



HAL
open science

Molecular Recognition and Drug-Lead Identification: What Can Molecular Simulations Tell Us?

Morra Giulia, Alessandro Genoni, Neves Marco A. C., Jr. Kenneth M. Merz,
Colombo Giorgio

► **To cite this version:**

Morra Giulia, Alessandro Genoni, Neves Marco A. C., Jr. Kenneth M. Merz, Colombo Giorgio. Molecular Recognition and Drug-Lead Identification: What Can Molecular Simulations Tell Us?. Current Medicinal Chemistry, 2010, 17 (1), pp.25-41. 10.2174/092986710789957797. hal-02196509

HAL Id: hal-02196509

<https://hal.univ-lorraine.fr/hal-02196509>

Submitted on 27 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Molecular Recognition and Drug-Lead Identification:

What Can Molecular Simulations Tell Us?

Giulia Morra,¹ Alessandro Genoni,^{1,2} Marco A.C. Neves^{1,3} Kenneth M. Merz, Jr.² and Giorgio Colombo^{1,*}

1) Istituto di Chimica del Riconoscimento Molecolare, CNR

Via Mario Bianco 9, 20131 Milano (Italy)

*Corresponding Author: giorgio.colombo@icrm.cnr.it

2) Department of Chemistry and Quantum Theory Project, University of Florida, P.O. Box 118435, Gainesville, FL 32611-8435.

3) Faculdade de Farmacia, Laboratorio de Quimica Farmaceutica, Centro de Estudos Farmaceuticos, Universidade de Coimbra, Coimbra, Portugal.

Abstract

Molecular recognition and ligand binding involving proteins underlie the most important life processes within the cell, such as substrate transport, catalysis, signal transmission, receptor trafficking, gene regulation, switching on and off of biochemical pathways.

Despite recent successes in predicting the structures of many protein-substrate complexes, the dynamic aspects of binding have been largely neglected by computational/theoretical investigations. Recently, several groups have started tackling these problems with the use of experimental and simulation methods and developed models describing the variation of protein dynamics upon complex formation, shedding light on how substrate or inhibitor binding can alter protein flexibility and function. The study of ligand-induced dynamic variations has also been exploited to review the concept of allosteric changes, in the absence of major conformational changes.

In this context, the study of the influence of protein motions on signal transduction and on catalytic activities has been used to develop pharmacophore models based on ensembles of protein conformations. These models, taking flexibility explicitly into account, are able to distinguish active inhibitors versus nonactive drug-like compounds, to define new molecular motifs and to preferentially identify specific ligands for a certain protein target.

The application of these methods holds great promise in advancing structure-based drug discovery and medicinal chemistry in general, opening up the possibility to explore broader chemical spaces than is normally done in an efficient way.

In this review, examples illustrating the extent to which simulations can be used to understand these phenomena will be presented along with examples of methodological developments to increase physical understanding of the processes and improve the possibility to rationally design new molecules.

Keywords: Molecular Dynamics, Molecular Recognition, Flexibility, Drug Design, Allostery.

Introduction.

Proteins are the main actuators of biological processes in cells and their malfunctioning is associated with the onset of disease conditions. Proteins are not static entities; rather their recognition and functional properties are strictly correlated to flexibility and dynamics.

Flexibility and dynamics are in fact fundamental in understanding protein mechanisms. Hemoglobin represents a paradigmatic example of these aspects: since the early days of protein crystallography and biochemical studies, it has been clear that the protein can be stabilized in two considerably different structures depending on its degree of oxygenation ^[1, 2].

Flexibility and conformational dynamics can also be exploited by enzymes to optimally arrange catalytic sites for reactivity. Dihydrofolate reductase (DHFR), for instance, catalyzes the reduction of dihydrofolate to tetrahydrofolate with the concurrent oxidation of the NADPH cofactor. In order to carry out the reaction, the catalytic residues of the enzyme, the substrate and the cofactor must be optimally arranged in space ^[3, 4]. DHFR is characterized by the presence of a flexible loop, the M20 loop, that can access three different conformations: closed, open and occluded ^[5]. When both the substrate and cofactor are bound, the closed conformation is favored over the others and the M20 loop is packed against the nicotinamide ring of NADPH. The closed form is the only conformation of the enzyme in which the substrate and the cofactor are optimally aligned for the reaction.

Flexibility and dynamics also play a primary role in protein-protein interactions. Protein-protein complex formation often involves structural changes that may extend well beyond local scale rearrangements, and that cannot be understood in terms of the lock-and-key model ^[6], or described by simple models based only on surface accessibility. In forming a complex with a given partner, proteins must adopt specific structures. These may be pre-existing in the accessible conformational space of the protein and the conformational equilibrium may be shifted towards them upon binding. Otherwise

specific structures are induced after the formation of a complex with the binding partner. In this context, it is important to underline that proteins very seldom function in isolation. Rather, they typically work within an ensemble, so that they must be sufficiently flexible to interact with multiple partners and carry out diverse tasks^[7].

Finally, the ability to visit alternate, closely related free-energy minima on the energy landscape and to trap specific states through ligand binding at sites distal from the active site, known as allosteric regulation, represents one of the most common and powerful means of regulating protein function (allosteric regulation) in the context of cellular metabolism. Allostery traditionally indicates the coupling of conformational changes between two different sites of a protein, often with an inhibitory effect.^[8] The Monod–Wyman–Changeux (MWC)^[9] and Koshland–Nemethy–Filmer (KNF)^[10] models for allostery were both proposed in the 1960s to explain the regulation upon ligand binding seen in hemoglobin and in several metabolic regulatory enzymes. Although originally introduced for oligomeric proteins, allostery does not require the protein to be multimeric and the modulation of protein function by means of interactions located at long distance from the active site is now recognized as a general feature of many monomeric proteins, as pointed out in^[11]. For instance, myoglobin can be allosterically regulated^[11].

Allosteric regulation does not either necessarily imply major conformational changes. Recent thermodynamic and conformational analyses have shown that allosteric communications can be transmitted solely by a variation in protein motions, at the backbone level, which do not determine a macroscopic change of shape in the protein.

From the beginning it was suggested that allostery may act as a general form of protein regulation, as it constitutes a direct and efficient mechanism for the modulation of cellular function in response to changes in concentration of small molecules. This led to the suggestion that allosteric sites might be even more useful as drug targets than active sites (See^[12] and references therein). The main advantage of exploiting allosteric effects in drug design is the specificity of the targets. While active sites are

overall very conserved within whole protein families, such that it is difficult to target selected members of the family, allosteric sites are not evolutionarily constrained and may be much more specific to single proteins ^[13]. The identification and targeting of allosteric sites have become the focus of both basic research and several drug design efforts. From the basic point of view, the study of allosteric sites and signal transduction mechanisms will provide precious insights into how nature has evolved control strategies in protein architectures. From the pharmaceutical point of view, these sites offer unique opportunities to discover new chemical entities for enzymes and other targets for which the discovery of active-site inhibitors has proved challenging.

While the presence of relevant ordered motions has been long recognized as a fundamental characteristic of functional proteins, the definition of their relevance to drug discovery/design has been more difficult.

From the examples and considerations illustrated above, it appears clear that the study of the influence of protein motions on aspects ranging from function to molecular recognition and binding can be better understood at the level of ensembles of conformations, rather than based on single static representations. Computational and theoretical approaches offer a unique means to investigate these aspects at different levels of resolution, and to include them into the processes of rational design and discovery of new drug-like molecules. Including target dynamics appears to hold a great deal of promise in advancing structure based molecular design and medicinal chemistry in general, allowing to optimize molecules targeting active sites and to determine the presence of allosteric sites for which inhibitors and/or modulators can be developed. The application of these methods will broaden the chemical space of available active molecules providing access to new chemotypes binding known receptors and possible new proteins.

In this review, we will present examples illustrating the extent to which simulations can be used to understand dynamic and allosteric phenomena in proteins along with examples of how to include flexibility in the docking and rational discovery/design of (new) molecules. Finally we will present examples coming from the study of Cytochrome type enzymes illustrating the fundamental importance of protein motions in the recognition and positioning of ligands.

Theory

Protein motions in the conformational space, such as local folding/unfolding rearrangements, ligand binding and also allosteric changes, have a cooperative character. The protein behaves like a cohesive unit, where any change results from the concerted contribution of all parts, and is not only the sum of their properties. In the case of allostery, cooperativity explains why local phenomena such as phosphorylation, ATP binding or hydrolysis can switch-on large-scale conformational transitions.

Any conformational change involves a free energy variation, resulting from a net combination of enthalpy and entropy changes. For instance, a favorable enthalpy change upon ligand binding is due to the formation of tighter attractive interactions, which in turn may reduce the protein entropy, due to a loss of flexibility of the molecule, or increase the solvent entropy by means of the hydrophobic effect. The balance between all these contributions yields the net free energy change that makes the binding favorable or unfavorable. In the case of allostery, at least two such binding events occur, showing cooperativity. This cooperativity can be positive, when the second binding event becomes more favorable in the presence of the first ligand (allosteric activator), or negative, when the second binding event is more costly due to the previous binding of the first ligand (allosteric inhibitor).

The current view of protein allostery is connected to the energy landscape model of protein folding, defining the native state of a protein as a conformational ensemble. Since proteins are not rigid, the native state exists as statistical ensemble with a number of local minima, separated by locally unfolded

regions given by transiently populated higher free energy states. Large scale conformational changes, occurring on timescales of micro- to milliseconds, involve energy barriers of several kT and correspond to the migration from one basin to another. Local motions on timescales of pico- to nanoseconds, including single residues fluctuations and loop rearrangements, are represented by motions on the rugged surface of the native basin ^[14-17]. The rough profile at the bottom of the protein folding funnel defines the range of the allosterically accessible folded substates. Binding the allosteric effector induces a shift in the population of states, by stabilizing one conformational state at the expenses of others, thus reducing the heterogeneity of the native ensemble. The stabilized conformation will be in turn more or less prone to binding the second ligand ^[18].

This view is related to the conformational selection model ^[19], which re-interprets the relationship between ligand binding and conformational change in an enzyme (traditionally referred to as the ‘induced fit’ mechanism) ^[20]. The conformational selection model is based on the assumption that structural changes observed in the presence of different substrates are already accessible in the absence of the ligand (see ^[21] and references therein) ^[22]. Although ligand-competent structures might be poorly populated and rather unstable in the absence of the ligand, as shown for the maltose-binding protein in ^[23], evidence of a pre-existing equilibrium between conformational states was provided by several experiments and also by computational studies ^[24]. An early indication of ‘pre-sampling’ of the conformational change came from molecular dynamics studies on myoglobin, which showed that the largest fluctuations are located in the same protein regions that experience the largest conformational change between the limiting static X-ray structures ^[25]. More importantly, not only structural changes, but also motions involved in biophysical (e.g. substrate recognition and binding), biochemical (e.g. catalysis) or biological (e.g. signaling, transcription) activities are observed to be pre-sampled before protein activity. This broadens the view of conformational changes upon binding, and also of allosteric changes, to include different dynamical states. The pre-existence of motions characteristic of catalytic

function was detected by the Kern group in the case of cyclophilin ^[16].

Computational studies based on molecular dynamics simulations on the N-terminal domain of human Heat Shock Protein 90 (Hsp90) pointed out, in agreement with the conformational selection model, that different functional motions, as detected by essential dynamics analysis, and elicited by different ligands, are basically all present in the free state of the molecule ^[26].

The view of allostery in terms of conformational and dynamic selection offers a generalization in the same spirit of the MWC model to a wider set of conformations and dynamical states than two single discrete structures, like the T and R conformations of hemoglobin.

The role of entropy and dynamics in allosteric communication between distant binding sites has been recently reviewed by Kern ^[27] and by Nussinov ^[7, 28]. Given the thermodynamic nature of allostery, communication across the protein can be mediated not only by enthalpy, by means of protein backbone rearrangements, but also by entropy, via changes in the dynamic fluctuations around the mean structure.

An entropic effect in positive cooperativity might generally be present, as recognized by Cooper and Dryden in the 80s ^[29]. They formalized an entropic component to the allosteric interaction free energy, which can be estimated by few Kcal/moles as a function of tiny rms changes of atoms and is due to the propagation of protein rigidity ^[29]. Namely, binding increases packing and locally rigidifies the protein structure. In an allosteric protein containing two or more binding sites, each site may cause partial rigidification upon individual binding. The global stiffening rigidification of the protein due to the binding of both ligands might partly be accomplished by either of the ligand binding steps, inducing positive cooperativity for the binding of the second ligand.

Another important role for driving the allosteric transition may be played by solvent that interacts with the exposed surface area of proteins and may modulate the rate of conformational changes upon ligand binding. This was recognized in the case of haemoglobin, for which hydration changes accompanying

structural differences between allosteric states were detected ^[30].

In the next sections, we will build on the concepts of protein flexibility to discuss their possible applications in drug discovery/design.

Including Dynamics in Docking and Drug Design.

Nowadays, the most sophisticated and rational procedures to discover novel drugs consist in computer-based approaches. All these strategies can be generally classified into two main categories: *de novo* drug-design methods and database virtual screening techniques. The former look for new possible ligands of therapeutic relevance that fit in a complementary way the electrostatic and hydrophobic properties of the target receptor, whereas the latter apply one or more filters to decide which compounds collected in given databases can be considered potential drug leads.

