Published in final edited form as:
*J Med Chem*. 2008 October 23; 51(20): 6237–6255. doi:10.1021/jm800562d.

# Target Flexibility:

## An Emerging Consideration in Drug Discovery and Design†

**Pietro Cozzini**[*,1,2], **Glen E. Kellogg**[*,3], **Francesca Spyrakis**[1,2], **Donald J. Abraham**[1], **Gabriele Costantino**[4], **Andrew Emerson**[5], **Francesca Fanelli**[6], **Holger Gohlke**[7], **Leslie A. Kuhn**[8], **Garrett M. Morris**[9], **Modesto Orozco**[10], **Thelma A. Pertinhez**[11], **Menico Rizzi**[12], and **Christoph Sotriffer**[13]

*Department of General and Inorganic Chemistry, University of Parma, Via G.P. Usberti 17/A 43100, Parma, Italy; National Institute for Biosystems and Biostructures, Rome, Italy; Department of Medicinal Chemistry and Institute for Structural Biology & Drug Discovery, Virginia Commonwealth University, Richmond, Virginia 23298-0540 USA; Department of Pharmaceutics, University of Parma, Via GP Usberti 27/A, 43100 Parma, Italy; High Performance Systems, CINECA Supercomputing Centre, Casalecchio di Reno, Bologna, Italy; Dulbecco Telethon Institute, Department of Chemistry, University of Modena and Reggio Emilia, Via Campi 183, 41100 Modena, Italy; Department of Mathematics and Natural Sciences, Pharmaceutical Institute, Christian-Albrechts-University, Gutenbergstr. 76, 24118 Kiel, Germany; Departments of Biochemistry & Molecular Biology, Computer Science & Engineering, and Physics & Astronomy, Michigan State University, East Lansing 48824-1319 USA; Department of Molecular Biology, MB-5, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037-1000 USA; Molecular Modeling and Bioinformatics Unit, Institute of Biomedical Research, Scientific Park of Barcelona, Department of Biochemistry and Molecular Biology, University of Barcelona, Josep Samitier 1-5, Barcelona 08028, Spain; Department of Experimental Medicine, University of Parma, Via Volturno, 39, 43100, Parma, Italy; Department of Chemical, Food, Pharmaceutical and Pharmacological Sciences, University of Piemonte Orientale "Amedeo Avogadro", Via Bovio 6, 28100 Novara, Italy; Institute of Pharmacy and Food Chemistry, University of Würzburg, Am Hubland, D-97074 Würzburg, Germany.*

1*Department of General and Inorganic Chemistry, University of Parma.*

2*National Institute for Biosystems and Biostructures (INBB).*

3*Virginia Commonwealth University.*

4*Department of Pharmaceutics, University of Parma.*

5*CINECA Supercomputing Centre.*

6*University of Modena and Reggio Emilia.*

7*Christian-Albrechts-University Kiel.*

8*Michigan State University.*

9*The Scripps Research Institute.*

10*University of Barcelona.*

---

†Consensus report of "From Structural Genomics to Drug Discovery: Modeling the Flexibility", September 20-21, 2007, Parma, Italy. International course and report were conceived by Pietro Cozzini and Glen E. Kellogg.

*To whom correspondence should be addressed. For G.E.K.: Department of Medicinal Chemistry, Virginia Commonwealth University, Box 980540, Richmond, VA 23298-0540; (phone) 804-828-6452; (fax) 804-827-3664; (e-mail) glen.kellogg@vcu.edu. For P.C.: Department of General and Inorganic Chemistry, University of Parma, Via G.P. Usberti 17/A 43100, Parma, Italy; (phone) +39-0521-905669; (fax) +39-0521-905556; (e-mail) pietro.cozzini@unipr.it.

11*Department of Experimental Medicine, University of Parma.*

12*University of Piemonte Orientale "Amedeo Avogadro"*

13*University of Würzburg.*

## Introduction

Structure-based drug discovery has played an important role in medicinal chemistry[1] beginning nearly when the first X-ray crystal structure of the myoglobin and hemoglobin proteins at near-atomic resolution were described by Perutz, Kendrew and colleagues.[2-5] Even though only static structures were (and still generally are) used for most Structure-Based Drug Design (SBDD), and indeed most molecular modeling, the importance of flexibility was recognized immediately: hemoglobin has two rather different structures, "tense" and "relaxed", depending on its oxygenation, although in recent years a family of relaxed hemoglobin structures with different tertiary structure conformations have been reported.[6] In fact, all proteins are inherently flexible systems. This flexibility is frequently essential for function (e.g., as in hemoglobin). Proteins have an intrinsic ability to undergo functionally relevant conformational transitions under native state conditions,[7,8] on a wide range of scales, both in time and space.[9] In adenylate kinase large conformational changes due to movements of the nucleotide 'lids' – rate-limiting for overall catalytic turnover[10,11] – are 'linked' with relatively small-amplitude atomic fluctuations on the ps timescale such that changes in the local backbone conformation are required for lid closure.[12] Nuclear receptors are modular proteins where a significant degree of conformational flexibility is essential to biological function. Most of the pharmacology of nuclear receptor ligands has been discussed on the basis of their ability to stabilize (or displace) a short α-helix segment (known as H12 or AF-2) localized at the carboxy terminus of the receptor in (or from) its conformation in the protein "active" form.[13-15] Available X-ray crystal structures show a surprisingly wide range of structural diversity in ligands binding to, and inhibiting, nuclear receptor proteins such as the farnesoid X-receptor (FXR).[16,17] Protein dynamics is also a key component of intramolecular and intermolecular communication/signaling mechanisms and an essential requirement for the function of G-protein coupled receptors (GPCRs), which are the largest known superfamily of membrane proteins. GPCRs regulate cell activity by transmitting extracellular signals to the inside of cells and respond to these signals by catalyzing nucleotide exchange in intracellular G-proteins.[18] Emerging evidence suggests that these receptors exist as homodimers, heterodimers and oligomers,[19] and act in multi-component units comprised of a variety of signaling and scaffolding molecules.[20] Thus, regulated protein-protein interactions are key features of GPCR function, and understanding these interactions in the dynamic cellular environment is a major research goal that may lead to new therapeutic approaches for this important target family.

In terms of medicinal chemistry and drug discovery, even the angiotensin-converting enzyme inhibitor captopril, credited as the first drug discovered using a protein binding site, binds to a protein — carboxypeptidase A — that was known to be highly flexible.[21] While most examples of successful modeling and computational drug design have been accomplished without *true* consideration of protein flexibility, this may be largely due to the fortunate result of relatively minor induced-fit adaptations of proteins upon ligand binding. Using flexibility as a criterion, we can classify three types of proteins: i) 'rigid' proteins, where ligand-induced changes are limited to relatively small side chain rearrangements, ii) flexible proteins, where relatively large movements around "hinge points" or at active site loops, with concomitant side chain motion, occur upon ligand binding, and iii) intrinsically unstable proteins, whose conformation is not defined until ligand binding. Currently, for technical reasons, the Protein Data Bank (PDB)[22] is artificially enriched in the first family, but genomic and proteomic projects have shown that the last two protein classes represent a very significant proportion of the proteome,[23] and probably include many important therapeutic targets. It is also noteworthy that the steadily increasing availability of experimentally-determined (X-ray/NMR/cryoEM) protein structures has not appeared to have resulted in a similar increase in the success rate of structure-based design approaches, although the significant time lag between availability of a structure and its exploitation may be masking the true success rate.

These facts clearly suggest that too much emphasis, or perhaps even hope, has been placed on rigid structures, regardless of their experimental or theoretical origin, and especially when dealing with proteins expected or known to undergo large conformational changes during their biological function. There are probably two main reasons that this short-sighted approach has evolved and continued: first, the impact of the static appearance of X-ray crystal structures (and their often beautiful images!) on the perception of protein structure, i.e., that the crystal structure is always the "correct" structure, and second, the conceptual (and technical) difficulties of dealing with moving targets. There is one extremely obvious cause of this rigid "bias" — to reduce disorder, and also to preserve precious crystals, crystallography data are now usually obtained at extremely low, non-biological, temperatures. Similarly, although proteins are somewhat solvated in their crystal lattice for crystallographic analysis, that is likely to be a flawed approximation of the true biological environment for many, if not most, proteins. In biological systems proteins express their functions in aqueous or semi-fluid environments and proteins in solution exist as an ensemble of energetically accessible conformations such that their three-dimensional structure is best described when **all** states are represented. For drug discovery, this is a radical paradigm shift from when captopril was invented, even if the active site flexibility of carboxypeptidase A *was* acknowledged at the time!

Thus, as protein conformational changes are initiated or stabilized by ligand binding and are essential to the protein's own function, the ability to measure or simulate dynamic changes taking place in proteins upon ligand binding is becoming a central issue in the design of bioactive compounds. The essence of the problem for drug discovery is that for a flexible target it is not known in advance which conformation the target will adopt in response to the binding of a particular ligand, or how to design such a ligand for an unknown conformation.

In the fall of 2007 the authors of this Perspective participated in a Course organized at the University of Parma (Italy) entitled "From Structural Genomics to Drug Discovery: Modeling the Flexibility" (www.course07.unipr.it). While we approach this problem from a wide variety of directions, both experimental and computational, there was a very strong agreement that the "Flexibility Era" for drug discovery was rapidly approaching and that there were a number of points of consensus among the authors of this contribution on how flexibility could be and should be exploited for drug discovery. While techniques such as fluorescence spectroscopy,[24] spin label electron paramagnetic resonance (EPR)[25] and Small Angle X-ray scattering[26] have significant applications in elucidating some aspects of protein flexibility, X-ray crystallography and Nuclear Magnetic Resonance (NMR) spectroscopy are the key technologies for characterizing molecular structure and will be the focus of this Perspective. As will be described below, these two tools have been evolving with respect to describing molecular flexibility. Molecular Dynamics (MD) and molecular docking have developed as separate sub-disciplines of computer-aided molecular design, but, at least for the purposes of drug discovery for flexible targets, these techniques and tools are complementary (see Figure 1). In this Perspective we will first describe some of the innovations in both experimental measurement and computational modeling of flexibility that we believe are going to impact strongly the future success of computer-aided molecular design. Thus, this Perspective will not exhaustively cover all methodological approaches that can be used to address the complex issues of protein dynamics, but instead reflects the themes of the course. We are excited about these prospects and offer some guidelines and forecasts for working under this new paradigm.

