_{i-1,j,k}, v^x_{i-\frac{1}{2},j,k} > 0 \\ c^n_{i,j,k}, v^x_{i-\frac{1}{2},j,k} < 0 \end{cases} \tag{6.34}$$

$$f^H(c^n_{i+\frac{1}{2},j,k}) = v^x_{i+\frac{1}{2},j,k}\left[\frac{1}{2}(c^n_{i,j,k} + c^n_{i+1,j,k}) - \frac{v^x_{i+\frac{1}{2},j,k}\Delta t}{2\Delta x}(c^n_{i+1,j,k} - c^n_{i,j,k})\right] \tag{6.35}$$

$$f^H(c^n_{i-\frac{1}{2},j,k}) = v^x_{i-\frac{1}{2},j,k}\left[\frac{1}{2}(c^n_{i,j,k} + c^n_{i-1,j,k}) - \frac{v^x_{i-\frac{1}{2},j,k}\Delta t}{2\Delta x}(c^n_{i,j,k} - c^n_{i-1,j,k})\right] \tag{6.36}$$

where the upscript on the velocity indicates the modulus along that direction (e.g. v^x is the velocity modulus along the x direction). $g^{L,H}$ and $g^{L,H}$, are expressed in the same way but acting respectively on the j (y) and k (z) indexes (directions). The velocities at the edge position are calculated using the average between the velocities at subsequent grid position, for example along x direction as: $v^x_{i\pm\frac{1}{2},j,k} = \frac{v^x_{i\pm1,j,k}+v^x_{i,j,k}}{2}$. The superbee flux limiter function was defined as [205]

$$\phi_{sb}(r) = \text{max}\ [0, \text{min}\ (2r, 1), \text{min}\ (r, 2)];\ \lim_{r\to\infty}[\phi_{sb}(r) = 2] \tag{6.37}$$

r is the ratio of successive gradients on the concentration grid given as

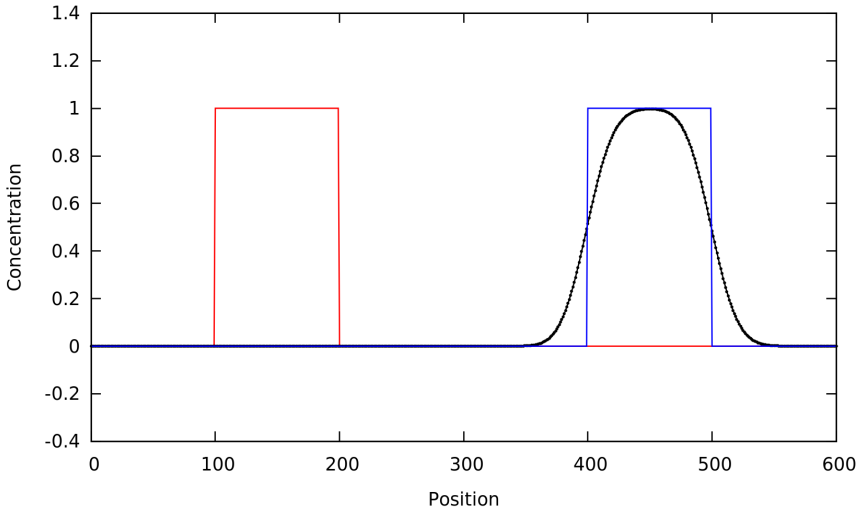
$$r_{i\pm\frac{1}{2}} = \frac{\pm c_{I\pm1} \mp c_I}{\pm c_{i\pm1} \mp c_i} \tag{6.38}$$

where I depends on the module of the velocity in the respective directions

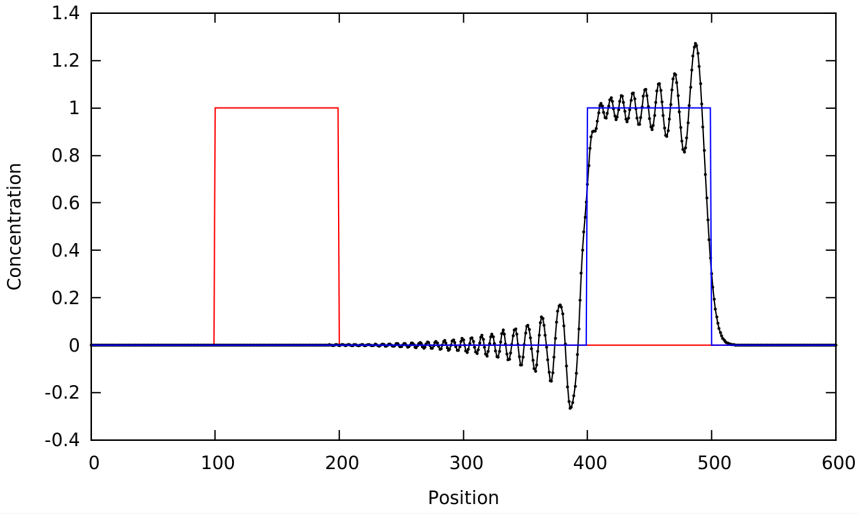
$$I = \begin{cases} i-1, v^x_{i\pm\frac{1}{2},j,k} > 0 \\ i+1, v^x_{i\pm\frac{1}{2},j,k} < 0 \end{cases} \tag{6.39}$$

The same function is applied along y and z using the j and k indexes. The superbee limiter function rely on the upper limit of the Sweby diagram [207] which describe a region that guarantees the scheme to be second-order and total variation diminishing (i.e. it preserves monotonicity of the concentration). The local truncation error of the method is $\mathcal{O}[(\Delta x)^2, (\Delta y)^2, (\Delta z)^2, (\Delta t)^2]$, so this is a second order scheme.

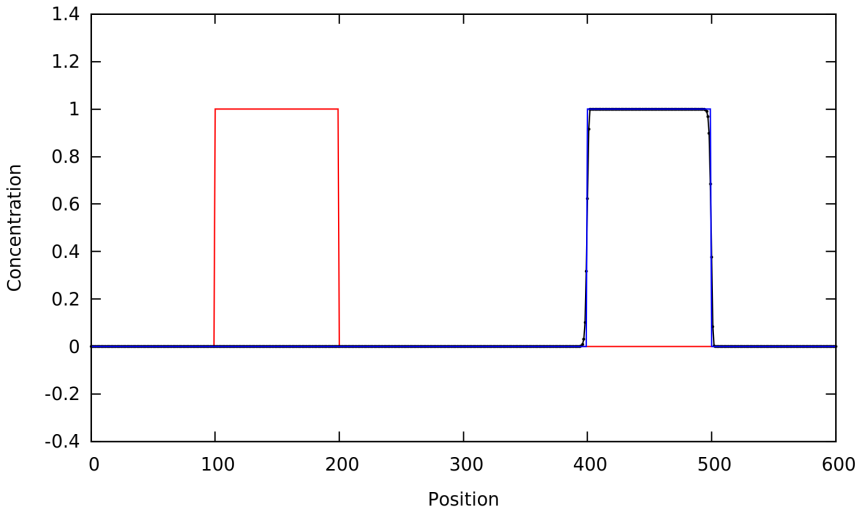
All these formulas refer to the x direction but the expressions for the y and z directions may be easily obtained swapping the i (x), j (y) and k (z) subscripts (superscripts). Usually a method with 3 (5) points stencil requires the Courant number to be less than or equal to 1 (2). The flux-limiter method we used, has a 5 points stencil as shown in Fig.6.4 but there is no relaxation in the CFL condition because the additional information obtained with 2 additional points, are only used to limit the higher order flux, thus the Courant number is Cr=1 [206]. The advection scheme is explicit therefore the concentration at the timestep n+1 need a calculation that involves only the concentration at time step n.



(a) Upwind numerical scheme (black)



(b) Lax-Wendroff numerical scheme (black)



(c) Flux-limited numerical scheme (black)

Figure 6.3: Advection of a step function (red) from left to right calculated exactly (blue) and with numerical schemes (black).

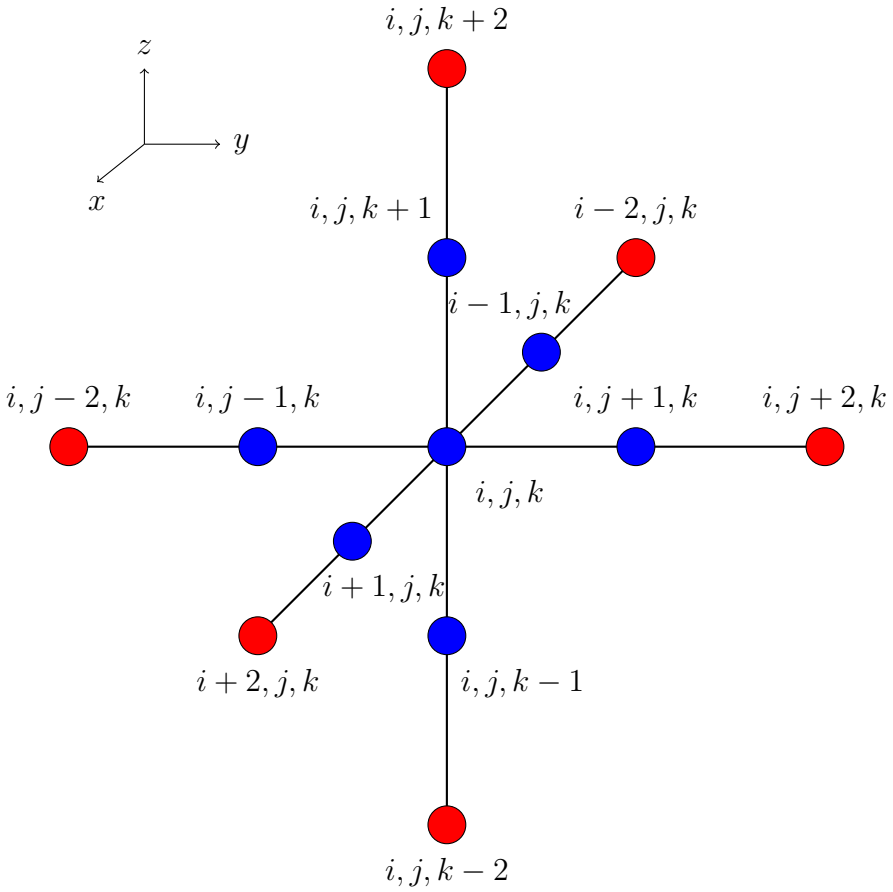


Figure 6.4: Stencil used for the advection method; the red circles represent the additional points required for the flux limiter.

The CFL conditions required by the scheme in order to be stable are the following [204]

$$\frac{v^{\alpha}\Delta t}{\Delta \alpha} \leq Cr \tag{6.40}$$

where $\alpha = x, y, z$ and $Cr = 1$ and in our case velocities didn't change in time so we determined the optimal timestep after the NS problem resolution, at the beginning of the ADR procedure. We used no-flux boundary conditions at walls, influx boundary conditions at inlets and extrapolations of interior grid points at the outlets boundaries, given by the following formula [206] [208]

$$c^{n}_{N_i+1,j} = 2c^{n}_{N_i,j} - c^{n}_{N_i-1,j} \tag{6.41}$$

where N_{i+1} is a point outside the domain.
Thanks to the fact that the scheme is explicit it may be parallelized easily using shared memory paradigm (OpenMP).

Diffusion

The diffusion operator was defined using the alternate direction implicit (ADI) method developed by Douglas [209], which is itself an operator splitting method and may be used to describe other physical phenomena like heat conduction [210]. This method is unconditionally stable and not computationally expensive because, it only requires the solution of a tridiagonal systems of equations, easily performed thanks to the tridiagonal matrix algorithm (TDMA also known as Thomas algorithm). The equation for the diffusive transport is obtained starting from the continuity equation with the diffusive flux expressed using Fick's first law $\mathbf{j}_{\mathrm{D}} = -D\nabla c$, that substituted in (6.23) gives the parabolic PDE

$$\frac{\partial c}{\partial t} - D\nabla^2 c = 0 \tag{6.42}$$

The Douglas method is a generalization of the Crank-Nicolson scheme[211] that is also unconditionally stable thus no time-step fulfillments are necessary[209]. If we use substitute (6.17) in (6.42) with the $n, n+1$ averaged right side, we derive the 3D analog of the Crank-Nicolson scheme given as

$$\frac{c^{n+1} - c^{n}}{\Delta t} = D\left[\frac{\delta^2_x(c^{n+1} + c^{n})}{(2(\Delta x)^2} + \frac{\delta^2_y(c^{n+1} + c^{n})}{2(\Delta y)^2} + \frac{\delta^2_z(c^{n+1} + c^{n})}{2(\Delta z)^2}\right] \tag{6.43}$$

that requires to solve a tridiagonal system with fringes. Instead of solving (6.43), we apply the Douglas [209] splitting approach and subdivide the problem in 3 different subproblems

$$\frac{c^{n+1} - c^{n}}{\Delta t} = D\left[\frac{\delta^2_x(c^{n+1/3} + c^{n})}{2(\Delta x)^2} + \frac{\delta^2_y c^{n}}{(\Delta y)^2} + \frac{\delta^2_z c^{n}}{(\Delta z)^2}\right] \tag{6.44}$$

$$\frac{c^{n+1}-c^n}{\Delta t} = D\left[\frac{\delta_x^2(c^{n+1/3}+c^n)}{2(\Delta x)^2} + \frac{\delta_y^2(c^{n+2/3}+c^n)}{2(\Delta y)^2} + \frac{\delta_z^2 c^n}{(\Delta z)^2}\right] \tag{6.45}$$

$$\frac{c^{n+1}-c^n}{\Delta t} = D\left[\frac{\delta_x^2(c^{n+1/3}+c^n)}{2(\Delta x)^2} + \frac{\delta_y^2(c^{n+2/3}+c^n)}{2(\Delta y)^2} + \frac{\delta_z^2(c^{n+1}+c^n)}{2(\Delta z)^2}\right] \tag{6.46}$$

In the (6.44) the x direction is implicit while y, z are explicit, the (6.45) is implicit with respect to y and explicit in the x, z directions and in the (6.46) the x, y directions are explicit and the z is implicit. If we rearrange (6.44) and subtract the (6.44) from (6.45) and (6.45) from (6.46) we obtain

$$\left(1-\frac{\nu}{2}\delta_x^2\right)c^{n+1/3} = \left(1+\frac{\nu}{2}\delta_x^2+\nu\delta_y^2+\nu\delta_z^2\right)c^n \tag{6.47}$$

$$\left(1-\frac{\nu}{2}\delta_y^2\right)c^{n+2/3} = c^{n+1/3}-\frac{\nu}{2}c^n \tag{6.48}$$

$$\left(1-\frac{\nu}{2}\delta_z^2\right)c^{n+1} = c^{n+2/3}-\frac{\nu}{2}c^n \tag{6.49}$$

where $\nu = D\Delta t/(\Delta x)^2 = D\Delta t/(\Delta y)^2 = D\Delta t/(\Delta z)^2$ because we use regular a Cartesian grid with $\Delta x = \Delta y = \Delta z$ and D is the translational diffusion coefficient of the considered chemical substance. Each of the δ_z^2 operators acts on a three point stencil, along the three directions, as represented in Fig.6.5

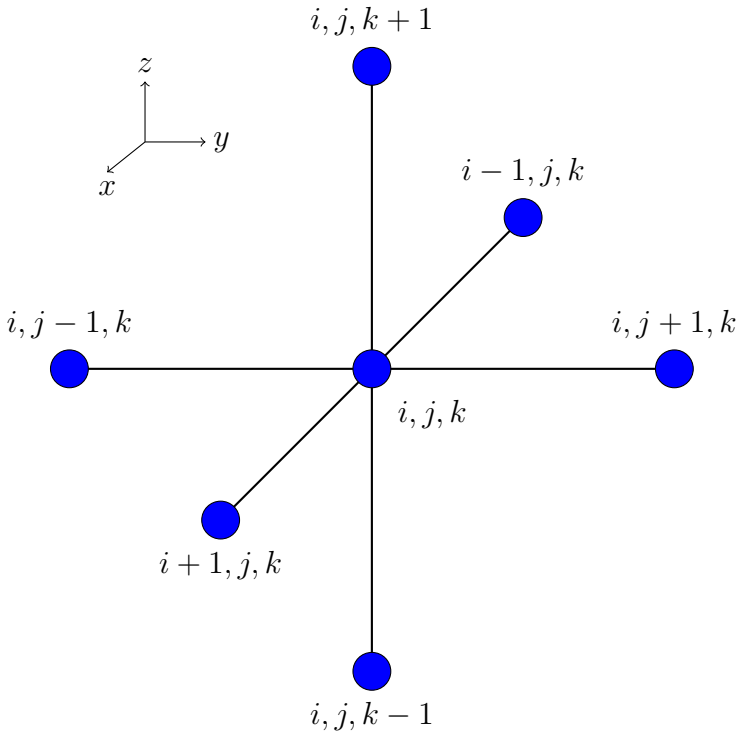


Figure 6.5: Stencil used for the ADI method.

No-flux boundary conditions are applied on the walls and inlet/outlet because it is assumed that advection is the dominating transport process across the inlet/outlet. This condition is imposed using Neumann boundary condition $(\mathbf{j}\cdot\mathbf{n}) = 0$, where **n** is the normal outgoing direction

across the boundary. For our Cartesian cell centered grid this condition is expressed as [212, 213]

$$\left.\frac{\partial c}{\partial x}\right|_{x=0} := -\frac{c_{1,j} - c_{0,j}}{\Delta x} + \mathcal{O}[(\Delta x)^2] = 0 \Rightarrow c_{0,j} = c_{1,j} \tag{6.50}$$

$$\left.\frac{\partial c}{\partial x}\right|_{x=L} := \frac{c_{N_i+1,j} - c_{N_i,j}}{\Delta x} + \mathcal{O}[(\Delta x)^2] = 0 \Rightarrow c_{N_i+1,j} = c_{N_i,j} \tag{6.51}$$

for the boundary at $x = 0$ and $x = L$ where the negative sign in formula (6.50) is due to the direction of the outgoing normal vector with respect to the x axis direction. Points outside the domain and along the y, z directions may be obtained similarly. To performs the update we need to solve 3 tridiagonal systems and this is easily done using the TDMA that requires $\mathcal{O}(N)$ operations, instead of the solution of the system generated by (6.43), which would require $\mathcal{O}(N^3)$ operations. For what concern the memory usage similar advantages can be achieved because the ADI method has linear memory requirements instead of superlinear ones necessary to employ Cholesky decomposition. If we eliminate (6.49),(6.48) and truncate terms depending from y, z in (6.47) we recover the Crank-Nicolson[211] scheme for the 1-dimensional diffusion problem. This method employs 3 points stencil along each direction (Fig. 6.5) and after the first step where the concentration vector elements are ordered along the x direction, for the second and third steps the elements must be ordered along the y and z directions respectively in order to apply diffusion along y and z. The ADI schemes is second order in both time and space with the local truncation error given as $\mathcal{O}[(\Delta x)^2, (\Delta y)^2, (\Delta t)^2]$. Even tough this is an implicit scheme, which are usually difficult to parallelize because the iterations in the TDMA require data at previous steps, some attempts of parallelizzation for this methods on GPU have been done [214–216].

Reaction

Contrary to the previous operators, the reaction term does not present spatial derivatives and requires the solution of a system of coupled ODE in each grid point without spatial coupling with other adjacent points; this is something like using a 0-dimensional model for each spatial grid point. The ODE problem may be solved integrating the system of equations in time but, depending on the kinetic rates involved in the reactions, a stiff behavior may be exhibited, which lead to unphysical results and numerical instabilities. When the kinetics were slow enough we integrated the system using an explicit Runge-Kutta (RK) 3/8 method, that requires a lower computational effort, otherwise we have to use a much more complex method that is based on the backward differentiation formula (BDF). The reaction process is a nonlinear system that couples together all the chemical species and describes how reactant transform to give intermediates and products. We can express the nonlinear ODE system as $dy/dt = f(y)$ where f stands for the reactions and y is not the direction but our function, as usually expressed in ODE equations.