In spite of the subdivision into the two main groups described above, most of the computational drug-design strategies use a rigid structure for the target protein, although the importance of the receptor flexibility has been very well known for a long time. In fact, proteins are flexible systems that can undergo quite large conformational changes and it is worthwhile to observe that these rearrangements are often strongly associated with fundamental protein functions. Accordingly, proteins can be subdivided into three categories: (1) “rigid” proteins, which show only very small side chain movements upon ligand binding; (2) flexible proteins, whose ligand-induced changes are relatively large rearrangements in the neighborhood of hinge points or at the active site in conjunction with side chain motions; (3) unstable proteins, which are characterized by a dynamic equilibrium between several conformations often structurally similar and at the same time energetically different ^[31]. This classification is strongly connected with the three different models that have been proposed in order to explain the ligand-protein interactions. The simplest one is the well-known “Lock and Key” model

introduced by Fisher ^[6]. It describes the chemical complementarity between a receptor and the corresponding ligand as the complementarity between a lock and a key and, of course, it works when the protein is strictly rigid. Unfortunately, when the receptor flexibility plays a fundamental role in the molecular recognition process, it is necessary to consider the more advanced “Induced Fit” description proposed by Koshland ^[32] according to which both the protein and ligand structures change during the binding process as a result of the introduction of the substrate in the chemical and structural environment of the receptor. Finally, for really unstable proteins, we have to take into account a more modern model ^[33] that describes the molecular recognition as a process in which the substrate selects the most suited receptor conformation among an ensemble of metastable states, shifting the dynamic population equilibrium towards the configuration adopted by the bound protein.

These models clearly point out that both experimentally and theoretically derived rigid structures are not suitable when dealing with very flexible receptors, whose conformational changes are driven by ligand binding and are crucial to express protein functions. The reason why most up to the recent years, drug-discovery strategies have not taken into account this aspect could be ascribed to the pragmatism that led us to consider the crystal structure as the real structure and to the conceptual and technical problems of working with non-defined targets. In this respect, although both experimentalist and theoreticians have recognized the role of flexibility for a long time, limitations in computer power and inefficiencies in simulations algorithm have hampered the wide applications of these concepts. Fortunately, for many targets, considerable progress could be made by keeping proteins rigid, so that pragmatism actually paid off. The need to devise new methods able to take into account and evaluate dynamic rearrangements in molecular recognition processes is however ever more compelling and, in this paragraph, we will show some of the current efforts made by the drug-design community to accomplish this task. It is important to note that, among them, there are also noteworthy techniques that efficiently couple Molecular Dynamics (MD) or Monte Carlo (MC) simulations with docking

experiments. We believe that computer simulations strategies are probably one of the best tools to get a more complete set of receptor conformations that can be afterwards used in docking or virtual screening methods, especially if we also want to consider high-energy configurations that cannot be detected by current experimental techniques, such as X-Ray crystallography and Nuclear Magnetic Resonance. Nevertheless, it is extremely worth noting that, due to some non-negligible drawbacks such as the stochastic nature of Molecular Dynamics and the rather limited length of simulations, we often sample only a small subset of the protein conformational space.

According to the classification introduced by Teodoro and Kavraki ^[34, 35], the strategies that consider receptor flexibility in modeling of protein-ligand interactions can be subdivided into five categories, which will be discussed with more details in this paper: (1) soft potential methods, (2) techniques that take into account a reduced number of degrees of freedom, (3) strategies which exploit multiple receptor structures, (4) modified computer simulations (MD or MC) methods and (5) strategies that use collective degrees of freedom in order to represent very large protein conformational rearrangements. In general, the accuracy provided by the different approaches is directly proportional to the associated computational complexity and, for the sake of completeness, it is also important to stress that the proposed subdivision is not so strict because many of the strategies that we will describe may belong to more than one category.

Soft potential methods are probably the simplest techniques that take into account a part of receptor flexibility and they essentially consist in a reduction of the Van der Waals contributions to the total energy score. This allows relaxing the high-energy penalty associated with clashes between ligand and receptor atoms and this is the reason why larger ligands are able to fit in binding sites suited for smaller molecules. It is worth to note that the “Induced Fit” model underlies this kind of approaches because the receptor is allowed to undergo some changes, both in backbone and side-chains, in order to respond to the presence of the new ligand. These strategies are very useful for their easy implementation and

computational speed, but great attention has to be paid in the definition of the “soft region” because, if it is too large (and this is usually the case when we want to account for large protein rearrangements), errors can arise. Among this group of methods we believe particularly interesting the approach developed by Apostolakis and coworkers^[36] At the initial stage, they apply the softening of the Van der Waals terms in order to avoid high-energy gradients due to steric clashes and then, gradually restoring the standard values for the non-bonded energy contributions, a minimization process is performed to slowly adapt each other the receptor and the ligand.

As mentioned above, another possibility to deal with protein flexibility consists in choosing those fundamental degrees of freedom, usually torsions in active-site side chains, that we want to model explicitly. Selecting is probably the trickiest part of these techniques and, to accomplish this task, both experimental and theoretical (e.g., from MD or MC simulations) *a priori* knowledge about alternative binding modes for the receptor in exam is usually necessary. To solve this problem, Anderson *et al.*^[37] have developed an interesting algorithm that, at first, starting from only a single protein structure, identifies the most flexible zones of the receptor (namely, those that are the most involved in the binding process) and that, afterwards, using rotamer libraries, predicts the most likely conformations for the previously detected regions. In this way the bias due to structures of proteins co-crystallized with inhibitors is sensitively reduce in following docking experiments.

The first attempt of modeling the receptor flexibility using a well determined subset of degrees of freedom was made by Leach^[38]. Exploring the protein and ligand conformational spaces by means of the Dead End Elimination (DDE)^[39] and the A* algorithm^[40], he was able to find, for each orientation of the ligand relative to the receptor, potential binding structures corresponding to combinations of residues side chains and ligand conformations whose energy is lower than a predetermined threshold. In this context, Schaffer and Verkhivker^[41] predicted the structures of complexes between two HIV-1 protease inhibitors and two mutants of this protein. In this case they performed flexible docking

computations that combine the MC simulated annealing technique to determine the ligand bound conformation and the DDE for the side chain optimization of the receptor binding site residues.

Moreover, it is worthwhile to note that some popular Docking programs as AutoDock^[42, 43] and GOLD^[44] have introduced the possibility to explicitly deal with the side chains flexibility. In particular, GOLD uses a genetic algorithm that allows describing not only ligand positions and conformations, but also hydrogen-bonding networks in the active site. Another GOLD-based strategy is the one devised by Verdonk *et al.*^[45], where only a small subset of protein side chains and terminal rotating hydrogen atoms are explicitly modeled in order to optimize the hydrogen bond interactions.

In this group of techniques we can also consider SLIDE^[46]. This method, after looking for the best matching between ligand and receptor using as criteria the best steric complementarity and the best matching between hydrophobic and hydrogen-bonding sites, resolves “Van der Waals collisions” between atoms of the two molecules performing a minimal number of side chain rotations, with the cost of a side chain movement evaluated as the product of the rotation angle and the number of atoms moved. A similar approach is SPECITOPE^[47] that uses minimal angle side chain rotations at the end of the docking procedure to remove steric clashes between the ligand and the protein. Unfortunately, although this algorithm is able to resolve many cases of overlap, it does not sample the conformational space completely and is not able to find the minimum energy conformation. In this context, it is worth to discuss the ROSETTA-LIGAND technique recently proposed by Davis and Baker^[48]. This new method, which uses a Monte Carlo sampling and the Rosetta full-atom energy function^[49] for scoring, does not account only for the receptor flexibility of the side chains in the binding pocket, but it also fully models the backbone flexibility. The preliminary results are promising since the new technique provides a larger set of docked complexes and reduces the average RMSD of the best-scoring docked poses compared to the rigid backbone results.

Among this group of strategies, the SCARE approach recently devised by Abagyan and coworkers^[50]

is very interesting. The algorithm scans every pairs of neighboring side chains in the receptor binding site and replaces them with alanine residues. Afterwards, the ligands in exam are docked to the previously determined “gapped” binding pockets and all the poses are scored. In the final stage, the original side chains are reintroduced and, after a proper optimization, each pose is scored again. Although only recently developed, the method is promising due to the fact that the number of corrected predictions greatly increased with respect to highly optimized docking strategies that use a single pocket conformation. These results are even more noteworthy considering that the technique proposed by Abagyan is fully automated and that it does not need to know *a priori* the ligand binding position or any information about the sites of potential variability of the receptor pocket.

Although the strategies described above can be considered as a step forward in dealing with receptor flexibility, the choice of selecting only a subset of degrees of freedom is not always the best one and, of course, it depends on the system under examination. For instance, the HIV-1 protease conformational space can be enough sampled modeling some side chain movements and a water molecule ^[43], but, on the other hand, there are many systems that undergo remarkable rearrangements during the ligand binding process. In these situations the “Induced Fit” model ^[32] is no longer a good description for molecular recognition and it is necessary to turn to the more advanced representation already mention above, according to which the ligand selects the optimal protein conformation among a conformational ensemble ^[33]. The simplest way to realize it computationally consists in developing strategies that consider simultaneously multiple receptor conformations (MRC) obtained from experimental data (i.e., from X-Ray or NMR techniques) or computational (i.e., from MD or MC simulations) sampling procedures.

The main advantage of this group of methods is that it is possible to model the protein flexibility not only in a specific small region. This also means that the MRC strategies allow describing the full receptor flexibility at a much lower computational cost compared to the one associated with possible

techniques that include all the receptor degrees of freedom. Nevertheless, flexibility is only modeled implicitly, which means that only a small fraction of protein conformational space is really taken into account. The effect of this poor sampling, particularly in case of experimental sampling, is that MRC docking results are mainly successful when the available receptor conformers derive from complexes with ligands similar to the one in exam. Finally, a second important problem is the need of a reliable energy function that enables to distinguish the small group of real low-energy structures among the hypothetical conformations generated during the sampling. This is a crucial aspect because an optimal energy function allows avoiding unpleasant drawbacks as incorrect pose predictions and false positives in virtual screening outputs.

Among the MRC techniques, FLEXE^[51] is probably one of the most remarkable examples. After superimposing different target structures, it merges the similar parts and considers the dissimilar ones as possible alternatives. Therefore, other than the receptor conformations in the starting set, new structures are generated and a more complete “ensemble” is taken into account during the docking experiments. Nevertheless, a recent study has pointed out that the method is not able to deal with large loop rearrangements and that, furthermore, it shows a worse performance in terms of enrichment factor than running multiple docking calculations for each receptor structure^[52]. The FLEXE philosophy also underlies the FITTED technique^[53] in the representation of receptor flexibility. In particular, when flexibility is fully considered, a genetic algorithm is used to combine different side-chain rotamers and backbone conformations included in the starting set of structures. In this context, Zavodszky *et al.* proposed an alternative approach^[54]. At first, to form the ensemble of receptor conformations, they analyzed protein flexibility using a graph-theory based algorithm and performed a random walk sampling. Afterwards, they applied the SLIDE method^[46] to the set of structures previously obtained. FLIPDock^[55] is another strategy recently developed to simultaneously take into account a great number of protein conformations. In order to accomplish this task, it exploits the sophisticated

“Flexibility Tree” algorithm, which, describing protein flexibility in terms of a hierarchical classification of movements, greatly reduces the computational cost associated with the description of receptor flexibility in docking computations. Shoichet and coworkers have also devised an interesting technique^[56] that is based on the DOCK program^[57-60]. This method initially considers an ensemble of predetermined protein structures and then, moving independently the flexible regions of the receptor and recombining their conformations, it implicitly takes into account a much greater number of structures. However, the most relevant feature is the fact that, for a given pose of the ligand, the algorithm chooses the best conformations of each part of the protein and this allows a sensitive reduction of the computational effort. Another strategy to examine is the one proposed by Sherman *et al.*^[61] that performed an iterative combination of rigid receptor docking and protein structure prediction using Glide^[62] and Prime^[63, 64], respectively. In particular, the method essentially consists in four main steps: (1) docking the ligands into the rigid receptor structure using a softened potential energy; (2) sampling and minimizing the best protein-ligand complexes obtained at the previous step; (3) docking the ligands into the best refined protein structures from step 2 using a hard energy function; (4) ranking the complexes using a composite scoring function that accounts for both the receptor-ligand interaction and for strain and solvation.

In the framework of the MRC methods, grid based approaches are particularly noteworthy. As well known, to efficiently evaluate the interaction energy between ligands and receptor, many docking strategies use interaction energy grids, which are calculated placing probe atoms at discrete points in the space around a target protein and assigning to these grid points the value of the interaction energy between the probes and the protein. The peculiarity of grid based MRC techniques is that, instead of considering one grid for each conformation in the ensemble, only one grid that represents all the structures is determined and used. The “*in situ* cross docking” approach^[65, 66] is probably the simplest example and it consists in joining together the grids computed for each protein binding-site. Despite its

simplicity, this method is characterized by a sensitive speedup in terms of computation time and it can simultaneously deal with a wide range of protein conformations, although the number of structures that can be examined in each computation is limited. More advanced grid based MRC strategies exploit grids averaging techniques. As for the *in situ* docking, the main advantage is the improved computational speed, whereas the main drawback is that the average structure represented by the grid may not correspond to a real target and, therefore, the resulting ligand poses may be significant only for the average representation.

According to Osterberg *et al.* ^[43], the grid averaging methods can be classified into four main categories. The first two are very simple and consist in creating (1) a new grid of mean values calculated all over the grids and (2) a new grid that considers only the minimum values all over the grids. Of course, as one should expect, these techniques perform poorly. More advanced averaging strategies are (3) the one proposed by Knegtel and coworkers ^[67], which consists either in an energy-weighted average or in a geometry-weighted average, and (4) a simple Boltzmann-weighted averaging of the interaction energies. The last two techniques provide better results in docking calculations, even if they sometimes may introduce potential dangerous artifacts.