## Characterizing Flexibility in Biomacromolecules

High-resolution experimental techniques such as X-ray crystallography and NMR usually only provide snapshots of one or just some of the conformations that are accessible to proteins. However, synchrotron X-ray sources have recently opened up the possibility of time-resolved measurements on single crystals. With the superior resolution of these sources it is also possible to evaluate the electronic oscillations around single atoms and obtain the probability density for the atom. While NMR has a number of limitations, it does have the seemingly large advantage of being performed under conditions and in solutions that mimic the biological environment. Generally the results from NMR are ensembles of low energy conformations that satisfy the coupling energy constraints displayed in the multi-dimensional spectra. Importantly, as the field strength of NMR spectrometers increases, not only is the size range of proteins amenable for study increased, but the resolutions of the spectra are enhanced, leading to the identification of more conformers. Also, new and more sophisticated NMR pulse sequences have allowed the extraction of more detailed 3D structural data.

Molecular Dynamics simulation is at present the best means to obtain a more complete set of protein conformers, especially for those at relatively high energy and not detectable with the current set of experimental tools. While the starting conformations for MD are, by necessity, crystallographic, NMR or built computationally models, the resulting trajectories of conformations or atomic motion can be thought of as “movies” showing the somewhat random motion of a protein at a given temperature. Sampling of these conformations can provide a set of unbiased structures for analysis with docking, virtual screening or other computational approaches. However, the stochastic nature of Molecular Dynamics, along with the rather limited timescales of simulations (typically only a few tens of nanoseconds or less), and the presence of high energetic and entropic barriers in flexible molecules, can render many of the conformational models uninteresting or redundant because only a small subset of the available models are obtained.

On the other hand, docking and virtual screening has progressed from placing rigid ligands into rigid sites in its earliest incarnations,[27] to placing flexible ligands in rigid sites,[28-31] to the current state-of-the-art of placing flexible ligands in semi-flexible sites.[32-38] In all of these cases the initial site model usually has its origin in experimental structure data. The challenge for drug discovery, as in docking or virtual screening, is to model the receptor plasticity that enables binding partners to conformationally adapt to one another. First, one needs to understand what can move and how; second, this knowledge needs to be transformed into a useful and reliable docking algorithm. The predisposition of proteins to undergo functionally relevant conformational transitions provides a route to this requirement, as it implies the pre-existence of conformations even in the absence of triggering events like ligand binding. The ligand “selects” the proper conformation from the ensemble of rapidly interconverting species.[7-9] In this approach, computational investigations of conformational fluctuations of unbound receptors with MD should reveal conformational states adopted by the analogous bound receptor. Alternatively, the problem of modeling conformational changes can be simplified by focusing what has been learned recently about protein side-chain motions that occur upon natural ligand or inhibitor binding.[39-42] Thus, the active site can be rationally adapted to the incoming ligand, i.e., by “induced fit”, which may also take advantage of low-energy conformational changes. The opposite can also be true: consider that the “soaking” method of inserting ligands into pre-formed protein crystal lattices to form crystals of the complex, which works for some but certainly not all cases, suggests that static lattices may, in fact, select ligands.

## X-Ray Crystallography

X-ray crystallography is generally considered the gold standard technique in producing experimental structural models of biological macromolecules. In quantitative terms (i.e., the number of structures deposited at the Protein Data Bank) X-ray crystallography has been the most productive structure elucidation tool for biomacromolecules. Indeed, a highly detailed picture emerges from X-ray diffraction analysis of a crystal that contains an active biomacromolecule. The standard use of macromolecular crystallography results in a static-, time- and space-averaged structure that unfortunately only poorly (if at all) represents the *ensemble* of conformations of the protein in action. Part of this may be attributed to crystallization being a "purification" in that only molecules with conformations compatible with the growing crystal lattice will be "frozen". The accuracy of a static crystal structure is defined by the resolution to which X-ray data have been measured. Numerous experimental factors contribute to the observed resolution, some related to the size and quality of the crystal itself, others to the brightness of the X-ray source employed and to experimental conditions.

While X-ray diffraction is probing electron density, neutron diffraction (available from nuclear reactor or spallation sources) probes the atomic nuclei and can thus experimentally locate hydrogen positions because of inherently high resolution; the downside is that rather large crystals are required. Typically, a resolution of less than 2 Å is generally considered good crystallographic data for a protein structure, while a resolution of around 1 Å is termed "atomic". Another set of quality metrics for crystallography, the R-factor or free R-factor, are measures of how well the refined structure fits the observed data or, in other words, the percentage difference between the real, i.e., measured electron density and that of the refined model.[43] Finally, each atom in a reported crystal structure will have a B-factor (or temperature factor) that in its most basic sense represents the atom's individual uncertainty in position, whether due to thermal motion, occupancy, experimental and modeling artifacts or other effects.[44]

In recent years however, mainly due to the availability of powerful synchrotron X-ray sources, several new approaches have been developed that go beyond the classical static crystallographic analysis. Two major contributions can be attributed to these advances: i) time-resolved measurements either on a series of crystals or using Laue diffraction for fast data collection can detect transient chemical states in the crystal and ii) it is possible to more precisely evaluate the oscillations of an atom around its position, thereby providing information on the dynamic properties of a protein in a crystal lattice. The latter approach is based on the modeling of an atomic displacement parameter, i.e., the thermal parameter or B-factor described above, which can be thought of as a probability density function for the location of each atom in the protein. At low resolution, modeling of this parameter is restricted to a spherical or 'isotropic' shape but at atomic resolution, an ellipsoid or 'anisotropic' model can be applied.[45,46] This anisotropic model provides both the magnitudes and directions of movement of each atom and its inclusion in a model determined at high resolution allows a *dynamic* description of the protein structure. Thus, at atomic resolution, analysis of the anisotropic thermal parameters allows extraction of the direction of motion, possibly leading to better detection of both subtle changes as well as larger scale motions.[47 and references therein] Figure 2 illustrates the high resolution (1.05 Å) X-ray crystal structure for *Mycobacterium tuberculosis* FprA,[48] which was solved with anisotropic thermal parameters, and revealed an unexpected chemical conversion of NADP+ to NADPO.

The dynamics of a biological system or process can be re-constructed in the form of a "movie" prepared by assembling a series of static structures corresponding to various states along a reaction pathway. Such an approach has been employed to visualize transitions from one type of folding motif to another, conformational changes associated with different types of ligand-induced structural adaptations, alternate motions between well-defined distinct conformations, changes in quaternary structure, subtle side-chain movements in the interior or on the surface of a protein and other structural rearrangements.[49] By exploiting synchrotron radiation and a multi-wavelength data collection technique known as the Laue method, "kinetic crystallography" can be performed to obtain the necessary set of static structures on relevant reaction pathways. When biological turnover is initiated in the crystal with light or other radiation, the formed transient structural species can be seen together with the associated structural rearrangements.[50,51 and references therein] Although movements occurring on a time scale that is faster than the time required to determine the structure by X-ray crystallography cannot be detected, time-resolved crystallography, when experimentally feasible, can provide an important contribution in describing protein flexibility. There is an ever-increasing synergy between kinetic crystallography and *in crystallo* UV/visible absorption and fluorescence, fluorescence lifetime, and Raman spectroscopies, in combination with various evolving physical trapping (e.g., temperature) and chemical trapping (e.g., adjusting solvent, pH, etc. to manipulate concentrations of chemical intermediates) strategies. These and other combinations of methods will be increasingly employed in the future to provide a dynamic view of biological processes carried out by proteins, and to give insight on how to influence these processes with chemical agents. Indeed, a stunning example was published very recently on superoxide reductase, an iron containing enzyme that neutralizes the highly cytotoxic superoxide radical produced by oxygen metabolism.[52] By combining kinetic crystallography with Raman spectroscopy, the authors have been able to film the enzyme in action,[53] thus providing an invaluable contribution for the understanding of its catalytic mechanism at molecular level.

## Nuclear Magnetic Resonance Spectroscopy

Nuclear Magnetic Resonance spectroscopy is an alternative method used to determine the three-dimensional structure of proteins. In contrast to crystallography, NMR experiments are performed in solution and thus can allow direct observation of the physical flexibility and the dynamics of their interactions with other molecules. Also in contrast to X-ray crystallography, NMR studies of biomacromolecules provide an ensemble of low-energy conformations of the molecule that satisfy specific geometric criteria determined by the experimental protocol. While each conformation alone can be thought of as a static snapshot of the molecule, together they provide a dynamic representation of the protein. To obtain structural data for biological-scale molecules, multidimensional multinuclear NMR is used where the off-diagonal cross peaks encode information regarding the interaction between nuclei in the molecule, which in turn relates to the distances between atoms. It is key to extract as much information as possible about these interatomic distances; thus each peak must be assigned to its particular nucleus in the biomolecule. Spectral analysis is primarily focused on the association of the position of the individual NMR lines in the spectrum (chemical shift) to a specific nucleus ($^{1}H$, $^{15}N$ or $^{13}C$) of the protein. Although the chemical shift for an atom is primarily determined by the atomic connectivity of the amino acid residue, it can also be affected by the interactions with the solvent and by the involvement of that residue in the protein's secondary and/or tertiary structure. Various strategies of applying pulse sequences, magnetization transfer and isotopic labeling ($^{15}N$ and $^{13}C$) are employed to resolve the peak assignments. Due to dramatic advances of the technique, both in terms of hardware and software, the range of protein size amenable for NMR structure solution has been significantly extended up to 80-100kD.[54,55] High resolution structure solutions of proteins embedded in membrane systems have also become possible.[56] Recently, a new method of determining the structure of complexes in solution based on the changes in chemical shift that occur when a ligand binds to a receptor has been described.[57]

The strategy used to derive 3D structure from NMR data is to start with a randomly-folded structure derived from the primary sequence. Then the structure is optimized using either a

Molecular Dynamics/simulated annealing protocol or distance geometry against the NMR-derived distance and torsion angle data (constraints) combined with empirical data (e.g., known bond lengths and angles) to reach a minimum potential energy. In general, there will be a family of structures satisfying the NMR constraints. Interestingly, the RMSD (root-mean-squared deviation from the minimum energy structure) turns out to be different for different regions in the structure. For example, flexible regions without secondary structure, e.g., loops, show a relatively larger deviation because these regions have fewer constraints.