The explicit four stage RK 3/8 integration scheme is defined as follows

$$y^{n+1} = y^n + \frac{\Delta t}{8}(k_1 + 3k_2 + 3k_3 + k_4) \tag{6.52}$$

where

$$k_1 = f(y^n) \tag{6.53}$$

$$k_2 = f\left(y^n + \frac{k_1}{3}\right) \tag{6.54}$$

$$k_3 = f\left(y^n - \frac{k_1}{3} + k_2\right) \tag{6.55}$$

$$k_4 = f(y^n + k_1 - k_2 + k_3) \tag{6.56}$$

It can be noticed that f is not dependent on the time because the reaction only depends on y. The RK 3/8 instead requires computations at 4 stages and it's a fourth order method with local truncation error $O[(\Delta t)^5]$. All the numerical simulations were executed until steady state for concentration was reached. To get rid of this problem we adopted the sundials libraries [217], that implement the backward differentiation formula in a variable order variable step multistep method, extremely powerful to treat stiff problems given as follows

$$\sum_{i=0}^{q} \alpha_{n,i} y^{n-i} + h_n \beta_{n,0} \dot{y}^n = 0 \tag{6.57}$$

with order q varying between 1 and 5, $h_n = t_n - t_{n-1}$, the coefficients are determined by the method order, the recent history of the step-sizes and $\alpha_{n,0} = -1$. The following non-linear system

$$G(y^n) \equiv y^n - h_n \beta_{n,0} f(t_n, y^n) - a_n = 0 \tag{6.58}$$

where $a_n \equiv \sum_{i>0}(\alpha_{n,i} y^{n-i} + h_n \beta_{n,i} \dot{y}^{n-i})$ must be solved (approximately) at each integration step with a Newton iteration which requires the solution of linear systems

$$M[y^{n(m+1)} - y^{n(m)}] = -G(y^{n(m)}) \tag{6.59}$$

with $M \approx I - \gamma J, J = \partial f / \partial y, \gamma = h_n \beta_{n,0}$. The initial guess is a predicted value of $y^{n(0)}$ explicitly computed from the history data. The Newton iteration is performed using a dense linear solver and the Jacobian must be provided. This multistep BDF method has a local truncation order given as $\mathcal{O}[(\Delta t)^{q+1}]$. In order to test this powerful solver we used a well known stiff ODE problem, known as Robertson reaction (see App. F)

6.3 Genetic algorithm and optimization of the reaction process

After the developement of the methodology to solve the ADR equations, we adopted the GA in order to maximize the reaction yield. These algorithms were originally created by Holland [218] and nowadays are parrt of a wider class of methods, called evolutionary computing or evolutionary algorithms. These robust optimization techniques permit to optimize the outcome of a complex problem by modifying several parameters at once in order to maximize a given objective function, e.g. in the case of a chemical reaction, the yield or the outflow of the desired product. The strength of these algorithms is given by the fact that they do not need information about the gradient of the objective function, which itself may also be discontinuous or non-differentiable; moreover they are ideally suitable to large dimensional problems and are easily implemented. In there cent past, they have been used to explore a class of acceptor compounds based on the 9,9'-bifluorenylidene in conjunction with the poly(3-hexylthiophene), to find candidates as acceptora for organic photovoltaic solar cells applications [219]. To the best of our knowledge this kind of techniques have been applied to solve shape/trajectory optimization problems within 2D/3D ADR problems context [220–222] but have not been yet used in outflow optimization problems involving several reactions and reactants.

We chose to limit the explored geometries to collection of neighboring squares that, also known as "polyominoes", well studied belong in the field of combinatorial problems. The CFD simulations were used first to estimate the outflow and this parameter, was then increased in order to improve the microchip by modifying the geometrical properties. The procedure explores a search space in order to find an optimal solution to the problem by simulating several devices (individuals). The methodology was applied to devices with a millimeter/sub-millimeter spatial scale but in general it may be employed for any kind of spatial scale that guarantee steady states for the fluid motion (low Reynolds number).

The canonical GA with élitism was employed to simulate an ensemble (population) constituted by eight different devices (individuals) and each simulation task was assigned to one physical core of our computing cluster, thanks to the MPI parallelization paradigm. The whole execution is controlled by the rank 0 process, that beyond the execution of the operations as the other processes, it also collects, distributes the data and runs the GA; thus a greater population may be simulated simply by requesting more physical cores. All these objects had a fixed number of inlets/outlets together with their positions, while the microchannels were freely modified with the only condition that inlet and outlet must be connected to the microcircuit in order to *a priori* exclude a chessboard-like pattern. Each device corresponds to a possible solution to the problem and its complete information (genotype) was coded inside an ensemble of data (chromosome) that completely defined the corresponding individual. The chromosome was given by a two-dimensional 11x11 array of bits (genes) that correspond to a total number of 121 squares, contained in a square lattice that we used to define the geometry of each device as depicted in Fig. 6.6. The squares number is directly proportional to the search space and

this kind of structures are similar to polyominoes[6]. To code these objects in the simplest way, we used a two-dimensional array that allowed to define the microchannels between inlets and outlets following a path generated by contiguous bits with value equal to 1.

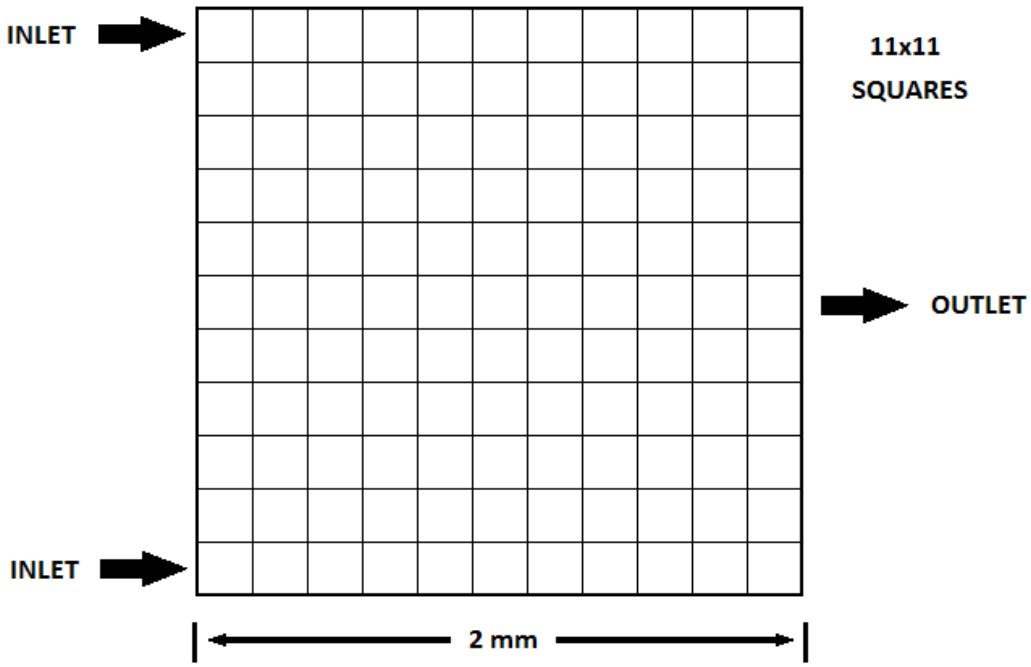


Figure 6.6: Structure of a microfluidic chip with 2 inlets and 1 outlet.

Given a fixed number of individuals (population), the GA starts randomly or deterministically by generating the starting individuals and subsequently it calculates the fitness function for each element. After that, it applies a set of genetic operators that modify the population, according to the idea of the natural selection, and attempt to improve the goodness of the individuals at every iteration cycle. The first genetic operator applied was the **roulette wheel selection**, that following a stochastic criterion, chooses a set of individuals from the current generation and put them in the mating pool, constituted by an equivalent number of individuals. The probability of selecting the individuals is proportional to the corresponding fitness, and each element may be extracted several times. Subsequently the **block crossover** operator recombines the individuals, swapping a portion of the information contained in the chromosomes. This operation was performed between consecutive couple of individuals and with a given execution probability (90%), choosing a rectangular block whose size and location were randomly generated; the information of each region of the chromosome was swapped between each couple of individuals that belonged to the mating pool. This crossover method has proved to be superior to other methods especially for the Ising problem [224]. After this modification of the mating pool, the **mutation**

[6]A polyomino is a finite subset of regular square tiling, connected by one or more of their sides. By joining the squares edge a plane geometric figure is formed. Nowadays only the possible number of polyominoes with 56 cells is exactly known, while in general, no precise laws have been discovered in order to determine the number of any possible polyomino that may be generated using a given number of squares but only estimates are provided [223].

operator changed the value of the genes from 0 to 1 or viceversa with a given probability (10%). Finally the **élitism** operator avoid losses of the two best individuals found in the the previous generation by simply recopying them unaltered in the current generation. This last operation also guaranteed the monotonicity of the maximum value of the fitness function of the whole population during the execution. The algorithm execution stopped when a convergence criteria or a given number of iterations were reached. All the execution of the code is summarized in the algorithm 5.

Algorithm 5 Molecular Dynamics

```
Generate initial population
do
    procedure FITNESS EVALUATION
        Solve NS equations
        do
            Solve ADR equations
        while Steady state reached
        Evaluate fluxes/yields
    end procedure
    Select parent individuals      ▷ selection of the mating pool using roulette wheel
    Crossover                      ▷ Mix the cromosome of two individuals
    Mutation                       ▷ Randomly flip a gene value
    élitism                        ▷ Restore previous best individual
while Stop criteria met
```

6.3.1 Reaction 1

We started our exploration studying the most simple bimolecular irreversible reaction, given as

$$A + B \xrightarrow{k} C \tag{6.60}$$

with the kinetic constant k=1 mmol^{-1} dm^3 s^{-1} and the translational diffusion coefficients equal to 5·10^{-10} m^2 s^{-1}. A end B were inserted at inlets 1 and 2 respectively with the concentration of 5 mmol dm^3, using a constant fluid velocity of 2 mm/s. This speed is in agreement with the operative flow rate of a syringe pump injector within the context of microfluidic devices. The fitness of the devices was defined as the outflow of the product C at the outlet

$$f = |j^C_{\text{outlet}}| \tag{6.61}$$

In Fig. 6.7 is reported the fitness optimized thanks to the GA which raise up respect to the fitness of the serpentine device. In Fig. 6.8 it can be noticed that the best device found thanks to the GA has a ladder pattern that maintains a thinner profile of the concentration with respect to the device with a serpentine pattern. The outflow of the optimized device is about 3.5 times the outflow of the serpentine.

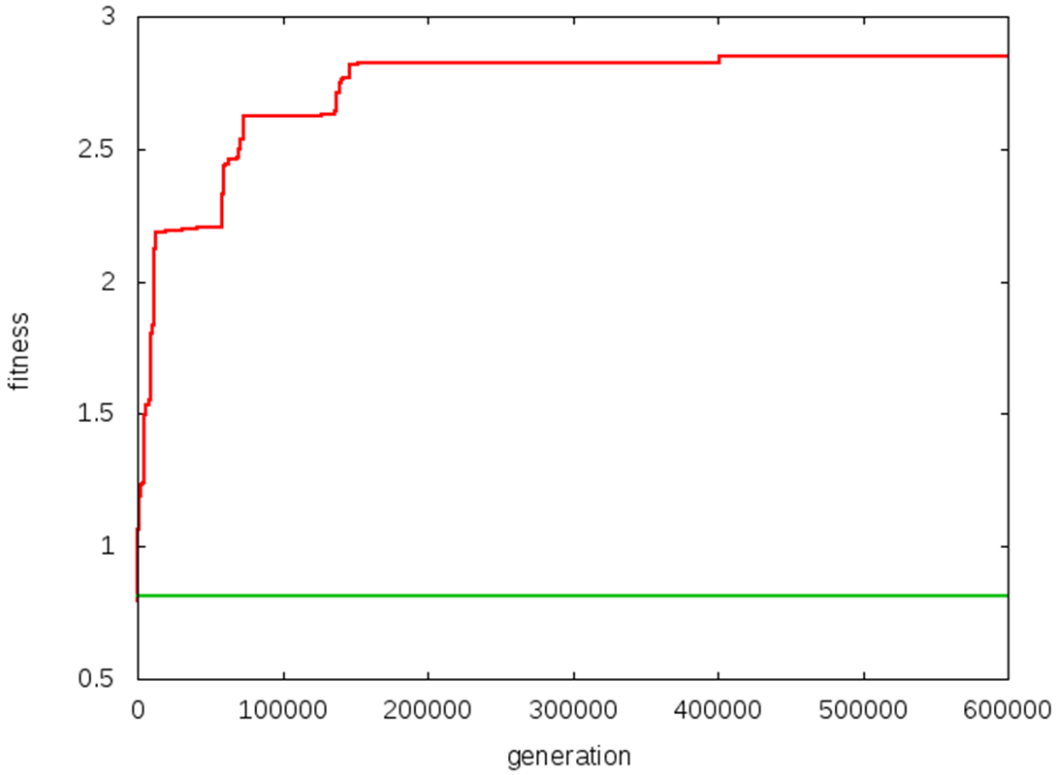


Figure 6.7: Fitness value for the reaction scheme (6.60). Red line is the fitness found with GA and green line represents the fitness value for the serpentine individual

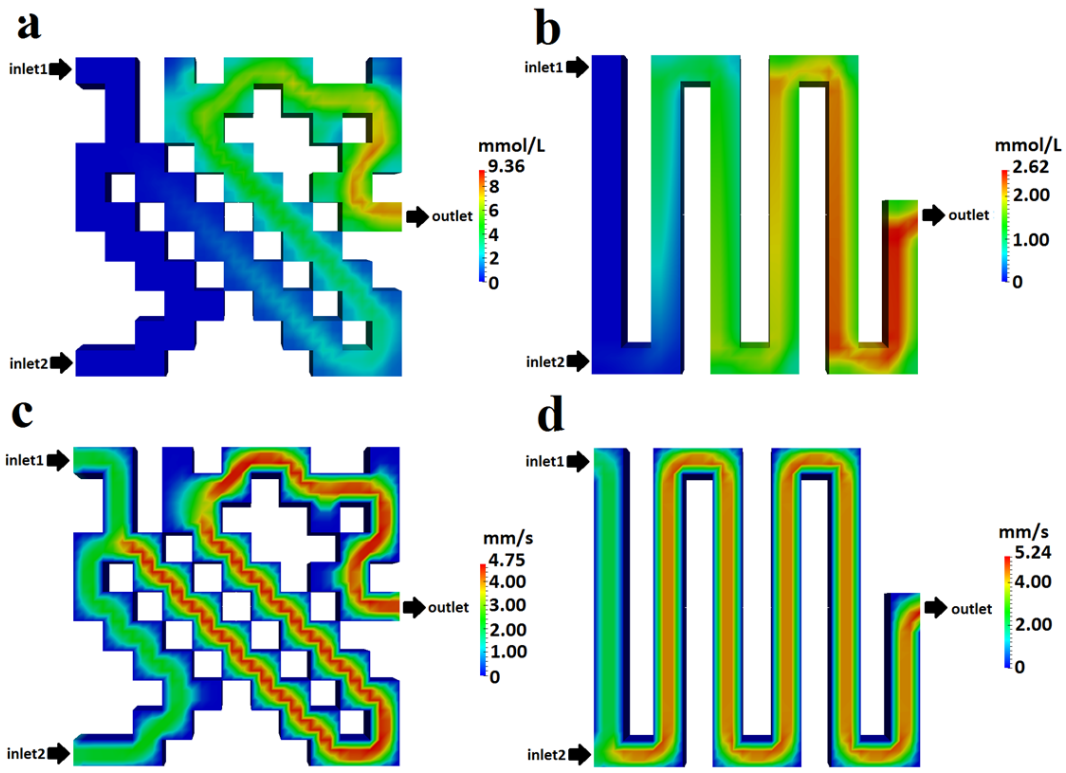


Figure 6.8: a,c (b,d) represent respectively the plot of the concentration of the product C and the modulus of the flow velocity for the best individual found with GA (serpentine) for the reaction scheme (6.60).

6.3.2 Reaction 2

We also studied a more complex mechanism with eight reactants defined as follows

$$A + B \xrightarrow{k} C \tag{6.62}$$

$$C + D \xrightarrow{k} E + F$$

$$F + G \xrightarrow{k} H$$

where k=1 mmol^{-1} dm^3 s^{-1} and the diffusion coefficient is the same for all the species and equal to 4·10^{-9} m^2 s^{-1}. A in inserted at inlet 1, B and D at inlet 3 and G at inlet 2 with a concentration of 5 mmol dm^{-3}. The fitness for this reaction was given as

$$f = |j^H_{\text{outlet2}}| \tag{6.63}$$

where H is the product of the third stage of the mechanism. As can be noticed from Fig. 6.9 the best individual found with the GA performs better than the serpentine-shaped microdevice. In particular the outflow of H is 2.3 times higher in the best device found with the GA than in the serpentine microcircuit.

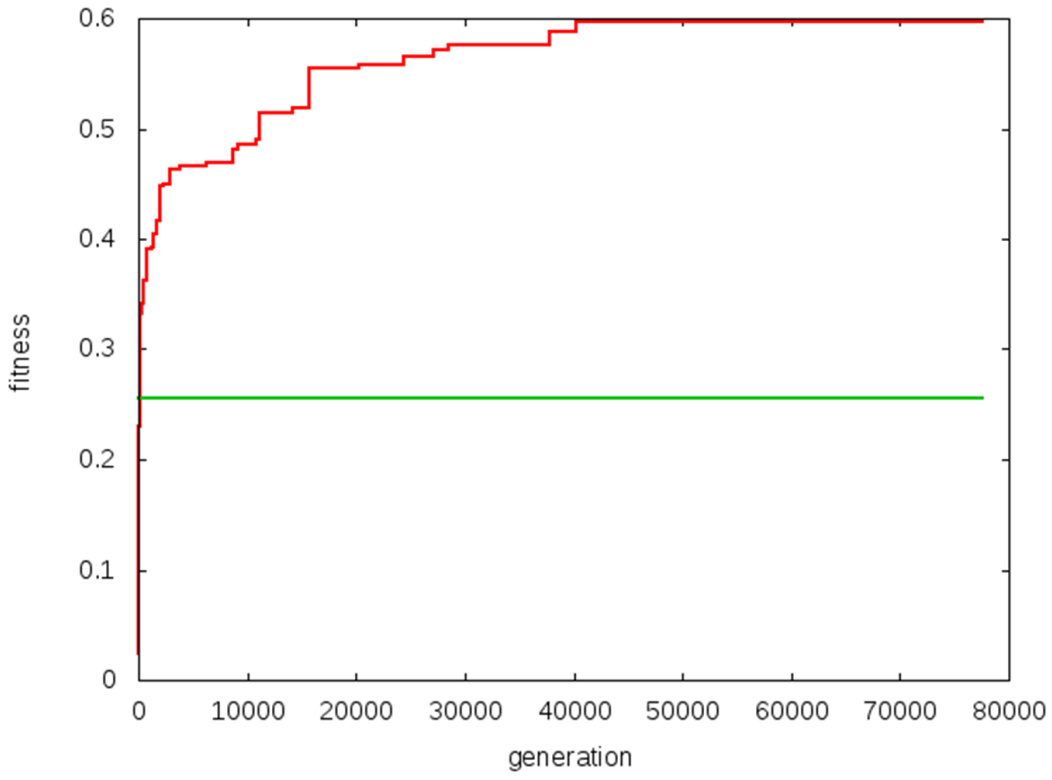


Figure 6.9: Fitness value for the reaction scheme (6.62) where red and green lines stands for the maximum and average fitness respectively, blue and magenta lines represent the fitness value for the best individual found with GA and the serpentine individual

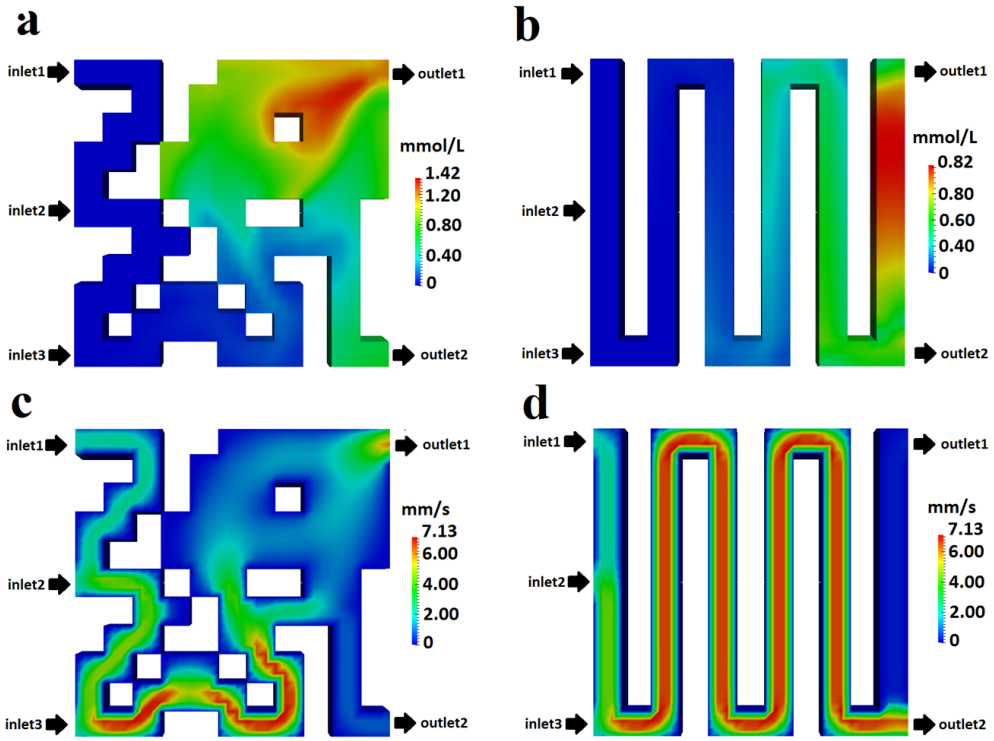


Figure 6.10: a,c (b,d) represent respectively the plot of the concentration of the desired product and the modulus of the flow velocity for the best individual found with GA (serpentine) for the reaction (6.62)

Although preliminary in nature, this study shows that in both reactions the GA may be employed to design microreactors optimized for the production of a given chemical, at a relatively modest computational cost.

