

Recent Advances on Computational Proteomics

Sérgio F. Sousa, Nuno M. F. S. A. Cerqueira, Marta A. S. Perez, Irina S. Moreira, António J. M. Ribeiro, Ana R. A. P. Neves, Maria J. Ramos and Pedro A. Fernandes

Abstract—In this work we report the recent progresses that have been achieved by our group in the last half decade on the field of computational proteomics. Specifically, we discuss the application of Molecular Dynamics Simulations and Electronic Structure Calculations in drug design, in the clarification of the structural and dynamic properties of proteins and enzymes and in the understanding of the catalytic and inhibition mechanism of cancer-related enzymes. A set of examples illustrate the concepts and help to introduce the reader into this important and fast moving field.

Keywords—Enzyme, Molecular Dynamics, Protein, Quantum Mechanics.

I. INTRODUCTION

The application of computational methods in the study of problems related to proteins and enzymes and to their interaction with other molecules is a powerful and vast field of research, which encompasses several different disciplines, normally grouped under the generic designation of Computational Proteomics. In this work, we trace and up-to-date analysis of this rich field of research, by presenting a brief introduction to its current standing and in particular to its most recent developments, illustrated through a selection of studies from our research group on 6 important thematic areas. These are the molecular dynamics simulations, the study of enzymatic mechanisms, benchmarking of DFT functionals,

S. F. Sousa is with REQUIMTE, Departamento de Química, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre, 687, 4169-007 Porto, Portugal.

N. M. F. S. A. Cerqueira is with REQUIMTE, Departamento de Química, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre, 687, 4169-007 Porto, Portugal.

M. A. S. Perez is with REQUIMTE, Departamento de Química, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre, 687, 4169-007 Porto, Portugal.

A. J. M. Ribeiro is with REQUIMTE, Departamento de Química, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre, 687, 4169-007 Porto, Portugal.

A. R. A. P. Neves is with REQUIMTE, Departamento de Química, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre, 687, 4169-007 Porto, Portugal.

I. S. Moreira was with REQUIMTE, Departamento de Química, Faculdade de Ciências, Universidade do Porto. She is now with the Weill Cornell Medical College, Department of Physiology and Biophysics, 1300 York Avenue, New York, NY 10065-4896, USA.

M. J. Ramos is with REQUIMTE, Departamento de Química, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre, 687, 4169-007 Porto, Portugal.

P. A. Fernandes is with REQUIMTE, Departamento de Química, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre, 687, 4169-007 Porto, Portugal, (phone: +351220402667; fax: +351220402659; e-mail: pafeman@fc.up.pt).

alanine scanning mutagenesis, drug design and flexible receptor docking.

II. MOLECULAR DYNAMICS SIMULATIONS

Molecular Dynamics (MD) simulations are now a standard tool in the study of biomolecules. Following the publication of the first study describing an MD simulation of a protein (the bovine pancreatic trypsin inhibitor) already 30 years ago [1], this field has been one of strong and enthusiastic development, taking advantage of the astonishing computational progress that has characterized the past decades and, more recently, of the good parallelization efficiency of most modern MD algorithms. MD is hence nowadays a rapidly developing field of science, with MD simulations being routinely applied to investigate the structure, dynamics and thermodynamics of biological molecules and their complexes. In fact, the dynamic properties of a protein have a profound effect on its functional behavior. MD simulations allow the study of the dynamic properties of a system. They enable the complex and dynamic processes that take place in biological systems to be analyzed and provide the ultimate detail concerning the individual particle motion as a function of time. Examples of application include the study of phenomena such as protein stability, molecular recognition, conformational changes, protein folding, and ion transport in biological systems [2], [3]. In addition, MD simulations are also frequently applied in drug design and in NMR and X-ray structure determination.

Basically, molecular dynamics can be seen as a type of deterministic computer simulation in which the time evolution of a set of interacting atoms is followed by the numeric integration of their equations of motion. Hence, molecular dynamics follows the laws of classical mechanics and most notably Newton's law. The resolution of the classical equation of motion allows kinetic and thermodynamic properties on the systems of interest to be derived. The interaction between different atoms and molecules are described by simple approximations of known physics, included in the force field used. The force field is comprised by two parts: one is the energy expression, a particular mathematical form that describes the interaction between all the particles in the system; the other part is a set of parameters for the several terms in the force field equation, fitted to reproduce experimental data or high-level ab initio molecular orbital calculations. Several different force fields exist, differing in terms of scope, accuracy and computational cost associated.

Popular biomolecular force fields include AMBER [4]-[6], CHARMM [7]-[9], and OPLS [10]-[12].

During the last decade we have tried to take advantage of the strengths of this methodology (alone or in combination with other computational methods) to help unraveling some of the secrets behind the activity of a significant number of biologically critical enzymes. Examples include the amyloid-beta binding alcohol dehydrogenase (ABAD) enzyme [13], the carbohydrate-binding modules (CBMs) from family 11 [14], [15], thioredoxin [16], reverse transcriptase [17], and Farnesyltransferase enzyme [18]-[20].

ABAD is a mitochondrial enzyme that catalyzes the NAD-dependent reversible oxidation and/or reduction of alcohol groups in a wide variety of substrates, including linear alcohols, steroids, branched fatty acids, 2-methyl-3-hydroxybutyryl-coA and β -hydroxybutyrate. This broad substrate specificity of ABAD enables it to participate in several metabolic processes within the mitochondria, namely, the oxidation of fatty acids, degradation of isoleucine, sex steroids metabolism and oxidation of steroid modulators of GABA_A receptors. To understand the activity of this enzyme, we have performed 10 ns MD simulations with the AMBER force field on the homotetramer of the ABAD enzyme, and on the structural units that comprise this tetramer, i.e. dimer and monomer, in the presence and absence of a NAD-inhibitor adduct [13]. The results showed that the tetramer, dimer and monomer all have a comparable stability, with the tetramerization process leading to an additional stabilization of regions on the enzyme that were exposed to the solvent in the dimer and monomer structures. This MD study also provided important atomistic insight on several important features: the binding of the cofactor and inhibitor stabilizes the protein, particularly at the substrate binding loop. In the absence of the ligand, this region was found to have a much higher flexibility and to adopt an open conformation. An additional interesting result emerging from this work was the conformational flexibility exhibited by the azepane and benzene rings of the inhibitor moiety of the adduct, which appears to be influenced by the mobility of the substrate binding loop. Globally, this study provided important clues for the design of new ABAD inhibitors, by showing the need to incorporate the flexibility of the substrate binding loop in the process.

CBMs are noncatalytic modules of a high-hierarchy multi-subunit complex called cellulosome, which are known to be crucial for the efficient degradation of polysaccharides, a process that is of high economical interest. We have used MD simulations [14], [15], with both CHARMM and AMBER force fields in an integrated approach with the recently developed MADAMM docking protocol [21] to understand how substrates bind to CBMs from family 11. Our results have shown that the binding interface of the CBM11 can bind only one single polysaccharide chain. The central binding site has affinity for polysaccharides with more than four subunits, while there are four main residues that have a central role in this interaction: Asp99, Arg126, Asp128, and Asp146. The

results further show that there are three tyrosine residues, Tyr22, Tyr53, and Tyr129 that are crucial for the guiding and packing of the polysaccharide to the charged regions, providing important insights into substrate binding by this important class of proteins.