Protein functionality can usually be associated with backbone and side-chain dynamics. Figure 3 illustrates backbone dynamics data derived from $^{15}N$-$^{1}H$ NMR experiments on mutant Sm14-M20(C62V),[58,59] a fatty acid binding protein found in *Schistosoma mansoni. S. mansoni* is a significant parasite of humans and one of the major agents of schistosomiasis. These data show a direct correlation between the decrease of the protein flexibility, mainly in the loop regions, and ligand binding. There are larger differences in the NMR-derived generalized order parameter $S^2$ between the *apo* and *holo* (bound) forms of the protein in these regions and near the residues involved in binding (see Figure 3).[60] It is worth emphasizing that protein movements span a broad range of timescales and NMR is the only currently available technique that can monitor and discriminate these molecular processes. Nuclear spin relaxation rate measurements report on fast (< ns) and slow (μs to ms) internal motions and can allow unraveling of the intermingled effects of static disorder, coherent intramolecular motions and chemical exchange processes, but also can determine molecular rotational diffusion (5–50 ns). The determination of the rates of magnetization transfer among protons with different chemical shifts and of proton/deuterium exchange, instead, report on very slow movements of protein domains (ms to days) and provide insight into conformational exchange processes.[61] Based on the results obtained from these NMR experiments it is possible to characterize both the thermodynamic and kinetic features of interactions with other molecules (either macromolecules or low molecular weight ligands). Eisenmesser *et al.*[62] demonstrated that dynamics can be monitored during enzyme catalysis at multiple sites by means of newly reported NMR relaxation dispersion experiments that probe molecular motions in the μs to ms timescale with higher sensitivity than other relaxation experiments.[63] Using this technique, the motions in free cyclophilin A were compared with those during turnover; this comparison showed that the motions are collective, propagating from the active site to remote sites.

Some proteins, such as partially folded polypeptide chains, are difficult to crystallize and, even if crystals can be obtained, the chain segments that are disordered in solution will either be ordered by intermolecular contacts in the crystal lattice, or will remain disordered in the crystal. In these cases, NMR is capable of providing structural information and an indication of the rate of the processes that mediate the transitions between the discrete structured states present in the conformational space spanned by the ensemble of NMR conformers.[64]

While there are intrinsic experimental difficulties and limitations with NMR structure determination, its ability to reveal flexibility and the fact that it is obtained in solution at conditions more similar to the conditions *in vivo* where biomolecules are active, are compelling. Damm and Carlson[65] showed, in a comparison between 28 NMR structures and 90 crystal structures of HIV-1 protease using a method termed Multiple Protein Structure (MPS), that NMR-MPS is better able to represent protein flexibility because the ensemble of NMR-derived conformers possesses a greater structural variation.

## Computational Molecular Dynamics

The experimental data derived from X-ray crystallography and Nuclear Magnetic Resonance provides a framework for understanding the structure and flexibility of proteins and other biomacromolecules, but can not illuminate all of the details of the motions that these molecules undergo. As stated above only relatively long-lived and more populated (i.e., lower energy) states will be observed and recorded by these methods. It is not unlikely that some ligands may bind to and stabilize higher energy states that are, for example, transitional between lower energy conformations. In addition, several experimental limits, i.e., protein molecular weight, protein solubility, time required for the analyses, crystallization difficulties, etc., affect either or both of these experimental techniques and prevent them from being applicable to all biomacromolecules of interest. In particular, the membrane-bound proteins, e.g., G-Protein Coupled Receptors, have proven very difficult to crystallize for examination by X-ray crystallography and are often too insoluble for NMR analysis (vide infra).

Computational approaches, and in particular, Molecular Dynamics (MD), can generate large numbers of protein conformations to be used in docking analyses and virtual screening experiments.[66,67] The impact of flexibility on docking will be discussed in the next section, but it is clear that just considering ligand or protein side chain flexibility may be inadequate for some (or many) cases and that coupling Molecular Dynamics with docking in some manner is better able to represent the accessible conformational space.[68-71]

Molecular Dynamics approaches are often characterized by scale, depending on the nature and size of the system to be analyzed, with reference to the ubiquitous compromise between speed and accuracy (or level of detail). Both coarse-grained and atom-level simulations will be discussed below, followed by a description of a relevant and current problem in drug discovery requiring multiscale MD analysis. Figure 4 illustrates for the Estrogen Receptor α (ERα) how more than one of these scales can be relevant within the same modeling system. We must acknowledge a key technological advance that underlies all aspects of this research, but particularly MD. It is interesting to note that the first MD simulation of a biomolecule is usually attributed to McCammon, Gelin and Karplus who generated a 9 ps trajectory of the bovine pancreatic trypsine inhibitor protein (886 atoms) in 1977.[72] Recently, a 50 ns dynamics simulation of an entire virus containing one million atoms was reported.[73] In simple terms this is about a $10^7$-fold performance increase in about 30 years due to advances in parallel computer architectures coupled with algorithms and software able to exploit them efficiently. This trend is expected to continue for the foreseeable future.

## Coarse-grained approaches

Coarse-grained models, where some of the fine atomistic details are usually smoothed over or averaged out, were developed for investigating longer time-scale dynamics. Information on general protein flexibility can be obtained by using simple normal mode analysis with simple Hooke's Law-like potentials as shown in eq. 1, where it is assumed that biologically important deformations (including those resulting from ligand binding) follow one or several of the natural deformation modes of the protein

$$E=\sum_{i,j} k_d\left(d_{ij}-d_{ij}{}^0\right)^2 \tag{1}$$

where $k_d$ is a distance-dependent (or distant-independent) force constant, $d_{ij}$ is the distance between residues i and j and $d_{ij}{}^0$ is the distance between residues i and j in the experimental structure. Potentials similar to those in eq. 1 can be easily implemented into Brownian dynamics algorithms, which present several advantages with respect to normal mode analysis, for example: i) the effects of time can be incorporated, ii) several protein molecules can be considered simultaneously, opening the possibility to study flexibility related to protein-protein interactions, iii) ligand-protein interactions can be generated during the trajectory and finally iv) solvent effects can be introduced by adding suitable residue-based potentials. An alternative to Brownian dynamics is discrete dynamics, where harmonic potentials are replaced by square wells delimited by finite or infinite energy walls. The advantage of discrete dynamics is that no integrations of Newton's equations of motions are required, since the residues move in a space of constant velocity until they hit an energy wall, at which time an elastic collision is assumed.

While the coarse-grained dynamic techniques can provide surprisingly accurate information on general protein dynamics, including deformation leading to the formation or reshaping of binding pockets, the finer details are lost, which can lead to erroneous results when the resulting models are used as targets in structure based drug discovery.

## Atomic-Level Molecular Dynamics

Atomic level-of-detail Molecular Dynamics simulations allow the simultaneous representation of both small atomic fluctuations and large protein movements. The main limitation of this technique is its computational cost which has limited its general application. A critical issue in this approach is the extent to which the MD simulation is able to sample the conformational states accessible to the protein (or protein/ligand complex). Indeed, despite vast improvements in computer power (and the parallelization of many MD codes), the majority of published MD simulations are still limited to just a few ns, where the probability of sampling multiple biologically-significant minima is rather low. The true extent of the achieved sampling should be evaluated in this case. Several computationally inexpensive techniques can be employed once the MD trajectories have been collected. For example, essential dynamics allows filtering of noise from essential motions, and cosine analysis of the eigenvectors may allow estimating the separation of 'real' motions from random diffusion. Even simpler is comparing the experimentally obtained crystallographic B-factors (vide supra) with the RMSD fluctuation per residue as extracted from the MD trajectory. A qualitative correlation between B-factors and RMSD fluctuation is a sign that the MD simulation has only wandered around the crystallographic minimum and not effectively sampled conformational space. In other words, the MD is producing little more than "wiggles" and not generating new energetically accessible conformations that can be used as targets for docking or virtual screening.

The very clear necessity of documenting protein flexibility and the reliability of MD simulations has recently resulted in the MODEL (Molecular Dynamics Extended Library; http://mmb.pcb.ub.es/MODEL) project. MODEL is a massive plan to provide the community with a database of protein flexibility covering all unique proteins in the PDB. At present it contains information (ca. 10 ns simulation time) on the dynamics behavior in water for around 1500 proteins, covering most of Cluster-90 in the PDB, i.e., it is representative of all structurally known proteins. The MODEL relational database contains more than 12 terabytes of data and provides atomic level-of-detail samplings, usable to improve ligand docking, to analyze the existence of hinge points, to study ligand diffusion to binding sites and to predict potential deformation movements that can alter the protein structure. MODEL is now being extended to cover protein-protein complexes and to analyze ligand-induced changes in structure and dynamics for a subset of proteins for which both bound and unbound structures are known. This project has only been possible with the newest generation of supercomputers and enhancements in MD algorithms that can now produce state-of-the-art trajectories approaching biologically relevant time scales (μs up to 1 ms). In fact, systematic analysis of the dynamics of the entire proteome is now possible. For applications in rational drug design, the accessibility of an ensemble of structures for a given protein instead of only a single conformer multiplies the possibility of success in docking and virtual screening procedures.