Chapter 7

Final remarks

Computational chemistry is a powerful and interdisciplinary tool that may be applied to a wide range of problems and systems. In this work, we have explored a representative subset of methods covering the whole extension of approaches available to the modern chemist, from quantum mechanics to classic molecular dynamics to continuum methods. Using quantum chemistry we systematically studied the energetics of all the possible dioxopeptides and their thioamidated analogs (800 molecules), in particular the formation energies revealed that the thiopeptides are the most stable in gas-phase. An ongoing study on the same molecules foresees the investigation of the flexibility and the conformation assumed by the oxo-/thio-peptide backbone, which seems to be more likely to stay in a fully extended conformation for the thioamidated peptides; in this case the molecules under exam were the homopeptides of the glycine and alanine, varying the number of constituent amino acids from 2 up to 6.

We also explored systematically another class of molecules interesting because of their GPx mimics nature, together with a peroxidase-like efficiency: the diphenyldichalcogenides. The study was conducted using both quantum chemistry and molecular dynamics methods; we attached functional groups with different chemical nature in the para position (amine, methoxy, methyl, clorine, hydrogen, nitro) and we noticed that the phenyls' rotational barriers, calculated at the quantum level, are ruled by the eletron donor/withdrawing character of the substituents. The more electron donor the substituent is the higher the barrier is, in agreement mostly with the Hammett series. Furthermore we derived the force field for these molecules (non bonded and bonded parameters), validating the results with molecular dynamics simulations in water and chloroform.

A portion of the channel protein connexin 26 was studied using an ONIOM multiscale method, in order to identify the energy involved in the binding of the calcium ion with some specific residues. Furthermore the full protein has been considered using the molecular dynamics in order to inspect interesting properties for the closing/opening mechanism mediated by the calcium ions. The γ-carboxylation of the glutammate residue have also been included, by parametrizing the force field parameters for this non-standard residue. Some preliminary results of this ongoing work showed that, the presence of the γ-carboxylated glutammate, seem to stiffen the

protein and also inhibit the closing/opening calcium mediated mechanism.

Finally we developed a complete methodology for the simulation of transport and reaction phenomena of chemical substances, which occurs in liquid phase, for a specific ensemble of geometric configurations. The method was applied to a microfluidic framework, where the small spatial scale allows to achieve fast mixing regime, together with small volumes and well controlled temperature of the reactions. The modeling of these devices was coupled with heuristic optimization of the channels geometry in order to maximize the reaction yield.

Computational chemistry allows to study systems going from small molecules, active sites, macromolecules up to macroscopic systems, provided that one choses wisely the appropriate methodology. This may be not an easy task because, in order to exploit efficiently the available (and usually limited) hardware/software resources, one needs to know at least the working principles of high performance computing (HPC) for different approaches valid at different length-scales. Hardware architecture must be coupled with the best algorithmic solutions, and this is also not a simple task, especially considering their fast evolution. A tangible representation of this technology and how fast it evolves, was the heterogeneous CINECA cluster Eurora, at the top of the GREEN500 chart (the world's most energy-efficient supercomputer) in June 2013 and dismissed only two years later because already obsolete. Brute force computations based on a single methodological approach are therefore not always the best choice for a chemist, who need to deal, characteristically, with multiscale problems, and needs to be able to combine complementary computational approaches and implement them creatively adopting suitable approximations.

Appendix A

Electron repulsion integrals

As we have seen in the HF method the ERI are the most computationally expensive part to be computed. This is also true for the DFT calculations that uses hybrid exchange-correlation functionals (e.g. B3LYP, PBE0) which include HF exchange. The properties of GTOs have been exploited by developing many different methods for computing these integrals and the most known are the following

- Pople and Henre (PH) [225] algorithm was very efficient but specific to basis functions with low angular momentum and high degree of contraction;
- Dupuis, Rys and King (DRK) [226] was based on the Rys polynomials and was used for small contraction coefficients and high angular momentum;
- McMurchie and Davidson (MD) [227] algorithm was a more general algorithm with respect to PH and served as the starting point for the latest developed PRISM algorithm;
- Obara and Saika (OS) [228] provided an interesting scheme, used as starting point for the more sophisticated methods developed by Pople and coworkers. The OS scheme provided a eight-term vertical recursion rule to obtain higher angular momentum primitive ERI from simpler and auxiliary integrals;
- Head, Gordon and Pople (HGP) [229] starting from the interesting improvements and recursion rules found by Obara and Saika introduced a new recursion rule to contracted integrals and a well defined algorithm that used a semplified five-term special case of the OS eight-term vertical recursion rule;
- Gill and Pople (PRISM) [230] generalized the MD and HGP schemes and is one the most efficient algorithm for calculating ERIs. Thanks to its 20 distinct paths, chosen on the base of the contraction degree and the angular momentum of the ERI that are to be calculated, it dynamically achieves the best efficiency in terms of floating point operations.

During last years new algorithms have been developed, based on the the “accompanying coordinate expansion formula”, that seem to faster than the PRISM [231–234].

Appendix B

Energetics of oxo/thio-dipeptides: additional information

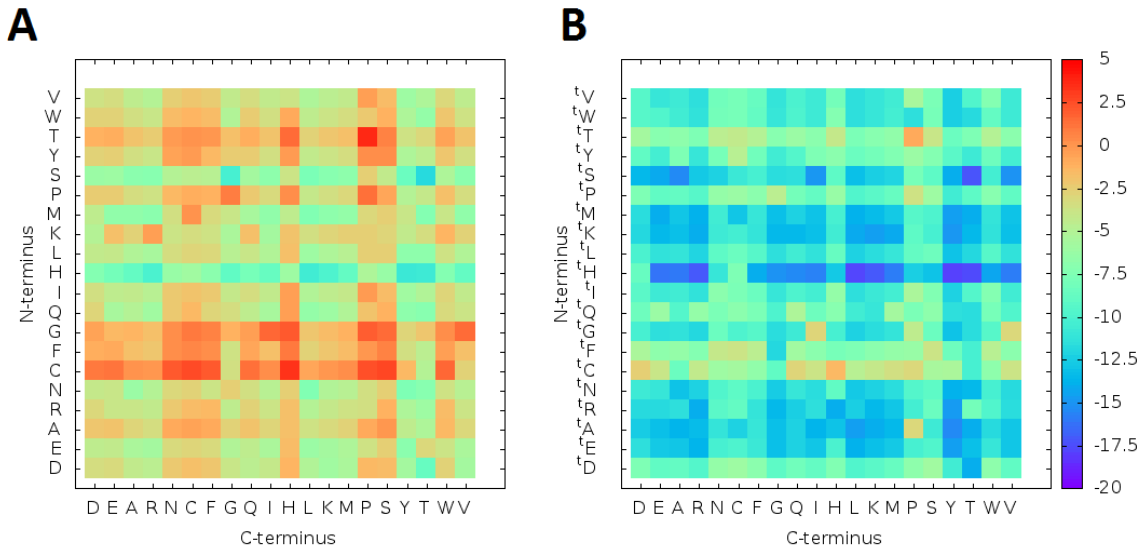


Figure B.1: Maps of $\Delta_r E$ (kcal/mol) calculated at PBE0/6-311G(d,p) level of theory for oxo- (A) and thio-dipeptides (B).

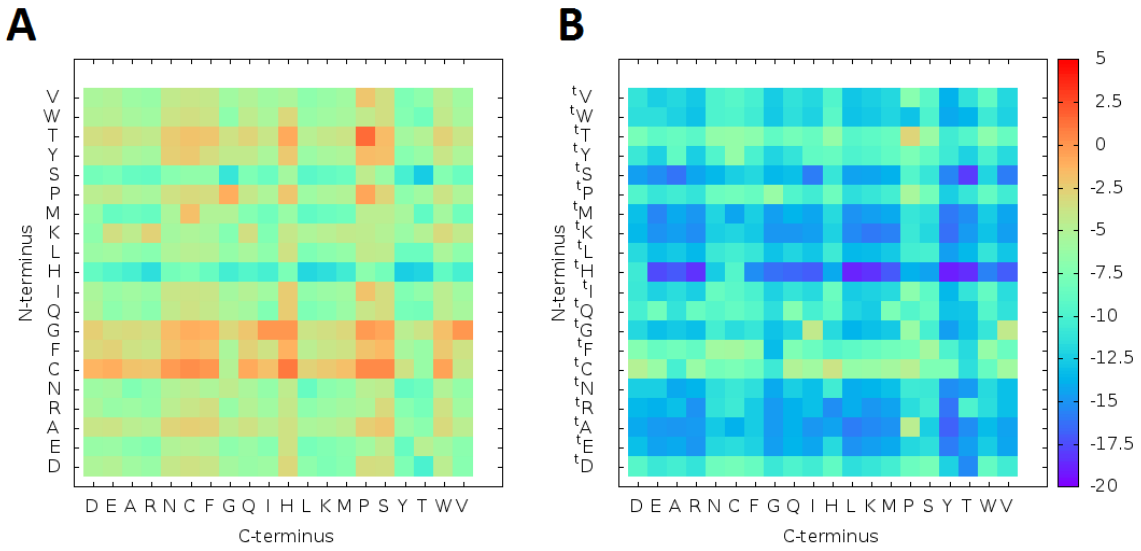


Figure B.2: Maps of $\Delta_r E$+ZPC calculated at PBE0/6-311G(d,p) level of theory for oxo- (A) and thio-dipeptides (B).

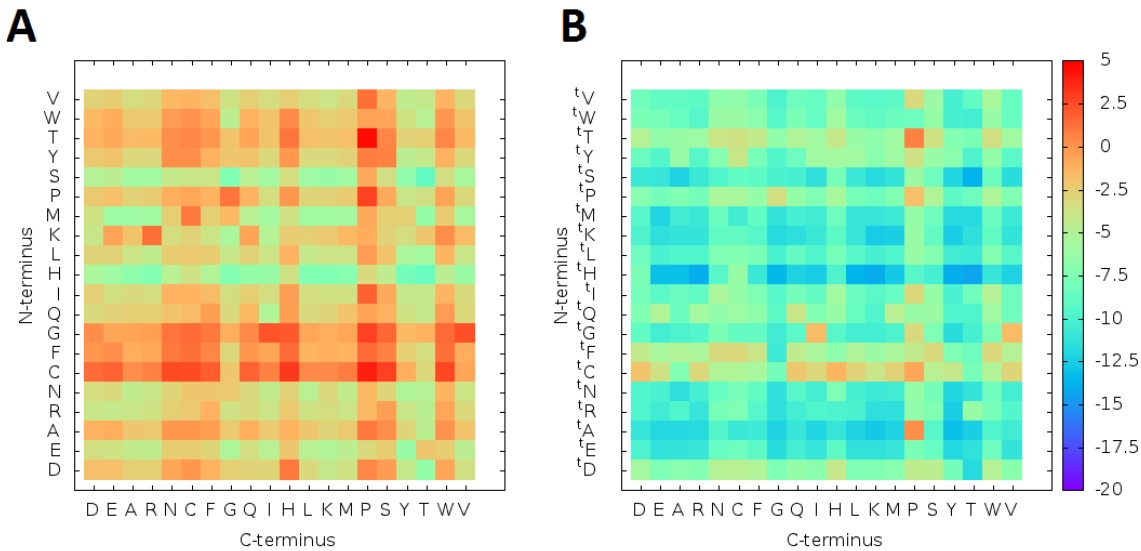


Figure B.3: Maps of $\Delta_r G$ calculated at PBE0/6-311G(d,p) level of theory for oxo- (A) and thio-dipeptides (B).

Appendix C

Energetics and calcium binding of a portion of connexin 26: additional information

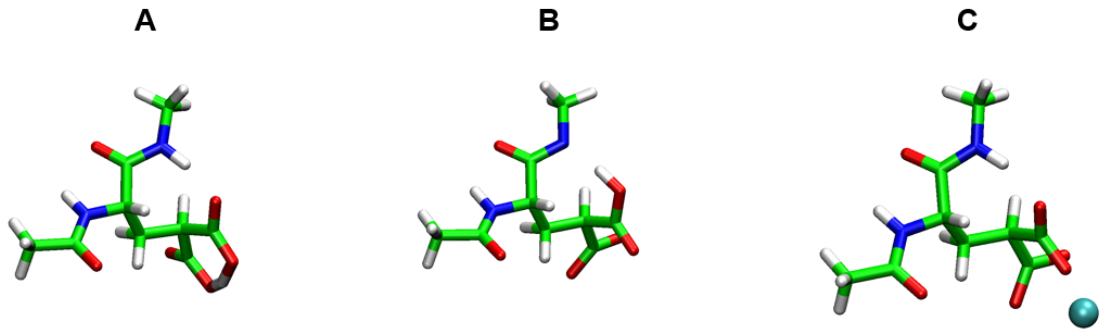


Figure C.1: Maps of $\Delta_r E$ (kcal/mol) calculated at PBE0/6-311G(d,p) level of theory for oxo- (A) and thio-dipeptides (B).

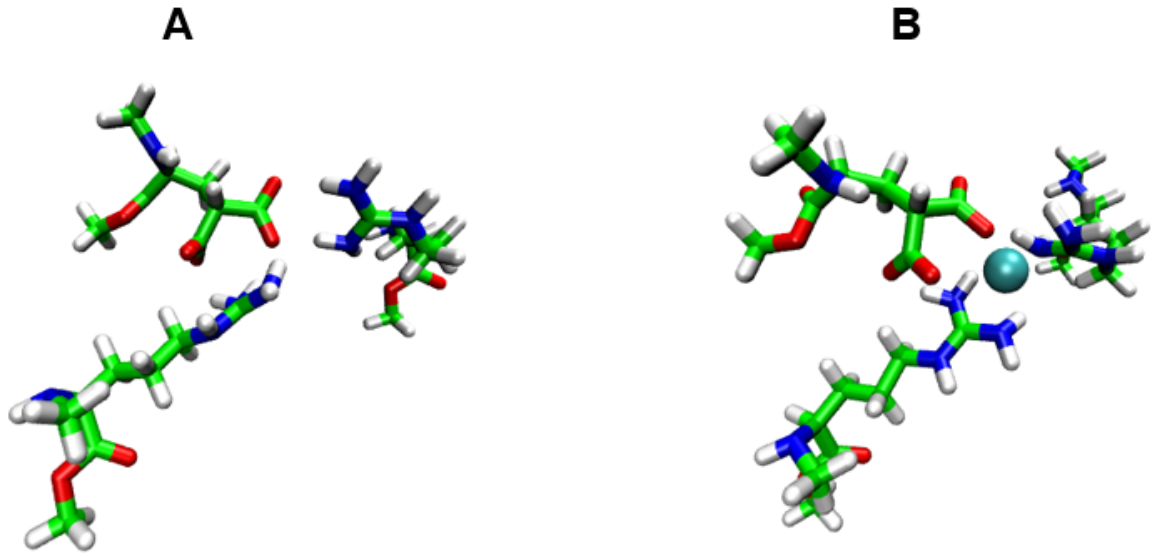


Figure C.2: Maps of $\Delta_r E$+ZPC calculated at PBE0/6-311G(d,p) level of theory for oxo- (A) and thio-dipeptides (B).

Appendix D

General amber force field for diphenyldichalcogenides: additional information

(p-NH_2–PhX)$_2$											
Bonds		Angles		Dihedrals							
r_{eq}	b	θ_{eq}	a	IDIVF[a]	$V_1/2$	$V_2/2$	$V_3/2$	$V_4/2$	γ
x-x		x-x-ca		ca-x-x-ca					
2.330	109.0	104.53	67.13	1	-1.85	-1.62	-0.66	0.00	180
(2.743)	(71.6)	(100.92)	(40.47)	1	(-1.47)	(-2.71)	(-0.59)	(0.02)	180
x-ca		x-ca-ca		x-x-ca-ca					
1.948	143.4	124.14	86.16	1	1.43	-0.19	-1.11	0.13	180
(2.145)	(122.8)	(119.94)	(78.24)	1	(0.90)	(-0.20)	(-0.69)	(0.09)	180
ca-ca		ca-ca-ca		x/ca-ca-ca-ca					
1.387	478.4	119.97	67.18	x/ca/ha-ca-ca-ha					
ca-ha		ca-ca-ha		ca/ha-ca-ca-nh					
1.087	344.3	120.01	48.46	4	–	14.5	–	–	180
hn-nh		ca-ca-nh		ca-ca-nh-hn					
1.014	401.2	120.13	69.34	4	–	4.2	–	–	180
ca-nh		hn-nh-hn							
1.364	449.0	114.85	40.05						
		ca-nh-hn							
		116.13	49.08						

Table D.1: Force-field parameters; values for Te analogue are given in brackets. [a] This is the factor which divides the barriers.

(p-OCH_3–PhX)$_2$									
Bonds		Angles		Dihedrals					
r_{eq}	b	θ_{eq}	a	IDIVF[a]	$V_1/2$	$V_2/2$	$V_3/2$	$V_4/2$	γ
x-x		x-x-ca		ca-x-x-ca					
2.330	109.0	104.53	67.13	1	-1.85	-1.62	-0.66	0.00	180
(2.743)	(71.6)	(100.92)	(40.47)	1	(-1.47)	(-2.71)	(-0.59)	(0.02)	180
x-ca		x-ca-ca		x-x-ca-ca					
1.948	143.4	124.14	86.16	1	-7.70	-0.11	-1.40	0.21	180
(2.145)	(122.8)	(119.94)	(78.24)	1	(1.26)	(-0.18)	(0.90)	(0.09)	180
ca-ca		ca-ca-ca		x/ca-ca-ca-ca					
1.387	478.4	119.97	67.18	x/ca/ha-ca-ca-ha					
ca-ha		ca-ca-ha		ca/ha-ca-ca-os					
1.087	344.3	120.01	48.46	4	–	14.5	–	–	180
ca-os		ca-ca-os		ca-ca-os-c3					
1.373	372.4	119.20	69.79	2	–	1.8	–	–	180
c3-os		h1-c3-os		h1-c3-os-ca					
1.439	301.5	108.82	50.84	3	–	–	1.15	–	0
c3-h1		c3-os-ca							
1.093	335.9	117.97	62.27						
		h1-c3-h1							
		109.55	39.18						

Table D.2: Force-field parameters; values for Te analogue are given in brackets. [a] This is the factor which divides the barriers.