Thioredoxins (Trx) are enzymes with a characteristic CXYC active-site motif that catalyze the reduction of disulfide bonds in other proteins. Following a DFT study on the reduction of disulfide bonds by this enzyme, in both the gas-phase and in water, considering a small active-site model, 1 ns of MD simulations on the full enzyme were performed for the CGPC, CGGC, wild-type Trx, and P34G Trx systems [16]. The results from the MD simulations provided important insight into the activity of this enzyme by showing that the activity of the thioredoxin active-site motif (CXYC) is determined not only by the structural rigidity associated with the particular variable residues (XY) but also by the number of amide N-H groups. The latter are involved in the stabilization of the Cys-32 thiolate and thus affect the acidity and nucleophilicity of this residue, a conclusion that opens new doors in understanding the activity of this enzyme.

Reverse transcriptase (RT) is a critical enzyme for the replication process of the human immunodeficiency virus (HIV). RT synthesizes a double-stranded DNA from the viral (+)-RNA genome suitable for integration in the host genome. Several nucleotide reverse transcriptase inhibitors (NRTIs) are currently used to inhibit retroviral replication. However, the development of resistance to NRTIs therapy through the accumulation of mutations in RT is an important limitation in the treatment of HIV/acquired immunodeficiency syndrome (AIDS), with the pattern of mutations developed greatly depending on the drugs administered during the antiretroviral therapy. We have employed a series of MD simulations using the AMBER force field to investigate the structural determinants in the NRTIs that cause RT resistance [17]. In particular, 1 ns of MD simulations were performed on the enzyme complexes with the natural substrate (dNPT) and with two very similar inhibitors that are known to follow different mechanism of drug resistance development - phosphorylated zalcitabine (ddCTP) and phosphorylated stavudine (d4TTP). The results allowed us to propose that the different resistance profiles characteristic of different inhibitors are a result of the different conformations adopted by the inhibitors at the N site. While d4TTP adopts an ideal conformation for catalysis forming an ion-dipole intramolecular interaction with the β -phosphate oxygen of the triphosphate, very similar to the one adopted by the natural substrate, ddCTP adopts a very different non-catalytic conformation. More studies and required to fully validate this hypothesis.

Farnesyltransferase enzyme is an important metalloenzyme that catalyzes the addition of a farnesyl group from farnesyl diphosphate (FPP), to a cysteine residue of a protein substrate containing a typical -CAAX motif at the carboxyl terminus, where C represents the cysteine residue that is farnesylated, A is an aliphatic amino acid, and X represents the terminal amino

acid residue [22]. Among the CAAX substrates for FTase are a number of biologically very relevant protein targets, known to be implicated in cancer development (like the Ras family of proteins) a feature that has turned the spotlight of the pharmaceutical industry to this important enzyme. Presently more than 500 patents describing FTase inhibitors have been developed [23]. We have performed 10 ns MD simulations on the 4 main catalytic states of this enzyme to understand how this enzyme works and also the effect of the solvent on the activity of this enzyme [18], [19]. Atomistic details on aspects like amino acid flexibility, radial distribution functions of water molecules surrounding the catalytically relevant Zn atom, the conformations of the substrate and product molecules, critical hydrogen bonds and some catalytically relevant distances were obtained providing important information for the subsequent modelling of the catalytic mechanism of this enzyme [24] and validating proposals that have arisen from previous first principles and hybrid QM/MM mechanistic calculations [25]-[28].

III. ENZYMATIC MECHANISMS

Computational methods offer an unparalleled opportunity to dissect in minute detail important enzymatic reaction mechanisms, providing atomistic insight into very specific processes of highly biological significance. While a vast number of experimental techniques are normally employed in the study of reaction mechanism, from spectroscopic experiments and mutagenesis studies to kinetic evaluations, no experimental technique, by itself, can normally give a full view of an entire catalytic pathway. Alone, such methodologies typically provide clues, observations, indications, or eventually working hypothesis, that have to be complemented with results obtained from other methods to yield a mechanistic proposal. As a puzzle that is being built, piece by piece, such mechanistic proposals tend to have some pieces missing. In addition, as more and more experimental studies are performed, new results that do not seem to fit exactly with the mechanistic proposal that is being put forward tend to accumulate. However, in spite of vast, the range of experimental techniques that can be applied to a given biological problem is limited. Computational methods can be used to go beyond the traditional limitations of standard experimental studies, providing the missing pieces. In fact, a variety of computational methods, including first principles quantum mechanics (QM), and most notably hybrid quantum mechanical/molecular mechanics (QM/MM), are nowadays routinely used to assess the viability of mechanistic proposals, to rationalize and reconcile data previously regarded as apparently contradicting experimental evidence, to discard mechanistic alternatives and to propose new pathways that are consistent with the available experimental data [29]. From the large number of catalytic mechanisms that we have studied during the past decade, we would like to highlight two particularly challenging systems, for which the application of

computational methods proved to be crucial. These are the Zinc metalloenzyme Farnesyltransferase (FTase) and the radical enzyme Ribonucleotide reductase (RNR)

A. Farnesyltransferase

In spite of the relatively large number of kinetic studies on mutant species of FTase and of the existence of X-ray crystallographic structures for the main intermediate states in the catalytic pathway of this enzyme [30], [31], the very limited range of experimental techniques directly applicable to the study of Zn(II) biological systems limited in great part an understanding of the catalytic mechanism of this enzyme, in which this metal cation was known to play a critical role [22]. In particular, the nature of the Zn coordination sphere was the subject of several very pungent doubts, with the different possible coordination alternatives implying very different catalytic mechanisms. Spectroscopic EXAFS data demonstrated that at the resting state the metal coordination sphere of this enzyme was tetracoordinated with one Zn-S bond, and three Zn-O/N bonds [32], an observation that together with the X-ray and mutagenesis studies would suggest Asp297 β , Cys299 β , and His362 β as Zinc ligands, but that would leave room for many doubts regarding the identity of the fourth ligand. Two strong alternatives had been proposed from experimental studies. These were a water molecule and the second oxygen atom of Asp297 β .

We used first principles quantum mechanics to solve this dilemma by performing quantum calculations (B3LYP) on small active site models, comprising the metal coordination sphere and the first shell of amino acid residues for both coordination alternatives [25]. Surprisingly, we found out that both alternatives yielded bond-lengths consistent with the EXAFS data and were located in very close energetic proximity. In addition, we demonstrated that the conversion between both states could take place with a very small energetic barrier at room temperature. Together these observations justified the existence of both states at equilibrium, through a mechanism termed carboxylate-shift [27]. This conclusion, which was in agreement with the experimental data available allowed a new mechanistic paradigm for FTase to be proposed, and was a major hallmark in FTase research.

The same type of studies allowed us to clarify the metal coordination spheres on the other intermediate states in the catalytic pathway of this enzyme [26], [28], [33]. Larger QM models and in particular QM/MM methods were used to bridge the gap between the different intermediate states, and to clarify other more global mechanistic doubts, allowing us to draw a full mechanist portrait of this enzyme. This set of studies culminated with the determination of a detailed atomistic quantum-chemical transition state model for the reaction catalyzed by this enzyme [24], a structure that can provide a blueprint for the design of more potent and specific FTase inhibitors, and that should be taken into consideration in

future studies on the inhibition of this important enzyme.

B. Ribonucleotide Reductase

Ribonucleotide Reductase is an enzyme that catalyzes the conversion of ribonucleotides into deoxyribonucleotides. Since the exploit and maintenance of the DNA information is dependent on the availability of deoxyribonucleotides through DNA replication and repair, this enzyme has become an attractive target for antitumor, antiviral and antibacterial therapies [34]. Several studies have been addressed to this enzyme in the past 20 years. However, and in spite of the fact that two compounds have been recently approved by the EMEA and FDA for the treatment of cancer (gemzar® and tezacitabine®), the catalytic mechanism and the inhibitory mechanism of this enzyme were barely understood five years ago, when our group started to study this system.