Ensemble docking (vide infra) employs a small number of carefully chosen receptor conformations, obtained from X-ray, NMR or MD simulations. The goal is to increase the probability that the ligand will dock, because there are multiple, hopefully somewhat diverse, targets. While experimental conformations from multiple X-ray or NMR derived structures would be most desirable, Molecular Dynamics simulations are more accessible for most proteins and can generate protein "receptor ensembles" adequate for docking ligands for lead discovery or refinement.[67] Ensemble docking will be a topic for the next section, but there are a few MD considerations that should be addressed here. First, the extent of diversity in conformations required for successful docking may be modest, as long as a number of distinct conformational states are generated. The results thus obtained, however, must not be over-interpreted, especially if ligands with significantly different profiles are studied, and significant conformational changes are expected to take place. In this case, the extent of the achieved sampling should definitely be evaluated. A second potential element of bias may be due to the conditions under which the MD simulations are carried out. For example, an ensemble docking carried out on snapshots of an MD simulation of a protein complexed with a given chemotype, will more favorably score ligands structurally related to that chemotype compared to those of other chemotypes. This same bias also applies to experimentally determined conformations, again suggesting caution that the results not be over-interpreted.

## Multiscale MD Simulations: G-Protein Coupled Receptors

Multiscale simulations, i.e., involving both coarse-grained and atomic level-of-detail, Molecular Dynamics, are virtually the only way for structurally investigating the biological properties, function and molecular interactions of signaling proteins like GPCRs for which experimental structural data are not available. Here, GPCRs will be used to illustrate the vital structural and functional information that can be obtained when applying MD simulations. GPCRs are allosteric proteins that transform extracellular signals into promotion of nucleotide exchange in intracellular G proteins. As such, they are composed of regions of high stability (low flexibility) and regions of low stability (high flexibility) that communicate with each other by transmitting signals between the extracellular region and the distal intracellular region. GPCRs exist as complex statistical conformation ensembles.[66,74-76] Their functional properties are related to the distribution of states within the native ensemble, which is differently affected by ligands, interacting proteins, the lipid membrane environment and/or amino acid mutations.[46,66,74,77] Thus computational modeling of GPCR function requires effective integration of supramolecular modeling and multiscale simulations.

Difficulties in understanding GPCR mechanisms of function are perhaps primarily due to the lack of high resolution structural information on these proteins. The data currently available are for rhodopsin, the cornerstone of family A GPCRs in its dark (inactive) state,[78] and the human $\beta_2$-adrenergic receptor—T4 lysozyme fusion protein (at 2.4 Å resolution) bound to the partial inverse agonist carazolol.[79,80] These structural models, especially rhodopsin, are suitable templates for comparative modeling of the many homologous receptors.[66] Additionally, the structural model of a photo-activated deprotonated intermediate of bovine rhodopsin, reminiscent of metarhodopsin II (MII) (PDB code: 2I37, 4.15 Å resolution),[81] has been recently released. Interestingly, this structure revealed unexpected structural similarities between the dark and photo-activated structures, in contrast to earlier predictions of large activation-associated conformational rearrangements.[81]

Much effort has been applied in the last decade or so in elaborating a computational strategy to infer the mechanisms of intra- and inter-molecular communication in GPCRs of the rhodopsin family.[66,82] The computational approach consists of comparative or *ab initio* modeling, ligand-protein and protein-protein docking followed by comparative MD simulations and analyses. Extensive MD analyses are instrumental in inferring the most significant structural features that make the difference between free and bound states of the proteins. Reducing the system's degrees of freedom by employing implicit membrane models and intra-helix distance restraints facilitates detection of the essential motions or structural changes that may correlate with receptor and G protein functionality. Potential mechanisms of ligand- or mutation-induced receptor activation and of receptor-induced guanosine diphosphate (GDP) release from the G protein can be obtained by comparing MD-generated average structures that are representative of different receptor and G protein states.

MD simulations of the communication between the sites of activating/inactivating mutations or of ligand binding and the putative G protein coupling domains in members of the glycoprotein receptor subfamily,[66,82-84] as well as receptors for serotonin,[85] melanine-concentrating hormone,[86] and thromboxane A2 (TXA2)[87] suggest that activating ligands (agonists) or mutations communicate with a distal cytosolic receptor domain near the highly conserved "E/DRY" motif by inducing a perturbation in the interaction pattern of the E/DRY arginine. This increases the solvent accessibility of some amino acids compared to that in inactive forms.[66,82,83,87] This communication is two-way in the sense that perturbation in the cytosolic domains can also be associated with structural changes in the extracellular receptor portions at the agonist binding site. This has been demonstrated for the kappa opioid receptor.[88] Likewise, activating mutations in different sites of the receptor helix bundle have the same effect.[66,82-84] In spite of the tremendous structural diversity of the different GPCR agonists, a few critical interactions appear to be needed for activating ligands in establishing proper communication with the G protein coupling domains.[66,82,85,87] The role of ligand binding, independent of its activating or inhibitory effect, does not appear to be limited to conformation selection but, instead, promotes new conformational states unlikely to be explored by the unbound receptor forms.[66,82,85,87] Developing an understanding of this complex and dynamic interplay will ultimately lead to improved design criteria for drugs targeting the wide range of disease states involving GPCRs.

In a recent study of the TXA2 receptor, the increase in solvent accessibility around the E/DRY receptor motif in response to agonist binding appeared to favor the docking of the C-terminus of the $Gq_\alpha$ subunit of the G protein between the cytosolic ends of selected receptor helices.[87] The establishment of interactions between agonist-bound receptor and G protein is, in turn, instrumental in favoring the formation of a GDP exit route between selected portions of the α-helical and Ras (GTPase)-like domains (Figure 5).[87] This is the first example in which comparative MD analyses highlighted the potential players in the communication between the binding site of the receptor agonist and that of GDP, which are almost 70 Å apart.[87]

## Practical Considerations in Applying MD

Molecular Dynamics seems to represent the most affordable and accessible method to produce many protein conformations at reasonable cost. Indeed, thanks to the availability of software and adequate hardware, MD has become a very popular tool for SBDD and other modeling tasks. It may even seem that MD is being over-utilized when it is applied to problems where simple energy minimizations would suffice for structure optimization. Importantly, the set up of MD simulations is certainly far from trivial; nor is the interpretation of results.[89] There are a number of issues related to widespread usage of Molecular Dynamics that should be briefly described here as a caution to the casual MD user. As mentioned above, *MD remains an expert system*; medicinal chemists and other non-specialists should not view the computers or MD software as the legendary "blackbox".

First, MD simulations are often referred to as "computer experiments" and should be regarded as such. Second, the lack of standards in file formats for structures, force fields and trajectories is a major stumbling block to universal acceptance of the technology. Third, MD simulations of proteins or other biomolecules require many conformational degrees of freedom, which is further complicated when the environment of the molecules, e.g., solvent or membrane, is also modeled. Fourth, and this is of particular interest for application to docking to relevant target structures, is that there are technical difficulties in accurately sampling the conformational space of large macromolecules where the energetic and entropic barriers are high with respect to the thermal energy at physiological temperatures. Fifth, it is not obvious to the non-expert which forcefield and/or MD program should be used for a particular system. Commonly used force field families include AMBER,[90] CHARMM,[91] GROMOS,[92] and OPLS.[93] While studies have shown that in most cases all these force fields give qualitatively the same results, there can be subtle differences, and they are all approximations to the real potential energy surface. Common MD programs for simulating biomolecules include CHARMM,[91] AMBER,[90] GROMACS[94] and NAMD.[95] The factors which determine the choice of the program include the force field compatibility and also the operation of the code on the particular computer hardware. Two aspects of an MD program's suitability for a hardware configuration are considered: the "performance" and the "parallel scalability". The first is measured in terms of picoseconds of simulation time per day on a single processor, while the second is a function of the number of computing nodes or processors allocated, normally expressed as "speedup", which is the ratio of the performance for N processors with respect to that for one processor. Most MD programs show good parallel scalability only up to a certain number of processors, after which communication costs between nodes begin to degrade performance. GROMACS often is cited with the best performance for small numbers of nodes, while NAMD currently appears to have the best scalability and is more suitable for systems with many processors. Lastly, it is not just total CPU time that should be considered in MD, but also the "global elapsed time" that is "real" time used before final results are obtained.

## Molecular Docking and Virtual Screening

There are many aspects of molecular docking and its variations like virtual screening that deserve significant attention with respect to the evolution of SBDD, but most are beyond the scope of this Perspective and have been discussed elsewhere.[96-99] In particular, scoring function development, e.g., consensus methods,[100-102] QM/MM-based Free Energy Perturbation methods that include high quality quantum-derived parameters for novel small molecules as well as desolvation terms,[103-107] or empirical free energy functions,[108-111] have received recent attention. Here, while acknowledging the critical future role for better scoring functions, we are focusing on the interface between docking and flexibility.