As already mentioned, it is extremely important to note that almost all the MRC methods presented above can be applied to sets of conformations that derive both from experiments and from computer simulations. In this context, it is necessary to observe that experimental techniques, such as X-Ray crystallography and NMR, can help in describing the structures and flexibility of proteins, but they are not able to provide complete insights of all molecular rearrangements because only lower energy states will be detected, while higher energy conformations, which can be eventually stabilized by ligand binding, are usually neglected. Furthermore, other aspects, such as crystallization difficulties and protein molecular weight, sometimes prevent to apply experimental techniques to the bio-molecule that we want to study. The dimensions of the system under study still represent an obstacle also for

computational methods, where the efficiency of calculations rapidly decreases as the number of atoms increases. In this context, simulations of systems of several hundred-thousand particles (including the solvent) are becoming available and are starting to be routinely used for functional and small-molecule discovery studies ^[68-70]. Computational sampling methods, in particular the ones that use Molecular Dynamics, are becoming more and more essential to generate large ensemble of protein structures to be used in MRC strategies. Nevertheless, although nowadays computer power is rapidly increasing (and the parallelization of many MD codes has greatly enhanced this aspect), the critical issue associated with Molecular Dynamics simulations is their length. In fact, due to time scale limitations, also the MD trajectories are often not able to sample all the protein structures relevant from a biological point of view and, therefore, large conformational rearrangements are not always reproduced. However, these drawbacks can be partially overcome applying replica exchange methods (REMD) that consist in efficient algorithms involving parallel simulations of several replicas of the biomolecular system at different temperatures, occasionally exchanging according to a modified Metropolis transition probability ^[71, 72], and allow a better exploration of conformational space, or exploiting coarse-grained samplings ^[73] that describe large receptor motions and can be combined with a fine-scaled MD sampling to account for local transitions.

Pang and Kozikowski were probably the first to use structures generated by Molecular Dynamics in a MRC strategy ^[74]. In particular, studying the binding of huperzine A (HA) to acetylcholinesterase (AChE), they carried out a short MD simulation of AChE and extracted a group of conformations which were afterwards used to rigidly dock the HA ligand. McCammon and coworkers proposed developments of this pioneer approach introducing the “Relaxed Complex Method” ^[75-77]. The technique is based on longer and more accurate Molecular Dynamics samplings and it takes into account much more receptor structures, which can be selected at regular time intervals or considering their conformational variability to form a complete ensemble. Another interesting example of MRC

method that uses snapshots from a MD simulation is the one devised by Broughton ^[78] that, using a weighted average method, combined the interaction energy grids associated with the collected structures into a single grid.

One of the most interesting and outstanding approaches that consider multiple receptor conformations performing computer simulations has been proposed by Carlson and coworkers ^[79-92]. In this strategy the active site of each structure is flooded with hundreds of small molecular probes, namely benzene molecules to identify hydrophobic and aromatic regions, ethane molecules to distinguish hydrophobic interactions from the aromatic ones and methanol molecules to define hydrogen-bonding sites. These probes are simultaneously optimized by means of a low temperature Monte Carlo minimization, during which the protein is held fixed and probe-probe interactions are completely ignored. After the minimization step, significant clusters of probes (namely, clusters constituted by eight or more probes) can be easily detected and each of them is represented by a “parent”, which is the lowest-energy element in the group. At this point, the results obtained for the different conformations are superimposed and, if a “cluster parent” conserved over many of the receptor structures (namely, for more than 40% or 50%), it is possible to define a “consensus cluster” that is associated with a spherical pharmacophore element. Therefore, the Carlson method allows obtaining pharmacophore models that implicitly account for receptor flexibility, models that can be subsequently used to design new ligands/inhibitors or for databases virtual screenings. Of course, it is worth to observe that this technique can be also applied to set of protein conformations that are derived experimentally.

The philosophy underling the strategy described above is also the basis for the approach recently devised by Colombo and coworkers ^[70, 93, 94]. As in the Carlson method, the final aim is to construct a new pharmacophore, but, instead of optimizing the placement of small probe molecules, they performed MD simulations of receptors bound with already known peptide-ligands. Analyzing these simulations, it is possible to determine those ligand-protein interactions that conserve over the

simulations and that, for this reason, also define pharmacophore elements. Although this strategy is conceptually very similar to the previous one, it is worthwhile to observe that using known ligands instead of small molecular probes allows to account for useful experimental information, even if the ligands considered in the simulations are often small peptides, which unfortunately do not completely cover all the features of real drugs.

As just described above, MD simulations play a very important role in improving the MRC techniques. Nevertheless, in order to better describe the molecular recognition process, it is necessary to simulate the interaction between ligand and receptor considering all the possible degrees of freedom and going beyond the limitations associated with the flexibility models previously presented (e.g., MRC models). To accomplish this task, we should perform full Molecular Dynamics or Monte Carlo simulations, which are characterized by a high accuracy, but also by a large computational cost due to the fact that the binding processes between ligands and proteins are usually very slow. Therefore, despite the rapidly increasing computer power, nowadays it is not possible to carry out standard MD or MC simulations to screen large databases of compounds. Nevertheless, introducing some reasonable approximations, which make the calculations less expensive and, at the same time, less accurate, we can obtain fundamental information that would be unfortunately lost using less flexible receptor descriptions.

The easiest way to reduce the computational cost consists in restricting the full simulations only to the active site, to the ligand and, if necessary, also to regions near to them. Mangoni *et al.* ^[95] have proposed an interesting improvement of this technique where they limit again the simulation to a well-defined region, but allowing a faster exploration of the receptor conformational space. In particular, they simulated the ligand internal motions, the solvent and the receptor at room temperature, whereas they set a much higher temperature for the ligand translational modes of motion. Other interesting ways to speed up the sampling are the strategy devised by Wang and Pack ^[96], which applied a scaling

function to the equations of motion in order to promote the crossing of barriers, and the multicanonical algorithms, which have been developed in the framework of both Monte Carlo and Molecular Dynamics techniques ^[97-99] and consist in a smoothing of the potential energy surfaces. Finally, in the context of Monte Carlo methods, we can also consider Prodock ^[100, 101] that is a modified MC strategy, which is mainly characterized by local gradient-based minimizations after each random move and by the scaling of the potential energy terms during the docking.

Unfortunately, despite the development of modified MD or MC techniques as those just described, to represent large-scale receptor flexibility as large conformational rearrangements of protein domains, it is necessary to take into account strategies that use collective degrees of freedom. These methods are less computational expensive than the traditional computer simulations because the number of independent degrees of freedom is sensitively reduced, but, on the other hand, these degrees of freedom are not the native ones and this may affect the final results.

Typical methods in this framework are those that consider the receptor harmonic normal modes of vibration and, in particular, the low-frequency modes that are associated with very delocalized motions. Interesting applications of this kind of techniques to the problem of molecular recognition are the study by Zacharias and Sklenar ^[102] to deal with DNA flexibility for the binding of small molecules and the study by Keserû and Kolossváry ^[103, 104] to investigate the inhibitors binding to HIV integrase. Another group of methods that exploit the receptor collective degrees of freedom relies on techniques that reduce the dimensions of the system under examination and, in particular, on the Principal Component Analysis (PCA). As the low-frequency normal modes in the computation of harmonic normal modes, also the most significant principal components in PCA (namely, the ones associated with the highest eigenvalues) describe very large rearrangements and most of the conformational variations in large biomolecules. In this context, Teodoro *et al.* ^[34, 35] developed a strategy that enables to sensitively simplify the description of the protein-ligand interaction, significantly reducing the high-dimensional

representation of the protein flexibility. Finally, the approach proposed by Tatsuni *et al.* ^[105] is particularly interesting. At first, they performed a Principal Component Analysis to determine the receptor global conformational changes and, afterwards, they coupled harmonic dynamics with molecular dynamics to take into account protein large rearrangements and local side chain flexibility, respectively. From preliminary studies on the HIV-1 protease and its ligands, it seems that the new strategy is able to efficiently reproduce the formation of molecular complexes.

Considering all the techniques presented so far, it is obvious that the ones that heavily use MD or MC simulations are the most accurate. Nevertheless, they are also the most computational demanding and, therefore, as already stressed above, they cannot be used when we have to perform screenings of very large databases. Hence, to overcome this problem and to preserve some accuracy, several research groups have recently developed hierarchical methods that combine different strategies, using less accurate and faster approaches for preliminary screenings and higher-level of theory and expansive strategies for later stages. Machicado *et al.* ^[106] proposed a three-step virtual screening strategy that enables to identify active ligands for buried protein cavities. It consists in applying some physicochemical filters, a fast docking procedure and, at the final stage, a finer flexible docking algorithm that exploits a Monte Carlo search technique and that uses a purposely defined binding free energy function to properly score the docked complexes. Another example is Extra Precision GLIDE ^[62, 107] that, at first, carries out docking calculations using relaxed criteria and then allows refining the results by means of a more sophisticated computation of the binding free energies. More recently, Lee and Sun have introduced another method ^[108] that actually consists in four different protocols that optimize the poses obtained from the classical docking program DOCK 4.0 ^[109] by means of proper successions of minimizations and molecular dynamics simulations followed by a final MM/GBSA (Molecular Mechanics / Generalized Born Surface Area) scoring. Finally, in this category of strategies,

it is worthwhile to consider also the approaches devised by Wang *et al.* ^[110] and by Graves *et al.* ^[111]. The former consists in the following four hierarchical filters to screen ligands databases: (1) pharmacophore model, (2) rigid docking, (3) solvation docking and (4) molecular dynamics simulation combined with the MM/PBSA (Molecular Mechanics / Poisson-Boltzmann Surface Area) method, which is exploited to compute the binding free energies of the thirty most promising hits that survive from the previous steps. It is extremely important to note that, while the first three filters take into account only the ligand flexibility, the last one consists in a MD simulation that samples a part of the conformational space of both the inhibitors and the receptor. The latter approach is based on the same philosophy, but the initial docking calculations are combined with MM/GBSA energy evaluations to rescore the best poses. Also in this case it is worth to stress that the MM/GBSA technique introduces a dynamic sampling of the protein-ligand complexes, although this is limited to the configurational space in the neighborhood of the starting docking poses. Analyzing the results obtained from preliminary tests performed on model cavity sites, it seems that this new strategy allows rescuing many docking false negatives, improves the binding geometry of most of the predicted structures and increases the diversity of the hit lists. Nevertheless, the rescoring technique introduces a great number of false positives, especially among the very top ranking ligands, and, unfortunately, this is probably due to the introduction of the protein flexibility by means of the MM/GBSA method.

Possible new Developments from the Inclusion of Flexibility in Drug Discovery: Targeting Allosteric Sites.

As we already stated in the introduction and theory section, allostery is now recognized as a general property of many monomeric proteins. Understanding the molecular determinants of information passage through allosteric modifications is now a fundamental goal of many research groups. Advances

in experiments and theory indicate that intramolecular signaling processes depend on the intrinsic dynamics of proteins and that signal transduction is performed by shifting among different fluctuating states in response to a certain input.

The exploitation of the molecular understanding of these phenomena for drug design or screening may have huge implications in pharmaceutical applications. Indeed, Allosteric sites offer opportunities for identifying new molecules and chemical structures, overcoming the hurdles posed by targeting active-site inhibitors with similarities among multiple proteins (or families), the development of drug-resistance etc. Moreover, the molecular understanding of signal transduction mechanisms and the discovery of related allosteric sites could open the possibility to design chemical modulators of protein functions.

Unfortunately, the rational discovery of allosteric sites is extremely challenging: the information and results of protein dynamics studies do not easily translate into medicinal chemistry strategies. As a consequence, high throughput screening technologies combined with X-ray crystallography have been applied to this goal in the context of industrial research ^[112]. Alternative strategies have focused on the use of peptide phage-display to identify peptides with affinity for previously unknown allosteric sites ^[113]. A site-directed strategy, called Tethering, to detect new sites takes advantage of native or introduced cysteines on protein surfaces to trap thiol-containing ligands. Ligands that bind to possible allosteric sites can be captured by the formation of a disulfide bond between the small-molecule and the cysteine near the putative site, which can be verified through subsequent mass-spectrometry and functional assays ^[114].

Rational strategies that lead to the identification of allosteric sites based on information on the dynamic coupling mechanisms between the active and a distant site would be highly desirable and could provide a general tool for new rational drug discovery. Computational and theoretical approaches offer the

unique possibility to study conformational changes and their couplings to alternative residue packing and interaction networks at atomic resolution. In the next section we describe some of the most interesting methods to detect allosteric sites and couplings in proteins. We envisage that coupling of these methods to the flexible docking/screening approaches described previously may constitute a viable strategy to discover new hits for new allosteric sites.

Molecular basis of allostery: Insights from computational methods.

Signal transduction along well defined paths connecting distant residues has been recognized as a common property of allosteric proteins. In the understanding of allosteric effects, theory aims at describing the site to site communication at the molecular or atomic level. A number of attempts have been carried out by means of sequence or structure-based computational methods, in order to predict allosteric sites and pathways in proteins.

Based on the sequence, the statistical site coupling approach (SCA) of Ranganathan allows the identification of evolutionarily conserved couplings between residue pairs^[115, 116]. The method is based on the assumption that, if the interaction between a pair of residues is relevant for protein function or folding, then the two residues should co-evolve, showing a relevant covariation in multiple sequence alignments (MSA) of a protein family. A statistical energy term can be derived from this covariance. This method was originally applied to the human tyrosine phosphatase PDZ domain family and led to the identification of residue patterns and networks linking active sites to distant regions, predicting a signaling pathway subsequently confirmed by NMR experiments^[117].

In the analysis of allosteric properties of PDZ domain, a structure-based approach making use of non-equilibrium MD simulations was introduced by Ota and Agard^[118]. The method, called anisotropic thermal diffusion, relies on the assumption that the communication properties of residues depend on the energetic couplings among them and signals are driven by the same pathways on which the kinetic

energy, originated when perturbing a site, dissipates through the structure. Therefore, a non-equilibrium MD simulation is set up, where a significant perturbation is created by artificially heating a single residue with a temperature bath of 300K, while the rest of the protein is thermalized at low temperature. The dissipation of the thermal fluctuation, tracked along the structure during dynamics, turns out to be highly anisotropic and the presence of preferred signaling pathways can be detected, in agreement with experimental data and the sequence-based computational methods of Ranganathan.