During this period, we have developed a comprehensive knowledge and thorough understanding of every process involved in the normal RNR functioning and on its inhibition mechanisms. For this purpose, computational methods, and in particular quantum chemical calculations have been used as an indispensable complement to the available structural biology studies and biochemical experiments. One advantage of such methods over the experimental techniques is that they can provide detailed structural and energetic information on transition states and reactive intermediates, not possible by normal physical methods due to their transient nature. Such methods are of particular importance in the case of RNR, since it is a radical enzyme, and these methods enable us to identify unstable intermediates to get a clear evaluation of the chemical mechanisms involved in these processes.

Our first approach was to explore different strategies for modeling the active site of this enzyme that is located in R1 monomer. For this effect we have built several models that ranging from 20 atoms [35], to 300 atoms [36], and more recently to the full R1 monomer (that contains 30000 atoms) [37]. For that hybrid QM/MM methodologies, and in particular the ONIOM approach were used allowing us to study the atoms directly involved in the reaction with high and computational expensive theoretical levels, while the remaining ones were described with lower and less expensive theoretical level. This approach has proven to be most successful, allowing us to tune the best computational approach for the study of this enzyme. This information was then used to explore all relevant chemical pathways that could be involved in the catalytic mechanism [35], [38] of RNR as well as in the inhibitory mechanism of several substrate analogues inhibitors that inactivate the function of the enzyme, either by destroying the radical that is essential for RNR activity, by covalent bonding of one intermediate to some residue that inactivates the active site or by inducing allosteric malfunction, or by promoting the dissociation of an intermediate that subsequently inactivates the enzyme [39]. The structure of these inhibitors is very similar to the one of the natural substrate, generally differing only on the atoms, or group of atoms, that are attached to carbons C-2', or C-3', of

the ribose ring. The inhibitors that were studied during this period were: the 2'-chloro-2'-deoxynucleoside-5'-diphosphates (Cl dNDP) [40], [41], 2'-azido-2'-deoxynucleoside-5'-diphosphates (N3 dNDP) [42], 2'-deoxy-2'-fluoro-nucleoside-5'-diphosphates (FdNDP) [40], [41], 2'-deoxy-2',2'-difluorocytidine-5'-diphosphate (F2 dNDP) [43], [44], 2'-mercapto-2'-deoxyarabi-nosylcytosine (SH dNDP) [44] and (E)-2'-Fluoromethylene-2'-deoxycytidine-5'-diphosphate (CH dNDP) [45]. Some of the results obtained during these years were pioneer in the RNR field and later confirmed and acknowledged by experimental findings. The collected information is now a valuable tool, representing a step forward in understanding all aspects related to the RNR mode of action. This can now be used in the investigation of the details that remain to be learned and that will help in finding new, more effective and safer ways of inhibiting the enzyme in a near future.

IV. BENCHMARKING OF DFT FUNCTIONALS

Density Functional Theory (DFT) has become one of the most used methods for calculating a variety of molecular properties, such as thermochemistry and thermochemical kinetics [46], [47]. The main reason for the popularity of DFT is the fact that these methods allow the inclusion of electron correlation without being as demanding computationally as other computational methods, such as post-Hartree-Fock methods. For this reason, DFT enables calculations on molecules of 100 or more atoms to be performed, a number which would be unacceptable with the other methods. Furthermore, DFT seems to give good results for systems involving *d*-block metals, what is not true for HF methods. Other reason for the success of DFT is that the density functionals include both exchange and correlation terms, which gives it an enormous edge in terms of computational diversity and efficiency over Hartree-Fock-based *ab initio* methods [46].

DFT credibility is being strengthened since 1964 when it was established by the two Hohenberg-Kohn theorems that the ground state electronic energy of a non-degenerate system (and, consequently, all other electronic properties) depends only on the electronic density and could be calculated by a density functional [48]. Such functional, however, was not known at the time and continues incognito at the present. Various attempts to obtain approximated functionals have been made, nevertheless, resulting in the appearance of dozens of possible alternatives [49], [50].

We can divide these density functionals into seven different categories: LSDA (Local Spin Density Approximation), GGA (Generalized Gradient Approximation), hybrid GGA, meta GGA, hybrid meta GGA, GGE (Generalized Gradient Exchange) and GGSC (Generalized Gradient with Scaled Correlation). The LSDA functionals depend only on the electron density, while GGA functionals depend explicitly on the gradient of electron density as well as on the density itself;

hybrid GGA functionals add a certain amount of Hartree–Fock (HF) exchange and meta GGA functionals depend on the kinetic energy density in addition to the electron density and its gradient. In that way, the hybrid meta GGA functionals depend on HF exchange, the electron density and its gradient, and also on the kinetic energy density. GGE functionals consist of GGA exchange functionals and local correlation functionals and GGSC functionals consist of GGA exchange functionals and local correlation functionals with a gradient correction [47].

Unlike *ab initio* methods, one cannot know how to systematically improve the accuracy of a DFT result. This makes the generation of functionals a somewhat subjective matter and the inclusion of certain empiric parameters usually necessary. Empiric calibration is also a drawback by itself, since functionals become less prone to be used outside the scope of its parameterization. For these reasons it becomes very difficult to rate DFT functionals or to assess which one is better for a particular system. For readers who are interested in using DFT methods, the key question is probably: “Which functional should I use for my study?”. It is difficult to recommend one DFT method because their number is overwhelming and each method has its own strengths and weaknesses. Therefore it is useful to identify a small set of functionals that perform well on the properties and type of system under study, as well as the availability and computational cost associated. The best way to choose the most suitable functional for a given system under study is to be always aware of the new and most recent benchmarking studies. In most of these studies, a model with a relatively small number of atoms is chosen, allowing a great number of calculations and a comparison with results from very high accuracy methods, such as QCISD or CCSD, which are computationally very demanding. Properties of the system are then estimated with various functionals and reference methods, and their values are compared. Experimental values should also be taken into account, if available. The rating is usually based on the Mean-Unsigned-Error (MUE) between each functional and the reference method.

Several benchmarking studies have tried to evaluate the performance of density functionals in a variety of aspects, including geometry [51], [52], kinetics barriers and thermochemistry [53], binding energies and dissociation energies [53], non-bonded interactions (hydrogen bonding, weak interactions, π - π interactions) [54]-[56], etc.

In our research group we are constantly developing benchmarking studies to help us find the best performing functional for the study of enzymatic catalysis and inhibition on each particular system that we study. We can take the enzyme beta-glucosidase as an example. This enzyme catalyses the hydrolysis of the extremely stable glycosidic bond, as well as the transglycosilation of sugar substrates. It has an enormous biotechnological interest for the food industry, where it is used to catalyse the large scale production of oligosaccharides. We are undertaking an atomic level study

of the catalytic mechanism of oligosaccharide synthesis, using DFT and Molecular Mechanics as theoretical levels. The objective is to understand the atomic-level factors that determine the outcome of the reaction (hydrolysis/transglycosilation) and the yield of each of the many transglycosilation products. Before embarking in such a complex study, which deals with a system including a large portion of the enzyme and substrates (up to thousands of atoms), and whose calculations can take several months, we have made a benchmarking study to compare a set of the most used DFT functionals against higher-level post-Hartree Fock calculations [57]. The target properties were the activation energies and the reaction energies of the catalytic cycle. A very small enzyme model was used, containing only the reactive amino acids and a small substrate, to allow for calculations of extreme accuracy, to test a large number of density functionals, and to confine the differences in the density functionals to the reactive region of the enzyme. The results have shown that the obtained energies indeed depend on the particular density functional employed. In our case the correct choice of the functional was crucial as the most widely used have resulted in activations energies that differed within a range of over 8 kcal/mol. An uncertainty of 8 kcal/mol (just due to the choice of the functional) is clearly not acceptable. Comparison with the very high level calculations allowed for the identification of the most accurate functional within the group (BB1K) for this particular reaction. Based on these conclusions we could start the very time-consuming calculations using an enzyme model without the risk of having inaccurate results.