(p-CH_3–PhX)$_2$									
Bonds		Angles		Dihedrals					
r_{eq}	b	θ_{eq}	a	IDIVF[a]	$V_1/2$	$V_2/2$	$V_3/2$	$V_4/2$	γ
x-x		x-x-ca		ca-x-x-ca					
2.330	109.0	104.53	67.13	1	-1.85	-1.62	-0.66	0.00	180
(2.743)	(71.6)	(100.92)	(40.47)	1	(-1.47)	(-2.71)	(-0.59)	(0.02)	180
x-ca		x-ca-ca		x-x-ca-ca					
1.948	143.4	124.14	86.16	1	-1.25	0.12	-1.43	0.16	180
(2.145)	(122.8)	(119.94)	(78.24)	1	(1.03)	(0.02)	(-0.71)	(0.08)	180
ca-ca		ca-ca-ca		ca-ca-c3-hc					
1.387	478.4	119.97	67.18	6	–	–	-2.91	–	180
ca-ha		ca-ca-ha		x/ca/c3-ca-ca-ca					
1.087	344.3	120.01	48.46	x/ca/ha/c3-ca-ca-ha					
c3-ca		ca-c3-hc		4	–	14.5	–	–	180
1.513	323.5	110.15	46.96						
c3-hc		c3-ca-ca							
1.092	337.3	120.63	63.84						
		hc-c3-hc							
		108.35	39.43						

Table D.3: Force-field parameters; values for Te analogue are given in brackets. [a] This is the factor which divides the barriers.

(p-Cl–PhX)$_2$									
Bonds		Angles		Dihedrals					
r_{eq}	b	θ_{eq}	a	IDIVF[a]	$V_1/2$	$V_2/2$	$V_3/2$	$V_4/2$	γ
x-x		x-x-ca		ca-x-x-ca					
2.330	109.0	104.53	67.13	1	-1.85	-1.62	-0.66	0.00	180
(2.743)	(71.6)	(100.92)	(40.47)	1	(-1.47)	(-2.71)	(-0.59)	(0.02)	180
x-ca		x-ca-ca		x-x-ca-ca					
1.948	143.4	124.14	86.16	1	-0.59	0.17	-1.08	0.15	180
(2.145)	(122.8)	(119.94)	(78.24)	1	(-1.54)	(0.45)	(0.85)	(0.10)	180
ca-ca		ca-ca-ca		x/ca-ca-ca-ca					
1.387	478.4	119.97	67.18	x/ca/ha-ca-ca-ha					
ca-ha		ca-ca-ha		ca-ca-ca-cl					
1.087	344.3	120.01	48.46	4	–	14.5	–	–	180
ca-cl		ca-ca-cl							
1.729	322.8	119.4	58.52						

Table D.4: Force-field parameters; values for Te analogue are given in brackets. [a] This is the factor which divides the barriers.

(p-NO_2–PhX)$_2$									
Bonds		Angles		Dihedrals					
r_{eq}	b	θ_{eq}	a	IDIVF[a]	$V_1/2$	$V_2/2$	$V_3/2$	$V_4/2$	γ
x-x		x-x-ca		ca-x-x-ca					
2.330	109.0	104.53	67.13	1	-1.85	-1.62	-0.66	0.00	180
(2.743)	(71.6)	(100.92)	(40.47)	1	(-1.47)	(-2.71)	(-0.59)	(0.02)	180
x-ca		x-ca-ca		x-x-ca-ca					
1.948	143.4	124.14	86.16	1	11.96	0.39	-0.33	0.02	180
(2.145)	(122.8)	(119.94)	(78.24)	1	(-1.21)	(0.40)	(1.69)	(0.00)	180
ca-ca		ca-ca-ca		x/ca-ca-ca-ca					
1.387	478.4	119.97	67.18	x/ca/ha-ca-ca-ha					
ca-ha		ca-ca-ha		ca/ha-ca-ca-no					
1.087	344.3	120.01	48.46	4	–	14.5	–	–	180
no-o		ca-ca-no		ca-ca-no-o					
1.219	761.2	119.54	66.88	4	–	2.4	–	–	180
ca-no		o-no-o							
1.468	322.6	125.13	77.15						
		ca-no-o							
		118.1	68.74						

Table D.5: Force-field parameters; values for Te analogue are given in brackets. [a] This is the factor which divides the barriers.

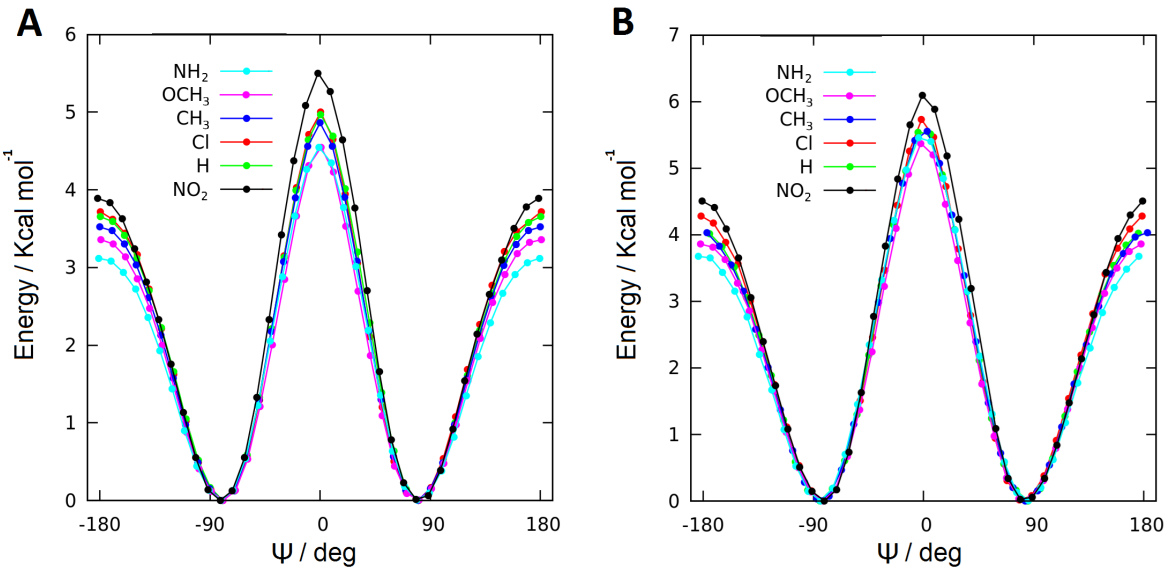


Figure D.1: Energy profiles at B3LYP/6-311G(d,p),SDD level of theory for the rotation of Ψ of all Se (A) and Te (B) diphenyl compounds.

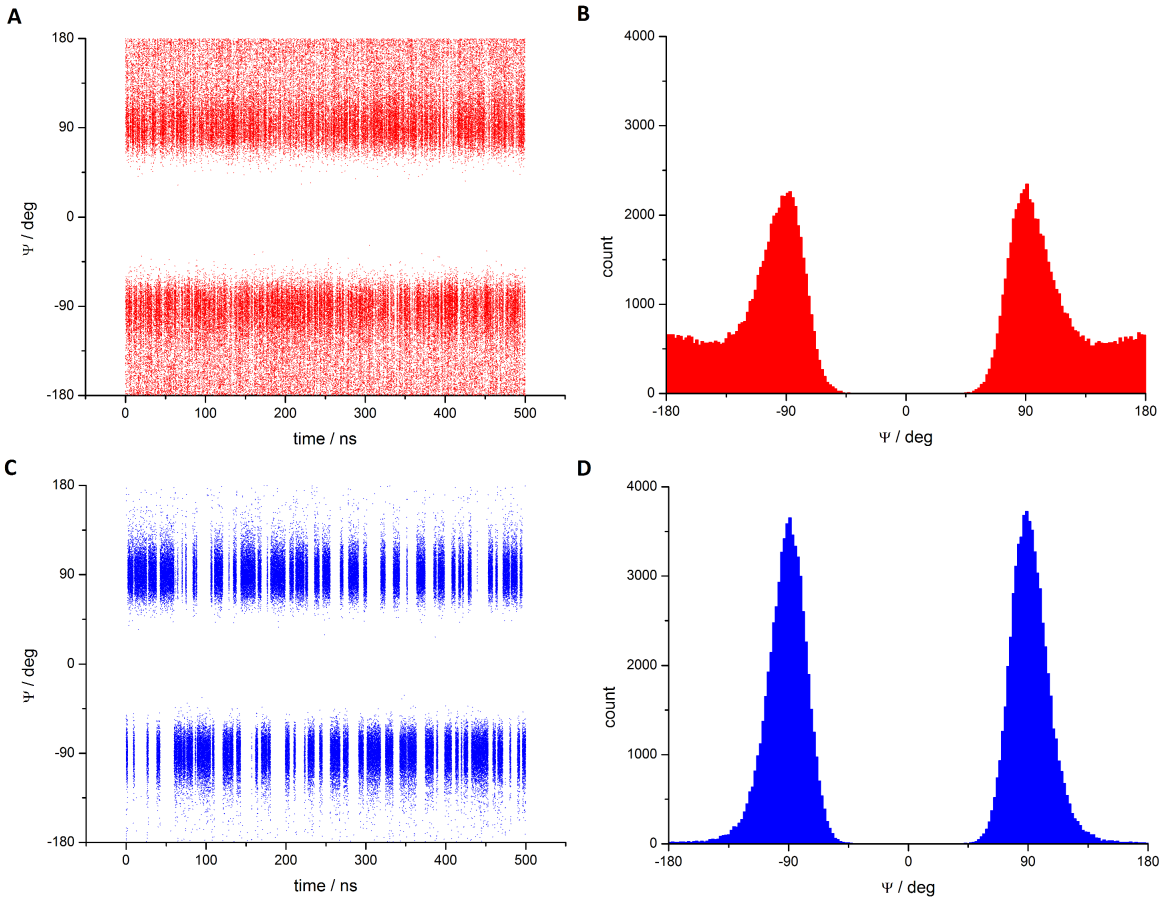


Figure D.2: Trajectories a (c) and distributions b (d) in chloroform of $(PhSe)_2$ (red) and $(PhTe)_2$ in blue.

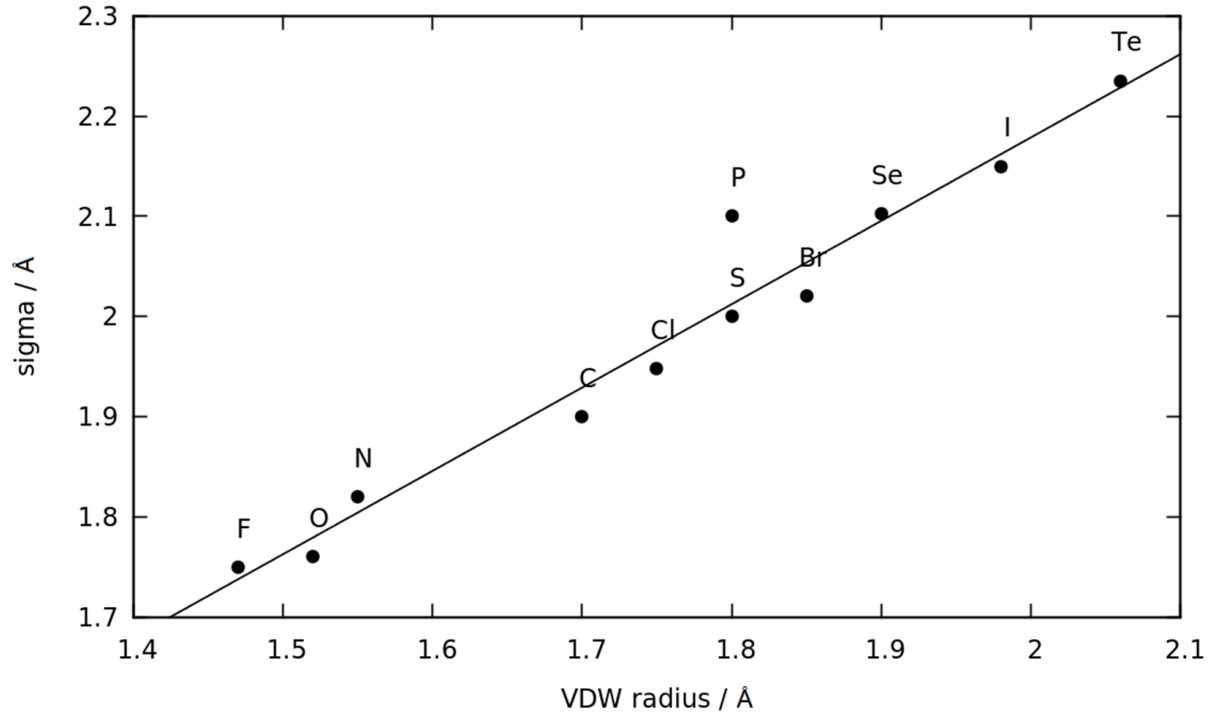


Figure D.3: σ_{GAFF} values vs vdW radii for C, N, O, P, S, F, Cl, Br, I atoms; For Se and Te the $\sigma_{UFF/2}$ values are taken.

Atom	vdW radius / Å	σ_{GAFF} / Å	$\sigma_{UFF/2}$ / Å [a]
C	1.70	1.90	1.9255
N	1.55	1.82	1.830
O [b]	1.52	1.76	1.750
F	1.47	1.75	1.682
P	1.80	2.10	2.087
S	1.80	2.00	2.0175
Cl	1.75	1.948	1.9735
Se	1.90	–	2.1025
Br	1.85	2.02	–
Te	2.06	–	2.235
I	1.98	2.15	–

Table D.6: σ_{GAFF} values vs vdW radii for C, N, O, P, S, F, Cl, Br, I atoms; values for Se and Te are taken as half σ_{UFF} values. [a] For GAFF parameters O is the TIP3P water oxygen. [b] Taken from Rappe, A., K.; Casewit, C., J.; Colwell, K., S.; Goddard III, W., A.; Skiff, W. M. UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations J. Am. Chem. Soc. 1992, 114, 10024–10035.

	CSD Identifier	X-X'	C_X-X C_X'-X'	C_X-X-X' C_X'-X'-X	C_X-X-X'-C_X' (Ψ)	C_1-C_X-X-X' C_1'-C_X'-X'-X $(\Phi_{1,2})$
(p-NH_2-PhSe)$_2$	—	2.52	1.94 1.94	103 103	-81	94 94
(p-NH_2-PhTe)$_2$	—	2.86	2.12 2.12	101 101	-84	92 92
(p-OCH_3-PhSe)$_2$	—	2.51	1.94 1.94	102 102	-79	92 92
(p-OCH_3-PhTe)$_2$	BAJHAV	2.85 (2.715)	2.13 (2.139) 2.13 (2.143)	101 (99.8) 101 (100.3)	-82 (-81)	93 (12) 93 (20)
(p-CH_3-PhSe)$_2$	QQQGBV01	2.50 (2.328)	1.95 (1.923) 1.95 (1.910)	102 (102.0) 102 (101.2)	80 (100)	87 (90) -96 (112)
(p-CH_3-PhTe)$_2$	DPTOTE	2.84 (2.696)	2.13 (2.125) 2.13 (2.131)	100 (101.5) 100 (100.0)	83 (86)	90 (37) 90 (37)
(PhSe)$_2$	YUXPIR	2.50 (2.307)	1.95 (1.946) 1.95 (1.934)	101 (102.4) 101 (103.5)	80 (85)	-96 (-23) 87 (0)
(PhTe)$_2$	DPHDTE01	2.84 (2.707)	2.14 (2.131) 2.14 (2.115)	100 (100.7) 100 (97.7)	-85 (-90)	-90 (-84) 92 (90)
(p-Cl-PhSe)$_2$	CLPHSE	2.50 (2.332)	1.95 (1.929) 1.95 (1.911)	101 (98.6) 101 (98.9)	-79 (-74)	-85 (-61) 98 (74)
(p-Cl-PhTe)$_2$	CLPHTE01	2.84 (2.680)	2.14 (2.144) 2.14 (2.160)	100 (99.6) 100 (101.6)	-82 (-89)	-88 (-13) -88 (-10)
(p-NO_2-PhSe)$_2$	DUWKEL	2.49 (2.302)	1.96 (1.921) 1.96 (1.921)	101 (103.2) 101 (103.2)	-82 (-88)	103 (23) 103 (23)
(p-NO_2-PhTe)$_2$	—	2.834	2.14 2.14	99.2 99.2	-82	98 98

Table D.7: Relevant interatomic distances (Å) and angles (°) of the fully optimized dichalcogenides; level of theory: B3LYP/6-311G(d,p),SDD. When available, parameters from the crystallographic structures taken from CSD are given in parentheses.

Appendix E

Advection-Diffusion-Reaction equation

The Advection-Diffusion-Reaction equation may be easily obtained using the continuity equation which is a very useful equation that describes the conservation of a physical quantity (e.g. charge, mass, probability density) and is expressed as

$$\frac{\partial \psi}{\partial t} + \nabla \cdot \mathbf{j} \tag{E.1}$$

where ψ is the physical quantity conserved and **j** its flux in the space. If we consider the variation of the number of moles n in time, given the concentration $c = n/V$ and the mass fluss $\mathbf{j}_{\text{mass}}$

$$\frac{dn}{dt} = \int_V \frac{\partial c}{\partial t} dV = -\oint_{\partial V} \mathbf{j}_{\text{mass}} \cdot d\mathbf{S} = -\int_V \nabla \cdot \mathbf{j}_{\text{mass}}\, dV \tag{E.2}$$

taking the integrand at the second and the fourth term we obtain the local form of the continuity equation

$$\frac{\partial c}{\partial t} + \nabla \cdot \mathbf{j}_{\text{mass}} = 0 \tag{E.3}$$

which in the case that c is not conserved during time (E.3) becomes

$$\frac{\partial c}{\partial t} + \nabla \cdot \mathbf{j}_{\text{mass}} = \mathbf{R} \tag{E.4}$$

where R is the term that describe the velocity of production or consumption of c. If the concentration of chemical substance is subject to advective and diffusive transport than we have to replace $\mathbf{j}_{\text{mass}}$ by the sum of these fluxes, but let's see first how they are expressed mathematically. The advective flux is the transport of the substance due to the bulk motion, so this will be proportional to the concentration and to the velocity of the fluid element, i.e.

$$\mathbf{j}_{\text{A}} = c\, \mathbf{v} \tag{E.5}$$

Instead the diffusive flux is described by one of the Fick's laws of diffusion, derived in 1855 by the German physician and physiologist Adolf Eugen Fick. The first laws stays that the matter moves

from zones of high to zones of low concentration, proportionally to the gradient (with opposite direction), so the flux is expressed as

$$\mathbf{j}_{\mathsf{D}} = -D\,\nabla c \tag{E.6}$$

where D is the proportionality coefficient known as diffusion coefficient and

$$\nabla = \left(\frac{\partial}{\partial x}\hat{\mathbf{u}}_x, \frac{\partial}{\partial y}\hat{\mathbf{u}}_y, \frac{\partial}{\partial z}\hat{\mathbf{u}}_z\right)^T \tag{E.7}$$

is the gradient. Fick's second law, may be obtained easily substituting the (E.6) in the (E.3) and describes the variation in time of the concentration. So to obtain the ADR equation we set $\mathbf{j}_{\mathsf{mass}} = \mathbf{j}_{\mathsf{A}} + \mathbf{j}_{\mathsf{D}}$ and the (E.4) may be rewritten as

$$\frac{\partial c}{\partial t} + \mathbf{v}\cdot\nabla c + c\nabla\cdot\mathbf{v} - \nabla\cdot(D\nabla c) = \mathbf{R} \tag{E.8}$$

If the velocity field is divergence free ($\nabla\cdot\mathbf{v} = 0$) and D is a constant, we obtain the following equation

$$\frac{\partial c}{\partial t} + \mathbf{v}\cdot\nabla c = D\nabla^2 c + \mathbf{R} \tag{E.9}$$

where ∇^2 is the Laplace operator, given as the divergence of the gradient $\nabla\cdot\nabla$. An interesting application of the diffusion-reaction model was for a study far from chemical descriptions, i.e. the sudy of the nucleation, spread and dissipation of crime hotspots [235].

Appendix F

Stiff problems: Robertson reaction

The Robertson reaction, also known as Rober problem[236], is a 3 stage reaction with rate constants that span 9 magnitude orders which is the reason for the stiffness. The reaction mechanism is the following

$$A \xrightarrow{k_1} B$$

$$B + B \xrightarrow{k_2} B + C$$

$$B + C \xrightarrow{k_3} A + C$$

where $k_1 = 0.04$, $k_2 = 3 \cdot 10^7$, $k_3 = 10^4$. The ODE system describing the rate of A, B, C is the following

$$\begin{cases} \frac{d[A]}{dt} = -k_1[A] + k_3[B][C] \\ \\ \frac{d[B]}{dt} = k_1[A] - k_2[B]^2 - k_3[B][C] \\ \\ \frac{d[C]}{dt} = k_2[B]^2 \end{cases} \tag{F.1}$$

with the initial conditions $[A](0) = 1$, $[B](0) = [C](0) = 0$. This reaction is often employed to test numerical integrators because of its "stiffness", which is unlikely solvable (at least for a long time) using explicit methods as Euler method [237]. The intermediate specie B is generated and disappears very quickly and A transforms into B as can be seen in Fig.F.1, which data were obtained with a serial program in 0.006 seconds. To obtain the same values with an explicit scheme an extremely small integration time-step should be used in order to prevent numerical oscillations, leading to a very long time computation.