It is well-accepted that proteins and in particular their active sites do not exist in single frozen conformations, perfectly sculpted to the shape of the incoming ligand. As described above, the beauty of X-ray crystal structures belies the uncertainties in both the technique and the molecular structure itself. However, despite this knowledge, docking experiments often begin with the false assumption that the protein can be represented by a single structure. This assumption *can* remain successful when there is no significant induced fit structural rearrangements upon ligand binding[30] (or when the active site has been pre-formed to recognize a particular class of ligand). Nonetheless, much effort over the past several years has been expended in developing new algorithms and docking programs that allow flexibility in fitting and scoring flexible ligands or ligand candidates in flexible binding sites. The first and considerably less time-consuming approaches included protein flexibility in docking experiments by simulating the possible movements of active site side-chains, e.g., version 4 of the popular program AutoDock[29-31] introduced the ability to include explicit protein sidechain flexibility.[34,112] GOLD uses a similar approach, by allowing a limited number of protein side chains to sample alternative rotameric conformations[37] and also rotating terminal hydrogen atoms to optimize hydrogen-bond interactions. FlexE selects from alternative side-chain conformations observed in other crystallographic structures to model flexibility.[32] SLIDE uses mean-field optimization to rotate protein or ligand side groups sufficiently to remove intermolecular van der Waals overlaps while docking.[40] Other approaches involve constrained geometric simulations,[113] invoking elasticity network theory,[114,115] or ensemble docking to structure families arising from dynamics, rotamer libraries, NMR or X-ray crystallography,[33,35,116-120] Monte Carlo methods,[121] protein structure prediction techniques,[122,123] or virtual alanine scanning and refinement.[124] DOCK3.5.54[38] evaluates complementarity of "components" or independently moving regions of the receptor. These components in combination give rise to a comprehensive ensemble of receptor conformations, yet the algorithm scales linearly with respect to the receptor's degrees of freedom. Other approaches use MD to effectively optimize the docked solutions obtained from rigid-receptor docking experiments in a post-processing step.[125] Finally, an algorithm employing "flexibility trees" has been recently described.[126] This approach greatly reduces the computational overhead for modeling receptor flexibility during docking.

### Side-chain flexibility

Kuhn and co-workers demonstrated the importance of including local flexibility with a study that docked the bound conformation of a series of known ligands to the ligand-free (unbiased) protein conformations. Only 9 out of 42 known thrombin ligands and 9 out of 15 glutathione S-transferase (GST) ligands could be docked without steric clashes when side-chain flexibility was neglected.[40] However, 90% of the same ligands could be docked to within 1.3 Å RMSD of the correct atom positions, on average, when small-scale side-chain flexibility was modeled using the docking and screening tool SLIDE.[127] Beyond the necessity of modeling flexibility to capture known inhibitors or substrates in protein complexes, accurate sampling of low-energy protein conformers allows these structures to be used as alternative targets for inhibitor design and screening.

Fortunately, the challenge of flexibility modeling is simplified enormously by the fact that most protein side chains undergo only small motions during ligand binding. In a further study of 63 protein-ligand crystallographic complexes, 83% of all active-site side-chain bonds that rotated in 32 thrombin-ligand complexes were observed to rotate 15 degrees or less relative to the ligand-free structure. The same was true for 91% of the GST side-chain rotations in 13 complexes, and 75% of side-chain rotations in 18 other, non-homologous complexes.[40] This is consistent with the finding that ligand binding often induces strain or non-rotamericity in side chains.[128] The dominance of small side-chain rotations is very good news, because relatively simple energy minimization or steric optimization procedures are often sufficient to model them.[129,130]

Why do active-site side chains typically move so little upon ligand binding? Studies performed on 30 low-homology protein complexes and on the corresponding ligand-free structures indicated that preservation of direct intra-protein hydrogen bonds is the main reason (see Figure 6a).[131] About 75% of all intra-protein hydrogen bonds in these sites are preserved upon ligand binding and the percentage of main-chain hydrogen bonds preserved is even higher (typically 85-100%), as indicated by the values for Ala, Ile, Gly, and other residues lacking side-chain hydrogen-bonding groups. However, the picture reverses when water-mediated hydrogen bonds in ligand-binding sites are considered: from the same 30 pairs of ligand-bound and free structures (Figure 6b) 50-80% of water-mediated intra-protein hydrogen bonds are observed to *break* upon ligand binding.[131] Thus, ligand-binding sites can be considered to be partitioned into pre-organized regions consisting of directly hydrogen-bonded groups within the protein, and other regions that are readily reorganized, due to the plasticity of their water-mediated hydrogen bonds. This allows a simplifying divide-and-conquer strategy in which most of the flexibility sampling effort in ligand binding sites can be focused on the hydrogen bonding groups that are not yet satisfied by intramolecular hydrogen bonds, exposure to solvent, or interaction with the ligand.

### Elastic Network Models and Constrained Geometric Simulations

While Molecular Dynamics and related modeling approaches are suitable for investigating subtle movements and conformational sub-states, larger conformational changes are not usually observed in the simulations unless particularly long time scales are studied. Proteins are often able and required to undergo significant movements to carry out their catalytic activity, or to interact with ligands and/or other macromolecules such as been demonstrated by ion channels, allosteric proteins or heat shock proteins.[132-134] In order to reproduce these significant conformational adjustments starting from unbound receptor structures, Gohlke and co-workers developed a two-step method based on recent developments in rigidity and elastic network theory.[135] In the first step, static properties of the macromolecule are determined by decomposing the molecule into rigid clusters using the graph-theoretical approach FIRST.[136] In the second step, the dynamic properties of the biomolecule are revealed by the rotations-translations of blocks approach[137] using an elastic network model representation of the coarse-grained protein.[138] On a data set of 10 proteins that show conformational changes upon ligand binding, the predicted directions and magnitudes of motions were shown to agree well with experimental observations, demonstrating that the motions presumed to be "ligand-induced" are already well-defined in the unbound receptor structure.

Although the constantly increasing availability of computational resources has mitigated this issue somewhat, MD is still extremely computationally expensive. The technology is also an expert system that is neither simple nor obvious in its application and sometimes prone to convergence problems. Computational time can be significantly reduced by performing constrained geometric simulations that use an efficient algorithm able to reproduce the motion of flexible and rigid parts by ghost template rearrangements.[139] Theassociated program, FRODA,[139] also uses natural coarse-graining for the treatment of rigid regions identified by FIRST.[136] When applied to the protein-protein interface of interleukin-2, FRODA highlighted transient pocket formation in agreement with experiments[140] and more elaborate MD studies (see Figure 7). In fact, when simulations were started from the unbound state, interface configurations were sampled by both methods that came as close as 1.0 Å RMSD to the bound conformation. These results strongly support the "conformation selection" model,[7] and the configurations may well be used in subsequent flexible docking approaches.

A number of docking programs, like AutoDock, use grid maps to represent the three-dimensional patterns of atomic affinity, electrostatic potential, desolvation free energy, etc. around the target molecule. These programs perform an interpolation for each atom in the ligand to estimate its interaction energy with the target, generally assuming the target is rigid. Some slight conformational flexibility in the target can be added by allowing some overlap in the radii of the grid points. However, a more sophisticated and novel paradigm for fully-flexible protein-ligand docking can be proposed, based on an *elastic* representation of potential grids in the binding pocket region of a receptor (Kazemi, Krüger, and Gohlke, unpublished results). Protein conformations can be sampled during docking without the need to re-calculate potential grids. Instead, grid points are moved along with the binding pocket region according to the laws of elasticity. This approach was tested on a comprehensive dataset of protein targets representing different classes of conformational changes during ligand binding. Notably, compared to docking to the *apo* conformation of the proteins, ligand binding mode predictions were greatly improved when grids deformed to a bound protein conformation were used instead. In addition, not only can sidechain and backbone movements be accounted for, but in principle, any pairwise scoring functions can be used.

### Docking with an Ensemble of Protein Conformations

While there are a small number of cases where experimental data from X-ray crystallography or NMR comprise an ensemble of protein structures (the "receptor ensemble") with enough range to fully explore conformational space, computational ensembles from Molecular Dynamics are most often used for generating protein ensembles adequate for lead docking. The caveats and limitations on MD described above must be considered. In particular, the scale of movement in the ensemble must be monitored to ensure that it is consistent with the structural diversity of ligands to be docked. This method of docking the set of ligands into all derived discrete conformations of the receptor is termed "ensemble docking".[117]

In the most simplistic variant, ensemble docking calculations are carried out sequentially, for one protein conformer after the other, which multiplies the required calculation time by the number of considered conformers. Alternatively, the protein conformers may be combined to an average representation, which is relatively straightforward with grid-based docking methods, i.e., combine two or more sets of grid maps from an ensemble of different conformations of the target. Two ways of combining these grid maps, an energy-weighted average and a geometric-weighted average of the interaction energy between the ligand and the receptor, were described by Knegtel et al.[141] These and three other methods of combining maps, simple mean, grid point minima, and simple Boltzmann-weighted average of the interaction energies, were evaluated and the Boltzmann-weighted average was shown to be best able to model both variations in conformation of the very plastic HIV-1 protease and the presence or absence of structural water molecules.[35] One considerable advantage of combining many different conformations of a target protein into a single grid is computational speed as compared to individual docking to multiple targets. There are, however, pitfalls to this approach: i) there are limits with respect to the tolerated structural differences among the conformers being averaged and ii) the average/composite structure as represented by the grid may not be a physically "real" target and the danger exists that "artificial" ligand poses are generated that are reasonable only for the averaged representation. As a further variant, there is an "in situ cross-docking" approach where multiple protein structures can be addressed simultaneously in a single (grid-based) docking run.[116,142,143] Although conceptually simple, in situ cross-docking can lead to significant speed-up over conventional serial cross-docking approaches, and the simultaneous optimization across multiple protein structures allows for a more direct selection of the optimal binding site. The method can be applied simultaneously to proteins with a wide range of structures, but there are limitations in the number of structures that can be examined in each calculation.

Given this evidence, does it actually make sense to address issues of flexibility and induced-fit by analyzing a set of pre-generated protein conformations? One can think of induced-fit as a process of preferential selection of conformations and corresponding shifts of equilibria.[144,145] In this view, states with an appropriately formed binding pocket, which in the absence of a ligand may be very weakly populated, are preferentially selected by the corresponding ligand because it stabilizes those states resulting in a net gain in free energy compared to other protein conformations in the ensemble. The conformational equilibrium is thus shifted towards the binding-competent conformations, and those conformational states become predominant and experimentally observable. Consequently, using pre-generated protein conformations to deal with protein flexibility appears reasonable, at least in principle. With access to the entire ensemble of low-energy protein conformers and to a reliable free energy function for calculating the affinity based on the protein-ligand interactions and the conformational contributions from the protein, predictions of the preferred geometry of the complex should then reduce to a "simple" optimization task. Unfortunately, neither of the two conditions are met in reality, as there is almost never access to a complete (or even sufficiently representative) ensemble of protein conformers, and current free energy functions used in docking are not sufficiently reliable to provide accurate scores for alternative binding modes for different protein conformers (especially if the free energy differences are small and/or dominated by entropic contributions). Accordingly, the success of induced fit docking with a structural ensemble of conformers is often limited to cases where the available protein conformers from complexes with particular ligands are indeed (and likely by chance) representative for the complexed state. This most often occurs when the ligands being investigated are similar to the ligands in the structural models comprising the ensemble. Also, the energetics must be such that the scoring function is competent to discriminate between alternative binding modes.