We believe that this kind of procedures should be standardized in electronic structure calculations.

V. ALANINE SCANNING MUTAGENESIS

The interactions of proteins are at the heart of almost every biological process. Their study is hence essential from a structural, energetic and therapeutic point of view [58]. Not only is important to unravel networks of protein-protein interactions in cells but also to identify and understand the residues directly involved in protein-protein binding: the hot spots [58]-[62]. Hot spots are usually defined as those residues that upon alanine mutation cause an increase in the binding free energy of more than 4.0 kcal/mol [63], [64]. Hot spot residues are highly correlated with structurally conserved residues, and are usually buried at the center of the interface surrounded by energetically less important residues that form a hydrophobic O-ring responsible for bulk solvent exclusion [58], [63]-[66]. In Figure 1 it is represented a complex between barnstar and barnase (PDBID: 1brs) in which the mutated residues are stressed by a vdW representation. Besides the hot spots, we have also highlighted the warm-spots that cause an increase in the binding energy between 2.0 and 4.0 kcal/mol, and the null-spots, which are the residues with binding free energy differences lower than 2.0 kcal/mol.

The experimental measure of the binding free energy is very costly, because it requires the production by mutagenesis of hundreds of variants that have to be analyzed by biophysically methods. Hence, a computational approach capable of predicting and understanding these residues would be particularly appealing [58]. A high number of different computational methods have been developed in recent years with different levels of accuracy and speed. All have specific advantages and limitations. In our research group we have developed an all-atom method that only requires a molecular dynamics simulation in implicit solvent (Generalized born solvation model) of the wild type structure [67]-[69]. A post-processing treatment of the complex allows not only the calculation of the wild-type complex energy, but as well the energies of the mutant complex and all the respective interacting monomers. Depending in which residue was mutated to an alanine; three different internal dielectric constant values were used: a value of 2 for the non-polar residues, a value of 3 for the polar residues and a value of 4 for the charged residues and histidine. This set of values take into account the different degree of relaxation of the interface upon alanine mutation. The stronger the interactions established by the amino-acid, the higher this value should be [67].

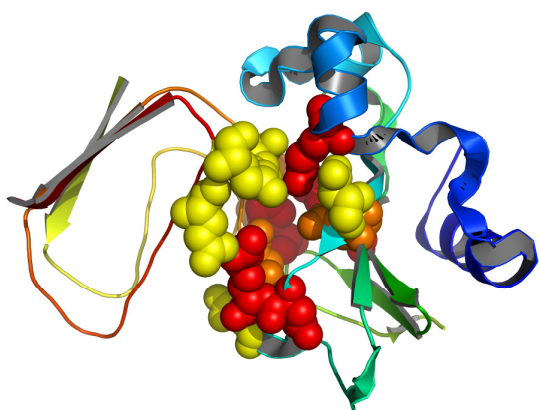


Figure 1: Complex formed between barnase and barnstar highlighting the mutated residues by a vdW representation. In yellow are represented the null-spots (relative binding energy < 2.0 kcal/mol), in orange the warm-spots (relative binding energy between 2.0 and 4.0 kcal/mol), and in red the hot-spots (residues with a relative binding energy higher than 4.0 kcal/mol)

This method was applied to different proteins with different sizes and presents an overall success rate (prediction of the binding affinity within an accuracy of 1 order of magnitude) of 80% and a 82% success rate for residues with free binding energy different higher than 2.0 kcal/mol. The mean and maximum unsigned errors obtained were of 0.80 and 4.52 kcal/mol, much lower than for other quantitative computational approaches. The performance of the method can also be calculated by using the F1 score, which is defined as a function of Precision (P, also called specificity) and Recall (R, also called sensitivity). F1 score, P and R can be defined as:

$$P = \frac{TP}{TP + FP} \quad R = \frac{TP}{TP + FN} \quad F1 = \frac{2PR}{P + R} \quad (1)$$

in which TP stands for true positive (predicted hot spot that it is an actual hot spot), FP stands for false positive (predicted hot spot that it is not an actual hot spot), and FN stands for false negative (non predicted hot spot that is an actual hot spot). For statistical purposes we have to lower the cut-off of the hot spot definition to residues that cause a binding free energy difference higher than 2.0 kcal/mol. In our database we have obtained a 0.53, 0.82 and 0.65 values for P, R, and F1 score, respectively [67], [69]. The F1 score is one of the best in the literature, with our method showing a remarkably higher sensitivity than others [70]-[72].

We developed a highly accurate quantitative atomistic method that it is still computationally affordable. Therefore, it can be applied to a large number of systems and mutations giving an accurate atomic-level description of biomolecules, a feature that is normally of great importance in elucidating their structures and functions.

VI. DRUG DESIGN

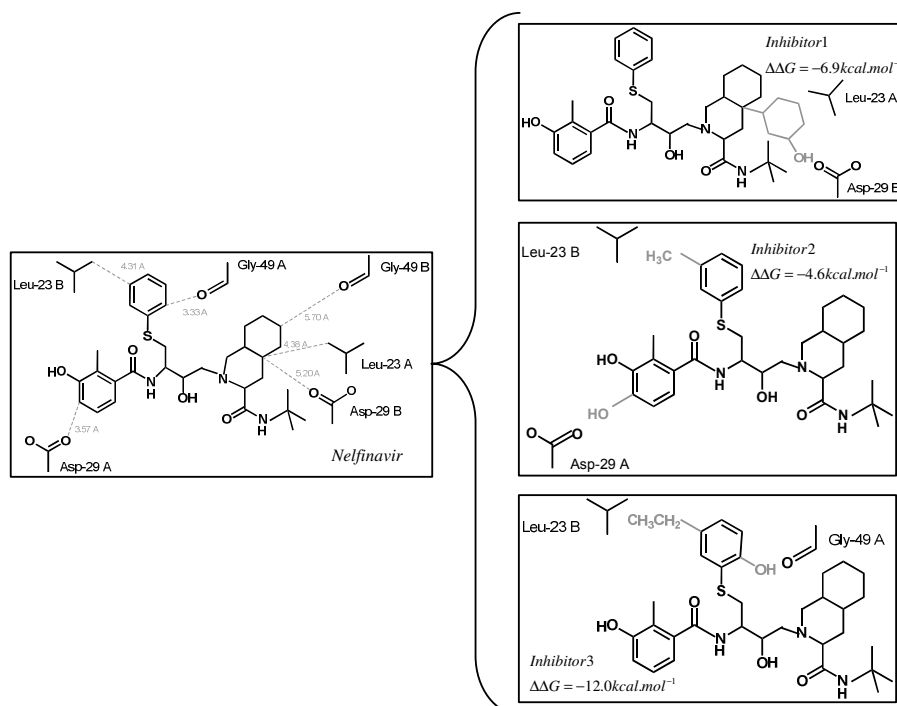
During the past few years we have modelled several new drugs by introducing changes in enzyme inhibitors previously approved for therapeutic use (lead inhibitors). These changes consisted in the rational addition/substitution of chemical groups to improve the interactions with conserved residues on the receptor, as to increase the inhibitory potency of the inhibitor. For this process, the three-dimensional structures of the complexes receptor:lead-inhibitor, available in the database *Protein Data Bank*, were used as the starting point. To study the behaviour of the complexes receptor:inhibitor, geometry optimizations and molecular dynamics simulations were made based on the standard molecular mechanics approximations [73]. To quantify the affinity of new inhibitors relatively to the lead inhibitor, we have calculated values of free energy of binding, using a complex computational method of high precision, called Thermodynamic Integration (TI) [73]-[76], and/or the more computationally accessible Molecular Mechanics Poisson-Boltzman Surface Area (MMPBSA) method [77]. TI is a rigorous method for calculating free energies, but requires extremely time consuming simulations, even with a large high-performance computer cluster. The faster and simpler, but still accurate, MMPBSA method is sometimes used as an alternative to TI.