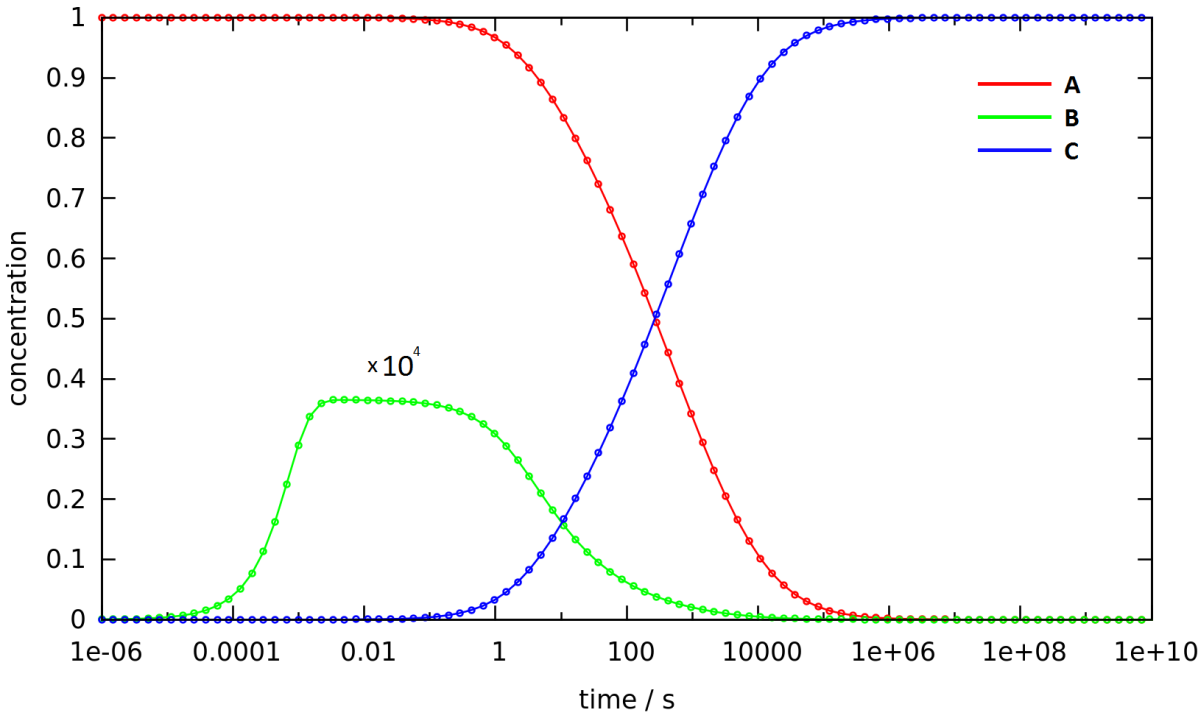


Figure F.1: Concentration values for the Robertson problems.

List of publications

[1] Mauro Torsello, Laura Orian, Marta De Zotti, Fernando Formaggio, and Antonino Polimeno. “Stability of secondary structure in oxo- and thio-dipeptides: a quantum chemistry analysis”. *in preparation* (2016).

[2] Mauro Torsello, Nicola Fortunati, Paolo Calligari, Laura Orian, and Antonino Polimeno. “The Ca2+-promoted gating mechanism in the wild type and the γ-carboxylated connexin 26: a molecular simulation study”. *in preparation* (2016).

[3] Mauro Torsello, Laura Orian, Marta De Zotti, Roberta Saini, Fernando Formaggio, and Antonino Polimeno. “Energetics of oxo- and thio-dipeptide formation via amino acid condensation: a systematic computational analysis”. *Phys. Chem. Chem. Phys.* 16 (33 2014), 17515–17522. URL: http://dx.doi.org/10.1039/C4CP02680B.

[4] Francesco Zonta, Fabio Mammano, Mauro Torsello, Nicola Fortunati, Laura Orian, and Antonino Polimeno. “Role of gamma carboxylated Glu47 in connexin 26 hemichannel regulation by extracellular Ca2+: Insight from a local quantum chemistry study”. *Biochemical and Biophysical Research Communications* 445.1 (2014), 10–15. URL: http://www.sciencedirect.com/science/article/pii/S0006291X14000874.

[5] Mauro Torsello, Antonio C. Pimenta, Lando P. Wolters, Irina S. Moreira, Laura Orian, and Antonino Polimeno. “General Amber Force Field Parameters for Diphenyldiselenides and Diphenylditellurides”. *in preparation* (2016).

Bibliography

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

[2] P. A. M. Dirac. *The Principles of Quantum Mechanics*. International series of monographs on physics. Clarendon Press, 1967. URL: https://books.google.it/books?id=piAIAQAAIAAJ.

[3] Christopher I. Bayly, Piotr Cieplak, Wendy D. Cornell, and Peter A. Kollman. "A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model". *The Journal of Physical ...* 97 (1993), 10269–10280. URL: http://pubs.acs.org/doi/abs/10.1021/j100142a004.

[4] "Hartree-Fock Calculator" (2015). URL: https://sites.google.com/site/maurotorsello/software/hafoc.

[5] J. A. Pople and R. K. Nesbet. "Self-Consistent Orbitals for Radicals". *The Journal of Chemical Physics* 22.3 (1954), 571–572. URL: http://scitation.aip.org/content/aip/journal/jcp/22/3/10.1063/1.1740120.

[6] Mireia Güell, Josep M. Luis, Miquel Solà, and Marcel Swart. "Importance of the Basis Set for the Spin-State Energetics of Iron Complexes". *The Journal of Physical Chemistry A* 112.28 (2008), 6384–6391. URL: http://dx.doi.org/10.1021/jp803441m.

[7] Stefan Grimme. “Semiempirical GGA-type density functional constructed with a long-range dispersion correction.” *Journal of computational chemistry* 27.15 (2006), 1787–99. URL: http://dx.doi.org/10.1002/jcc.20495.

[8] Stefan Grimme, Jens Antony, Stephan Ehrlich, and Helge Krieg. “A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu”. *The Journal of Chemical Physics* 132.15 (2010), 154104. URL: http://scitation.aip.org/content/aip/journal/jcp/132/15/10.1063/1.3382344.

[9] Stefan Grimme, Stephan Ehrlich, and Lars Goerigk. “Effect of the damping function in dispersion corrected density functional theory”. *Journal of Computational Chemistry* 32.7 (2011), 1456–1465. URL: http://doi.wiley.com/10.1002/jcc.21759.

[10] W. Kohn and L. J. Sham. “Self-Consistent Equations Including Exchange and Correlation Effects”. *Phys. Rev.* 140 (4A 1965), A1133–A1138. URL: http://link.aps.org/doi/10.1103/PhysRev.140.A1133.

[11] A. D. Becke. “Density-functional exchange-energy approximation with correct asymptotic behavior”. *Phys. Rev. A* 38 (6 1988), 3098–3100. URL: http://link.aps.org/doi/10.1103/PhysRevA.38.3098.

[12] Chengteh Lee, Weitao Yang, and Robert G. Parr. “Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density”. *Phys. Rev. B* 37 (2 1988), 785–789. URL: http://link.aps.org/doi/10.1103/PhysRevB.37.785.

[13] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch. “Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields”. *The Journal of Physical Chemistry* 98.45 (1994), 11623–11627. URL: http://dx.doi.org/10.1021/j100096a001.

[14] Axel D. Becke. “Density-functional thermochemistry. III. The role of exact exchange”. *The Journal of Chemical Physics* 98.7 (1993), 5648–5652. URL: http://scitation.aip.org/content/aip/journal/jcp/98/7/10.1063/1.464913.

[15] John P. Perdew, Matthias Ernzerhof, and Kieron Burke. “Rationale for mixing exact exchange with density functional approximations”. *The Journal of Chemical Physics* 105.22 (1996), 9982–9985. URL: http://scitation.aip.org/content/aip/journal/jcp/105/22/10.1063/1.472933.

[16] Carlo Adamo and Vincenzo Barone. “Toward reliable density functional methods without adjustable parameters: The PBE0 model”. *The Journal of Chemical Physics* 110.13 (1999), 6158–6170. URL: http://scitation.aip.org/content/aip/journal/jcp/110/13/10.1063/1.478522.

[17] A. Warshel and M. Levitt. “Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme”. *Journal of Molecular Biology* 103.2 (1976), 227–249. URL: http://www.sciencedirect.com/science/article/pii/0022283676903119.

[18] U. Chandra Singh and Peter A. Kollman. “A combined ab initio quantum mechanical and molecular mechanical method for carrying out simulations on complex molecular systems: Applications to the CH3Cl + Cl− exchange reaction and gas phase protonation of polyethers”. *Journal of Computational Chemistry* 7.6 (1986), 718–730. URL: http://dx.doi.org/10.1002/jcc.540070604.

[19] Martin J. Field, Paul A. Bash, and Martin Karplus. “A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations”. *Journal of Computational Chemistry* 11.6 (1990), 700–733. URL: http://dx.doi.org/10.1002/jcc.540110605.

[20] Lung Wa Chung, W. M. C. Sameera, Romain Ramozzi, Alister J. Page, Miho Hatanaka, Galina P. Petrova, Travis V. Harris, Xin Li, Zhuofeng Ke, Fengyi Liu, Hai-Bei Li, Lina Ding, and Keiji Morokuma. “The ONIOM Method and Its Applications”. *Chemical Reviews* 115.12 (2015), 5678–5796. URL: http://dx.doi.org/10.1021/cr5004419.

[21] Adri C. T. van Duin, Siddharth Dasgupta, Francois Lorant, and William A. Goddard. “ReaxFF: A Reactive Force Field for Hydrocarbons”. *The Journal of Physical Chemistry A* 105.41 (2001), 9396–9409. URL: http://dx.doi.org/10.1021/jp004368u.

[22] Dusanka Janezic and Bojan Orel. “Implicit Runge-Kutta method for molecular dynamics integration”. *Journal of Chemical Information and Computer Sciences* 33.2 (1993), 252–257. URL: http://dx.doi.org/10.1021/ci00012a011.

[23] Alexander Kyrychenko, Galina V. Karpushina, Sergey I. Bogatyrenko, Alexander P. Kryshtal, and Andrey O. Doroshenko. “Preparation, structure, and a coarse-grained molecular dynamics model for dodecanethiol-stabilized gold nanoparticles”. *Computational and Theoretical Chemistry* 977.1-3 (2011), 34–39. URL: http://dx.doi.org/10.1016/j.comptc.2011.09.003.

[24] Jia Qi Lin, Hong Wu Zhang, Zhen Chen, Yong Gang Zheng, Zhong Qiang Zhang, and Hong Fei Ye. “Simulation study of aggregations of monolayer-protected gold nanoparticles in solvents”. *Journal of Physical Chemistry C* 115.39 (2011), 18991–18998. URL: http://dx.doi.org/10.1021/jp204735d.

[25] Yinfeng Li, Hongyan Yuan, Annette Von Dem Bussche, Megan Creighton, Robert H. Hurt, Agnes B. Kane, and Huajian Gao. “Graphene microsheets enter cells through spontaneous membrane penetration at edge asperities and corner sites”. *Proceedings of the National Academy of Sciences* 110.30 (2013), 12295–12300. URL: http://www.pnas.org/content/110/30/12295.abstract.

[26] "Protein data bank" (2015). URL: http://www.rcsb.org/.

[27] Tom Darden, Darrin York, and Lee Pedersen. "Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems". *The Journal of Chemical Physics* 98.12 (1993), 10089–10092. URL: http://scitation.aip.org/content/aip/journal/jcp/98/12/10.1063/1.464397.

[28] A. Quarteroni. *Modeling the Heart and the Circulatory System*. MS&A. Springer International Publishing, 2015. URL: http://www.springer.com/us/book/9783319052298.

[29] P Chen and J Qu. "Backbone Modification of beta-Hairpin-Forming Tetrapeptides in Asymmetric Acyl Transfer Reactions". *Journal of Organic Chemistry* 76.9 (2011), 2994–3004. URL: http://pubs.acs.org/doi/abs/10.1021/jo200403g.

[30] Jan Helbing, Harald Bregy, Jens Bredenbeck, Rolf Pfister, Peter Hamm, Robert Huber, Josef Wachtveitl, Luca De Vico, and Massimo Olivucci. "A Fast Photoswitch for Minimally Perturbed Peptides: Investigation of the trans → cis Photoisomerization of N-Methylthioacetamide". *Journal of the American Chemical Society* 126.28 (2004), 8823–8834. URL: http://dx.doi.org/10.1021/ja049227a.

[31] H Bregy, H Heimgartner, and J Helbing. "A Time-resolved Spectroscopic Comparison of the Photoisomerization of Small beta-Turn-forming Thioxopeptides". *Journal of Physical Chemistry B* 113.6 (2009), 1756–1762. URL: http://pubs.acs.org/doi/abs/10.1021/jp8089402.

[32] M Hollosi, E Kollat, J Kajtar, M Kajtar, and G D Fasman. "Chiroptical Labeling of Folded Polypeptide Conformations - the Thioamide Probe". *Biopolymers* 30.11-12 (1990), 1061–1072. URL: http://onlinelibrary.wiley.com/doi/10.1002/bip.360301107/abstract.

[33] M De Zotti, B Biondi, C Peggion, M De Poli, H Fathi, S Oancea, C Toniolo, and F Formaggio. "Partial thioamide scan on the lipopeptaibiotic trichogin GA IV. Effects on folding and bioactivity". *Beilstein Journal of Organic Chemistry* 8 (2012), 1161–1171. URL: http://dx.doi.org/10.3762%2Fbjoc.8.129.

[34] M De Poli and J Clayden. "Thionoglycine as a multifunctional spectroscopic reporter of screw-sense preference in helical foldamers". *Organic & Biomolecular Chemistry* 12.5 (2014), 836–843. URL: http://pubs.rsc.org/en/content/articlelanding/2014/ob/c3ob42167h#!divAbstract.

[35] E J Petersson, J M Goldberg, and R F Wissner. "On the use of thioamides as fluorescence quenching probes for tracking protein folding and stability". *Physical Chemistry Chemical Physics* 16.15 (2014), 6827–6837. URL: http://pubs.rsc.org/en/content/articlelanding/2014/cp/c3cp55525a#!divAbstract.

[36] "Thioamide-Based Fluorescent Protease Sensors". *Journal of the American Chemical Society* 136.5 (2014), 2086–2093. URL: http://dx.doi.org/10.1021/ja412297x.

[37] D Wildemann, C Schiene-Fischer, T Aumuller, A Bachmann, T Kiefhaber, C Lucke, and G Fischer. “A nearly isosteric photosensitive amide-backbone substitution allows enzyme activity switching in ribonuclease S”. *Journal of the American Chemical Society* 129.16 (2007), 4910–4918. URL: http://pubs.acs.org/doi/abs/10.1021/ja069048o.

[38] Mike SCHUTKOWSKI, Klaus NEUBERT, and Gunter FISCHER. “Influence on proline-specific enzymes of a substrate containing the thioxoaminoacyl–prolyl peptide bond”. *European Journal of Biochemistry* 221.1 (1994), 455–461. URL: http://dx.doi.org/10.1111/j.1432-1033.1994.tb18758.x.

[39] Shao Yao, Reena Zutshi, and Jean Chmielewski. “Endothiopeptide inhibitors of HIV-1 protease”. *Bioorganic & Medicinal Chemistry Letters* 8.6 (1998), 699–704. URL: http://www.sciencedirect.com/science/article/pii/S0960894X98001000.

[40] Adriana L. Andrade, Kenneth Melich, G. Gregory Whatley, Sarah R. Kirk, and Jeffrey W. Karpen. “Cyclic Nucleotide-Gated Channel Block by Hydrolysis-Resistant Tetracaine Derivatives”. *Journal of Medicinal Chemistry* 54.13 (2011), 4904–4912. URL: http://dx.doi.org/10.1021/jm200495g.

[41] H A S Hansen, K Clausen, and T F M la Cour. “Structure of the dithiopeptide ({\it Z})-Glyt-Glyt-Gly-OBzl”. *Acta Crystallographica Section C* 43.3 (1987), 522–524. URL: http://dx.doi.org/10.1107/S0108270187095179.

[42] R A Shaw, E Kollat, M Hollosi, and H H Mantsch. “Hydrogen-Bonding and Isomerization in Thioamide Peptide Derivatives”. *Spectrochimica Acta Part a-Molecular and Biomolecular Spectroscopy* 51.8 (1995), 1399–1412. URL: http://www.sciencedirect.com/science/article/pii/0584853995014489.

[43] Fernando Formaggio, Marco Crisma, Claudio Toniolo, and Cristina Peggion. “All-Thioamidated Homo-α-Peptides: Synthesis and Conformation”. *European Journal of Organic Chemistry* 2013.17 (2013), 3455–3463. URL: http://dx.doi.org/10.1002/ejoc.201300050.

[44] N Mandel, G Mandel, B L Trus, J Rosenberg, G Carlson, and R E Dickerson. “Tuna cytochrome c at 2.0 A resolution. III. Coordinate optimization and comparison of structures.” *Journal of Biological Chemistry* 252.13 (1977), 4619–4636. URL: http://www.jbc.org/content/252/13/4619.abstract.

[45] Jane S. Richardson. “The Anatomy and Taxonomy of Protein Structure”. Vol. 34. Advances in Protein Chemistry. 1981, 167–339. URL: http://www.sciencedirect.com/science/article/pii/S0065323308605203.

[46] Michael J. Dudek and Jay W. Ponder. “Accurate modeling of the intramolecular electrostatic energy of proteins”. *Journal of Computational Chemistry* 16.7 (1995), 791–816. URL: http://dx.doi.org/10.1002/jcc.540160702.

[47] Charles L. Brooks and David A. Case. “Simulations of peptide conformational dynamics and thermodynamics”. *Chemical Reviews* 93.7 (1993), 2487–2502. URL: http://dx.doi.org/10.1021/cr00023a008.

[48] Amil G. Anderson and Jan Hermans. “Microfolding: Conformational probability map for the alanine dipeptide in water from molecular dynamics simulations”. *Proteins: Structure, Function, and Bioinformatics* 3.4 (1988), 262–265. URL: http://dx.doi.org/10.1002/prot.340030408.

[49] Douglas J. Tobias and Charles L. Brooks III. “Conformational equilibrium in the alanine dipeptide in the gas phase and aqueous solution: a comparison of theoretical results”. *The Journal of Physical Chemistry* 96.9 (1992), 3864–3870. URL: http://dx.doi.org/10.1021/j100188a054.

[50] Tran Trung Tran, Herbert Treutlein, and Antony W. Burgess. “Conformational analysis of thiopeptides: (φ,ψ) maps of thio-substituted dipeptides”. *Journal of Computational Chemistry* 22.10 (2001), 1026–1037. URL: http://dx.doi.org/10.1002/jcc.1063.

[51] V.G. Badelin, E.Yu. Tyunina, A.V. Krasnov, V.V. Tyunina, N.I. Giricheva, and A.V. Girichev. “Mass spectrometry study of the sublimation of aliphatic dipeptides”. *Russian Journal of Physical Chemistry A* 86.3 (2012), 457–462. URL: http://dx.doi.org/10.1134/S0036024412030065.

[52] K. Clausen, M. Thorsen, and S.-O. Lawesson. “Studies on amino acids and peptides—I : Synthesis of N-benzyloxycarbonylendo-thiodipeptide esters”. *Tetrahedron* 37.21 (1981), 3635–3639. URL: http://www.sciencedirect.com/science/article/pii/S0040402001988921.

[53] Thomas Hoeeg-Jensen, Carl Erik Olsen, and Arne Holm. “Thioacylation Achieved by Activation of a Monothiocarboxylic Acid with Phosphorus Reagents”. *The Journal of Organic Chemistry* 59.6 (1994), 1257–1263. URL: http://dx.doi.org/10.1021/jo00085a010.

[54] Yu Rao, Xuechen Li, Pavel Nagorny, Joji Hayashida, and Samuel J. Danishefsky. “A simple method for the conversion of carboxylic acids into thioacids with Lawesson’s reagent”. *Tetrahedron Letters* 50.48 (2009), 6684–6686. URL: http://www.sciencedirect.com/science/article/pii/S0040403909018139.