A pioneering development in modeling protein flexibility for docking calculations using MD to generate the conformation ensemble is the Relaxed Complex Method from the work of McCammon et al.[68-70] In this approach, Molecular Dynamics simulations of the targets are performed prior to docking, and different conformations are selected; the simplest is to choose conformations at regular time intervals, but it is also possible to select the most structurally diverse conformations from all conformations generated. It is then a matter of docking the ligand of interest to each of these different ‘snapshots’ from the Molecular Dynamics trajectory. At the end, a histogram of binding energies and one or more different binding modes are obtained. This approach, using AutoDock as the docking engine, is credited with the discovery of a novel binding trench in HIV integrase[71] which laid the groundwork for the development of the first clinically-approved integrase inhibitor Raltegravir.[146]

However, both the success and limits of ensemble docking can readily be illustrated with an example from the well-characterized aldose reductase system, an enzyme showing pronounced conformational adaptations upon ligand binding.[147] Based on dozens of high-resolution crystal structures and extensive Molecular Dynamics simulations, detailed knowledge of the binding site conformations is available.[148-150] The binding pocket is characterized by a very stable region surrounding the catalytic site and a highly mobile area close to the “specificity pocket”. Although very localized, this mobility is mediated by side-chain rotations and a stretch of flexible backbone. Essentially, three different protein conformers (and binding modes) are observed, with minor variants for two of the three major conformers. One of the conformers has to date only been observed in complex with the inhibitor tolrestat (**1**), thus representing a unique binding mode (Figure 8). Two recently synthesized tolrestat analogues (**2, 3**) with high affinity[151,152] were investigated to ascertain whether they adopted a similar binding mode. While docking of tolrestat to all three aldose reductase conformers correctly reproduced its binding mode in the open specificity pocket, docking of **2** and **3** predicted preferred binding to a conformer with a closed specificity pocket, which was confirmed by subsequent crystallographic analysis. Interestingly, the actual binding mode of **3** is very similar to the docking prediction, but compound **2** shows a hitherto unobserved binding mode with an entirely new conformation of the aldose reductase binding site not involving the specificity pocket (Figure 9)! This new conformation features the unexpected opening of a salt bridge involving a lysine side-chain that is not obviously compensated by a new protein-ligand contact. While this broken salt bridge likely explains the failure to properly dock **2** in aldose reductase even with explicit side chain flexibility, more importantly, this brings into focus the larger reality that, even with supposedly well-characterized systems like aldose reductase, new binding modes are possible and one can never be certain that a pre-calculated ensemble has exhaustively explored flexibility.

A slightly different approach was illustrated in a recent study carried out by Orozco and co-workers[153] on p38 MAP kinase, a serine/threonine kinase involved in major signal transduction pathways and a key factor in the modulation of the level of Tumor Necrosis Factor-α. This study was designed to understand the binding of a new pyridinylheterocycle family of inhibitors and is a very relevant example of combining results from X-ray crystallography, homology modeling, quantum chemistry, classical docking and Molecular Dynamics. Figure 10a illustrates the four potential binding modes of one member of this series, **4**, a lead compound with excellent inhibitory activity (161 nM)[153,154] that is presumed to bind at the ATP binding site of p38 MAP kinase. While a crystal structure for the MAP kinase complex with **4** is not available, even one at relatively high resolution would not necessarily resolve these 4 cases.[153] First, a high-level *ab initio* study of the N1-H/N2-H tautomerism in the pyrazolopyridine group suggested that the N1-H model (see Figure 10a) is energetically favored over the N2-H model. Docking analysis suggested two binding modes (Figure 10a, top and bottom row) with fairly minor differences in protein-ligand interactions. The two modes and two tautomers were thus subjected to 2 ns MD simulations where monitoring of the RMSD of **4** from the binding site, the ligand-protein interaction energy, and the key hydrogen bond between the pyridine nitrogen and the amino group of Met109 clearly favors the first binding mode (and N1-H) as illustrated in Figure 10b. This approach was also able to quantify the interactions of each protein residue with X (see Figure 10c).

## Summary and Perspectives

The history of developments in computer-aided drug discovery (CADD) was recently reviewed by John Van Drie[155] in the context of the Bezdek curve showing the progression of the technology from “naïve euphoria” to ”peak of hype” through the “depths of cynicism” to “true user benefits” approaching the “asymptote of reality”. The current environment in CADD is an apparently healthy mixture of multiple paradigms, each having gone through the hype and cynicism phases and rebounded with their somewhat diminished cadre of true believers to being generally beneficial. The technologies with long term benefit are generally those that have adapted by adding value from complementary techniques, much as combinatorial synthesis has found a niche by becoming target-specific. In a similar way, structure-based drug design has matured by taking advantage of emerging technologies over the years. It is not certain whether the first protein crystallographers, who were generally physicists, were thinking about SBDD as they were painstakingly collecting and processing their data. Certainly, however, it was only a few short years before the enormous potential of understanding biology through structure emerged.[1,156,157] Currently, through the structural biology consortia and their high-throughput crystallography, there is an explosion of data available as targets, but the function of some of these new proteins remain poorly understood. This is another challenge not in the scope of this Perspective. An emerging development that we do expect to have a profound impact on SBDD is the consideration of target flexibility as part of the design process. We have reviewed above many of the experimental and computational approaches that are currently in use or under development. We offer here some perspectives on what still needs to be accomplished and the resulting benefits of incorporating flexibility as a key component in drug discovery methodology.

First and foremost, there is the need for more and better quality experimental data on flexibility of protein systems. Indeed this is a major challenge for both X-ray crystallography and NMR spectroscopy. Two barriers exist: a) time resolution: NMR is able to resolve rapid small scale motions and with isotope exchange very slow vibrations, and X-ray crystallography, while it clearly cannot deal with large motions within a single crystal, can detect with synchrotron radiation some high frequency vibrations. Nonetheless these are lengthy, complex and sometimes expensive experiments; b) structure resolution: the introduction of very high field magnets (up to 22 T) and the development of cryo-probes coupled with innovations in data processing have improved considerably the NMR resolution power, however, molecular size is still the major limitation (50-60 kD is currently considered the limit for obtaining high resolution structures). Operating at very low temperature, thus dampening the vibrational motion, has been a major breakthrough for X-ray crystallography. However a consideration often lost in this regard is that free energies of association and binding are measured at room temperature or higher and the biological processes that are being simulated occur at biological temperature, while high-resolution crystallography is routinely performed at around liquid nitrogen temperature. This disconnect has not been fully resolved, but it is interesting to note that at *very* high resolutions alternate conformations for labile residues can be observed in X-ray structures, suggesting that rapidly freezing crystals does preserve some flexibility information.

In the absence of extensive experimental data, the technology that will make the largest impact on understanding and exploiting flexibility is Molecular Dynamics simulations. While MD is almost ubiquitously available, there are a number of issues regarding its widespread usage that suggest some degree of caution (vide supra). However, over the next few years, MD will very likely become even more accessible, and hopefully many of these issues will become transparent even to the less experienced users. But, looking past implementation issues, there is probably inadequate evidence that MD simulations are truly representative of molecular motions, i.e., are following real paths, especially for large and complex molecules, even if the sampled conformations do appear to represent realistic local minima. Conventional MD single trajectory simulation is usually not able to reproduce large conformational changes due to timescale limitations; the application of multiple-trajectory or Replica Exchange methods,[158] may lead to a better exploration of conformational space. Combining coarse-grained sampling to identify large motions with fine-scaled sampling methods to more accurately probe local transitions and energetics is another solution. However, there is absolutely no guarantee that submitting an unliganded (*apo*) structure to long MD simulations (of any type or combination of types) will generate conformations suitable for docking a set of ligands. While both protein (and ligands) undergo significant conformational adjustments upon binding, those motions are only a very small fraction of the motions simulated by MD. Ligand-induced MD, where the simulation is performed in the presence of a ligand may represent a possible solution, but the protein conformations thus generated will be greatly affected by the structure of that particular molecule. Identifying and focusing on the molecular motions *truly* critical to ligand binding and subsequent docking may end up in the same place as docking algorithms incorporating local flexibility, but from the opposite direction. To expand this notion, we believe that the best way to identify correctly docked ligand and receptor residue conformations is in proper estimation of the free energy of binding, and not necessarily on suitable overlap with reference crystal structures.[110] Docking a ligand against an ensemble of protein models, and normalizing the resulting set of binding energy predictions with Boltzmann statistics, including improved evaluation of solvent effects, is likely to be an upcoming evolutionary development in free energy scoring functions. One approach to this problem that dramatically illustrates the difficulties of scoring is the Computational Titration method that simultaneously optimizes protonation, solvent conformation and hydrogen bonding for ligand-protein complexes while calculating a Boltzmann-weighted free energy score for the ensemble.[159,160] At present this is applied only to static structures where the resulting ensemble is isocrystallographic.[160] The Multiple Protein Structure method of Carlson[65] and developments in ensemble docking also may reveal new technology in scoring functions.[116,117,161-163]