Another application of the statistical coupling analysis (SCA), in combination with a coarse grained Brownian dynamic simulation, was introduced by Chen et al. ^[119] to analyze the signaling network of E. coli dihydrifolate reductase. This enzyme undergoes an allosteric conformational transition involving the Met20 loop during the catalytic cycle. By means of the SCA, a number of highly covarying residues is hypothesized to be the minimal network signaling the kinetics of the conformational transition. A long term simulation of the protein a coarse-grained description is then produced to generate the covariance map, showing that correlated and anticorrelated motions on the microsecond time scale involve the same conserved residues, even if they are spatially separated.

The role of low frequency, collective modes in mediating the signal propagation along given pathways is emerging as a plausible mechanism to explain allosteric effects. Recently, Bahar et al introduced a novel information-theoretic approach called MAPS (MARKovian Propagation of Signals) to study the transition of GroEL from the T (apo) state into the R (ADP and ATP bound, respectively to cis and trans rings) state ^[120]. The protein structure is modeled as a network of interactions, where the interaction strength between two residues depends on the number of atom-atom contacts between the two. When a site is perturbed, the perturbation propagates through the network, being dissipated by means of coupled fluctuations in residue positions (fluctuation-dissipation) and eventually relaxing to a new equilibrium. It has been shown that signal propagation time is shorter when pairs of residues are subject to smaller fluctuations in their inter residue distance, which is the case for instance for tightly

interacting residues. By treating the network with a clustering hierarchical reduction, sites with a "high allosteric potential" can be identified while communication pathways can be defined by tracking the maximum-likelihood paths along the network. The analysis of the ADP-bound GroEL-GroES structure reveals strong inter-subunit couplings between the cis and the trans ring and with the cochaperonin, and a number of residues or chain regions are correctly identified as important for allosteric communication. Moreover, the global motions of the complex, described by the lowest eigenmode in a ANM representation, are responsible for modulating the interactions of the allosterically relevant residues at the interface of the subunits.

The analysis of the interresidue distance fluctuations has been applied also in the framework of molecular dynamics simulations to investigate the large-scale conformational rearrangements of the yeast Hsp90 dimer upon ATP binding and hydrolysis. The perspective of MD is useful in capturing the global changes occurring upon ligand binding, without neglecting the details of the side chain orientation at the local level. In fact, the subtle changes in side chain arrangement constitute the first step in response to a binding event and then accumulate to produce a global change in the structure. The communication propensity between the ATP binding site and the distant C-terminal interface (more than 80 Ångstrom apart), measured in terms of interresidue distance fluctuations, is increased in the presence of ATP or ADP and is switched off in the unbound complex. Moreover, ATP and ADP appear to communicate with distinct regions of the C-terminal area, indicating a specific allosteric effect of each ligand ^[121] (Figure 1).

Another network approach based on the topology of the protein, which also makes use of MD simulations to take side chain details into account, is the protein structure network PSN approach of Vishveshwara *et al.* ^[122, 123]. The method of Protein structure network analysis has been applied to an allosteric protein, member of the class of aminoacyl tRNA synthetases and responsible for picking up the cognate amino acid in the active site and cognate tRNA in the anticodon binding site. The method

consists of carrying out molecular dynamics simulations of the protein in different liganded states, and then, for each dynamic snapshot, constructing a network of interactions between the residues, where each residue is a node and an edge between two residues represents the interaction, defined by the fraction of atom pairs which are closer than 4.5 Ångstrom. These structure networks evolve during dynamics and can be used to find the communication pathways between two endpoints^[122] (selecting the shortest and most dynamically correlated path). The paths turn out to be inherently present in the enzyme. They establish efficient communication at distance of 70 Ångstrom between the anticodon region and the activation site, in the presence of both the tRNA and MetAMP that drive specific residue–residue contacts. The analysis of the graph in terms of cliques (sets of mutually connected nodes), communities of cliques a (non-identical cliques sharing nodes) and hubs (nodes making more than four edges) also reveals a different dynamic behavior for communication paths, which are flexible and robust. The network parameters describing the noncovalent interaction of side chains are able to discriminate subtle differences in the side chain interactions, taking place in response to different ligands^[123].

The knowledge of the allosteric pathways transmitting signals in a protein allows in principle to regulate its function by means of suitable small molecules that can bind to the allosteric sites, but also by designing ad hoc mutations able to interfere with the protein (mis)function. A suggestion in this sense comes from a recent study by Liu and Nussinov on the pVHL protein, a tumor suppressor protein. A naturally occurring mutant of this protein with reduced thermodynamic stability is associated to a number of cancer diseases^[124]. By means of MD simulations the region mainly responsible for the low stability of the protein is identified. In order to avoid mutations directly on this area, which would perturb the protein function, regions dynamically correlated with it are found by analyzing the covariance matrix and mutations are inserted there. The simulation of the new mutants suggests a repairing effect of the mutations, that appear to be able to restore the wild type stability.

An applicative example on Cytochrome P450 enzymes:

Insights from molecular dynamics simulations

The cytochrome P450 is a large family of heme-containing enzymes with a key role in oxidative transformations of endogenous (fatty acids, steroids, prostaglandins) and exogenous compounds (drugs and environmental chemicals) ^[125]. About 90% of all marketed drugs, clinical and preclinical drug candidates are substrates of human CYPs ^[126]. Therefore, any attempt to evaluate and predict drug metabolism will have to consider these enzymes.

In silico prediction of drug metabolism is a very challenging task due to a large number of variables involved, related to either the physico-chemical properties of the compounds and the nature of the enzymes ^[127]. Understanding how cytochrome P450 enzymes recognize multiple structurally diverse substrates has been limited for many years by the lack of knowledge about the three-dimensional structures of human cytochrome P450s. Homology models based on the crystal structures of bacterial and fungal enzymes ^[128-132] have been useful to assess possible protein-drug interactions ^[133-135]. However, one of the main achievements in this field was the structure determination of the first mammalian enzyme, the CYP2C5 from rabbit ^[136], with increased sequence identity to the target enzyme when compared to microbial CYPs, followed other mammalian and human P450 enzymes including CYP1A2 ^[137], CYP2A6 ^[138], CYP2A13 ^[139], CYP2C8 ^[140], CYP2C9 ^[141], CYP2D6 ^[142] and CYP3A4 ^[143]. All members of this large class of enzymes share a common octahedral heme iron covalently linked to the apoprotein by the sulfur atom of a cysteine residue. The protein chain of each individual P450 varies both in terms of molecular weight (45 - 60 kDa) and amino acid sequence ^{[144,}
^{145]}.

The crystal structures of cytochrome P450 enzymes revealed large conformational changes between ligand-free and ligand-bound states, and between cytochrome P450s bound to different molecules, which clearly demonstrated that CYPs are extremely flexible proteins. Protein movements range from relatively small induced fit displacements to accommodate small substrates within the binding pocket of the enzyme, such as with diclofenac bound to CYP2C5 ^[146], up to dramatic conformational changes induced by large substrates such as ketoconazole and erythromycin bound to CYP3A4 ^[147]. On the other hand, the active site of cytochrome P450 enzymes is located inside the protein, above the heme cofactor, accessible through several flexible channels which can be adjusted by the physico-chemical properties of the ligands ^[148] (Figure 2). For this reason, the dynamics of P450 enzymes must be taken into account in order to accurately predict drug metabolism by human cytochromes. In this section examples are presented that illustrate recent applications of molecular dynamics to get further insight on structural and dynamical aspects of cytochrome P450s, with particular attention to the recently elucidated human enzymes.

Mechanism of substrate specificity

CYP3A4 is one of the most extensively studied cytochromes by classical atomistic simulations. This enzyme has a broad substrate specificity being responsible for the metabolism of more than 50% of all marketed drugs ^[149]. The structural and dynamical basis of CYP3A4 promiscuity was recently investigated by two independent groups. Suh and coworkers used B-factors derived from MD simulations to identify a highly-flexible loop (residues 214-216) close to a substrate access channel ^[150]. A similar approach was used by Otyepka and coworkers in a comparative study with three P450 enzymes, CYP3A4, CYP2C9 and CYP2A6 [29]. Good correlation was found between substrate specificity and malleability of these enzymes. CYP3A4, the most promiscuous enzyme, was also the

most flexible, whereas CYP2A6, the most selective enzyme, was the most rigid. Furthermore, distinct domains were identified in terms of flexibility, namely, a relatively rigid heme-binding pocket, a more flexible proximal domain responsible for the interaction with redox partners, and a highly malleable distal domain responsible for the interaction with membranes and substrate channeling. Taken together, these differences highlight the importance of flexibility for substrate entrance and product release, as well as interaction with membranes and redox partners. On the other hand, high regio and stereoselectivity of CYP reactions requires a relatively rigid binding pocket.

In another study the molecular basis of broad substrate specificity and high regioselectivity, a typical feature of most cytochrome P450 enzymes, was studied through multiple molecular dynamics simulations of warfarin bound to CYP2C9 [151]. While the large binding cavity was very flexible, adopting different shapes and allowing several substrate binding modes, access to the active site was limited by a narrow, rigid and hydrophobic channel with a major impact on regioselectivity.

Active site access channels

Due to the buried nature of cytochrome P450 active sites, substrate specificity may be influenced not only by molecular recognition at the binding pocket but also by the opening of channels and their physicochemical properties. Current nanosecond time scale molecular dynamics simulations not always provide a complete representation of ligand migration pathways, particularly when opening of channels not seen in crystal structures is required. This limitation was overcome by the use of enhanced sampling methods with lower computational power requirements. Steered molecular dynamics (SMD) and random expulsion molecular dynamics (REMD), for example, accelerate the speed of ligand movements by applying an artificial force in addition to the standard molecular mechanics force field

[152, 153]. These techniques were initially applied to bacterial cytochromes such as P450cam, P450BM-3 and P450eryF, identifying several ligand access pathways and opening mechanisms involved [154]. Egress between the F-G loop, the β_1 sheet and either the B' helix (P450cam, P450BM-3) or the B-B' loop (P450eryF) was common to all enzymes, however, different ligand access mechanisms were identified, involving more or less pronounced backbone conformational changes, aromatic side chain rotations and long-range electrostatic interactions. Moreover, the crystal structures of mammalian cytochrome P450 enzymes have different visible channels when compared to bacterial P450s [155]. The preferred channel of bacterial CYPs is mostly closed in crystal structures of mammalian enzymes, however, alternative pathways are open, suggesting adaptation to different cellular environments and substrate specificity profiles [156, 157]. The involvement of channeling residues in the catalytic efficiency of cytochrome P450 enzymes was confirmed experimentally by combining enzymatic assays, mutagenesis studies and steered molecular dynamics simulations [158].

Prediction of enzyme-substrate interactions

Defining the most probable oxidative pathways involved in drug metabolism requires accurate calculation of free energies of binding to cytochrome P450 enzymes. Static approaches based on rigid protein-ligand interactions such as docking and scoring strategies, are able to predict binding modes and identify high affinity compounds from large electronic databases [103, 159]. However, currently available scoring functions show little or no correlation with experimental binding affinities. On the other hand, molecular dynamics simulations account for both the structure and the dynamics of the proteinligand interactions, providing better estimates when used to predict free energy differences between complexes with different ligands and proteins. Successful free energy calculations have been reported using several methods such as linear interaction energy calculations (LIE) [160] [48], used to

study the metabolic activation of chemical carcinogens by human CYP1A1 ^[161], thermodynamic integration (TI) ^[162], used to rationalize the stereospecific metabolism by polymorphic forms of CYP2D6 ^[163], and a combination of MD simulations, entropy estimates and Poisson-Boltzmann calculations (MM-PBSA) ^[164] used to study ligand binding to CYP2B4 ^[165].

Besides electronic criteria such as the enthalpy change involved in formation of transition state structures ^[166], the most likely positions for metabolism in a drug depend on the interactions between the enzyme and the substrate, affecting their relative orientation to the heme iron. Therefore, the probability of reaction at a given substrate position is highly related to the distances and angles between the heme iron and the reactive atoms involved. Taking these measures into account using docking and molecular dynamics simulations, provides good qualitative predictions of regio- and stereo-selective transformations, as demonstrated in recent *in silico* metabolism predictions with CYP1A1/2 ^[167, 168], CYP2D6 ^[169, 170], CYP2E1 ^[171] and CYP3A4 ^[172]. Moreover, MD simulations were used to elucidate the mechanism of cooperative binding to cytochrome P450 enzymes, responsible for the atypical, non-Michaelis-Menten kinetics ^[173-175] observed in members of the P450 family with large binding pockets, such as CYP3A4, allowing simultaneous fit of at least two substrate molecules ^[137, 140, 176]. Significant differences between MD simulations with substrate-free, single-bounded and double-bounded enzyme demonstrated that key residues for the mechanism of cooperative binding are stabilized by the second substrate, inducing a favorable orientation of the molecule that undergoes oxidation and decreasing the distance between the site of metabolism and the heme iron.

Another important application of molecular dynamics simulations investigated substrate binding modes at positions distant from the heme iron, which are inconsistent with the mechanism of catalysis, such as those identified in crystal structures of Npalmitoylglycine bound to P450BM-3 ^[177], warfarin bound to CYP2C9 ^[141], progesterone bound to CYP3A4 ^[143] and palmitic acid bound to CYP2C8 ^[140]. Replica

exchange molecular dynamics (REMD) were used to study the equilibrium between distal and proximal binding. Simulations demonstrated that while the distal state is the most populated at low temperatures, at room temperature, the majority of the population is in the proximal state ^[178]. Therefore, it has been hypothesized that unproductive binding modes arise from low energy conformations adopted by the binding site at the low temperatures used for crystallization, and that at biologically relevant temperatures the binding pocket of P450 enzymes suffers several conformational changes leading to ligand displacement and initiation of the enzymatic mechanism.