[55] John P. Perdew, Kieron Burke, and Matthias Ernzerhof. “Generalized Gradient Approximation Made Simple”. *Phys. Rev. Lett.* 77 (18 1996), 3865–3868. URL: http://link.aps.org/doi/10.1103/PhysRevLett.77.3865.

[56] John P. Perdew, Matthias Ernzerhof, and Kieron Burke. “Rationale for mixing exact exchange with density functional approximations”. *The Journal of Chemical Physics* 105.22 (1996), 9982–9985. URL: http://scitation.aip.org/content/aip/journal/jcp/105/22/10.1063/1.472933.

[57] John P. Perdew, Kieron Burke, and Matthias Ernzerhof. “Generalized Gradient Approximation Made Simple [Phys. Rev. Lett. 77, 3865 (1996)]”. *Phys. Rev. Lett.* 78 (7 1997), 1396–1396. URL: http://link.aps.org/doi/10.1103/PhysRevLett.78.1396.

[58] Jacopo Tomasi, Benedetta Mennucci, and Roberto Cammi. “Quantum Mechanical Continuum Solvation Models”. *Chemical Reviews* 105.8 (2005), 2999–3094. URL: http://dx.doi.org/10.1021/cr9904009.

[59] Stefan Grimme, Jens Antony, Stephan Ehrlich, and Helge Krieg. “A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu”. *The Journal of Chemical Physics* 132.15 (2010). URL: http://scitation.aip.org/content/aip/journal/jcp/132/15/10.1063/1.3382344.

[60] Stefan Grimme, Stephan Ehrlich, and Lars Goerigk. “Effect of the damping function in dispersion corrected density functional theory”. *Journal of Computational Chemistry* 32.7 (2011), 1456–1465. URL: http://dx.doi.org/10.1002/jcc.21759.

[61] F. Matthias Bickelhaupt and Evert Jan Baerends. “Kohn-Sham Density Functional Theory: Predicting and Understanding Chemistry”. *Reviews in Computational Chemistry*. John Wiley & Sons, Inc., 2007, 1–86. URL: http://dx.doi.org/10.1002/9780470125922.ch1.

[62] Tom Ziegler and Arvi Rauk. “On the calculation of bonding energies by the Hartree Fock Slater method”. *Theoretica chimica acta* 46.1 (1977), 1–10. URL: http://dx.doi.org/10.1007/BF02401406.

[63] Tom Ziegler and Arvi Rauk. “A theoretical study of the ethylene-metal bond in complexes between copper(1+), silver(1+), gold(1+), platinum(0) or platinum(2+) and ethylene, based on the Hartree-Fock-Slater transition-state method”. *Inorganic Chemistry* 18.6 (1979), 1558–1565. URL: http://dx.doi.org/10.1021/ic50196a034.

[64] R. Parthasarathy. “Crystal structure of glycylglycine hydrochloride”. *Acta Crystallographica Section B* 25.3 (1969), 509–518. URL: http://dx.doi.org/10.1107/S0567740869002482.

[65] H. A. S. Hansen, K. Clausen, and T. F. M. la Cour. “Structure of the thiopeptide (*Z*)-Glyt-Gly-OBzl”. *Acta Crystallographica Section C* 43.3 (1987), 519–522. URL: http://dx.doi.org/10.1107/S0108270187095180.

[66] E.L. Diaz, E.S. Domalski, and J.C. Colbert. “Enthalpies of combustion of glycylglycine and dl-alanyl-dl-alanine”. *The Journal of Chemical Thermodynamics* 24.12 (1992), 1311–1318. URL: http://www.sciencedirect.com/science/article/pii/S0021961405802721.

[67] Hugh M. Huffman. “Thermal Data. XV. The heats of combustion and free energies of some compounds containing the peptide bond”. *The Journal of Physical Chemistry* 46.8 (1942), 885–891. URL: http://dx.doi.org/10.1021/j150422a011.

[68] L.A. Kochergina, O.N. Krutova, and V.G. Badelin. “Standard enthalpies of formation of D,L-α-alanyl-glycine and the products of its dissociation in aqueous solutions”. *Russian Journal of Physical Chemistry A* 85.7 (2011), 1145–1148. URL: http://dx.doi.org/10.1134/S0036024411070144.

[69] Song Nguon Ngauv, Raphael Sabbah, and Marc Laffitie. “Thermodynamique de composes azotes III. Etude Thermochimique de la glycine et de la l-α-alanine”. *Thermochimica Acta* 20.3 (1977), 371–380. URL: http://www.sciencedirect.com/science/article/pii/0040603177850910.

[70] V.P. Vasilev, V.A. Borodin, and S.B Kopnyshev. “Standard enthalpies of formation of L-histidine and L-proline”. *Russ. J. Phys. Chem. (Engl. Transl.)* 63 (1989), 891–892. URL: http://webbook.nist.gov/cgi/cbook.cgi?Source=1989VAS%2FBOR891-892&Mask=1E9F.

[71] Christopher J. Dinsmore and Douglas C. Beshore. “Recent advances in the synthesis of diketopiperazines”. *Tetrahedron* 58.17 (2002), 3297–3312. URL: http://www.sciencedirect.com/science/article/pii/S0040402002002399.

[72] G. Söhl and K. Willecke. “Gap junctions and the connexin protein family”. *Cardiovascular Research* 62.2 (2004), 228–232. URL: http://cardiovascres.oxfordjournals.org/content/62/2/228.

[73] D.A. Goodenough and D.L. Paul. “Gap junctions.” *Cold Spring Harbor perspectives in biology* 1.1 (2009), a002576. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2742079/.

[74] W.H. Evans, E. De Vuyst, and L. Leybaert. “The gap junction cellular internet: Connexin hemichannels enter the signalling limelight”. *Biochemical Journal* 397.1 (2006), 1–14. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1479757/.

[75] J.C. Sáez, V.M. Berthoud, M.C. Brañes, A.D. Martínez, and E.C. Beyer. “Plasma membrane channels formed by connexins: Their regulation and functions”. *Physiological Reviews* 83.4 (2003), 1359–1400. URL: http://physrev.physiology.org/content/83/4/1359.

[76] Anna Pfenniger, Annelise Wohlwend, and Brenda R. Kwak. “Mutations in connexin genes and disease”. *European Journal of Clinical Investigation* 41.1 (2011), 103–116. URL: http://dx.doi.org/10.1111/j.1365-2362.2010.02378.x.

[77] H. Li, T.-F. Liu, A. Lazrak, C. Peracchia, G.S. Goldberg, P.D. Lampe, and R.G. Johnson. “Properties and regulation of gap junctional hemichannels in the plasma membranes of cultured cells”. *Journal of Cell Biology* 134.4 (1996), 1019–1030. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2120949/.

[78] A. Pfahnl and G. Dahl. “Gating of cx46 gap junction hemichannels by calcium and voltage”. *Pflugers Archiv European Journal of Physiology* 437.3 (1999), 345–353. URL: http://link.springer.com/article/10.1007%2Fs004240050788.

[79] J.M. Gómez-Hernández, M. De Miguel, B. Larrosa, D. González, and L.C. Barrio. “Molecular basis of calcium regulation in connexin-32 hemichannels”. *Proceedings of the National Academy of Sciences of the United States of America* 100.26 (2003), 16030–16035. URL: http://www.pnas.org/content/100/26/16030.abstract.

[80] Arnulfo Torres, Fushun Wang, Qiwu Xu, Takumi Fujita, Radoslaw Dobrowolski, Klaus Willecke, Takahiro Takano, and Maiken Nedergaard. “Extracellular Ca2+ Acts as a Mediator of Communication from Neurons to Glia”. *Science Signaling* 5.208 (2012), ra8–ra8.

[81] D.J. Müller, G.M. Hand, A. Engels, and G.E. Sosinsky. “Conformational changes in surface structures of isolated connexin 26 gap junctions”. *EMBO Journal* 21.14 (2002), 3598–3607. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC126111/.

[82] Ilaria Fasciani, Ana Temperán, Leonel F. Pérez-Atencio, Adela Escudero, Paloma Martínez-Montero, Jesús Molano, Juan M. Gómez-Hernández, Carlos L. Paino, Daniel González-Nieto, and Luis C. Barrio. “Regulation of connexin hemichannel activity by membrane potential and the extracellular calcium in health and disease”. *Neuropharmacology* 75 (2013), 479–490. URL: http://www.sciencedirect.com/science/article/pii/S0028390813001366.

[83] F. Zonta, G. Polles, G. Zanotti, and F. Mammano. “Permeation pathway of homomeric connexin 26 and connexin 30 channels investigated by molecular dynamics”. *Journal of Biomolecular Structure and Dynamics* 29.5 (2012), 985–998. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3500697/.

[84] T. Kwon, Q. Tang, and T.A. Bargiello. “Voltage-dependent gating of the C×32*43E1 hemichannel: Conformational changes at the channel entrances”. *Journal of General Physiology* 141.2 (2013), 243–259. URL: http://jgp.rupress.org/content/141/2/243.full.

[85] D. Locke, S. Bian, H. Li, and A.L. Harris. “Post-translational modifications of connexin26 revealed by mass spectrometry”. *Biochemical Journal* 424.3 (2009), 385–398. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4640697/.

[86] S. Maeda, S. Nakagawa, M. Suga, E. Yamashita, A. Oshima, Y. Fujiyoshi, and T. Tsukihara. “Structure of the connexin 26 gap junction channel at 3.5 Å resolution”. *Nature* 458.7238 (2009), 597–602. URL: http://www.nature.com/nature/journal/v458/n7238/full/nature07869.html.

[87] T. Kwon, A.L. Harris, A. Rossi, and T.A. Bargiello. “Molecular dynamics simulations of the Cx26 hemichannel: Evaluation of structural models with Brownian dynamics”. *Journal of General Physiology* 138.5 (2011), 475–493. URL: http://jgp.rupress.org/content/138/5/475.long.

[88] V. Piazza, M. Beltramello, M. Menniti, E. Colao, P. Malatesta, R. Argento, G. Chiarella, L.V. Gallo, M. Catalano, N. Perrotti, F. Mammano, and E. Cassandro. “Functional analysis of R75Q mutation in the gene coding for Connexin 26 identified in a family with non-syndromic hearing loss”. *Clinical Genetics* 68.2 (2005), 161–166. URL: http://dx.doi.org/10.1111/j.1399-0004.2005.00468.x.

[89] G. Richard, T.W. White, L.E. Smith, R.A. Bailey, J.G. Compton, D.L. Paul, and S.J. Bale. “Functional defects of Cx26 resulting from a heterozygous missense mutation in a family with dominant deaf-mutism and palmoplantar keratoderma”. *Human Genetics* 103.4 (1998), 393–399. URL: http://link.springer.com/article/10.1007%2Fs004390050839.

[90] T. Thomas, T. Aasen, M. Hodgins, and D.W. Laird. “Transport and function of Cx26 mutants involved in skin and deafness disorders”. *Cell Communication and Adhesion* 10 (2003), 353–358. URL: http://www.ncbi.nlm.nih.gov/pubmed/14681041.

[91] Y. Chen, Y. Deng, X. Bao, L. Reuss, and G.A. Altenberg. “Mechanism of the defect in gap-junctional communication by expression of a connexin 26 mutant associated with dominant deafness”. *FASEB Journal* 19.11 (2005), 1516–1518. URL: http://www.fasebj.org/content/early/2005/08/27/fj.04-3491fje.long.

[92] Sabrina W. Yum, Kleopas A. Kleopa, Susan Shumas, and Steven S. Scherer. “Diverse Trafficking Abnormalities of Connexin32 Mutants Causing {CMTX}”. *Neurobiology of Disease* 11.1 (2002), 43–52. URL: http://www.sciencedirect.com/science/article/pii/S0969996102905450.

[93] Françoise Denoyelle, Dominique Weil, Marion A. Maw, Stephen A. Wilcox, Nicholas J. Lench, Denise R. Allen-Powell, Amelia H. Osborn, Hans-Henrik M. Dahl, Anna Middleton, Mark J. Houseman, Catherine Dodé, Sandrine Marlin, Amel Boulila-ElGaïed, Mohammed Grati, Hammadi Ayadi, Saïda BenArab, Pierre Bitoun, Geneviève Lina-Granade, Jacqueline Godet, Mirna Mustapha, Jacques Loiselet, Élie El-Zir, Anne Aubois, Alain Joannard, Jacqueline Levilliers, Éréa-Noël Garabédian, Robert F. Mueller, R. J. McKinlay Gardner, and Christine Petit. “Prelingual Deafness: High Prevalence of a 30delG Mutation in the Connexin 26 Gene”. *Human Molecular Genetics* 6.12 (1997), 2173–2177. URL: http://hmg.oxfordjournals.org/content/6/12/2173.abstract.

[94] Christoph Hamelmann, Geoffrey K. Amedofu, Katrin Albrecht, Birgit Muntau, Annette Gelhaus, George W. Brobby, and Rolf D. Horstmann. “Pattern of connexin 26 (GJB2) mutations causing sensorineural hearing impairment in Ghana”. *Human Mutation* 18.1 (2001), 84–85. URL: http://dx.doi.org/10.1002/humu.1156.

[95] Christoph Hamelmann, Geoffrey K. Amedofu, Katrin Albrecht, Birgit Muntau, Annette Gelhaus, George W. Brobby, and Rolf D. Horstmann. “Pattern of connexin 26 (GJB2) mutations causing sensorineural hearing impairment in Ghana”. *Human Mutation* 18.1 (2001), 84–85. URL: http://dx.doi.org/10.1002/humu.1156.

[96] Linda Jo Bone, Suzanne M. Deschênes, Rita J. Balice-Gordon, Kenneth H. Fischbeck, and Steven S. Scherer. "Connexin32 and X-linked Charcot–Marie–Tooth Disease". *Neurobiology of Disease* 4.3–4 (1997), 221–230. URL: http://www.sciencedirect.com/science/article/pii/S0969996197901522.

[97] Francesco Zonta, Guido Polles, Maria Sanasi, Mario Bortolozzi, and Fabio Mammano. "The 3.5 angstrom Xray structure of the human connexin26 gap junction channel is unlikely that of a fully open channel". *Cell Communication and Signaling* 11.1 (2013), 15. URL: http://www.biosignaling.com/content/11/1/15.

[98] S.F. Boys and F. Bernardi. "The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors". *Molecular Physics* 19.4 (1970), 553–566. URL: http://dx.doi.org/10.1080/00268977000101561.

[99] Stefan Dapprich, István Komáromi, K.Suzie Byun, Keiji Morokuma, and Michael J Frisch. "A new {ONIOM} implementation in Gaussian98. Part I. The calculation of energies, gradients, vibrational frequencies and electric field derivatives1". *Journal of Molecular Structure: {THEOCHEM}* 461–462 (1999), 1–21. URL: http://www.sciencedirect.com/science/article/pii/S0166128098004758.

[100] James J. P. Stewart. "Optimization of parameters for semiempirical methods II. Applications". *Journal of Computational Chemistry* 10.2 (1989), 221–264. URL: http://dx.doi.org/10.1002/jcc.540100209.

[101] Marilena Di Valentin, Elena Meneghin, Laura Orian, Antonino Polimeno, Claudia Büchel, Enrico Salvadori, Christopher W.M. Kay, and Donatella Carbonera. "Triplet–triplet energy transfer in fucoxanthin-chlorophyll protein from diatom Cyclotella meneghiniana: Insights into the structure of the complex". *Biochimica et Biophysica Acta (BBA) - Bioenergetics* 1827.10 (2013), 1226–1234. URL: http://www.sciencedirect.com/science/article/pii/S0005272813001205.

[102] Marilena Di Valentin, Claudia E. Tait, Enrico Salvadori, Laura Orian, Antonino Polimeno, and Donatella Carbonera. "Evidence for water-mediated triplet–triplet energy transfer in the photoprotective site of the peridinin–chlorophyll a–protein". *Biochimica et Biophysica Acta (BBA) - Bioenergetics* 1837.1 (2014), 85–97. URL: http://www.sciencedirect.com/science/article/pii/S0005272813001229.

[103] Todor Dudev and Carmay Lim. "Monodentate versus Bidentate Carboxylate Binding in Magnesium and Calcium Proteins: What Are the Basic Principles?" *The Journal of Physical Chemistry B* 108.14 (2004), 4546–4557. URL: http://dx.doi.org/10.1021/jp0310347.

[104] Atsunori Oshima, Tomoko Doi, Kaoru Mitsuoka, Shoji Maeda, and Yoshinori Fujiyoshi. “Roles of Met-34, Cys-64, and Arg-75 in the Assembly of Human Connexin 26: IMPLICATION FOR KEY AMINO ACID RESIDUES FOR CHANNEL FORMATION AND FUNCTION”. *Journal of Biological Chemistry* 278.3 (2003), 1807–1816. URL: http://www.jbc.org/content/278/3/1807.abstract.

[105] Xuhui Tong, William Lopez, Yu Liu, Andrew L. Harris, and Jorge E. Contreras. “Disruption of Salt Bridge Interactions Modifies Gating Kinetics of Connexin Hemichannels”. *Biophysical Journal* 104.2 (2013), 631a–632a. URL: http://dx.doi.org/10.1016/j.bpj.2012.11.3492.

[106] Vytas K. Verselis, Maria P. Trelles, Clio Rubinos, Thaddeus A. Bargiello, and Miduturu Srinivas. “Loop Gating of Connexin Hemichannels Involves Movement of Pore-lining Residues in the First Extracellular Loop Domain”. *Journal of Biological Chemistry* 284.7 (2009), 4484–4493. URL: http://www.jbc.org/content/284/7/4484.abstract.

[107] N M Kumar and N B Gilula. “The gap junction communication channel”. *Cell* 84 (1996), 381–388. URL: http://www.sciencedirect.com/science/article/pii/S0092867400812829.

[108] V M Unger, N M Kumar, N B Gilula, and M Yeager. “Three-dimensional structure of a recombinant gap junction membrane channel.” *Science* 283.5405 (1999), 1176–1180. URL: http://classic.sciencemag.org/content/283/5405/1176.long.

[109] Sarel J Fleishman, Vinzenz M Unger, Mark Yeager, and Nir Ben-Tal. “A Calpha model for the transmembrane alpha helices of gap junction intercellular channels.” *Molecular cell* 15.6 (2004), 879–88. URL: http://www.ncbi.nlm.nih.gov/pubmed/15383278.

[110] Atsunori Oshima, Kazutoshi Tani, Yoko Hiroaki, Yoshinori Fujiyoshi, and Gina E Sosinsky. “Three-dimensional structure of a human connexin26 gap junction channel reveals a plug in the vestibule.” *Proceedings of the National Academy of Sciences of the United States of America* 104 (2007), 10034–10039. URL: http://www.pnas.org/content/104/24/10034.full.

[111] Francesco Zonta, Guido Polles, Giuseppe Zanotti, and Fabio Mammano. “Permeation Pathway of Homomeric Connexin 26 and Connexin 30 Channels Investigated by Molecular Dynamics”. *Journal of Biomolecular Structure and Dynamics* 29.5 (2012), 985–998. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3500697/.

[112] Francesco Zonta, Guido Polles, Maria Federica Sanasi, Mario Bortolozzi, and Fabio Mammano. “The 3.5 ångström X-ray structure of the human connexin26 gap junction channel is unlikely that of a fully open channel.” *Cell communication and signaling : CCS* 11.1 (2013), 15. URL: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3599431{\&}tool=pmcentrez{\&}rendertype=abstract.

[113] Francesco Zonta, Damiano Buratto, Chiara Cassini, Mario Bortolozzi, and Fabio Mammano. “Molecular dynamics simulations highlight structural and functional alterations in deafness-related M34T mutation of connexin 26”. *Frontiers in Physiology* (2014), 1–9. URL: http://journal.frontiersin.org/article/10.3389/fphys.2014.00085/abstract.

[114] Francesco Zonta, Giorgia Girotto, Damiano Buratto, Giulia Crispino, Anna Morgan, Khalid Abdulhadi, Moza Alkowari, Ramin Badii, Paolo Gasparini, and Fabio Mammano. “The p.Cys169Tyr variant of connexin 26 is not a polymorphism”. *Human Molecular Genetics* 24.9 (2015), 2641–2648. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4383868/.

[115] Francesco Zonta, Fabio Mammano, Mauro Torsello, Nicola Fortunati, Laura Orian, and Antonino Polimeno. “Role of gamma carboxylated Glu47 in connexin 26 hemichannel regulation by extracellular Ca2+: Insight from a local quantum chemistry study”. *Biochemical and Biophysical Research Communications* 445.1 (2014), 10–15. URL: http://www.sciencedirect.com/science/article/pii/S0006291X14000874.