We have been especially interested in developing drugs in areas of lethal diseases, such as Cancer and AIDS. Examples include the development of new inhibitors for HIV-1 Protease based on lead inhibitors (Nelfinavir, Amprenavir, etc) [78]. Nelfinavir (Viracept®, Pfizer) is a potent, orally bioavailable inhibitor of the enzyme HIV-1 protease, which has been developed through structure-based drug design projects, and that has been approved worldwide for the treatment of HIV

infected patients. However, the virus has shown the ability of fixing mutations which decrease the affinity of the Nelfinavir for the binding pocket of Protease, compromising the long term efficacy of the drug. More potent Protease drugs should interact more favourably with well-conserved residues, Leu23, Ala28, Gly49, Arg87, and Asp29 [79], which cannot mutate, as the mutants would render the enzyme catalytically inactive. We have excluded two of them, Ala28 and Arg87. Ala28 was excluded due to its hydrophobic nature and to the fact that it is close to two other hydrophobic residues of high variability. Arg87 was excluded because it is located at a distance of 8Å from the inhibitor in a position which can be perceived as a "second shell" of residues. Scheme 1 shows a schematic representation of the binding region for Nelfinavir with which the substitutions introduced in the inhibitors are meant to interact. The dashed lines give an idea of the location of the

empty pockets between the well-conserved amino acids and the inhibitors. Scheme 1 shows also three examples for which significant increases in affinity can still be achieved without changing the overall structure, molecular mass and hydrophobicity of the inhibitors, thus preserving their very favourable ADME properties. To evaluate the inhibitory efficiency, the values of the binding free energy are calculated and compared with the ones obtained for Nelfinavir using the MMPBSA. For the MMPBSA calculations, we have made minimizations and molecular dynamics simulations with the Amber software package [73]. The values for the binding free energy difference presented (scheme 1) are validated by the correct reproduction of the experimental binding free energy for Nelfinavir.

Scheme 1- Three new inhibitors, based on Nelfinavir as lead, which have a higher affinity *in silico* for Protease than Nelfinavir (the negatives binding free energy difference $\Delta\Delta G$ reflect this).



This example shows how powerful computer-aided drug design can be for rational lead optimisation, avoiding handling dangerous and toxic materials and greatly reducing experimental costs. There have been dramatic successes with drug design.

VII. FLEXIBLE RECEPTOR DOCKING

Several docking algorithms have been developed in the last two decades to predict protein-ligand binding poses. Initially,

molecular docking was used to predict and reproduce protein-ligand complexes. Successes of these early studies led to the exploration of molecular docking as a tool in drug discovery to identify hit/lead compounds, often by database screening. Since the development of combinatorial chemistry, molecular docking is also applied to aid in the design of combinatorial libraries and to pre-screen *real* or *virtual* compound databases *in silico*.

Any docking application includes two major algorithms: the search algorithm and the scoring or evaluation function.

The search algorithm explores the potential energy landscape of the complex in enough detail to find the global energy minimum. The scoring function has to be realistic enough to assign the most favorable scores to the solutions created by the search algorithm. Different simplifications are used to make molecular docking attractive for specific applications. To have reliable protein-ligand docking results, there are at least two fundamental prerequisites: a reliable scoring function that must be able to estimate ligand-protein interactions, and a proper treatment of ligand and protein flexibility. Moreover, it is also desirable that these algorithms should be fast, accurate and be able to reproduce experimental results.

From the computational point of view this is a complex task as a very large number of degrees of freedom must be evaluated, which requires huge computational resources. To reduce the dimension of the problem, most docking procedures treat the receptor as a rigid body, exploring only the translational, rotational, and torsional degrees of freedom of the ligand. This procedure has shown to be quite successful in the cases where the active site is rigid, and does not change meaningfully between the bound and the unbound forms [80]-[83]. However, there are many examples in the literature where the protein conformation changes between these two states and the application of "the rigid receptor-based" docking methods was shown to be less successful.

In spite of the computational advances observed in the past 10 years, dealing with receptor flexibility in docking methodologies is still a cumbersome issue [83]. The main reason behind this difficulty is the large number of degrees of freedom that have to be considered in this kind of calculation, which turns these approaches extremely time-consuming and therefore unattractive, in particular for virtual screening purposes. The brute force strategy of exploring systematically all the degrees of freedom is simply not feasible from a computational point of view. The most critical issues in the development of these algorithms are the scoring function used to evaluate the complex resulting from the docking process [84] and the capability of the search algorithm to predict the conformational movements of a protein that are relevant for the docking process, such as loop movements, domain movements, or simply side chain rotamer exchange. In spite of these difficulties, it is expected that in a near future, all docking algorithms will take the protein flexibility into account. The question that remains to answer is when, how it will be accomplished and what is the best way of doing it? With the exponential growth of the computational resources, we believe that it may be sooner than we expect, but the attractiveness of the algorithm (speed/accuracy) will dictate if these approaches will be successful or not in the near future.

In order to overcome all these difficulties and to propose a method capable of introducing a significant degree of flexibility into the protein during the docking procedure, our group has developed a new docking protocol called MADAMM (Multi stAged Docking with an Automated Molecular Modeling protocol) [21]. The algorithm is based on

the observation that in a wide set of proteins in bound and unbound forms, the movements of the backbone structure are almost negligible ($<2 \text{ \AA}$), and that in 85% of the ligand-protein complexes, only a small number of side chains (three or less) undergo conformational changes upon ligand binding [15]. Taking this into account, the MADAMM algorithm includes a side chain rotamer library routine that introduces a certain degree of flexibility in the active site region of the receptor where the ligand is supposed to bind. The complexes generated in this routine are then subjected to a set of minimization steps that improve the interactions between the receptor and the ligand using standard molecular mechanics force fields.

To demonstrate the potential and capabilities of the MADAMM protocol we show in the next figure our attempts to dock progesterone to the active site of monoclonal antiprogestosterone antibody DB3. The unbound (Figure 2, top left, dark gray) and bound structure (Figure 2, top left, light gray) of the receptor have a particular difference in the active site structure, which is related with the conformation rearrangement of Trp100. In the unbound form it is facing to the center of the active site, while in the bound form it is displaced to the right hand side. This small, but crucial conformation rearrangement is sufficient to generate a faulty result when we try to dock the ligand using standard docking software to the unbound structure (Figure 2, top right). This result is not dependent on the limitations of the search algorithm or scoring function of the docking algorithm, but simply because some volume of the active site was occupied in the unbound form and it is required for the correct binding of the ligand (as it can be observed analysing the conformation of Thr100 in the bound structure).

The MADAMM protocol allowed us to flexibilize Thr100 together with other residues from the active site during the docking procedure, by generating an ensemble of receptor structures in which the rotamers of the flexibilized residues are taken from an extensive rotamer library. Afterwards, each complex was subjected to a set of minimization steps and the final solutions re-scored using a scoring function based on the binding energy of the complex. The best score solution (Figure 2, bottom left) obtained with MADAMM is very similar to the one that is observed the unbound X-ray structure, showing with only a small RMSD of 0.13 \AA .

These results allow to conclude that the orientation of particular residues at the active site cavity has a crucial influence on the way receptor-ligand interact, and if neglected can be responsible for a large number of the false positive results which are obtained with the available rigid-based docking software.