Role of active site water molecules

High resolution crystal structures of cytochrome P450s have revealed conserved water molecules within the binding pocket of the enzymes, establishing a link between the heme iron and key amino acids involved in the mechanism of catalysis ^[129, 159]. These waters play an essential role in the mechanism of reaction by providing protons needed for the formation of the final oxidative specie ^[179]. On the other hand, water molecules can favor ligand binding free energy, either entropically, by displacement of ordered waters to the bulk solvent, or enthalpically, for example, by mediating hydrogen bond bridges. These effects have been studied in cytochrome P450 enzymes using molecular dynamics simulations ^[180, 181]. However, P450 hydration is often only incompletely revealed in the crystal structures due to large motions within the binding pocket, as well as inadequate resolution of the diffraction technique. Molecular dynamics simulations provide an alternative strategy to study the mobility of water molecules within the binding pocket of human P450 enzymes, identifying ordered water molecules and determining exchange rates with bulk solvent through several channels ^[182].

Acknowledgements: The authors gratefully acknowledge the FP6 project BacAbs (Contract nr. LSHB-CT-2006-037325) and Associazione Italiana per la Ricerca sul Cancro for funding. We thank Prof. Dr.

Maria Luisa Vaz Sa e Melo for support to MN. MN acknowledges FCT for a Ph.D. Grant SFRH/BD/17624/2004.

Captions for Figures.

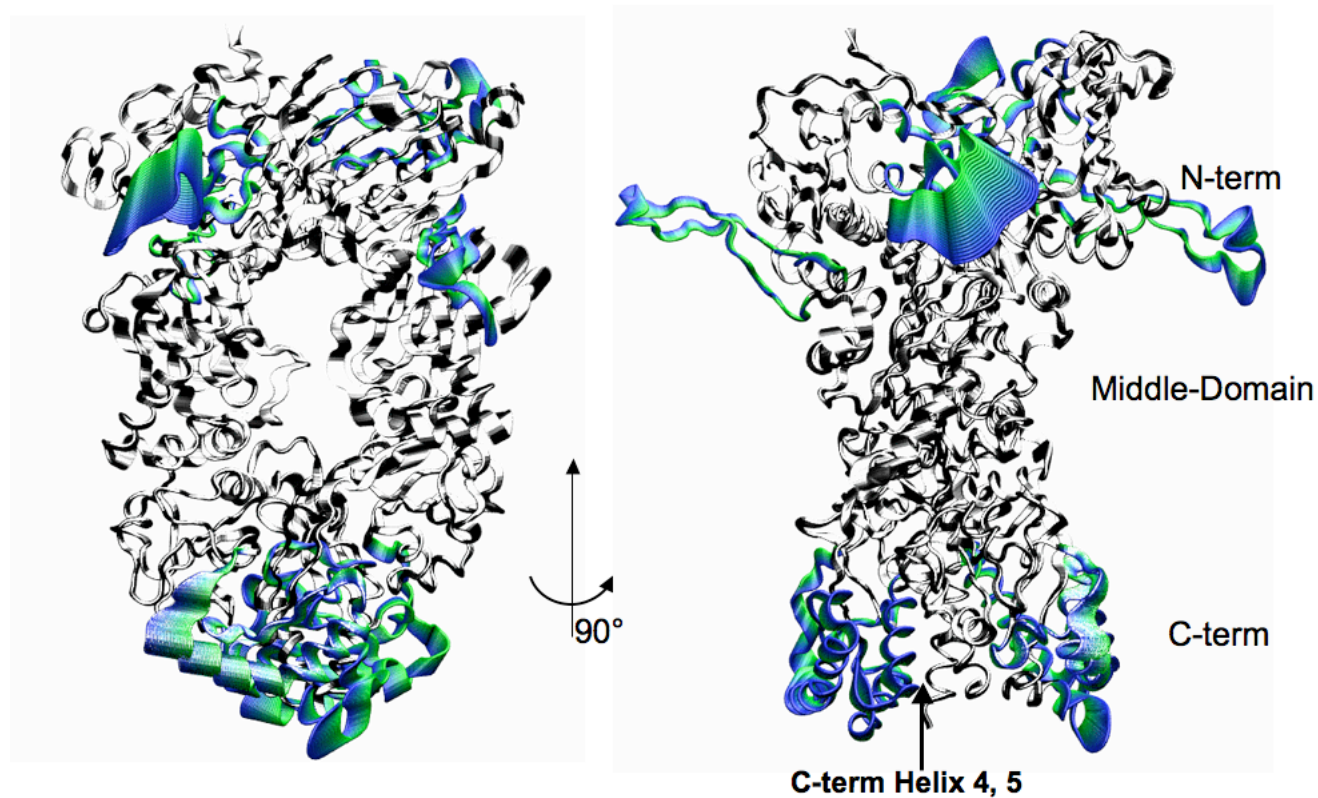


Figure 1. Consecutive snapshots from the MD simulation of the Hsp90 ATP-complex: Regions undergoing relevant correlated motions are highlighted in colors; they involve the ATP lid of both protomers and helices 3 at the CTDs. The color code from green to blue reflects the time evolution of the trajectory.

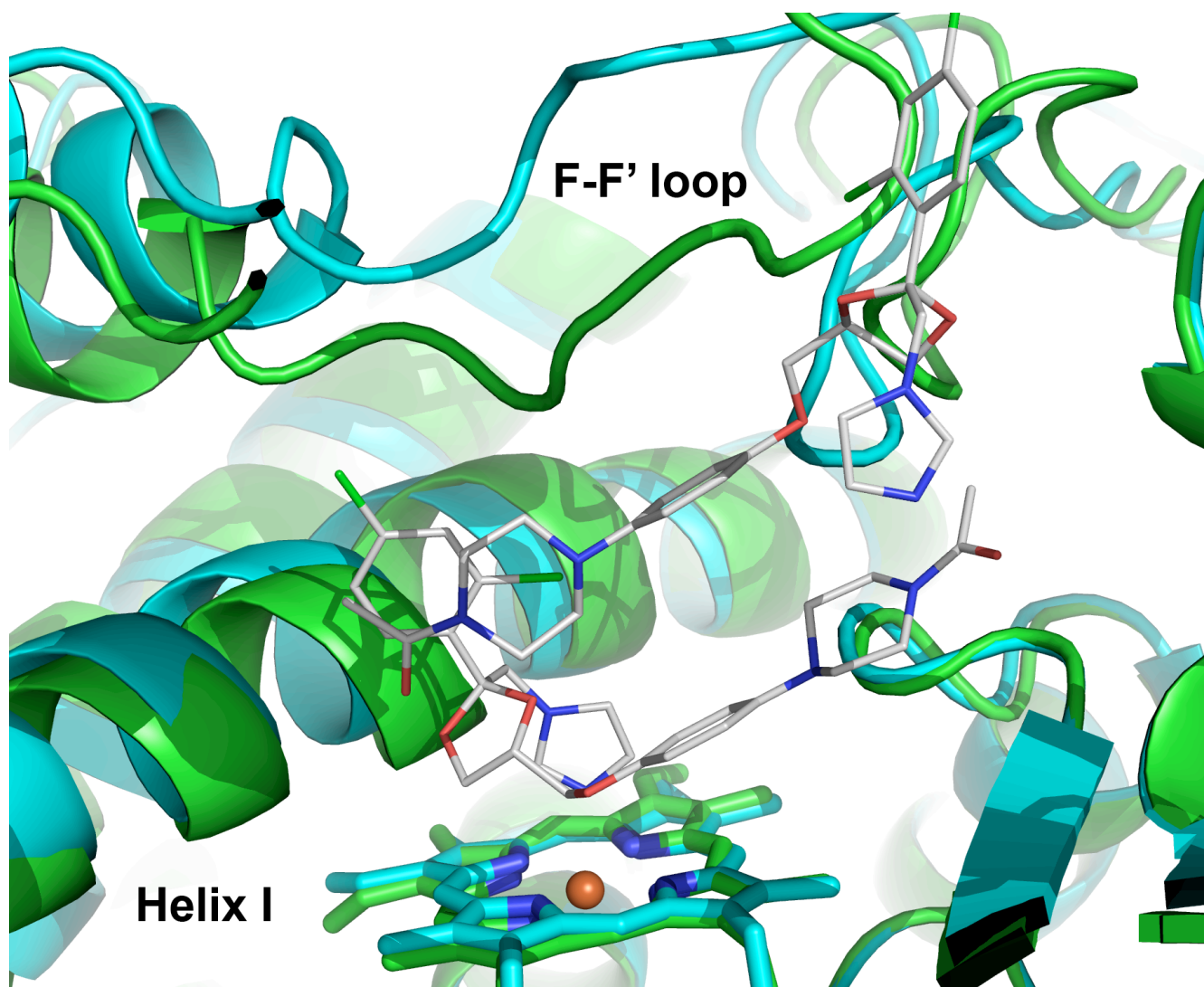


Figure 2. Backbone superimposition of ligand-free (green, pdb entry 1TQN) and ketoconazole-bound CYP3A4 (cyan, pdb entry 2V0M). A significant F-F' loop displacement occurs upon ligand binding.

References

- [1] Bolton, W.; Perutz, M.F. Three dimensional Fourier synthesis of horse deoxyhemoglobin at 2.8 angstrom resolution. *Nature*, **1970**, *228*, (551-552).
- [2] Fermi, G.; Perutz, M.F.; Shaanan, B.; Fourme, R. The crystal structure of human deoxyhaemoglobin at 1.74 Å resolution. *J. Mol. Biol.*, **1984**, *175*, 159-174.
- [3] Agarwal, P.K.; Billeter, S.R.; Rajagopalan, P.T.R.; Benkovic, S.J.; Hammes-Schiffer, S. Network of coupled promoting motions in enzyme catalysis *Proc. Natl. Acad. Sci. USA*, **2002**, *99*, 2794-2799.
- [4] Boehr, D.D.; McElheny, D.; Dyson, H.J.; Wright, P.E. The dynamic energy landscape of Dihydrofolate reductase catalysis *Science*, **2006**, *313*, 1638-1642.
- [5] Osborne, M.J.; Schnell, J.; Benkovic, S.J.; Dyson, H.J.; Wright, P.E. Backbone Dynamics in Dihydrofolate Reductase Complexes: Role of Loop Flexibility in the Catalytic Mechanism *Biochemistry*, **2001**, *40*, 9846-9859.
- [6] Fischer, E. *Ber. Dtsch. Chem. Ges.*, **1894**, *27*, 2985-2993.
- [7] Tsai, C.J.; del Sol, A.; Nussinov, R. Allostery: absence of a change in shape does not imply that allostery is not at play *J. Mol. Biol.*, **2008**, *378*, (1), 1-11.
- [8] Monod, J.; Jacob, F. *Cold Spring Harbor Symp. Quant. Biol.*, **1961**, *26*, 389-401.
- [9] Monod, J.; Wyman, J.; Changeux, J.P. On the nature of allosteric transitions—a plausible model. *J. Mol. Biol.*, **1965**, *12*, 88.
- [10] Koshland, D.E.; Nemethy, G.; Filmer, D. Comparison of experimental binding data and theoretical models in proteins containing subunits. *Biochemistry*, **1966**, *5*, 365.
- [11] Gunasekaran, K.; Ma, B.Y.; Nussinov, R. Is allostery an intrinsic property of all proteins? *Proteins: Struct. Funct. Genet.*, **2004**, *57*, 433–443.
- [12] Hardy, J.A.; Wells, J.A. Searching for new allosteric sites in enzymes. *Curr. Opin. Struct. Biol.*, **2004**, *14*, 706-715.
- [13] Swain, J.F.; Gierasch, L.M. The changing landscape of protein allostery *Curr. Opin. Struct. Biol.*, **2006**, *16*, (1), 102-108.
- [14] Elber, R.; Karplus, M. Multiple conformational states of proteins: a molecular dynamics analysis of myoglobin *Science* **1987**, *318*, 318-321.
- [15] Henzler-Wildman, K.A.; Lei, M.; Thai, V.; Kerns, S.J.; Karplus, M.; Kern, D. A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature*, **2007**, *450*, (6), 913-918.
- [16] Eisenmesser, E.Z.; Millet, O.; Labeikovsky, W.; Korzhnev, D.M.; Wolf-Watz, M.; Bosco, D.A.; Skalicky, J.J.; Kay, L.E.; Kern, D. Intrinsic dynamics of an enzyme underlies catalysis *Nature*, **2005**, *438*, (7064), 117-121.
- [17] Henzler-Wildman, K.A.; Thai, V.; Lei, M.; Ott, M.; Wolf-Watz, M.; Fenn, T.; Pozharski, E.; Wilson, M.A.; Petsko, G.A.; Karplus, M.; Hubner, C.G.; Kern, D. Intrinsic motions along an enzymatic reaction trajectory *Nature*, **2007**, *450*, 838-844.
- [18] Cui, Q.; Karplus, M. Allostery and Cooperativity revisited *Protein Science*, **2008**, *17*, 1295.
- [19] Lange, O.F.; Lakomek, N.A.; Fares, C.; Schroder, G.F.; Walter, K.F.A.; Becker, S.; Meiler, J.; Grubmüller, H.; Griesinger, C.; de Groot, B.L. Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. *Science*, **2008**, *320*, (5882), 1471-1475.
- [20] Caflisch, A.; Verkhivker, G.M. Folding for binding or binding for folding? *Trends in Biotechnology*, **2003**, *21*, (423-425).
- [21] Bahar, I.; Chennubhotla, C.; Tobi, D. Intrinsic dynamics of enzymes in the unbound state and, relation to allosteric regulation *CURRENT OPINION IN STRUCTURAL BIOLOGY*, **2007**, *17*, (6), 633-640
- [22] Weber, G. Ligand binding and internal equilibrium in proteins. *Biochemistry*, **1972**, *11*, 864-

878.