To conclude, perhaps the major challenge for most practitioners of drug discovery is that modeling flexibility requires a change in mind-set. The comforting idea that there is one ligand perfectly adapted for **one** static protein "structure" is outdated. Even the approach of searching through an ensemble of conformations to find one matching and accommodating the ligand of interest is an inexact approximation of biological reality even though it is pragmatic and successful. It is also the current state-of-the-art. However, the future may provide us with tools able to visualize and score flexible biological molecules even as they move and change conformation. The ability to accurately predict the free energy of binding for proposed protein-ligand complexes remains an unresolved problem. However, the recent work of Gilson and coworkers in characterizing the binding in host-guest complexes,[164,165] and the extension to ligand-protein complexes,[166] is providing invaluable information for designing new methods of estimating the free energy of binding. Scoring functions for docking and virtual screening have recently been reviewed.[96] Certainly, the additional dimension of protein flexibility, whether it arises from localized site adaptations or from an ensemble of test conformations, further complicates the development of reliable scoring functions. Enthusiasm for expanding the set of conformations in creating more exhaustive ensembles must be tempered with reality in that the conformations must be energetically accessible and meaningful, which of course reiterates the need for more experimental data on flexibility in biomacromolecules. The right solution, as always, resides in a real understanding of the biological system of interest, which allows us to use the proper tools. Thus, the benefits of including flexibility in docking studies that were illustrated in the sample cases above were largely for systems where extensive experimental data was available. Much of this work has been performed with paradigms, algorithms and software that are still at the cutting edge and definitely not mainstream. However, as with most developments in CADD, once the technology is proven and the market assessed, turnkey systems will start becoming available. Clearly, virtual screening campaigns that incorporate more target flexibility will identify more putative ligands worthy of closer examination than those with a single static target, but will also, unfortunately, probably generate more false positives.[60,167] New virtual screening approaches will certainly emerge, probably using an efficient hybridization of static, averaged and ensemble targets. Because recent innovations in experimental and computational biology and medicinal chemistry have now coalesced around flexibility and dynamics of structure, the authors have collaborated in producing this consensus Perspective. We are, despite the remaining hurdles, very enthusiastic about the future of drug discovery in the upcoming “Flexibility Era”, and look forward to the innovations that will arise as this paradigm takes root in the broader drug discovery community.

## Acknowledgements

The support of our institutions and extramural funding are gratefully acknowledged: Italian National Council of Research, National Institute for Biosystems and Biostructures (INBB), CNR-Bioinformatics Project (P.C. and F.S.), Italian Ministry of University and Scientific Research - FIRB RBNE0157EH (F.S.); U.S. NIH GM71894 (G.E.K.); U.S. NIH GM67249 and AI53877 (L.A.K.); U.S. NIH GM69832 (G.M.M), Marie Curie Incoming International Fellowship MIF1-CT-2005-022050 (T.A.P.); International project funded by MIUR about the development of PLP dependent inhibitors in collaboration between University of Oklahoma, Virginia Commonwealth University and University of Parma (2006-2008). We also wish to thank Drs. J. C. Hackett, P. D. Mosier, M. K. Safo and J. N. Scarsdale (VCU), G. Ingletto and A. Mozzarelli (U. Parma) and P. G. De Benedetti (U. Modena and Reggio Emilia) for providing materials and for critical reading of the manuscript.

## Biographies

**Pietro Cozzini** received his Laurea degree in Chemistry at the University of Parma in 1978 and has been involved in computational chemistry research throughout his career. He is presently Senior Researcher in the Department of General Chemistry, University of Parma and teaches chemistry, molecular modelling and database courses in several degree programs. Dr. Cozzini’s current research interests focus on the development of computational methods for exploring energetics in biomolecular associations (with particular consideration of the role of water) and the discovery of new lead compounds, identification of new food pollutants and the design of new chemosensors using *in silico* screening.

**Glen E. Kellogg** received his Ph.D. degree from the University of Arizona in 1985 under the direction of Dennis Lichtenberger in Inorganic Chemistry with research involving photoelectron spectroscopy, quantum mechanical calculations and organometallic synthesis. After a postdoctoral at Northwestern University with Tobin Marks, he joined the Medicinal Chemistry department at Virginia Commonwealth University where he has focused on computational chemistry and molecular modeling. Dr. Kellogg is co-author of the HINT program that uses the empirical data from solvent partitioning as the basis of a free energy forcefield. He is currently Associate Professor and Assistant Department Chair of the VCU Department of Medicinal Chemistry and Director of the Molecular Modeling Facility. His research interests have focused on understanding biomolecular interactions by developing and applying new computational tools to important biological problems.

**Francesca Spyrakis** received her Ph.D. in Biochemistry and Molecular Biology in 2007 at the University of Parma in the laboratories of Pietro Cozzini and Andrea Mozzarelli. She is now a postdoctoral supported by the Italian Institute of Biomolecules and Biostructures in the University of Parma General Chemistry Department. Her research is largely devoted to testing and developing alternative scoring methods for the analysis and prediction of the binding free energies of biomolecular interactions. This is made possible by collaborations with Virginia Commonwealth University, the University of Modena and Reggio Emilia and the University of Barcelona.

**Don J. Abraham** is an Emeritus Professor in the Department of Medicinal Chemistry, School of Pharmacy and Emeritus Director of the Institute for Structural Biology and Drug Discovery at Virginia Commonwealth University. Dr. Abraham's interdisciplinary research focuses on structure based drug design including X-ray crystallography, molecular modeling, synthetic medicinal chemistry, and structure - function studies involving allosteric proteins. His research group has advanced three molecules (vanillin, RSR13, and-5HMF) into two phase one studies for treatment of Sickle Cell Anemia (vanillin and 5-HMF) and eight phase one/two clinical trials, and two phase three trials for treatment of metastatic brain tumors receiving cranial radiation. He has been issued 26 patents with 11 pending and has published over 175 research articles. Dr. Abraham is currently co-chief editor of the seventh edition of Burger's Medicinal Chemistry and chair (2008) of the ACS Division of Medicinal Chemistry.

**Gabriele Costantino** received his Ph.D. in Chemistry in 1992 from the University of Perugia. From 1994 to 2006 he was Assistant Professor and then Associate Professor in the Department of Medicinal Chemistry at the University of Perugia. Since December 2006 he is Professor of Medicinal Chemistry at the University of Parma. He is the author of more than 80 papers and his research interests cover the application of computational methods to the design and synthesis of biologically active compounds, with special interest devoted to CNS-acting ligands.

**Andrew Emerson** is a computational chemist at the CINECA supercomputer center in Bologna, Italy.

**Francesca Fanelli** received her PhD in Computational Biophysical Chemistry in 1994 from the University of Modena.. She is now an Associate Scientist of the Dulbecco Telethon Institute and an Associate Professor of Molecular Engineering at the University of Modena and Reggio Emilia. Her research since 1991 has been almost entirely devoted to developing computational protocols and molecular models to unravel the chemical communication mechanisms involving G protein-Coupled Receptors (GPCRs). Dr. Fanelli has published nearly 100 articles with the majority of them concerning computational modeling of GPCRs and their interacting partners.

**Holger Gohlke** received his Ph.D. in Pharmaceutical Chemistry from the University of Marburg under the supervision of Gerhard Klebe, with an interim research stay in Francois Diederich's lab at ETH Zurich. Following his degree, he pursued postdoctoral studies in the group of David Case at The Scripps Research Institute, La Jolla, CA, where he focused on understanding and predicting the energetics and dynamics of protein-protein interactions. Subsequently, he became Assistant Professor of Molecular Bioinformatics at the University of Frankfurt. Since 2008, he is Professor of Pharmaceutical and Medicinal Chemistry at the University of Kiel and head of the Computational Pharmaceutical Chemistry and Molecular Bioinformatics research group. His research is currently directed at understanding, predicting, and modulating interactions involving biomacromolecules from a theoretical perspective.

**Leslie Kuhn** received her Ph.D. in Biophysics at the University of Pennsylvania in 1989 on the prediction of membrane protein structure with John S. Leigh, Jr. Postdoctoral research on structural prediction of bioactive motifs and protein:ligand interactions was carried out with Elizabeth Getzoff and John Tainer at The Scripps Research Institute. In 1994, Dr. Kuhn moved to Michigan State University as an Assistant Professor, progressing to become the founding Director of MSU's Center for Biological Modeling, Professor in Biochemistry & Molecular Biology and Computer Science & Engineering, and Adjunct Professor in Physics & Astronomy. Her laboratory focuses on modeling protein conformational change to improve the affinity and specificity of designed ligands.

**Garrett M. Morris** obtained his B.A. in Chemistry from Jesus College, Oxford, in 1988, working for his Part II (final year) of his undergraduate degree in the laboratory of W. Graham Richards at the Physical Chemistry Laboratory, Oxford, where he also obtained his D.Phil. in computational chemistry in 1991. The same year, he joined the laboratory of Prof. Arthur J. Olson at The Scripps Research Institute to pursue a post doctorate in ligand-protein docking. His research interests include the development of improved scoring functions for lead identification and optimisation, novel search algorithms, and flexibility in ligand-protein docking.

**Modesto Orozco** received his Ph.D. in Chemistry at Universitat Autònoma de Barcelona in 1990. From 1991-2002 he was Associate Professor of Biochemistry at Universitat de Barcelona, where he is now Professor. Between 1991-1993 he was also an Associate Researcher at Yale University. Currently Dr. Orozco is Group Leader at the Institut de Recerca Biomèdica at the Barcelona Science Park, Director of the Structural Bioinformatic Unit of the National Institute of Bioinformatics, Director of the Life Science Department at the Barcelona Supercomputing Center and Director of the Joint IRB-BSC Program on Conputational Biology. He is the author of nearly 300 peer review papers in international journals and the member of many journal advisory/editorial boards such as Journal of Computational Chemistry, Theoretical Chemistry Accounts, Journal of Chemical Theory and Computation and Physical Chemistry Chemical Physics.