Other successful examples using MADAMM protocol have been obtained in the last year [14], [15], [21]. These studies show that when protein flexibility plays a relevant role in the receptor:ligand interaction, standard docking procedures normally fail and generate faulty results. On the other hand, MADAMM has shown to be a powerful tool for investigating

ligand binding poses, allowing for the researcher to understand the importance of protein flexibility during the binding process. This methodology is therefore extremely useful to predict the binding pose of a ligand into a specific receptor where no experimental data is available.

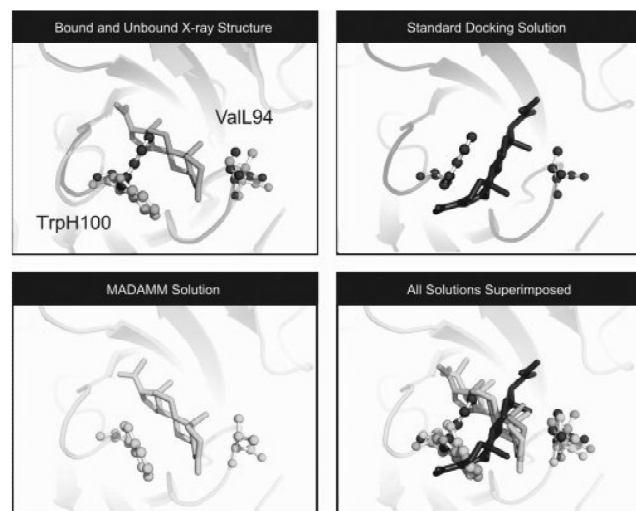


Figure 2: *Top left, dark gray:* X-ray structure of the unbound form of the protein. *Top left, light gray:* cocrystallized X-ray structure with the bound substrate. *Top right:* standard docking solution (GOLD). *Bottom left:* MADAMM result. *Bottom right:* all solutions superimposed.

ACKNOWLEDGMENT

The authors would like to thank the financial support of Fundação para a Ciência e a Tecnologia (Project PTDC/QUI/68302/2006). M. A. S. Perez thanks Fundação para a Ciência e a Tecnologia for a PhD grant (SFRH/BD/43600/2008).

REFERENCES

- [1] J. A. McCammon, B. R. Gelin, and M. Karplus, "Dynamics of folded proteins," *Nature*, vol. 267, pp. 585-590, 1977.
- [2] M. Karplus and J. A. McCammon, "Molecular dynamics simulations of biomolecules," *Nat. Struct. Biol.*, vol. 9, pp. 646-652, 2002.
- [3] A. Warshel, "Molecular dynamics simulations of biological reactions," *Acc. Chem. Res.*, vol. 35, pp. 385-395, 2002.
- [4] S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, et al., "A new force-field for molecular mechanical simulation of nucleic-acids and proteins," *J. Am. Chem. Soc.*, vol. 106, pp. 765-784, 1984.
- [5] S. J. Weiner, P. A. Kollman, D. T. Nguyen, and D. A. Case, "An all-atom force field for simulations of protein and nucleic acids," *J. Comput. Chem.*, vol. 7, pp. 230-252, 1986.
- [6] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, et al., "A 2nd generation force-field for the simulation of proteins, nucleic-acids, and organic-molecules," *J. Am. Chem. Soc.*, vol. 117, pp. 5179-5197, 1995.
- [7] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, "CHARMM - a program for macromolecular energy, minimization, and dynamics calculations," *J. Comput. Chem.*, vol. 4, pp. 187-217, 1983.
- [8] A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field et al., "All-atom empirical potential for molecular modeling and dynamics studies of proteins," *J. Phys. Chem. B*, vol. 102, pp. 3586-3616, 1998.
- [9] A. D. MacKerell, N. Banavali, and N. Foloppe, "Development and current status of the CHARMM force field for nucleic acids," *Biopolymers*, vol. 56, pp. 257-265, 2000.
- [10] W. L. Jorgensen and J. Tirado-Rives, "The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin," *J. Am. Chem. Soc.*, vol. 110, pp. 1657-1666, 1988.
- [11] J. Pranata, S. G. Wierschke, and W. L. Jorgensen, "Opls Potential Functions for Nucleotide Bases - Relative Association Constants of Hydrogen-Bonded Base-Pairs in Chloroform," *J. Am. Chem. Soc.*, vol. 113, pp. 2810-2819, 1991.
- [12] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, "Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids," *J. Am. Chem. Soc.*, vol. 118, pp. 11225-11236, 1996.
- [13] A. T. Marques, P. A. Fernandes, and M. J. Ramos, "Molecular dynamics simulations of the amyloid-beta binding alcohol dehydrogenase (ABAD) enzyme," *Bioorgan. Med. Chem.*, vol. 16, pp. 9511-9518, 2008.
- [14] N. F. Bras, N. M. F. S. Cerqueira, P. A. Fernandes, and M. J. Ramos, "Carbohydrate-binding modules from family 11: Understanding the binding mode of polysaccharides," *Int. J. Quantum Chem.*, vol. 108, pp. 2030-2040, 2008.
- [15] A. Viegas, N. F. Bras, N. M. F. S. Cerqueira, P. A. Fernandes, J. A. M. Prates, C. M. G. A. Fontes, et al., "Molecular determinants of ligand specificity in family 11 carbohydrate binding modules - an NMR, X-ray crystallography and computational chemistry approach," *FEBS J.*, vol. 275, pp. 2524-2535, 2008.
- [16] A. T. P. Carvalho, M. Swart, J. N. P. van Stralen, P. A. Fernandes, M. J. Ramos, and F. M. Bickelhaupt, "Mechanism of thioredoxin-catalyzed disulfide reduction. Activation of the buried thiol and role of the variable active-site residues," *J. Phys. Chem. B*, vol. 112, pp. 2511-2523, 2008.
- [17] A. T. P. Carvalho, P. A. Fernandes, and M. J. Ramos, "Insights on resistance to reverse transcriptase: The different patterns of interaction of the nucleoside reverse transcriptase inhibitors in the deoxyribonucleotide triphosphate binding site relative to the normal substrate," *J. Med. Chem.*, vol. 49, pp. 7675-7682, 2006.
- [18] S. F. Sousa, P. A. Fernandes, and M. J. Ramos, "Enzyme Flexibility and the Catalytic Mechanism of Farnesyltransferase: Targeting the Relation," *J. Phys. Chem. B*, vol. 112, pp. 8681-8691, 2008.
- [19] S. F. Sousa, P. A. Fernandes, and M. J. Ramos, "Molecular Dynamics Simulations on the Critical States of the Farnesyltransferase Enzyme," *Bioorg. Med. Chem.*, pp. (10.1016/j.bmc.2009.03.055), 2009.
- [20] S. F. Sousa, P. A. Fernandes, and M. J. Ramos, "Molecular dynamics analysis of farnesyltransferase: A closer look into the amino acid behavior," *Int. J. Quant. Chem.*, vol. 108, pp. 1939-1950, 2008.
- [21] N. M. F. S. Cerqueira, N. F. Bras, P. A. Fernandes, and M. J. Ramos, "MADAMM: A multistaged docking with an automated molecular modeling protocol," *Proteins*, vol. 74, pp. 192-206, 2009.
- [22] S. F. Sousa, P. A. Fernandes, and M. J. Ramos, "Unraveling the mechanism of the farnesyltransferase Enzyme," *J. Biol. Inorg. Chem.*, vol. 10, pp. 3-10, 2005.
- [23] S. F. Sousa, P. A. Fernandes, and M. J. Ramos, "Farnesyltransferase inhibitors: A detailed chemical view on an elusive biological problem," *Curr. Med. Chem.*, vol. 15, pp. 1478-1492, 2008.
- [24] S. F. Sousa, P. A. Fernandes, and M. J. Ramos, "The Search for the Mechanism of the Reaction Catalyzed by Farnesyltransferase," *Chemistry*, 2009.
- [25] S. F. Sousa, P. A. Fernandes, and M. J. Ramos, "Farnesyltransferase--New Insights into the Zinc-Coordination Sphere Paradigm: Evidence for a Carboxylate-Shift Mechanism," *Biophys. J.*, vol. 88, pp. 483-494, 2005.