- [23] Millet, O.; Hudson, R.P.; Kay, L.E. The energetic cost of domain reorientation in maltose-binding protein as studied by NMR and fluorescence spectroscopy *Proc. Natl. Acad. Sci. USA*, **2003**, *100*, 12701-12705.
- [24] Wang, J.; Verkhivker, G.M. Energy landscape theory, funnels, specificity, and optimal criterion of biomolecular binding *Phys. Rev. Lett.*, **2003**, *90*, (18), 188181.
- [25] Frauenfelder, H.; Parak, F.; Young, R.D. Conformational substates in proteins *Annu. Rev. Biophys. Biophys. Chem.*, **1988**, *17*, 451-479.
- [26] Colombo, G.; Morra, G.; Meli, M.; Verkhivker, G.M. Understanding ligand-based modulation of the Hsp90 molecular chaperone dynamics at atomic resolution. *Proc. Natl. Acad. Sci. USA*, **2008**, *105*, (23), 7676-7681.
- [27] Kern, D.; Zuiderweg, E.R.P. The role of dynamics in allosteric regulation. *Curr. Opin. Struct. Biol.*, **2003**, *13*, 748-757.
- [28] Tsai, C.J.; del Sol, A.; Nussinov, R. Protein allostery, signal transmission and dynamics: a classification scheme of allosteric mechanisms. *Mol. BioSyst.*, **2008**, *5*, (3), 207-216.
- [29] Cooper, A.; Dryden, D.T. Allostery without conformational change. A plausible model. *Eur. Biophys. J.*, **1984**, *11*, 103-109.
- [30] Colombo, R.F.; Rau, D.C.; Parsegian, V.A. Protein solvation in allosteric regulation: a water effect on hemoglobin *Science*, **1992**, *256*, 665.
- [31] Verkhivker, G.M.; Bouzida, D.; Gehlhaar, D.K.; Rejto, P.A.; Freer, S.T.; Rose, P.W. Complexity and simplicity of ligand-macromolecule interactions: the energy landscape perspective. *Curr. Opin. Struct. Biol.*, **2002**, *12*, 197-203.
- [32] Koshland, D.E. Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proc. Natl. Acad. Sci. USA* **1958**, *44*, (2), 98-104.
- [33] Ma, B.; Kumar, S.; Tsai, C.J.; Nussinov, R. Folding funnels and binding mechanisms *Protein Eng.*, **1999**, *12*, (9), 713-720.
- [34] Teodoro, M.L.; Kavraki, L.E. Conformational flexibility models for the receptor in structure based drug design *Current Pharmaceutical Design*, **2003**, *9*, (20), 1635-1648.
- [35] Teodoro, M.L.; Phillips, G.N.; Kavraki, L.E. Understanding protein flexibility through dimensionality reduction *Journal of Computational Biology*, **2003**, *10*, (3-4), 617-634.
- [36] Apostolakis, J.; Pluckthun, A.; Caflisch, A. Docking small ligands in flexible binding sites *Journal of Computational Chemistry*, **1998**, *19*, (1), 21-37.
- [37] Anderson, A.C.; O'Neil, R.H.; Surti, T.S.; Stroud, R.M. Approaches to solving the rigid receptor problem by identifying a minimal set of flexible residues during ligand docking *Chemistry & Biology*, **2001**, *8*, (5), 445-457.
- [38] Leach, A.R. Ligand Docking to Proteins with Discrete Side-Chain Flexibility *Journal of Molecular Biology*, **1994**, *235*, (1), 345-356.
- [39] Desmet, J.; Demaeyer, M.; Hazes, B.; Lasters, I. The Dead-End Elimination Theorem and Its Use in Protein Side-Chain Positioning *Nature*, **1992**, *356*, (6369), 539-542.
- [40] Hart, P.E.; Nilson, N.J.; Raphael, B. *IEET T. Syst. Sci. Cyb.*, **1968**, *4*, 100.
- [41] Schaffer, L.; Verkhivker, G.M. Predicting structural effects in HIV-1 protease mutant complexes with flexible ligand docking and protein side-chain optimization *Proteins-Structure Function and Genetics*, **1998**, *33*, (2), 295-310.
- [42] Morris, G.M.; Goodsell, D.S.; Halliday, R.S.; Huey, R.; Hart, W.E.; Belew, R.K.; Olson, A.J. Automated Docking Using a Lamarckian Genetic Algorithm and an Empirical Binding Free Energy Function *J. Comp. Chem.*, **1998**, *19*, 1639-1662.
- [43] Osterberg, F.; Morris, G.M.; Sanner, M.F.; Olson, A.J.; Goodsell, D.S. Automated docking to multiple target structures: Incorporation of protein mobility and structural water heterogeneity in

- AutoDock *Proteins-Structure Function and Bioinformatics*, **2002**, *46*, (1), 34-40.
- [44] Jones, G.; Willett, P.; Glen, R.C.; Leach, A.R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking *Journal of Molecular Biology*, **1997**, *267*, (3), 727-748.
- [45] Verdonk, M.L.; Cole, J.C.; Hartshorn, M.J.; Murray, C.W.; Taylor, R.D. Improved protein-ligand docking using GOLD *Proteins-Structure Function and Genetics*, **2003**, *52*, (4), 609-623.
- [46] Zavodszky, M.I.; Kuhn, L.A. Side-chain flexibility in protein-ligand binding: The minimal rotation hypothesis *Protein Science*, **2005**, *14*, (4), 1104-1114.
- [47] Schnecke, V.; Swanson, C.A.; Getzoff, E.D.; Tainer, J.A.; Kuhn, L.A. Screening a peptidyl database for potential ligands to proteins with side-chain flexibility *Proteins-Structure Function and Genetics*, **1998**, *33*, (1), 74-87.
- [48] Davis, I.W.; Baker, D. ROSETTALIGAND Docking with Full Ligand and Receptor Flexibility *J. Mol. Biol.*, **2009**, *385*, (2), 381-392.
- [49] Kuhlman, B.; Dantas, G.; Ireton, G.C.; Varani, G.; Stoddard, B.L.; Baker, D. Design of a novel globular protein fold with atomic-level accuracy *Science*, **2003**, *302*, (5649), 1364-1368.
- [50] Bottegoni, G.; Kufareva, I.; Totrov, M.; Abagyan, R. A new method for ligand docking to flexible receptors by dual alanine scanning and refinement (SCARE) *Journal of Computer-Aided Molecular Design*, **2008**, *22*, (5), 311-325.
- [51] Claussen, H.; Buning, C.; Rarey, M.; Lengauer, T. FlexE: Efficient molecular docking considering protein structure variations *Journal of Molecular Biology*, **2001**, *308*, (2), 377-395.
- [52] Polgar, T.; Keseru, G.M. Ensemble docking into flexible active sites. Critical evaluation of FlexE against JNK-3 and beta-secretase *Journal of Chemical Information and Modeling*, **2006**, *46*, (4), 1795-1805.
- [53] Corbeil, C.R.; Englebienne, P.; Moitessier, N. Docking ligands into flexible and solvated macromolecules. 1. Development and validation of FITTED 1.0 *Journal of Chemical Information and Modeling*, **2007**, *47*, (2), 435-449.
- [54] Zavodszky, M.I.; Ming, L.; Thorpe, M.F.; Day, A.R.; Kuhn, L.A. Modeling correlated main-chain motions in proteins for flexible molecular recognition *Proteins-Structure Function and Bioinformatics*, **2004**, *57*, (2), 243-261.
- [55] Zhao, Y.; Sanner, M.F. FLIPDock: Docking flexible ligands into flexible receptors *Proteins-Structure Function and Bioinformatics*, **2007**, *68*, (3), 726-737.
- [56] Wei, B.Q.; Weaver, L.H.; Ferrari, A.M.; Matthews, B.W.; Shoichet, B.K. Testing a flexible-receptor docking algorithm in a model binding site *Journal of Molecular Biology*, **2004**, *337*, (5), 1161-1182.
- [57] Lorber, D.M.; Udo, M.K.; Shoichet, B.K. Protein-protein docking with multiple residue conformations and residue substitutions *Protein Science*, **2002**, *11*, (6), 1393-1408.
- [58] Su, A.I.; Lorber, D.M.; Weston, G.S.; Baase, W.A.; Matthews, B.W.; Shoichet, B.K. Docking molecules by families to increase the diversity of hits in database screens: Computational strategy and experimental evaluation *Proteins-Structure Function and Genetics*, **2001**, *42*, (2), 279-293.
- [59] Lorber, D.M.; Shoichet, B.K. Flexible ligand docking using conformational ensembles *Protein Science*, **1998**, *7*, (4), 938-950.
- [60] Meng, E.C.; Shoichet, B.K.; Kuntz, I.D. Automated Docking with Grid-Based Energy Evaluation *Journal of Computational Chemistry*, **1992**, *13*, (4), 505-524.
- [61] Sherman, W.; Day, T.; Jacobson, M.P.; Friesner, R.A.; Farid, R. Novel procedure for modeling ligand/receptor induced fit effects *Journal of Medicinal Chemistry*, **2006**, *49*, (2), 534-553.
- [62] Friesner, R.A.; Banks, J.L.; Murphy, R.B.; Halgren, T.A.; Klicic, J.J.; Mainz, D.T.; Repasky, M.P.; Knoll, E.H.; Shelley, M.; Perry, J.K.; Shaw, D.E.; Francis, P.; Shenkin, P.S. Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy *Journal of Medicinal Chemistry*, **2004**, *47*, (7), 1739-1749.

- [63] Jacobson, M.P.; Kaminski, G.A.; Friesner, R.A.; Rapp, C.S. Force field validation using protein side chain prediction *Journal of Physical Chemistry B*, **2002**, *106*, (44), 11673-11680.
- [64] Jacobson, M.P.; Friesner, R.A.; Xiang, Z.X.; Honig, B. On the role of the crystal environment in determining protein side-chain conformations *Journal of Molecular Biology*, **2002**, *320*, (3), 597-608.
- [65] Zentgraf, M.; Fokkens, J.; Sotriffer, C.A. Addressing protein flexibility and ligand selectivity by "in situ cross-docking" *Chemmedchem*, **2006**, *1*, (12), 1355-+.
- [66] Sotriffer, C.A.; Dramburg, I. "In situ cross-docking" to simultaneously address multiple targets *Journal of Medicinal Chemistry*, **2005**, *48*, (9), 3122-3125.
- [67] Knegtel, R.M.A.; Kuntz, I.D.; Oshiro, C.M. Molecular docking to ensembles of protein structures *Journal of Molecular Biology*, **1997**, *266*, (2), 424-440.
- [68] Klepeis, J.L.; Lindorff-Larsen, K.; Dror, R.O.; Shaw, D.E. Long-timescale molecular dynamics simulations of protein structure and function *Curr. Op. Struct. Biol.*, **2009**, *19*, (2), 120-127.
- [69] Shaw, D.E.; Deneroff, M.M.; Dror, R.O.; Kuskin, J.S.; Larson, R.H.; Salmon, J.K.; Young, C.; Batson, B.; Bowers, K.J.; Chao, J.C.; Eastwood, M.P.; Gagliardo, J.; Grossman, J.P.; Ho, C.R.; Ierardi, D.J.; Kolossvary, I.; Klepeis, J.L.; Layman, T.; McLeavey, C.; Moraes, M.A.; Mueller, R.; Priest, E.C.; Shan, Y.B.; Spengler, J.; Theobald, M.; Towles, B.; Wang, S.C. Anton, a special-purpose machine for molecular dynamics simulation *Communications of the Acm*, **2008**, *51*, (7), 91-97.
- [70] Meli, M.; Pennati, M.; Curto, M.; Daidone, M.G.; Plescia, J.; Toba, S.; Altieri, D.C.; Zaffaroni, N.; Colombo, G. Small-Molecule Targeting of Heat Shock Protein 90 Chaperone Function: Rational Identification of a New Anticancer Lead *J. Med. Chem.*, **2006**, *49*, (26), 7721-7730.
- [71] Sugita, Y.; Okamoto, Y. Replica-exchange multicanonical algorithm and multicanonical replica-exchange method for simulating systems with rough energy landscape. *Chem. Phys. Lett.*, **2000**, *329*, 261-270.
- [72] Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics method for protein folding *Chemical Physics Letters*, **1999**, *314*, (1-2), 141-151.
- [73] Cavasotto, C.N.; Kovacs, J.A.; Abagyan, R.A. Representing receptor flexibility in ligand docking through relevant normal modes *Journal of the American Chemical Society*, **2005**, *127*, (26), 9632-9640.
- [74] Pang, Y.P.; Kozikowski, A.P. Prediction of the Binding-Sites of Huperzine-a in Acetylcholinesterase by Docking Studies *Journal of Computer-Aided Molecular Design*, **1994**, *8*, (6), 669-681.
- [75] Lin, J.H.; Perryman, A.L.; Schames, J.R.; McCammon, J.A. The relaxed complex method: Accommodating receptor flexibility for drug design with an improved scoring scheme *Biopolymers*, **2003**, *68*, (1), 47-62.
- [76] Lin, J.H.; Perryman, A.L.; Schames, J.R.; McCammon, J.A. Computational drug design accommodating receptor flexibility: The relaxed complex scheme *Journal of the American Chemical Society*, **2002**, *124*, (20), 5632-5633.
- [77] Kua, J.; Zhang, Y.K.; McCammon, J.A. Studying enzyme binding specificity in acetylcholinesterase using a combined molecular dynamics and multiple docking approach *Journal of the American Chemical Society*, **2002**, *124*, (28), 8260-8267.
- [78] Broughton, H.B. A method for including protein flexibility in protein-ligand docking: Improving tools for database mining and virtual screening *Journal of Molecular Graphics & Modelling*, **2000**, *18*, (3), 247-+.
- [79] Lerner, M.G.; Bowman, A.L.; Carlson, H.A. Incorporating dynamics in E. coli dihydrofolate reductase enhances structure-based drug discovery *Journal of Chemical Information and Modeling*, **2007**, *47*, (6), 2358-2365.
- [80] Lerner, M.G.; Meagher, K.L.; Carlson, H.A. Automated clustering of probe molecules from

solvent mapping of protein surfaces: new algorithms applied to hot-spot mapping and structure-based drug design *J. Comp. Aided Mol. Des.*, **2008**, 22, (10), 727-736.

[81] Bowman, A.L.; Nikolovska-Coleska, Z.; Zhong, H.Z.; Wang, S.M.; Carlson, H.A. Small molecule inhibitors of the MDM2-p53 interaction discovered by ensemble-based receptor models *J. Am. Chem. Soc.*, **2007**, 129, (42), 12809-12814.