[116] M. Torsello, Antonio C. Pimenta, Lando P. Wolters, Irina S. Moreira, Laura Orian, and Antonino Polimeno. “General Amber Force Field Parameters for Diphenyldiselenides and Diphenylditellurides”. *in preparation* (2016).

[117] Viktor Hornak, Robert Abel, Asim Okur, Bentley Strockbine, Adrian Roitberg, and Carlos Simmerling. “Comparison of multiple Amber force fields and development of improved protein backbone parameters”. *Proteins: Structure, Function, and Bioinformatics* 65.3 (2006), 712–725. URL: http://onlinelibrary.wiley.com/doi/10.1002/prot.21123/abstract.

[118] Brent H. Besler, Kenneth M. Merz, and Peter A. Kollman. “Atomic charges derived from semiempirical methods”. *Journal of Computational Chemistry* 11.4 (1990), 431–439. URL: http://dx.doi.org/10.1002/jcc.540110404.

[119] Robin M. Betz and Ross C. Walker. “Paramfit: Automated optimization of force field parameters for molecular dynamics simulations”. *Journal of Computational Chemistry* 36.2 (2015), 79–87. URL: http://dx.doi.org/10.1002/jcc.23775.

[120] Min-yi Shen and ANDREJ ŠALI. “Statistical potential for assessment and prediction of protein structures”. *Protein Science: A Publication of the Protein Society* 15.11 (2006), 2507. URL: /pmc/articles/PMC2242414/?report=abstract.

[121] A Sali and T L Blundell. “Comparative protein modelling by satisfaction of spatial restraints”. *Journal of Molecular Biology* 234.3 (1993), 779–815. URL: http://www.sciencedirect.com/science/article/pii/S0022283683716268.

[122] Sunhwan Jo, Taehoon Kim, Vidyashankara G Iyer, and Wonpil Im. “CHARMM-GUI: A web-based graphical user interface for CHARMM”. *Journal of Computational Chemistry* 29.11 (2008), 1859–1865. URL: http://onlinelibrary.wiley.com/doi/10.1002/jcc.20945/abstract.

[123] Sunhwan Jo, Xi Cheng, Shahidul M Islam, Lei Huang, Huan Rui, Allen Zhu, Hui Sun Lee, Yifei Qi, Wei Han, Kenno Vanommeslaeghe, Alexander D MacKerell, Benoît Roux, and Wonpil Im. “CHARMM-GUI PDB Manipulator for Advanced Modeling and Simulations of Proteins Containing Nonstandard Residues.” *Advances in protein chemistry and structural biology* 96 (2014), 235–65. URL: http://www.sciencedirect.com/science/article/pii/S1876162314000030.

[124] Sunhwan Jo, Taehoon Kim, Wonpil Im, and Adam Yuan. “Automated Builder and Database of Protein/Membrane Complexes for Molecular Dynamics Simulations”. *PLoS ONE* 2.9 (2007), e880. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1963319/.

[125] Sunhwan Jo, Joseph B. Lim, Jeffery B. Klauda, and Wonpil Im. “CHARMM-GUI membrane builder for mixed bilayers and its application to yeast membranes”. *Biophysical Journal* 97.1 (2009), 50–58. URL: http://dx.doi.org/10.1016/j.bpj.2009.04.013.

[126] Emilia L. Wu, Xi Cheng, Sunhwan Jo, Huan Rui, Kevin C. Song, Eder M. Dávila-Contreras, Yifei Qi, Jumin Lee, Viviana Monje-Galvan, Richard M. Venable, Jeffery B. Klauda, and Wonpil Im. “CHARMM-GUI Membrane Builder toward realistic biological membrane simulations”. *Journal of Computational Chemistry* (2014), 1997–2004. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4165794/.

[127] Jumin Lee, Xi Cheng, Jason M. Swails, Min Sun Yeom, Peter K. Eastman, Justin A. Lemkul, Shuai Wei, Joshua Buckner, Jong Cheol Jeong, Yifei Qi, Sunhwan Jo, Vijay S. Pande, David A. Case, III Charles L. Brooks, Jr. Alexander D. MacKerell, Jeffery B. Klauda, and Wonpil Im. “CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field”. *Journal of Chemical Theory and Computation* 12.1 (2016), 405–413. URL: http://dx.doi.org/10.1021/acs.jctc.5b00935.

[128] William L. Jorgensen, Jayaraman Chandrasekhar, Jeffry D. Madura, Roger W. Impey, and Michael L. Klein. “Comparison of simple potential functions for simulating liquid water”. *The Journal of Chemical Physics* 79.2 (1983), 926. URL: http://scitation.aip.org/content/aip/journal/jcp/79/2/10.1063/1.445869$\backslash$nhttp://link.aip.org/link/?JCPSA6/79/926/1.

[129] In Suk Joung and Thomas E Cheatham. “Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations.” *The journal of physical chemistry. B* 112.30 (2008), 9020–41. URL: http://www.pubmedcentral.nih.

gov/articlerender.fcgi?artid=2652252{\&}tool=pmcentrez{\&}rendertype=abstract.

[130] In Suk Joung and III Thomas E. Cheatham. "Molecular Dynamics Simulations of the Dynamic and Energetic Properties of Alkali and Halide Ions Using Water-Model-Specific Ion Parameters". *The Journal of Physical Chemistry B* 113.40 (2009), 13279–13290. URL: http://dx.doi.org/10.1021/jp902584c.

[131] D.A. Case, J.T. Berryman, R.M. Betz, D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke, A.W. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T.S. Lee, S. LeGrand, P. Li, T. Luchko, R. Luo, B. Madej, K.M. Merz, G. Monard, P. Needham, H. Nguyen, H.T. Nguyen, I. Omelyan, A. Onufriev, D.R. Roe, A. Roitberg, R. Salomon-Ferrer, C.L. Simmerling, W. Smith, J. Swails, R.C. Walker, J. Wang, R.M. Wolf, X. Wu, D.M. York, and P.A. Kollma. *AMBER 14*. 2015. URL: http://ambermd.org.

[132] Ulrich Essmann, Lalith Perera, Max L Berkowitz, Tom Darden, Hsing Lee, and Lee G Pedersen. "A smooth particle mesh Ewald method". *J Chem Phys* 103.1995 (1995), 8577–8593. URL: http://scitation.aip.org/content/aip/journal/jcp/103/19/10.1063/1.470117.

[133] M F. Crowley, T a. Darden, T E. Cheatham III, and D W. Deerfield II. "Adventures in improving the scaling and accuracy of a parallel molecular dynamics program". *The Journal of Supercomputing* 11 (1997), 255. URL: http://link.springer.com/article/10.1023/A:1007907925007.

[134] Celeste Sagui and Thomas A. Darden. "P3M and PME: A comparison of the two methods". *AIP Conference Proceedings* 492.1 (1999), 104–113. URL: http://scitation.aip.org/content/aip/proceeding/aipcp/10.1063/1.1301523.

[135] William Humphrey, Andrew Dalke, and Klaus Schulten. "VMD – Visual Molecular Dynamics". *Journal of Molecular Graphics* 14 (1996), 33–38. URL: http://www.ncbi.nlm.nih.gov/pubmed/8744570#.

[136] Sander Pronk, Szilárd Páll, Roland Schulz, Per Larsson, Pär Bjelkmar, Rossen Apostolov, Michael R. Shirts, Jeremy C. Smith, Peter M. Kasson, David Van Der Spoel, Berk Hess, and Erik Lindahl. "GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit". *Bioinformatics* 29.7 (2013), 845–854. URL: http://bioinformatics.oxfordjournals.org/content/29/7/845.long.

[137] Jan Brezovsky, Eva Chovancova, Artur Gora, Antonin Pavelka, Lada Biedermannova, and Jiri Damborsky. "Software tools for identification, visualization and analysis of protein tunnels and channels". *Biotechnology Advances* 31.1 (2013), 38–49. URL: http://dx.doi.org/10.1016/j.biotechadv.2012.02.002.

[138] Atsunori Oshima. “Structure and closure of connexin gap junction channels”. *Febs Letters* 588 (2014), 1230–1237. URL: http://www.sciencedirect.com/science/article/pii/S0014579314000714.

[139] T G Back. *Organoselenium Chemistry: A Practical Approach*. 1999. URL: http://books.google.ru/books?id=dpE59twLcOEC.

[140] J Drabowicz and T Wirth. *Organoselenium Chemistry: Modern Developments in Organic Synthesis*. 2000. URL: http://books.google.ru/books?id=thY6OO5cX44C.

[141] Diana M Freudendahl, Stefano Santoro, Sohail A Shahzad, Claudio Santi, and Thomas Wirth. “Green chemistry with selenium reagents: development of efficient catalytic reactions.” *Angewandte Chemie (International ed. in English)* 48.45 (2009), 8409–11. URL: http://onlinelibrary.wiley.com/doi/10.1002/anie.200903893/abstract.

[142] Luca Sancineto, Caterina Tidei, Luana Bagnoli, Francesca Marini, Eder J. Lenardão, and Claudio Santi. “Selenium Catalyzed Oxidation of Aldehydes: Green Synthesis of Carboxylic Acids and Esters”. *Molecules* 20.6 (2015), 10496. URL: http://www.mdpi.com/1420-3049/20/6/10496.

[143] Cristina W. Nogueira, Gilson Zeni, and João B. T. Rocha. “Organoselenium and Organotellurium Compounds: Toxicology and Pharmacology”. *Chemical Reviews* 104.12 (2004), 6255–6286. URL: http://dx.doi.org/10.1021/cr0406559.

[144] Edward R. T. Tiekink. “Therapeutic potential of selenium and tellurium compounds: Opportunities yet unrealised”. *Dalton Trans.* 41 (21 2012), 6390–6395. URL: http://dx.doi.org/10.1039/C2DT12225A.

[145] RobertJ. Hondal and ErikL. Ruggles. “Differing views of the role of selenium in thioredoxin reductase”. *Amino Acids* 41.1 (2011), 73–89. URL: http://dx.doi.org/10.1007/s00726-010-0494-6.

[146] David C. Turner and Thressa C. Stadtman. “Purification of protein components of the clostridial glycine reductase system and characterization of protein A as a selenoprotein”. *Archives of Biochemistry and Biophysics* 154.1 (1973), 366–381. URL: http://www.sciencedirect.com/science/article/pii/0003986173900696.

[147] L Flohe, W a Günzler, and H H Schock. “Glutathione peroxidase: a selenoenzyme.” *FEBS letters* 32 (1973), 132–134. URL: http://www.sciencedirect.com/science/article/pii/0014579373807550.

[148] J T Rotruck, A L Pope, H E Ganther, A B Swanson, D G Hafeman, and W G Hoekstra. “Selenium: Biochemical Role as a Component of Glutathione Peroxidase”. *Science* 179.73 (1973), 588–590. URL: http://science.sciencemag.org/content/179/4073/588.abstract.

[149] F Ursini, M Maiorino, and C Gregolin. “The selenoenzyme phospholipid hydroperoxide glutathione peroxidase”. *Biochimica et biophysica acta* 839.1 (1985), 62–70. URL: http://dx.doi.org/10.1016/0304-4165(85)90182-5.

[150] J W Forstrom, J J Zakowski, and a L Tappel. “Identification of the catalytic site of rat liver glutathione peroxidase as selenocysteine.” *Biochemistry* 17.13 (1978), 2639–44. URL: http://www.ncbi.nlm.nih.gov/pubmed/678534.

[151] W A Gunzler, G J Steffens, A Grossmann, S M Kim, F Otting, A Wendel, and L Flohe. “The amino-acid sequence of bovine glutathione peroxidase”. *Hoppe Seylers Z Physiol Chem* 365.2 (1984), 195–212. URL: http://www.ncbi.nlm.nih.gov/pubmed/6714945.

[152] A. Wendel, B. Kerner, and K. Graupe. “The Selenium Moiety of Glutathione Peroxidase”. *Functions of Glutathione in Liver and Kidney*. Ed. by Helmut Sies and Albrecht Wendel. Proceedings in Life Sciences. 1978, 107–113. URL: http://dx.doi.org/10.1007/978-3-642-67132-6_13.

[153] Rajeev Prabhakar, Thom Vreven, Keiji Morokuma, and Djamaladdin G. Musaev. “Elucidation of the Mechanism of Selenoprotein Glutathione Peroxidase (GPx)-Catalyzed Hydrogen Peroxide Reduction by Two Glutathione Molecules: A Density Functional Study†”. *Biochemistry* 44.35 (2005), 11864–11871. URL: http://dx.doi.org/10.1021/bi050815q.

[154] Laura Orian, Pierluigi Mauri, Antonella Roveri, Stefano Toppo, Louise Benazzi, Valentina Bosello-Travain, Antonella De Palma, Matilde Maiorino, Giovanni Miotto, Mattia Zaccarin, Antonino Polimeno, Leopold Flohé, and Fulvio Ursini. “Selenocysteine oxidation in glutathione peroxidase catalysis: an MS-supported quantum mechanics study”. *Free Radical Biology and Medicine* 87 (2015), 1–14. URL: http://www.sciencedirect.com/science/article/pii/S0891584915002798.

[155] Govindasamy Mugesh, Wolf Walther Du Mont, and Helmut Sies. “Chemistry of biologically important synthetic organoselenium compounds”. *Chemical Reviews* 101.7 (2001), 2125–2179. URL: http://pubs.acs.org/doi/abs/10.1021/cr000426w.

[156] Krishna P. Bhabak and Govindasamy Mugesh. “Functional Mimics of Glutathione Peroxidase: Bioinspired Synthetic Antioxidants”. *Accounts of Chemical Research* 43.11 (2010), 1408–1419. URL: http://dx.doi.org/10.1021/ar100059g.

[157] Tankred Schewe. “Molecular actions of Ebselen—an antiinflammatory antioxidant”. *General Pharmacology: The Vascular System* 26.6 (1995), 1153–1169. URL: http://www.sciencedirect.com/science/article/pii/030636239500003J.

[158] R. Lesser and R. Weiß. “Über selenhaltige aromatische Verbindungen (VI)”. *Berichte der deutschen chemischen Gesellschaft (A and B Series)* 57.7 (1924), 1077–1082. URL: http://dx.doi.org/10.1002/cber.19240570703.

[159] Albrecht Wendel, Martina Fausel, Hasan Safayhi, Gisa Tiegs, and Rainer Otter. “A novel biologically active seleno-organic compound—II”. *Biochemical Pharmacology* 33.20 (1984), 3241–3245. URL: http://www.sciencedirect.com/science/article/pii/0006295284900844.

[160] A Müller, E Cadenas, P Graf, and H Sies. “A novel biologically active seleno-organic compound–I. Glutathione peroxidase-like activity in vitro and antioxidant capacity of PZ 51 (Ebselen).” *Biochemical pharmacology* 33.20 (1984), 3235–9. URL: http://www.ncbi.nlm.nih.gov/pubmed/6487370.

[161] Grégory Durand. “Synthetic Antioxidants”. *Molecular Basis of Oxidative Stress. Chemistry, Mechanisms, and Disease Pathogenesis*. 2013, 377–406. URL: http://dx.doi.org/10.1002/9781118355886.ch15.

[162] Laura Orian and Stefano Toppo. “Organochalcogen peroxidase mimetics as potential drugs: a long story of a promise still unfulfilled”. *Free Radical Biology and Medicine* 66 (2014), 65–74. URL: http://www.sciencedirect.com/science/article/pii/S0891584913001007.

[163] Michael J. Parnham and Helmut Sies. “The early research and development of ebselen”. *Biochemical Pharmacology* 86.9 (2013), 1248–1253. URL: http://www.sciencedirect.com/science/article/pii/S0006295213005339.

[164] Lando P. Wolters and Laura Orian. “Peroxidase Activity of Organic Selenides: Mechanistic Insights from Quantum Chemistry”. *Current Organic Chemistry* 20.2 (2016), 189–197. URL: http://www.eurekaselect.com/node/133476/article.

[165] Mohammad Ibrahim, Waseem Hassan, DaianeFrancine Meinerz, Matheus dos Santos, Claudia V. Klimaczewski, Anna M. Deobald, MariciliaS. Costa, CristinaW. Nogueira, NildaB.V. Barbosa, and JoaoB.T. Rocha. “Antioxidant properties of diorganoyl diselenides and ditellurides: modulation by organic aryl or naphthyl moiety”. *Molecular and Cellular Biochemistry* 371.1–2 (2012), 97–104. URL: http://dx.doi.org/10.1007/s11010-012-1426-4.

[166] IgeJ. Kade, MarcioW. Paixão, OscarE.D. Rodrigues, NildaB.V. Barbosa, AntonioL. Braga, DaianaS. Ávila, CristinaW. Nogueira, and JoãoB.T. Rocha. “Comparative Studies on Dicholesteroyl Diselenide and Diphenyl Diselenide as Antioxidant Agents and their Effect on the Activities of Na+/K+ ATPase and δ-Aminolevulinic acid Dehydratase in the Rat Brain”. *Neurochemical Research* 33.1 (2008), 167–178. URL: http://dx.doi.org/10.1007/s11064-007-9432-8.

[167] I. J. Kade and J. B. T. Rocha. “Comparative study on the influence of subcutaneous administration of diphenyl and dicholesteroyl diselenides on sulphydryl proteins and antioxidant parameters in mice”. *Journal of Applied Toxicology* 30.7 (2010), 688–693. URL: http://dx.doi.org/10.1002/jat.1542.

[168] C. M. Andersson, A. Hallberg, R. Brattsand, I. A. Cotgreave, L. Engman, and J. Persson. "Glutathione peroxidase-like activity of diaryl tellurides". *Bioorganic and Medicinal Chemistry Letters* 3.12 (1993), 2553–2558. URL: http://www.sciencedirect.com/science/article/pii/S0960894X01807150.

[169] Lars Engman, David Stern, Mikael Pelcman, and Carl M. Andersson. "Thiol Peroxidase Activity of Diorganyl Tellurides". *The Journal of Organic Chemistry* 59.8 (1994), 1973–1979. URL: http://dx.doi.org/10.1021/jo00087a008.

[170] Waseem Hassan and Joao Batista Teixeira Rocha. "Interaction Profile of Diphenyl Diselenide with Pharmacologically Significant Thiols". *Molecules* 17.10 (2012), 12287. URL: http://www.mdpi.com/1420-3049/17/10/12287.

[171] Daiane Francine Meinerz, Josiane Allebrandt, Douglas O.C. Mariano, Emily P. Waczuk, Felix Antunes Soares, Waseem Hassan, and João Batista T. Rocha. "Differential genotoxicity of diphenyl diselenide $(PhSe)_2$ and diphenyl ditelluride $(PhTe)_2$". *PeerJ* 2 (2014), e290. URL: https://doi.org/10.7717/peerj.290.

[172] Bani Kanta Sarma, and G. Mugesh. "Glutathione Peroxidase (GPx)-like Antioxidant Activity of the Organoselenium Drug Ebselen: Unexpected Complications with Thiol Exchange Reactions". *Journal of the American Chemical Society* 127.32 (2005), 11477–11485. URL: http://dx.doi.org/10.1021/ja052794t.

[173] Thomas G. Back, and Ziad Moussa. "Remarkable Activity of a Novel Cyclic Seleninate Ester as a Glutathione Peroxidase Mimetic and Its Facile in Situ Generation from Allyl 3-Hydroxypropyl Selenide". *Journal of the American Chemical Society* 124.41 (2002), 12104–12105. URL: http://dx.doi.org/10.1021/ja028030k.

[174] Thomas G. Back, and Ziad Moussa. "Diselenides and Allyl Selenides as Glutathione Peroxidase Mimetics. Remarkable Activity of Cyclic Seleninates Produced in Situ by the Oxidation of Allyl ω-Hydroxyalkyl Selenides". *Journal of the American Chemical Society* 125.44 (2003), 13455–13460. URL: http://dx.doi.org/10.1021/ja0357588.

[175] Thomas G. Back, Ziad Moussa, and Masood Parvez. "The Exceptional Glutathione Peroxidase-Like Activity of Di(3-hydroxypropyl) Selenide and the Unexpected Role of a Novel Spirodioxaselenanonane Intermediate in the Catalytic Cycle". *Angewandte Chemie International Edition* 43.10 (2004), 1268–1270. URL: http://dx.doi.org/10.1002/anie.200353128.