- [26] S. F. Sousa, P. A. Fernandes, and M. J. Ramos, "Farnesyltransferase: Theoretical Studies on Peptide Substrate Entrance - Thiol or Thiolate Coordination?," *J. Mol. Struct. (Theochem)*, vol. 729, pp. 125-129, 2005.
- [27] S. F. Sousa, P. A. Fernandes, and M. J. Ramos, "The carboxylate-shift in zinc enzymes: a computational study," *J. Am. Chem. Soc.*, vol. 129, pp. 1378-1385, 2007.
- [28] S. F. Sousa, P. A. Fernandes, and M. J. Ramos, "Theoretical Studies on Farnesyltransferase: The Distances Paradox Explained," *Proteins*, vol. 66, pp. 205-218, 2007.
- [29] M. J. Ramos and P. A. Fernandes, "Computational enzymatic catalysis," *Acc. Chem. Res.*, vol. 41, pp. 689-698, 2008.
- [30] S. B. Long, P. J. Casey, and L. S. Beese, "Reaction path of protein farnesyltransferase at atomic resolution," *Nature*, vol. 419, pp. 645-650, 2002.
- [31] H. W. Park, S. R. Boduluri, J. F. Moomaw, P. J. Casey, and L. S. Beese, "Crystal Structure of Protein Farnesyltransferase at 2.25 Angstrom Resolution," *Science*, vol. 275, pp. 1800-1804, 1997.
- [32] D. A. Tobin, J. S. Pickett, H. L. Hartman, C. A. Fierke, and J. E. Penner-Hahn, "Structural characterization of the zinc site in protein farnesyltransferase," *J. Am. Chem. Soc.*, vol. 125, pp. 9962-9969, 2003.
- [33] S. F. Sousa, P. A. Fernandes, and M. J. Ramos, "Theoretical Studies on Farnesyltransferase: Evidence for Thioether Product Coordination to the Active-Site Zinc Sphere," *J. Comput. Chem.*, vol. 28, pp. 1160-1168, 2007.
- [34] N. M. F. S. A. Cerqueira, S. Pereira, P. A. Fernandes, and M. J. Ramos, "Overview of ribonucleotide reductase inhibitors: an appealing target in anti-tumour therapy," *Curr. Med. Chem.*, vol. 12, pp. 1283-1294, 2005.
- [35] N. M. F. S. A. Cerqueira, P. A. Fernandes, L. A. Eriksson, and M. J. Ramos, "New insights into a critical biological control step of the mechanism of ribonucleotide reductase," *J. Mol. Struct. (Theochem)*, vol. 709, pp. 53-65, 2004.
- [36] N. M. F. S. A. Cerqueira, P. A. Fernandes, L. A. Eriksson, and M. J. Ramos, "Ribonucleotide activation by enzyme ribonucleotide reductase: understanding the role of the enzyme," *J. Comput. Chem.*, vol. 25, pp. 2031-2037, 2004.
- [37] N. M. F. S. A. Cerqueira, P. A. Fernandes, L. A. Eriksson, and M. J. Ramos, "Dehydration of ribonucleotides catalysed by ribonucleotide reductase: the role of the enzyme," *Biophys. J.*, vol. 90, pp. 2109-2119, 2006.
- [38] P. A. Fernandes, L. A. Eriksson, and M. J. Ramos, "The reduction of ribonucleotides catalyzed by the enzyme ribonucleotide reductase," *Theor. Chem. Acc.*, vol. 108, pp. 352-364, 2002.
- [39] S. Pereira, N. M. F. S. A. Cerqueira, P. A. Fernandes, and M. J. Ramos, "Computational studies on class I ribonucleotide reductase: Understanding the mechanism of action and inhibition of a cornerstone enzyme for the treatment of cancer," *Eur. Biophys. J.*, vol. 35, pp. 125-135, 2006.
- [40] P. A. Fernandes and M. J. Ramos, "Theoretical studies on the mode of inhibition of ribonucleotide reductase by 2'-substituted substrate analogues," *Chem. Eur. J.*, vol. 9, pp. 5916-5925, 2003.
- [41] N. M. F. S. Cerqueira, P. A. Fernandes, and M. J. Ramos, "Enzyme ribonucleotide reductase: Unraveling an enigmatic paradigm of enzyme inhibition by furanone derivatives," *J. Phys. Chem. B*, vol. 110, pp. 21272-21281, 2006.
- [42] S. Pereira, P. A. Fernandes, and M. J. Ramos, "Theoretical study of ribonucleotide reductase mechanism-based inhibition by 2'-azido-2'-deoxyribonucleoside 5'-diphosphates," *J. Comput. Chem.*, vol. 25, pp. 227-237, 2004.
- [43] S. Pereira, P. A. Fernandes, and M. J. Ramos, "Mechanism for ribonucleotide reductase inactivation by the anticancer drug gemcitabine," *J. Comput. Chem.*, vol. 25, pp. 1286-1294, 2004.
- [44] S. Pereira, P. A. Fernandes, and M. J. Ramos, "Theoretical study on the inhibition of ribonucleotide reductase by 2'-mercapto-2'-deoxyribonucleoside-5'," *J. Am. Chem. Soc.*, vol. 127, pp. 5174-5179, 2005.
- [45] P. A. Fernandes and M. J. Ramos, "Theoretical studies on the mechanism of inhibition of Ribonucleotide Reductase by (E)-2'-Fluoromethylene-2'-deoxycytidine-5'-diphosphate," *J. Am. Chem. Soc.*, vol. 125, pp. 6311-6322, 2003.
- [46] A. Ghosh, "Just how good is DFT?," *J. Biol. Inorg. Chem.*, vol. 11, pp. 671-673, 2006.
- [47] N. E. Schultz, Y. Zhao, and D. G. Truhlar, "Density functionals for inorganometallic and organometallic chemistry," *J. Phys. Chem. A*, vol. 109, pp. 11127-11143, 2005.
- [48] P. Hohenberg and W. Kohn, "Inhomogeneous Electron Gas," *Phys. Rev. B*, vol. 136, pp. B864, 1964.
- [49] Y. Zhao and D. G. Truhlar, "Density functionals with broad applicability in chemistry," *Acc. Chem. Res.*, vol. 41, pp. 157-167, 2008.
- [50] S. F. Sousa, P. A. Fernandes, and M. J. Ramos, "General Performance of Density Functionals," *J. Phys. Chem A*, vol. 111, pp. 10439-10452, 2007.
- [51] S. F. Sousa, P. A. Fernandes, and M. J. Ramos, "Comparative assessment of theoretical methods for the determination of geometrical properties in biological zinc complexes," *J. Phys. Chem. B*, vol. 111, pp. 9146-9152, 2007.
- [52] P. Jurecka, J. Cerny, P. Hobza, and D. R. Salahub, "Density functional theory augmented with an empirical dispersion term. Interaction energies and geometries of 80 noncovalent complexes compared with ab initio quantum mechanics calculations," *J. Comput. Chem.*, vol. 28, pp. 555-569, 2007.
- [53] A. Karton, A. Tamopolsky, J. F. Lamere, G. C. Schatz, and J. M. L. Martin, "Highly Accurate First-Principles Benchmark Data Sets for the Parametrization and Validation of Density Functional and Other Approximate Methods. Derivation of a Robust, Generally Applicable, Double-Hybrid Functional for Thermochemistry and Thermochemical Kinetics," *J. Phys. Chem. A*, vol. 112, pp. 12868-12886, 2008.
- [54] Y. Zhao and D. G. Truhlar, "Density functionals for noncovalent interaction energies of biological importance," *J. Chem. Theor. Comput.*, vol. 3, pp. 289-300, 2007.
- [55] T. van Mourik, "Assessment of Density Functionals for Intramolecular Dispersion-Rich Interactions," *J. Chem. Theor. Comput.*, vol. 4, pp. 1610-1619, 2008.
- [56] R. Peverati and K. K. Baldrige, "Implementation and Performance of DFT-D with Respect to Basis Set and Functional for Study of Dispersion Interactions in Nanoscale Aromatic Hydrocarbons," *J. Chem. Theor. Comput.*, vol. 4, pp. 2030-2048, 2008.
- [57] N. F. Bras, S. A. Moura-Tamames, P. A. Fernandes, and M. J. Ramos, "Mechanistic Studies on the Formation of Glycosidase-Substrate and Glycosidase-Inhibitor Covalent Intermediates," *J. Comput. Chem.*, vol. 29, pp. 2565-2574, 2008.
- [58] I. S. Moreira, P. A. Fernandes, and M. J. Ramos, "Hot spots-A review of the protein-protein interface determinant amino-acid residues," *Proteins*, vol. 68, pp. 803-812, 2007.
- [59] S. Jones and J. M. Thornton, "Principles of protein-protein interactions," *Proc. Natl. Acad. Sci. USA*, vol. 93, pp. 13-20, 1996.
- [60] A. A. Bogan and K. S. Thorn, "Anatomy of hot spots in protein interfaces," *J. Mol. Biol.*, vol. 280, pp. 1-9, 1998.
- [61] T. Clackson and J. A. Wells, "A Hot-Spot of Binding-Energy in A Hormone-Receptor Interface," *Science*, vol. 267, pp. 383-386, 1995.
- [62] W. L. Delano, "Unraveling hot spots in binding interfaces: progress and challenges," *Curr. Opin. Struct. Biol.*, vol. 12, pp. 14-20, 2002.
- [63] I. S. Moreira, P. A. Fernandes, and M. J. Ramos, "Detailed microscopic study of the full ZipA : FtsZ interface," *Proteins*, vol. 63, pp. 811-821, 2006.
- [64] I. S. Moreira, P. A. Fernandes, and M. J. Ramos, "Unraveling the importance of protein-protein interaction: Application of a computational alanine-scanning mutagenesis to the study of the IgG1 streptococcal protein G (C2 fragment) complex," *J. Phys. Chem. B*, vol. 110, pp. 10962-10969, 2006.
- [65] I. S. Moreira, P. A. Fernandes, and M. J. Ramos, "Hot spot computational identification: Application to the complex formed between the hen egg white lysozyme (HEL) and the antibody HyHEL-10," *Int. J. Quant. Chem.*, vol. 107, pp. 299-310, 2007.
- [66] I. S. Moreira, P. A. Fernandes, and M. J. Ramos, "Hot spot occlusion from bulk water: A comprehensive study of the complex between the lysozyme HEL and the antibody FVD1.3," *J. Phys. Chem. B*, vol. 111, pp. 2697-2706, 2007.
- [67] I. S. Moreira, P. A. Fernandes, and M. J. Ramos, "Unravelling Hot Spots: a comprehensive computational mutagenesis study," *Theor. Chem. Acc.*, vol. 117, pp. 99-113, 2007.