[82] Carlson, H.A.; Masukawa, K.M.; Rubins, K.; Bushman, F.D.; Jorgensen, W.L.; Lins, R.D.; Briggs, J.M.; McCammon, J.A. Developing a dynamic pharmacophore model for HIV-1 integrase *J. Med. Chem.*, **2000**, 43, (11), 2100-2114.

[83] Carlson, H.A.; Smith, R.D.; Khazanov, N.A.; Kirchoff, P.D.; Dunbar, J.B.; Benson, M.L. Differences between High- and Low-Affinity Complexes of Enzymes and Nonenzymes *Journal of Medicinal Chemistry*, **2008**, 51, (20), 6432-6441.

[84] Bowman, A.L.; Nikolovska-Coleska, Z.; Zhong, H.Z.; Wang, S.M.; Carlson, H.A. Small molecule inhibitors of the MDM2-p53 interaction discovered by ensemble-based receptor models *Journal of the American Chemical Society*, **2007**, 129, (42), 12809-12814.

[85] Damm, K.L.; Carlson, H.A. Exploring experimental sources of multiple protein conformations in structure-based drug design *Journal of the American Chemical Society*, **2007**, 129, (26), 8225-8235.

[86] Bowman, A.L.; Lerner, M.G.; Carlson, H.A. Protein flexibility and species specificity in structure-based drug discovery: Dihydrofolate reductase as a test system *Journal of the American Chemical Society*, **2007**, 129, (12), 3634-3640.

[87] Meagher, K.L.; Lerner, M.G.; Carlson, H.A. Refining the multiple protein structure pharmacophore method: Consistency across three independent HIV-1 protease models *Journal of Medicinal Chemistry*, **2006**, 49, (12), 3478-3484.

[88] Meagher, K.L.; Carlson, H.A. Incorporating protein flexibility in structure-based drug discovery: Using HIV-1 protease as a test case *Journal of the American Chemical Society*, **2004**, 126, (41), 13276-13281.

[89] Carlson, H.A. Protein flexibility and drug design: how to hit a moving target *Current Opinion in Chemical Biology*, **2002**, 6, (4), 447-452.

[90] Carlson, H.A. Protein flexibility is an important component of structure-based drug discovery *Current Pharmaceutical Design*, **2002**, 8, (17), 1571-1578.

[91] Carlson, H.A.; Masukawa, K.M.; Rubins, K.; Bushman, F.D.; Jorgensen, W.L.; Lins, R.D.; Briggs, J.M.; McCammon, J.A. Developing a dynamic pharmacophore model for HIV-1 integrase *Journal of Medicinal Chemistry*, **2000**, 43, (11), 2100-2114.

[92] Carlson, H.A.; McCammon, J.A. Accommodating protein flexibility in computational drug design *Molecular Pharmacology*, **2000**, 57, (2), 213-218.

[93] Plescia, J.; Salz, W.; Xia, F.; Pennati, M.; Zaffaroni, N.; Daidone, M.G.; Meli, M.; Dohi, T.; Fortugno, P.; Nefedova, Y.; Gabrilovich, D.I.; Colombo, G.; Altieri, D.C. Rational design of Shepherdin, a novel anticancer agent *Cancer Cell*, **2005**, 7, 457-467.

[94] Esteras-Chopo, A.; Morra, G.; Moroni, E.; Serrano, L.; de la Paz, M.L.; Colombo, G. A Molecular Dynamics Study of the Interaction of D-Peptide Amyloid Inhibitors with Their Target Sequence Reveals a Potential Inhibitory Pharmacophore Conformation *Journal of Molecular Biology*, **2008**, 383, (1), 266-280.

[95] Mangoni, R.; Roccatano, D.; Di Nola, A. Docking of flexible ligands to flexible receptors in solution by molecular dynamics simulation *Proteins-Structure Function and Bioinformatics*, **1999**, 35, (2), 153-162.

[96] Pak, Y.S.; Wang, S.M. Application of a molecular dynamics simulation method with a generalized effective potential to the flexible molecular docking problems *Journal of Physical Chemistry B*, **2000**, 104, (2), 354-359.

[97] Berg, B.A.; Neuhaus, T. Multicanonical Ensemble - a New Approach to Simulate 1st-Order

- Phase-Transitions *Physical Review Letters*, **1992**, *68*, (1), 9-12.
- [98] Nakajima, N.; Higo, J.; Kidera, A.; Nakamura, H. Flexible docking of a ligand peptide to a receptor protein by multicanonical molecular dynamics simulation *Chemical Physics Letters*, **1997**, *278*, (4-6), 297-301.
- [99] Nakajima, N.; Nakamura, H.; Kidera, A. Multicanonical ensemble generated by molecular dynamics simulation for enhanced conformational sampling of peptides *Journal of Physical Chemistry B*, **1997**, *101*, (5), 817-824.
- [100] Trosset, J.Y.; Scheraga, H.A. PRODOCK: Software package for protein modeling and docking *Journal of Computational Chemistry*, **1999**, *20*, (4), 412-427.
- [101] Trosset, J.Y.; Scheraga, H.A. Flexible docking simulations: Scaled collective variable Monte Carlo minimization approach using Bezier splines, and comparison with a standard Monte Carlo algorithm *Journal of Computational Chemistry*, **1999**, *20*, (2), 244-252.
- [102] Zacharias, M.; Sklenar, H. Harmonic modes as variables to approximately account for receptor flexibility in ligand-receptor docking simulations: Application to DNA minor groove ligand complex *Journal of Computational Chemistry*, **1999**, *20*, (3), 287-300.
- [103] Keseru, G.M.; Kolossvary, I. Fully flexible low-mode docking: Application to induced fit in HIV integrase *Journal of the American Chemical Society*, **2001**, *123*, (50), 12708-12709.
- [104] Keseru, G.M. A virtual high throughput screen for high affinity cytochrome P450cam substrates. Implications for in silico prediction of drug metabolism *Journal of Computer-Aided Molecular Design*, **2001**, *15*, (7), 649-657.
- [105] Tatsumi, R.; Fukunishi, Y.; Nakamura, H. A hybrid method of molecular dynamics and harmonic dynamics for docking of flexible ligand to flexible receptor *Journal of Computational Chemistry*, **2004**, *25*, (16), 1995-2005.
- [106] Machicado, C.; Lopez-Llano, J.; Cuesta-Lopez, S.; Bueno, M.; Sancho, J. Design of ligand binding to an engineered protein cavity using virtual screening and thermal up-shift evaluation *Journal of Computer-Aided Molecular Design*, **2005**, *19*, (6), 421-443.
- [107] Friesner, R.A.; Murphy, R.B.; Repasky, M.P.; Frye, L.L.; Greenwood, J.R.; Halgren, T.A.; Sanschagrin, P.C.; Mainz, D.T. Extra precision glide: Docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes *Journal of Medicinal Chemistry*, **2006**, *49*, (21), 6177-6196.
- [108] Lee, M.R.; Sun, Y.X. Improving docking accuracy through molecular mechanics generalized born optimization and scoring *Journal of Chemical Theory and Computation*, **2007**, *3*, (3), 1106-1119.
- [109] Ewing, T.J.A.; Makino, S.; Skillman, A.G.; Kuntz, I.D. DOCK 4.0: Search strategies for automated molecular docking of flexible molecule databases *Journal of Computer-Aided Molecular Design*, **2001**, *15*, (5), 411-428.
- [110] Wang, J.M.; Kang, X.S.; Kuntz, I.D.; Kollman, P.A. Hierarchical database screenings for HIV-1 reverse transcriptase using a pharmacophore model, rigid docking, solvation docking, and MM-PB/SA *Journal of Medicinal Chemistry*, **2005**, *48*, (7), 2432-2444.
- [111] Graves, A.P.; Shivakumar, D.M.; Boyce, S.E.; Jacobson, M.P.; Case, D.A.; Shoichet, B.K. Rescoring docking hit lists for model cavity sites: Predictions and experimental testing *Journal of Molecular Biology*, **2008**, *377*, (3), 914-934.
- [112] Oikonomakos, N.G.; Skamnaki, V.T.; Tsitsanou, K.E.; Gavalas, N.G.; Johnson, L.N. A new allosteric site in glycogen phosphorylase b as a target for drug interactions *Structure Fold. Des.*, **2000**, *8*, 575-584.
- [113] Dennis, M.S.; Eigenbrot, C.; Skelton, N.J.; Ultsch, M.H.; Santell, L.; Dwyer, M.A.; O'Connell, M.P.; Lazarus, R.A. Peptide exosite inhibitors of factor VIIa as anticoagulants. *Nature*, **2000**, *404*, 465-470.
- [114] Hardy, J.A.; Lam, J.; Nguyen, J.T.; O'Brien, T.; Wells, J. Discovery of an allosteric site in

- caspases *Proc. Natl. Acad. Sci. USA*, **2004**, *101*, 12461-12466.
- [115] Suel, G.M.; Lockless, S.W.; Wall, M.A.; Ranganathan, R. Evolutionary conserved networks of residues mediate allosteric communication in proteins. *Nat. Struct. Biol.*, **2003**, *10*, 59-69.
- [116] Shulman, A.I.; Larson, C.; Magelsdorf, D.J.; Ranganathan, R. Structural determinants of allosteric ligand activation in RXR heterodimers. *Cell*, **2004**, *116*, 417-429.
- [117] Fuentes, E.J.; Der, C.J.; Lee, A.L. Ligand dependent dynamics of intramolecular signaling in a PDZ domain *J. Mol. Biol.*, **2004**, *335*, 1105-1115.
- [118] Ota, N.; Agard, D.A. Intramolecular signaling pathways revealed by modeling anisotropic thermal diffusion. *J. Mol. Biol.*, **2005**, *351*, (2), 345-354.
- [119] Chen, J.; Dima, R.I.; Thirumalai, D. Allosteric communication in dihydrofolate reductase: Signaling network and pathways for closed to occluded transition and back *J. Mol. Biol.*, **2007**, *374*, (1), 250-266.
- [120] Chennubhotla, C.; Yang, Z.; Bahar, I. Coupling between global dynamics and signal transduction pathways: a mechanism of allostery for chaperonin GroEL. *Mol. BioSyst.*, **2008**, *4*, 287-292.
- [121] Morra, G.; Verkhivker, G.M.; Colombo, G. Modeling signal propagation mechanisms and ligand-based conformational dynamics of the Hsp90 molecular chaperone full length dimer. *PLOS Comp. Biol.*, **2009**, In Press.
- [122] Ghosh, A.; Vishveshwara, S. A study of communication pathways in methionyl-tRNA synthetase by molecular dynamics simulations and structure network analysis. *Proc. Natl. Acad. Sci. USA*, **2007**, *104*, (40), 15711-15716.
- [123] Ghosh, A.; Vishveshwara, S. Variations in Clique and Community Patterns in Protein Structures during Allosteric Communication: Investigation of Dynamically Equilibrated Structures of Methionyl-tRNA Synthetase Complexes *Biochemistry*, **2008**, *2008*, (47), 11398-11407.
- [124] Liu, J.; Nussinov, R. Allosteric effects in the marginally stable von Hippel-Lindau tumor suppressor protein and allostery-based rescue mutant design *Proc. Natl. Acad. Sci. USA*, **2008**, *105*, (3), 901-906.
- [125] Rendic, S. Summary of information on human CYP enzymes: Human P450 metabolism data *Drug Metabolism Reviews*, **2002**, *34*, (1-2), 83-448.
- [126] Testa, B.; Kramer, S.D. The biochemistry of drug metabolism - An introduction - Part 2. Redox reactions and their enzymes *Chemistry & Biodiversity*, **2007**, *4*, (3), 257-405.
- [127] Boobis, A.; Gundert-Remy, U.; Kremers, P.; Macheras, P.; Pelkonen, O. In silico prediction of ADME and pharmacokinetics - Report of an expert meeting organised by COST B15 *European Journal of Pharmaceutical Sciences*, **2002**, *17*, (4-5), 183-193.
- [128] Cupp-Vickery, J.R.; Poulos, T.L. Structure of cytochrome P450eryF involved in erythromycin biosynthesis *Nature Structural Biology*, **1995**, *2*, (2), 144-153.
- [129] Nagano, S.; Poulos, T.L. Crystallographic study on the dioxygen complex of wild-type and mutant cytochrome P450cam - Implications for the dioxygen activation mechanism *Journal of Biological Chemistry*, **2005**, *280*, (36), 31659-31663.
- [130] Poulos, T.L.; Finzel, B.C.; Gunsalus, I.C.; Wagner, G.C.; Kraut, J. The 2.6-Å crystal structure of *Pseudomonas putida* cytochrome P-450 *Journal of Biological Chemistry*, **1985**, *260*, (30), 6122-6130.
- [131] Hasemann, C.A.; Ravichandran, K.G.; Peterson, J.A.; Deisenhofer, J. Crystal structure and refinement of cytochrome P450terp at 2.3 Å resolution *Journal of Molecular Biology*, **1994**, *236*, (4), 1169-1185.
- [132] Ravichandran, K.G.; Boddupalli, S.S.; Hasemann, C.A.; Peterson, J.A.; Deisenhofer, J. Crystal structure of hemoprotein domain of P450BM-3, a prototype for microsomal P450's *Science*, **1993**, *261*,

(5122), 731-736.