[176] Zeyuan Dong, Junqiu Liu, Shizhong Mao, Xin Huang, Bing Yang, Xiaojun Ren, Guimin Luo, and Jiacong Shen. "Aryl Thiol Substrate 3-Carboxy-4-Nitrobenzenethiol Strongly Stimulating Thiol Peroxidase Activity of Glutathione Peroxidase Mimic 2, 2'-Ditellurobis(2-Deoxy-β-Cyclodextrin)". *Journal of the American Chemical Society* 126.50 (2004), 16395–16404. URL: http://dx.doi.org/10.1021/ja045964v.

[177] Bruno Cardey and Mironel Enescu. “A computational study of thiolate and selenolate oxidation by hydrogen peroxide.” *Chemphyschem : a European journal of chemical physics and physical chemistry* 6.6 (2005), 1175–80. URL: http://www.ncbi.nlm.nih.gov/pubmed/15883994.

[178] Bruno Cardey, and Mironel Enescu. “Selenocysteine versus Cysteine Reactivity: A Theoretical Study of Their Oxidation by Hydrogen Peroxide”. *The Journal of Physical Chemistry A* 111.4 (2007), 673–678. URL: http://dx.doi.org/10.1021/jp0658445.

[179] Steven M. Bachrach, Dustin W. Demoin, Michelle Luk, and James V. Miller Jr. “Nucleophilic Attack at Selenium in Diselenides and Selenosulfides. A Computational Study”. *The Journal of Physical Chemistry A* 108.18 (2004), 4040–4046. URL: http://dx.doi.org/10.1021/jp037972o.

[180] S M Bachrach, C J Walker, F Lee, and S Royce. “Effect ring strain on nucleophilic sub at selenium [calcd rxn cyclic diselenide & selenenylsulfide]”. *J. Org. Chem.* 72.14 (2007), 5174–5182. URL: http://pubs3.acs.org/acs/journals/doilookup?in{_}doi=10.1021/jo070578s.

[181] Gavin S. Heverly-Coulson, Russell J. Boyd, Otilia Mó, and Manuel Yáñez. “Revealing Unexpected Mechanisms for Nucleophilic Attack on S—S and Se—Se Bridges”. *Chemistry – A European Journal* 19.11 (2013), 3629–3638. URL: http://dx.doi.org/10.1002/chem.201203328.

[182] Craig A. Bayse and Sonia Antony. “Modeling the Oxidation of Ebselen and Other Organoselenium Compounds Using Explicit Solvent Networks”. *The Journal of Physical Chemistry A* 113.19 (2009), 5780–5785. URL: http://dx.doi.org/10.1021/jp901880n.

[183] Sonia Antony and Craig A. Bayse. “Modeling the Mechanism of the Glutathione Peroxidase Mimic Ebselen”. *Inorganic Chemistry* 50.23 (2011), 12075–12084. URL: http://dx.doi.org/10.1021/ic201603v.

[184] Jason K. Pearson, and Russell J. Boyd. “Effect of Substituents on the GPx-like Activity of Ebselen: Steric versus Electronic”. *The Journal of Physical Chemistry A* 112.5 (2008), 1013–1017. URL: http://dx.doi.org/10.1021/jp076404w.

[185] Krishna P. Bhabak and Govindasamy Mugesh. “A Simple and Efficient Strategy To Enhance the Antioxidant Activities of Amino-Substituted Glutathione Peroxidase Mimics”. *Chemistry – A European Journal* 14.28 (2008), 8640–8651. URL: http://dx.doi.org/10.1002/chem.200800963.

[186] Junmei Wang, Romain M. Wolf, James W. Caldwell, Peter A. Kollman, and David A. Case. “Development and testing of a general amber force field”. *Journal of Computational Chemistry* 25.9 (2004), 1157–1174. URL: http://dx.doi.org/10.1002/jcc.20035.

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

[188] S. H. Vosko, L. Wilk, and M. Nusair. "Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis". *Canadian Journal of Physics* 58.8 (1980), 1200–1211. URL: http://www.nrcresearchpress.com/doi/abs/10.1139/p80-159#.VpOdIirhChc.

[189] David Feller. "The role of databases in support of computational chemistry calculations". *Journal of Computational Chemistry* 17.13 (1996), 1571–1586. URL: http://dx.doi.org/10.1002/(SICI)1096-987X(199610)17:13<1571::AID-JCC9>3.0.CO;2-P.

[190] Karen L. Schuchardt, Brett T. Didier, Todd Elsethagen, Lisong Sun, Vidhya Gurumoorthi, Jared Chase, Jun Li, and Theresa L. Windus. "Basis set exchange: A community database for computational sciences". *Journal of Chemical Information and Modeling* 47.3 (2007), 1045–1052. URL: http://pubs.acs.org/doi/abs/10.1021/ci600510j.

[191] Kirk A. Peterson, Detlev Figgen, Erich Goll, Hermann Stoll, and Michael Dolg. "Systematically convergent basis sets with relativistic pseudopotentials. II. Small-core pseudopotentials and correlation consistent basis sets for the post-d group 16–18 elements". *The Journal of Chemical Physics* 119.21 (2003), 11113–11123. URL: http://scitation.aip.org/content/aip/journal/jcp/119/21/10.1063/1.1622924.

[192] Piotr Cieplak, James Caldwell, and Peter Kollman. "Molecular mechanical models for organic and biological systems going beyond the atom centered two body additive approximation: aqueous solution free energies of methanol and N-methyl acetamide, nucleic acid base, and amide hydrogen bonding and chloroform/water partition coefficients of the nucleic acid bases". *Journal of Computational Chemistry* 22.10 (2001), 1048–1057. URL: http://dx.doi.org/10.1002/jcc.1065.

[193] William L. Jorgensen, Jayaraman Chandrasekhar, Jeffry D. Madura, Roger W. Impey, and Michael L. Klein. "Comparison of simple potential functions for simulating liquid water". *The Journal of Chemical Physics* 79.2 (1983), 926–935. URL: http://scitation.aip.org/content/aip/journal/jcp/79/2/10.1063/1.445869.

[194] Frank H. Allen. “The Cambridge Structural Database: a quarter of a million crystal structures and rising”. *Acta Crystallographica Section B* 58 (2002), 380–388. URL: http://dx.doi.org/10.1107/S0108768102003890.

[195] F. Matthias Bickelhaupt, Miquel Solà, and Paul von Ragué Schleyer. “Theoretical investigation of the relative stabilities of XSSX and X2SS isomers (X = F, Cl, H, and CH3)”. *Journal of Computational Chemistry* 16.4 (1995), 465–477. URL: http://dx.doi.org/10.1002/jcc.540160410.

[196] Marcello Baldo, Alain Forchioni, Kurt J. Irgolic, and Giuseppe C. Pappalardo. “Carbon-13 spin-lattice relaxation and molecular motion of diphenyl dichalcogenides”. *Journal of the American Chemical Society* 100.1 (1978), 97–100. URL: http://dx.doi.org/10.1021/ja00469a016.

[197] A. K. Rappé, C. J. Casewit, K. S. Colwell, W. A. Goddard, and W. M. Skiff. “UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations”. *Journal of the American Chemical Society* 114.25 (1992), 10024–10035. URL: http://pubs.acs.org/doi/abs/10.1021/ja00051a040?journalCode=jacsat.

[198] Simone Silvestrini, Daniela Dalle Nogare, Tommaso Carofiglio, Enzo Menna, Paolo Canu, and Michele Maggini. “Continuous Flow Synthesis of Methanofullerenes in Microstructured Reactors: A Kinetic Study”. *European Journal of Organic Chemistry* 2011.28 (2011), 5571–5576. URL: http://dx.doi.org/10.1002/ejoc.201100993.

[199] Daniel Mark, Stefan Haeberle, Gunter Roth, Felix von Stetten, and Roland Zengerle. “Microfluidic lab-on-a-chip platforms: requirements, characteristics and applications”. *Chem. Soc. Rev.* 39 (3 2010), 1153–1182. URL: http://dx.doi.org/10.1039/B820557B.

[200] (2013). URL: http://www.math.kz/images/journal/2013-4/Otelbaev_N-S_21_12_2013.pdf.

[201] (2013). URL: http://www.openfoam.org.

[202] D Lanser and J.G Verwer. “Analysis of operator splitting for advection–diffusion–reaction problems from air pollution modelling”. *Journal of Computational and Applied Mathematics* 111.1–2 (1999), 201–216. URL: http://www.sciencedirect.com/science/article/pii/S0377042799001430.

[203] Gilbert Strang. “On the Construction and Comparison of Difference Schemes”. *SIAM Journal on Numerical Analysis* 5.3 (1968), 506–517. URL: http://dx.doi.org/10.1137/0705041.

[204] Milan Kuchařík, Richard Liska, Stanly Steinberg, and Burton Wendroff. “Optimally-stable Second-order Accurate Difference Schemes for Non-linear Conservation Laws in 3D”. *Appl. Numer. Math.* 56.5 (May 2006), 589–607. URL: http://dx.doi.org/10.1016/j.apnum.2005.04.004.

[205] P L Roe. “Characteristic-Based Schemes for the Euler Equations”. *Annual Review of Fluid Mechanics* 18.1 (1986), 337–365.

[206] Randall J. LeVeque. *Finite volume methods for hyperbolic problems*. Cambridge texts in applied mathematics. Cambridge, New York: Cambridge University Press, 2002. URL: http://opac.inria.fr/record=b1100566.

[207] P. K. Sweby. “High Resolution Schemes Using Flux Limiters for Hyperbolic Conservation Laws”. *SIAM Journal on Numerical Analysis* 21.5 (1984), 995–1011. URL: http://www.jstor.org/stable/2156939.

[208] J. Strikwerda. *Finite Difference Schemes and Partial Differential Equations, Second Edition*. Society for Industrial and Applied Mathematics, 2004. URL: http://epubs.siam.org/doi/abs/10.1137/1.9780898717938.

[209] Jr. Douglas Jim. “Alternating direction methods for three space variables”. English. *Numerische Mathematik* 4.1 (1962), 41–63. URL: http://dx.doi.org/10.1007/BF01386295.

[210] Ting-Yuan Wang and C.C. Chen. “3-D Thermal-ADI: a linear-time chip level transient thermal simulator”. *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on* 21.12 (2002), 1434–1445. URL: http://ieeexplore.ieee.org/xpls/icp.jsp?arnumber=1097863.

[211] J. Crank and P. Nicolson. “A practical method for numerical evaluation of solutions of partial differential equations of the heat-conduction type”. *Advances in Computational Mathematics* 6.1 (1996), 207–226. URL: http://dx.doi.org/10.1007/BF02127704.

[212] J.W. Thomas. *Numerical Partial Differential Equations: Finite Difference Methods*. Graduate Texts in Mathematics v. 1. Springer, 1995. URL: http://link.springer.com/book/10.1007%2F978-1-4899-7278-1.

[213] J.W. Thomas. *Numerical Partial Differential Equations: Finite Difference Methods*. Texts in Applied Mathematics. Springer New York, 2013. URL: http://www.springer.com/us/book/9780387983462.

[214] Zhangping Wei, Byunghyun Jang, Yaoxin Zhang, and Yafei Jia. “Parallelizing Alternating Direction Implicit Solver on {GPUs}”. *Procedia Computer Science* 18 (2013), 389–398. URL: http://www.sciencedirect.com/science/article/pii/S1877050913003451.

[215] V.M. Paskonov, S.B. Berezin, and E.S. Korukhova. “A dynamic visualization system for multiprocessor computers with common memory and its application for numerical modeling of the turbulent flows of viscous fluids”. *Moscow University Computational Mathematics and Cybernetics* 31.4 (2007), 133–142. URL: http://dx.doi.org/10.3103/S0278641907040012.

[216] David Michéa and Dimitri Komatitsch. “Accelerating a three-dimensional finite-difference wave propagation code using GPU graphics cards”. *Geophysical Journal International* 182.1 (2010), 389–402. URL: http://dx.doi.org/10.1111/j.1365-246X.2010.04616.x.

[217] Alan C. Hindmarsh, Peter N. Brown, Keith E. Grant, Steven L. Lee, Radu Serban, Dan E. Shumaker, and Carol S. Woodward. “SUNDIALS: Suite of Nonlinear and Differential/Algebraic Equation Solvers”. *ACM Trans. Math. Softw.* 31.3 (2005), 363–396. URL: http://doi.acm.org/10.1145/1089014.1089020.

[218] J.H. Holland. *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*. University of Michigan Press, 1975. URL: https://books.google.co.in/books?id=JE5RAAAAMAAJ.

[219] Silvia Carlotto, Laura Orian, and Antonino Polimeno. “Heuristic approaches to the optimization of acceptor systems in bulk heterojunction cells: a computational study”. *Theoretical Chemistry Accounts* 131.3 (2012), 1–7. URL: http://dx.doi.org/10.1007/s00214-012-1191-1.

[220] Susana Gomez, Benjamin Ivorra, and Angel Manuel Ramos. “Optimization of a pumping ship trajectory to clean oil contamination in the open sea”. *Mathematical and Computer Modelling* 54.1–2 (2011), 477–489. URL: http://www.sciencedirect.com/science/article/pii/S0895717711001403.

[221] David E. Hertzog, Benjamin Ivorra, Bijan Mohammadi, Olgica Bakajin, and Juan G. Santiago. “Optimization of a Microfluidic Mixer for Studying Protein Folding Kinetics”. *Analytical Chemistry* 78.13 (2006), 4299–4306. URL: http://dx.doi.org/10.1021/ac051903j.

[222] Benjamin Ivorra, Juana L. Redondo, Juan G. Santiago, Pilar M. Ortigosa, and Angel M. Ramos. “Two- and three-dimensional modeling and optimization applied to the design of a fast hydrodynamic focusing microfluidic mixer for protein folding”. *Physics of Fluids* 25.3 (2013). URL: http://scitation.aip.org/content/aip/journal/pof2/25/3/10.1063/1.4793612.

[223] Enrica Duchi, Simone Rinaldi, and Gilles Schaeffer. “The number of Z-convex polyominoes”. *Advances in Applied Mathematics* 40.1 (2008), 54–72. URL: http://www.sciencedirect.com/science/article/pii/S0196885806002004.

[224] Jennifer Ryan Charles A. Anderson Kathryn F. Jones. “A Two-Dimensional Genetic Algorithm for the Ising Problem”. *Complex Systems* 5.3 (1991), 327–333.

[225] John A Pople and Warren J Hehre. “Computation of electron repulsion integrals involving contracted Gaussian basis functions”. *Journal of Computational Physics* 27.2 (1978), 161–168. URL: http://www.sciencedirect.com/science/article/pii/0021999178900013.

[226] J. Rys, M. Dupuis, and H. F. King. “Computation of electron repulsion integrals using the rys quadrature method”. *Journal of Computational Chemistry* 4.2 (1983), 154–157. URL: http://dx.doi.org/10.1002/jcc.540040206.

[227] Larry E McMurchie and Ernest R Davidson. “One- and two-electron integrals over cartesian gaussian functions”. *Journal of Computational Physics* 26.2 (1978), 218–231. URL: http://www.sciencedirect.com/science/article/pii/002199917890092X.

[228] S. Obara and A. Saika. “Efficient recursive computation of molecular integrals over Cartesian Gaussian functions”. *The Journal of Chemical Physics* 84.7 (1986), 3963–3974. URL: http://scitation.aip.org/content/aip/journal/jcp/84/7/10.1063/1.450106.

[229] Martin Head-Gordon and John A. Pople. “A method for two-electron Gaussian integral and integral derivative evaluation using recurrence relations”. *The Journal of Chemical Physics* 89.9 (1988), 5777–5786. URL: http://scitation.aip.org/content/aip/journal/jcp/89/9/10.1063/1.455553.

[230] Peter M. W. Gill and John A. Pople. “The prism algorithm for two-electron integrals”. *International Journal of Quantum Chemistry* 40.6 (1991), 753–772. URL: http://dx.doi.org/10.1002/qua.560400605.

[231] Kazuhiro Ishida. “ACE algorithm for the rapid evaluation of the electron-repulsion integral over Gaussian-type orbitals”. *International Journal of Quantum Chemistry* 59.3 (1996), 209–218. URL: http://dx.doi.org/10.1002/(SICI)1097-461X(1996)59:3<209::AID-QUA4>3.0.CO;2-1.

[232] Kazuhiro Ishida. “Rapid algorithm for computing the electron repulsion integral over higher order Gaussian-type orbitals: Accompanying coordinate expansion method”. *Journal of Computational Chemistry* 19.8 (1998), 923–934. URL: http://dx.doi.org/10.1002/(SICI)1096-987X(199806)19:8<923::AID-JCC11>3.0.CO;2-8.

[233] Masato Kobayashi and Hiromi Nakai. “New recurrence relations for the rapid evaluation of electron repulsion integrals based on the accompanying coordinate expansion formula”. *The Journal of Chemical Physics* 121.9 (2004), 4050–4058. URL: http://scitation.aip.org/content/aip/journal/jcp/121/9/10.1063/1.1778712.

[234] Masao Hayami, Junji Seino, and Hiromi Nakai. “Extension of accompanying coordinate expansion and recurrence relation method for general-contraction basis sets”. *Journal of Computational Chemistry* 35.20 (2014), 1517–1527. URL: http://dx.doi.org/10.1002/jcc.23646.

[235] Martin B. Short, P. Jeffrey Brantingham, Andrea L. Bertozzi, and George E. Tita. “Dissipation and displacement of hotspots in reaction-diffusion models of crime”. *Proceedings of the National Academy of Sciences* 107.9 (2010), 3961–3965. URL: http://www.pnas.org/content/107/9/3961.abstract.

[236] H. H. Robertson. “The solution of a set of reaction rate equations”. *Academ. Press.* (1966), 178–182.

[237] Ernst Hairer and Gerhard Wanner. *Solving ordinary differential equations II: Stiff and differential-algebraic problems*. Vol. 14. 1991, xvi+601. URL: http://www.mendeley.com/research/solving-ordinary-differential-equations-ii-stiff-and-differentialalgebraic-problems-1/.

Acknowledgements

Vorrei innanzitutto ringraziare il mio supervisore, il prof. Antonino Polimeno, per avermi offerto l'opportunità di lavorare su tematiche estremamente affascinanti, dato spazio per pensare e crescere ma anche per aver creduto in me.

Ringrazio la dott.ssa Laura Orian che mi ha sostenuto durante questi tre anni, con cui ho condiviso gran parte del lavoro svolto e che mi ha spronato ad esprimere i miei pensieri, nonostante la mia timidezza.

Ringrazio la prof.ssa Camilla Ferrante per il bellissimo periodo trascorso a fare didattica in laboratorio con gli studenti, per la sua simpatia e la sua generosità.

Un ringraziamento va ai miei colleghi di dottorato Nicola, Marco e Maurizio. La vostra compagnia e simpatia è stata fondamentale per giungere fino a questo punto. Mi mancheranno le nostre mangiate e bevute assieme!

Ringrazio i colleghi d'ufficio Paolo e Mirco per avermi incoraggiato ed avermi sempre dato ottimi suggerimenti.

Ringrazio tutti i ragazzi del nostro gruppo di ricerca ma non solo: Francesco, Enrico, Alessandro, Marco con cui ho condiviso idee e opinioni, Marco Albertini per la sua simpatia e generosità ma anche Lando per la sua disponibilità e gentilezza.

Giulia un ringraziamento speciale va a te! Non sarei giunto fin qui se non fosse stato per la tua presenza, mia stella. Esserti vicino mi rende una persona migliore e mi fa sentire speciale.

Ringrazio di cuore tutta la mia famiglia: mia madre Enza, mio padre Attilio, la mia adorata sosò Ylenia e Giovanni, sapere che ci siete sempre ridimensiona le difficoltà e i problemi quotidiani. Qualunque ringraziamento vi faccia non sarà mai abbastanza e nonostante sia da un po' di tempo che siamo distanti, saper di poter sempre tornare a casa mi rincuora e mi fa sentire più vicino.

Ringrazio tutti i miei coinquilini Luigi, Davide, Miriana, Corinna e Michele per aver reso via San Pietro 55 la nostra Casa.

Ringrazio Riccardo, Eleonora e Alessandro con cui ho trascorso momenti stupendi facendo apnea. Spero di tornar presto ad essere dei vostri!

Infine ringrazio tutti gli amici "di giù": Leo, Fabio, Antonio, Sara e gli amici "di su": Matteo, Enrico, Alessandro e Stefano che mi hanno sostenuto durante tutti questi anni trascorsi nella bellissima Padova.