- [68] I. S. Moreira, P. A. Fernandes, and M. J. Ramos, "Accuracy of the numerical solution of the Poisson-Boltzmann equation," *J. Mol. Struct. (Theochem)*, vol. 729, pp. 11-18, 2005.
- [69] I. S. Moreira, P. A. Fernandes, and M. J. Ramos, "Computational alanine scanning mutagenesis - An improved methodological approach," *J. Comput. Chem.*, vol. 28, pp. 644-654, 2007.
- [70] T. Kortemme and D. Baker, "A simple physical model for binding energy hot spots in protein-protein complexes," *Proc. Natl. Acad. Sci. USA*, vol. 99, pp. 14116-14121, 2002.
- [71] E. Guney, N. Tuncbag, O. Keskin, and A. Gursoy, "HotSprint: database of computational hot spots in protein interfaces," *Nucleic Acids Res.*, vol. 36, pp. D662-D666, 2008.
- [72] S. J. Darnell, D. Page, and J. C. Mitchell, "An automated decision-tree approach to predicting protein interaction hot spots," *Proteins*, vol. 68, pp. 813-823, 2007.
- [73] D. A. Case, T. A. Darden, T. E. Cheatham III, C. L. Simmerling, J. Wang, R. E. Duke, et al., "AMBER 8," University of California, San Francisco, 2004.
- [74] M. R. Shirts, J. W. Pitera, W. C. Swope, and V. S. Pande, "Extremely precise free energy calculations of amino acid side chain analogs: Comparison of common molecular mechanics force fields for proteins," *J. Chem. Phys.*, vol. 119, pp. 5740-5761, 2003.
- [75] J. W. Pitera and W. F. van Gunsteren, "A comparison of non-bonded scaling approaches for free energy calculations," *Mol. Simul.*, vol. 28, pp. 45-65, 2002.
- [76] A. Blondel, "Ensemble variance in free energy calculations by thermodynamic integration: Theory, optimal "Alchemical" path, and practical solutions," *J. Comput. Chem.*, vol. 25, pp. 985-993, 2004.
- [77] P. A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. H. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D. A. Case, and T. E. Cheatham, "Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models," *Acc. Chem. Res.*, vol. 33, pp. 889-897, 2000.
- [78] M. A. S. Perez, P. A. Fernandes, and M. J. Ramos, "Drug design: New inhibitors for HIV-1 protease based on Nelfinavir as lead," *J. Mol. Graph. Model.*, vol. 26, pp. 634-642, 2007.
- [79] W. Wang and P. A. Kollman, "Computational study of protein specificity: The molecular basis of HIV-1 protease drug resistance," *Proc. Natl. Acad. Sci. USA*, vol. 98, pp. 14937-14942, 2001.
- [80] S. Gupta, L. M. Rodrigues, A. P. Esteves, A. M. F. Oliveira-Campos, M. S. J. Nascimento, N. Nazareth, H. Cidade, M. P. Neves, F. Fernandes, M. Pinto, N. M. F. S. A. Cerqueira, and N. Bras, "Synthesis of N-aryl-5-amino-4-cyanopyrazole derivatives as potent xanthine oxidase inhibitors," *Eur. J. Med. Chem.*, vol. 43, pp. 771-780, 2007.
- [81] M. L. Lamb and W. L. Jorgensen, "Computational approaches to molecular recognition," *Curr. Opin. Chem. Biol.*, vol. 1, pp. 449-457, 1997.
- [82] C. N. Cavasotto and N. Singh, "Docking and high throughput docking: Successes and the challenge of protein flexibility," *Curr. Comput. Aided Drug Des.*, vol. 4, pp. 221-234, 2008.
- [83] S. F. Sousa, P. A. Fernandes, and M. J. Ramos, "Protein-ligand docking: Current status and future challenges," *Proteins*, vol. 65, pp. 15-26, 2006.
- [84] N. Brooijmans and I. D. Kuntz, "Molecular recognition and docking algorithms," *Ann. Rev. Biophys. Biomol. Struct.*, vol. 32, pp. 335-373, 2003.