- [133] De Groot, M.J.; Kirton, S.B.; Sutcliffe, M.J. In silico methods for predicting ligand binding determinants of cytochromes P450 *Current Topics in Medicinal Chemistry*, **2004**, *4*, (16), 1803-1824.
- [134] Lewis, D.F.V. Molecular modeling of human cytochrome P450-substrate interactions *Drug Metabolism Reviews*, **2002**, *34*, (1-2), 55-67.
- [135] de Graaf, C.; Vermeulen, N.P.E.; Feenstra, K.A. Cytochrome P450 in silico: An integrative modeling approach *Journal of Medicinal Chemistry*, **2005**, *48*, (8), 2725-2755.
- [136] Williams, P.A.; Cosme, J.; Sridhar, V.; Johnson, E.F.; McRee, D.E. Microsomal cytochrome P4502C5: comparison to microbial P450s and unique features *Journal of Inorganic Biochemistry*, **2000**, *81*, (3), 183-190.
- [137] Sansen, S.; Yano, J.K.; Reynald, R.L.; Schoch, G.A.; Griffin, K.J.; Stout, C.D.; Johnson, E.F. Adaptations for the oxidation of polycyclic aromatic hydrocarbons exhibited by the structure of human P450 1A2 *Journal of Biological Chemistry*, **2007**, *282*, (19), 14348-14355.
- [138] Yano, J.K.; Hsu, M.H.; Griffin, K.J.; Stout, C.D.; Johnson, E.F. Structures of human microsomal cytochrome P450 2A6 complexed with coumarin and methoxsalen *Nature Structural & Molecular Biology*, **2005**, *12*, (9), 822-823.
- [139] Smith, B.D.; Sanders, J.L.; Porubsky, P.R.; Lushington, G.H.; Stout, C.D.; Scott, E.E. Structure of the human lung cytochrome P450 2A13 *Journal of Biological Chemistry*, **2007**, *282*, (23), 17306-17313.
- [140] Schoch, G.A.; Yano, J.K.; Wester, M.R.; Griffin, K.J.; Stout, C.D.; Johnson, E.F. Structure of human microsomal cytochrome P4502C8 - Evidence for a peripheral fatty acid binding site *Journal of Biological Chemistry*, **2004**, *279*, (10), 9497-9503.
- [141] Williams, P.A.; Cosme, J.; Ward, A.; Angova, H.C.; Vinkovic, D.M.; Jhoti, H. Crystal structure of human cytochrome P4502C9 with bound warfarin *Nature*, **2003**, *424*, (6947), 464-468.
- [142] Rowland, P.; Blaney, F.E.; Smyth, M.G.; Jones, J.J.; Leydon, V.R.; Oxbrow, A.K.; Lewis, C.J.; Tennant, M.G.; Modi, S.; Eggleston, D.S.; Chenery, R.J.; Bridges, A.M. Crystal structure of human cytochrome P450 2D6 *Journal of Biological Chemistry*, **2006**, *281*, (11), 7614-7622.
- [143] Williams, P.A.; Cosme, J.; Vinkovic, D.M.; Ward, A.; Angove, H.C.; Day, P.J.; Vonnrhein, C.; Tickle, I.J.; Jhoti, H. Crystal structures of human cytochrome P450 3A4 bound to metyrapone and progesterone *Science*, **2004**, *305*, (5684), 683-686.
- [144] Danielson, P.B. The cytochrome P450 superfamily: Biochemistry, evolution and drug metabolism in humans *Current Drug Metabolism*, **2002**, *3*, (6), 561-597.
- [145] Otyepka, M.; Skopalik, J.; Anzenbacherova, E.; Anzenbacher, P. What common structural features and variations of mammalian P450s are known to date? *Biochimica et Biophysica Acta-General Subjects*, **2007**, *1770*, (3), 376-389.
- [146] Wester, M.R.; Johnson, E.F.; Marques-Soares, C.; Dijols, S.; Dansette, P.M.; Mansuy, D.; Stout, C.D. Structure of mammalian cytochrome P450 2C5 complexed with diclofenac at 2.1 Å resolution: Evidence for an induced fit model of substrate binding *Biochemistry*, **2003**, *42*, (31), 9335-9345.
- [147] Ekroos, M.; Sjogren, T. Structural basis for ligand promiscuity in cytochrome P450 3A4 *Proceedings of the National Academy of Sciences of the United States of America*, **2006**, *103*, (37), 13682-13687.
- [148] Cojocar, V.; Winn, P.J.; Wade, R.C. The ins and outs of cytochrome P450s *Biochimica et Biophysica Acta-General Subjects*, **2007**, *1770*, (3), 390-401.
- [149] Wrighton, S.A.; Schuetz, E.G.; Thummel, K.E.; Shen, D.D.; Korzekwa, K.R.; Watkins, P.B. The human CYP3A subfamily: Practical considerations *Drug Metabolism Reviews*, **2000**, *32*, (3-4), 339-361.
- [150] Park, H.; Lee, S.; Suh, J. Structural and dynamical basis of broad substrate specificity, catalytic

- mechanism, and inhibition of cytochrome P450 3A4 *JOURNAL OF THE AMERICAN CHEMICAL SOCIETY*, **2005**, *127*, (39), 13634-13642.
- [151] Seifert, A.; Tatzel, S.; Schmid, R.D.; Pleiss, J. Multiple molecular dynamics simulations of human P450 monooxygenase CYP2C9: The molecular basis of substrate binding and regioselectivity toward warfarin *Proteins-Structure Function and Bioinformatics*, **2006**, *64*, (1), 147-155.
- [152] Liu, X.L.; Wang, X.C.; Jiang, H.L. A steered molecular dynamics method with direction optimization and its applications on ligand molecule dissociation *Journal of Biochemical and Biophysical Methods*, **2008**, *70*, (6), 857-864.
- [153] Ludemann, S.K.; Lounnas, V.; Wade, R.C. How do substrates enter and products exit the buried active site of cytochrome P450cam? 1. Random expulsion molecular dynamics investigation of ligand access channels and mechanisms *Journal of Molecular Biology*, **2000**, *303*, (5), 797-811.
- [154] Winn, P.J.; Ludemann, S.K.; Gauges, R.; Lounnas, V.; Wade, R.C. Comparison of the dynamics of substrate access channels in three cytochrome P450s reveals different opening mechanisms and a novel functional role for a buried arginine *Proceedings of the National Academy of Sciences of the United States of America*, **2002**, *99*, (8), 5361-5366.
- [155] Wade, R.C.; Winn, P.J.; Schlichting, E.; Sudarko. A survey of active site access channels in cytochromes P450 *Journal of Inorganic Biochemistry*, **2004**, *98*, (7), 1175-1182.
- [156] Schleinkofer, K.; Sudarko; Winn, P.J.; Ludemann, S.K.; Wade, R.C. Do mammalian cytochrome P450s show multiple ligand access pathways and ligand channelling? *Embo Reports*, **2005**, *6*, (6), 584-589.
- [157] Li, W.H.; Liu, H.; Scott, E.E.; Grater, F.; Halpert, J.R.; Luo, X.M.; Shen, J.H.; Jiang, H.L. Possible pathway(s) of testosterone egress from the active site of cytochrome P4502B1: A steered molecular dynamics simulation *Drug Metabolism and Disposition*, **2005**, *33*, (7), 910-919.
- [158] Scott, E.E.; Liu, H.; He, Y.Q.; Li, W.H.; Halpert, J.R. Mutagenesis and molecular dynamics suggest structural and functional roles for residues in the N-terminal portion of the cytochrome P4502B1 helix *Archives of Biochemistry and Biophysics*, **2004**, *423*, (2), 266-276.
- [159] Kemp, C.A.; Flanagan, J.U.; van Eldik, A.J.; Marechal, J.D.; Wolf, C.R.; Roberts, G.C.K.; Paine, M.J.I.; Sutcliffe, M.J. Validation of model of cytochrome P450 2D6: An in silico tool for predicting metabolism and inhibition *Journal of Medicinal Chemistry*, **2004**, *47*, (22), 5340-5346.
- [160] Aqvist, J.; Medina, C.; Samuelsson, J.E. A new method for predicting binding affinity in computer-aided drug design *Protein Engineering*, **1994**, *7*, (3), 385-391.
- [161] Szklarz, G.D.; Paulsen, M.D. Molecular modeling of cytochrome P450 1A1: Enzyme-substrate interactions and substrate binding affinities *Journal of Biomolecular Structure & Dynamics*, **2002**, *20*, (2), 155-162.
- [162] Beveridge, D.L.; Dicapua, F.M. Free energy via molecular simulation: applications to chemical and biomolecular systems *Annual Review of Biophysics and Biophysical Chemistry*, **1989**, *18*, 431-492.
- [163] de Graaf, C.; Oostenbrink, C.; Keizers, P.H.J.; Vugt-Lussenburg, B.M.A.; Commandeur, J.N.M.; Vermeulen, N.P.E. Free energies of binding of R- and S-propranolol to wild-type and F483A mutant cytochrome P450 2D6 from molecular dynamics simulations *European Biophysics Journal with Biophysics Letters*, **2007**, *36*, (6), 589-599.
- [164] Kollman, P.A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.H.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D.A.; Cheatham, T.E. Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models *Accounts of Chemical Research*, **2000**, *33*, (12), 889-897.
- [165] Harris, D.L.; Park, J.Y.; Gruenke, L.; Waskell, L. Theoretical study of the ligand-CYP2B4 complexes: Effect of structure on binding free energies and heme spin state *Proteins-Structure Function and Bioinformatics*, **2004**, *55*, (4), 895-914.
- [166] Jones, J.P.; Mysinger, M.; Korzekwa, K.R. Computational models for cytochrome P450: A

predictive electronic model for aromatic oxidation and hydrogen atom abstraction *Drug Metabolism and Disposition*, **2002**, *30*, (1), 7-12.

[167] Liu, J.G.; Ericksen, S.S.; Sivaneri, M.; Besspiata, D.; Fisher, C.W.; Szklarz, G.D. The effect of reciprocal active site mutations in human cytochromes P450 1A1 and 1A2 on alkoxyresorufin metabolism *Archives of Biochemistry and Biophysics*, **2004**, *424*, (1), 33-43.

[168] Ericksen, S.S.; Szklarz, G.D. Regiospecificity of human cytochrome P450 1A1-mediated oxidations: The role of steric effects *Journal of Biomolecular Structure & Dynamics*, **2005**, *23*, (3), 243-256.

[169] Keizers, P.H.J.; de Graaf, C.; de Kanter, F.J.J.; Oostenbrink, C.; Feenstra, K.A.; Commandeur, J.N.M.; Vermeulen, N.P.E. Metabolic regio- and stereoselectivity of cytochrome P450 2D6 towards 3,4-methylenedioxy-N-alkylamphetamines: in silico predictions and experimental validation *Journal of Medicinal Chemistry*, **2005**, *48*, (19), 6117-6127.

[170] Venhorst, J.; ter Laak, A.M.; Commandeur, J.N.M.; Funae, Y.; Hiroi, T.; Vermeulen, N.P.E. Homology modeling of rat and human cytochrome P450 2D (CYP2D) isoforms and computational rationalization of experimental ligand-binding specificities *Journal of Medicinal Chemistry*, **2003**, *46*, (1), 74-86.

[171] Park, J.Y.; Harris, D. Construction and assessment of models of CYP2E1: Predictions of metabolism from docking, molecular dynamics, and density functional theoretical calculations *Journal of Medicinal Chemistry*, **2003**, *46*, (9), 1645-1660.

[172] Kuhn, B.; Jacobsen, W.; Christians, U.; Benet, L.Z.; Kollman, P.A. Metabolism of sirolimus and its derivative everolimus by cytochrome P450 3A4: Insights from docking, molecular dynamics, and quantum chemical calculations *Journal of Medicinal Chemistry*, **2001**, *44*, (12), 2027-2034.

[173] Ueng, Y.F.; Kuwabara, T.; Chun, Y.J.; Guengerich, F.P. Cooperativity in oxidations catalyzed by cytochrome P450 3A4 *Biochemistry*, **1997**, *36*, (2), 370-381.

[174] Atkins, W.M. Non-Michaelis-Menten kinetics in cytochrome P450-catalyzed reactions *Annual Review of Pharmacology and Toxicology*, **2005**, *45*, 291-310.

[175] Dabrowski, M.J.; Schrag, M.L.; Wienkers, L.C.; Atkins, W.M. Pyrene-pyrene complexes at the active site of cytochrome P450 3A4: Evidence for a multiple substrate binding site *JOURNAL OF THE AMERICAN CHEMICAL SOCIETY*, **2002**, *124*, (40), 11866-11867.

[176] Schoch, G.A.; Yano, J.K.; Sansen, S.; Dansette, P.M.; Stout, C.D.; Johnson, E.F. Determinants of cytochrome P4502C8 substrate binding - Structures of complexes with montelukast, troglitazone, felodipine, and 9-cis-retinoic acid *Journal of Biological Chemistry*, **2008**, *283*, (25), 17227-17237.

[177] Haines, D.C.; Tomchick, D.R.; Machius, M.; Peterson, J.A. Pivotal role of water in the mechanism of P450BM-3 *Biochemistry*, **2001**, *40*, (45), 13456-13465.

[178] Ravindranathan, K.P.; Gallicchio, E.; Friesner, R.A.; McDermott, A.E.; Levy, R.M. Conformational equilibrium of cytochrome P450BM-3 complexed with N-palmitoylglycine: A replica exchange molecular dynamics study *JOURNAL OF THE AMERICAN CHEMICAL SOCIETY*, **2006**, *128*, (17), 5786-5791.

[179] Meunier, B.; de Visser, S.P.; Shaik, S. Mechanism of oxidation reactions catalyzed by cytochrome P450 enzymes *Chemical Reviews*, **2004**, *104*, (9), 3947-3980.

[180] Helms, V.; Wade, R.C. Thermodynamics of water mediating protein-ligand interactions in cytochrome P450cam: A molecular dynamics study *Biophysical Journal*, **1995**, *69*, (3), 810-824.

[181] Helms, V.; Wade, R.C. Hydration energy landscape of the active site cavity in cytochrome P450cam *Proteins-Structure Function and Bioinformatics*, **1998**, *32*, (3), 381-396.

[182] Rydberg, P.; Rod, T.H.; Olsen, L.; Ryde, U. Dynamics of water molecules in the active-site cavity of human cytochromes P450 *Journal of Physical Chemistry B*, **2007**, *111*, (19), 5445-5457.