**Thelma A. Pertinhez** received her M.S. and PhD. degrees in Biochemistry at the University of São Paulo, Brazil in 1998. She spent two years at the Oxford Centre for Molecular Sciences (United Kingdom) under the direction of Christopher M. Dobson as a postdoctoral. In 2000 she returned to Brazil at the Brazilian Laboratory of Synchrotron Light where she has been involved in planning and implementing the multi-user BioNMR laboratory, and in research activity in structural biology. In 2005, she moved to the Department of Experimental Medicine at the University of Parma (Italy) as a researcher with a Marie Curie Incoming Fellowship where she continued her studies on the structure-function relationship of peptides and proteins using multinuclear and multidimensional Nuclear Magnetic Resonance.

**Menico Rizzi** received his degree in chemistry and his Ph. D. in protein crystallography from the University of Pavia, Italy. He is currently full professor of biochemistry at the University of Piemonte Orientale "A. Avogadro", where he coordinates a research group working in the field of structural enzymology. Presently, his research focuses on enzymes involved in NAD (P) homeostasis and tryptophan degradation in the human central nervous system, in the mosquito malaria vector *Anopheles gambiae* and in *Mycobacterium tuberculosis*, for the discovery of novel therapeutics against neurological disorders and poverty-linked diseases.

**Christoph A. Sotriffer** obtained his doctoral degree in Chemistry at the University of Innsbruck (Austria) under the supervision of Professor Klaus Liedl. Currently, he is Professor of Pharmaceutical Chemistry at the University of Würzburg, Germany. His research interests are focused on computational methods of structure-based drug design, in particular docking approaches, scoring functions, and molecular dynamics simulations. Before his current appointment (obtained in 2006) he worked as researcher and teaching assistant in the group of Gerhard Klebe at the University of Marburg, Germany. From 1999 to 2000 he carried out postdoctoral research at the University of California, San Diego, USA, in the group of Andrew McCammon.

## §Abbreviations

SBDD, structure-based drug design
FXR, farnesoid X-receptor
GPCRs, G-protein coupled receptors
PDB, protein data bank
NMR, nuclear magnetic resonance
MD, molecular dynamics
RMSD, root-mean-squared deviation
MPS, multiple protein structure
ERα, estrogen receptor α
MODEL, molecular dynamics extended library
MII, metarhodopsin II
GDP, guanosine diphosphate
TXA2, thromboxane A2
E/DRY motif, glutamic acid/aspartic acid-arginine-tyrosine motif
GST, glutathione S-transferase
MAP kinase, mitogen-activated protein kinase
CADD, computer-aided drug discovery

## References

1. Abraham, DJ. Comprehensive Medicinal Chemistry II. Elsevier; Oxford: 2007. Structure-based Drug Design - A Historical Perspective and the Future.
2. Bolton W, Perutz MF. Three dimensional fourier synthesis of horse deoxyhaemoglobin at 2.8 Angstrom units resolution. Nature 1970;228:551–552. [PubMed: 5472471]
3. Fermi G, Perutz MF, Shaanan B, Fourme R. The crystal structure of human deoxyhaemoglobin at 1.74 A resolution. J Mol Biol 1984;175:159–174. [PubMed: 6726807]
4. Kendrew JC, Bodo G, Dintzis HM, Parrish RG, Wyckoff H, et al. A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. Nature 1958;181:662–666. [PubMed: 13517261]
5. Watson H. The Stereochemistry of the Protein Myoglobin. The Stereochemistry of the Protein Myoglobin 1969;4:299–333.
6. Safo MK, Abraham DJ. The enigma of the liganded hemoglobin end state: a novel quaternary structure of human carbonmonoxy hemoglobin. Biochemistry 2005;44:8347–8359. [PubMed: 15938624]
7. Tsai CJ, Kumar S, Ma B, Nussinov R. Folding funnels, binding funnels, and protein function. Protein Sci 1999;8:1181–1190. [PubMed: 10386868]
8. Bahar I, Chennubhotla C, Tobi D. Intrinsic dynamics of enzymes in the unbound state and relation to allosteric regulation. Curr Opin Struct Biol 2007;17:633–640. [PubMed: 18024008]
9. Henzler-Wildman K, Kern D. Dynamic personalities of proteins. Nature 2007;450:964–972. [PubMed: 18075575]
10. Vonrhein C, Schlauderer GJ, Schulz GE. Movie of the structural changes during a catalytic cycle of nucleoside monophosphate kinases. Structure 1995;3:483–490. [PubMed: 7663945]
11. Wolf-Watz M, Thai V, Henzler-Wildman K, Hadjipavlou G, Eisenmesser EZ, et al. Linkage between dynamics and catalysis in a thermophilic-mesophilic enzyme pair. Nat Struct Mol Biol 2004;11:945–949. [PubMed: 15334070]
12. Henzler-Wildman KA, Lei M, Thai V, Kerns SJ, Karplus M, et al. A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. Nature 2007;450:913–916. [PubMed: 18026087]

13. Krauss, G. Biochemistry of signal transduction and regulation. Wiley-VCH; 2003. p. 541
14. Hellal-Levy C, Fagart J, Souque A, Wurtz JM, Moras D, et al. Crucial role of the H11-H12 loop in stabilizing the active conformation of the human mineralocorticoid receptor. Mol Endocrinol 2000;14:1210–1221. [PubMed: 10935545]
15. Hu X, Lazar MA. Transcriptional repression by nuclear hormone receptors. Trends Endocrinol Metab 2000;11:6–10. [PubMed: 10652499]
16. Mi LZ, Devarakonda S, Harp JM, Han Q, Pellicciari R, et al. Structural basis for bile acid binding and activation of the nuclear receptor FXR. Mol Cell 2003;11:1093–1100. [PubMed: 12718893]
17. Downes M, Verdecia MA, Roecker AJ, Hughes R, Hogenesch JB, et al. A chemical, genetic, and structural analysis of the nuclear bile acid receptor FXR. Mol Cell 2003;11:1079–1092. [PubMed: 12718892]
18. Pierce KL, Premont RT, Lefkowitz RJ. Seven-transmembrane receptors. Nat Rev Mol Cell Biol 2002;3:639–650. [PubMed: 12209124]
19. Terrillon S, Bouvier M. Roles of G-protein-coupled receptor dimerization. EMBO Rep 2004;5:30–34. [PubMed: 14710183]
20. Brady AE, Limbird LE. G protein-coupled receptor interacting proteins: emerging roles in localization and signal transduction. Cell Signal 2002;14:297–309. [PubMed: 11858937]
21. Smith CG, Vane JR. The discovery of captopril. Faseb J 2003;17:788–789. [PubMed: 12724335]
22. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, et al. The Protein Data Bank. Nucleic Acids Res 2000;28:235–242. [PubMed: 10592235]
23. Tompa P. Intrinsically unstructured proteins. Trends Biochem Sci 2002;27:527–533. [PubMed: 12368089]
24. Somogyi B, Lakos Z, Szarka A, Nyitrai M. Protein flexibility as revealed by fluorescence resonance energy transfer: an extension of the method for systems with multiple labels. J Photochem Photobiol B 2000;59:26–32. [PubMed: 11332886]
25. Hubbell WL, Cafiso DS, Altenbach C. Identifying conformational changes with site-directed spin labeling. Nat Struct Biol 2000;7:735–739. [PubMed: 10966640]
26. Lipfert J, Doniach S. Small-angle X-ray scattering from RNA, proteins, and protein complexes. Annu Rev Biophys Biomol Struct 2007;36:307–327. [PubMed: 17284163]
27. Kuntz ID, Blaney JM, Oatley SJ, Langridge R, Ferrin TE. A geometric approach to macromolecule-ligand interactions. J Mol Biol 1982;161:269–288. [PubMed: 7154081]
28. Rarey M, Kramer B, Lengauer T, Klebe G. A fast flexible docking method using an incremental construction algorithm. J Mol Biol 1996;261:470–489. [PubMed: 8780787]
29. Morris GM, Goodsell DS, Huey R, Hart W, Belew R, et al. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. J Comput Chem 1998;19:1639–1662.
30. Goodsell DS, Olson AJ. Automated docking of substrates to proteins by simulated annealing. Proteins 1990;8:195–202. [PubMed: 2281083]
31. Morris GM, Goodsell DS, Huey R, Olson AJ. Distributed automated docking of flexible ligands to proteins: parallel applications of AutoDock 2.4. J Comput Aided Mol Des 1996;10:293–304. [PubMed: 8877701]
32. Claussen H, Buning C, Rarey M, Lengauer T. FlexE: efficient molecular docking considering protein structure variations. J Mol Biol 2001;308:377–395. [PubMed: 11327774]
33. Cavasotto CN, Abagyan RA. Protein flexibility in ligand docking and virtual screening to protein kinases. J Mol Biol 2004;337:209–225. [PubMed: 15001363]
34. Huey R, Morris GM, Olson AJ, Goodsell DS. A semiempirical free energy force field with charge-based desolvation. J Comput Chem 2007;28:1145–1152. [PubMed: 17274016]
35. Osterberg F, Morris GM, Sanner MF, Olson AJ, Goodsell DS. Automated docking to multiple target structures: incorporation of protein mobility and structural water heterogeneity in AutoDock. Proteins 2002;46:34–40. [PubMed: 11746701]
36. Jain AN. Surflex-Dock 2.1: robust performance from ligand energetic modeling, ring flexibility, and knowledge-based search. J Comput Aided Mol Des 2007;21:281–306. [PubMed: 17387436]

37. Verdonk ML, Cole JC, Hartshorn MJ, Murray CW, Taylor RD. Improved protein-ligand docking using GOLD. Proteins 2003;52:609–623. [PubMed: 12910460]
38. Wei BQ, Weaver LH, Ferrari AM, Matthews BW, Shoichet BK. Testing a flexible-receptor docking algorithm in a model binding site. J Mol Biol 2004;337:1161–1182. [PubMed: 15046985]
39. Najmanovich R, Kuttner J, Sobolev V, Edelman M. Side-chain flexibility in proteins upon ligand binding. Proteins 2000;39:261–268. [PubMed: 10737948]
40. Zavodszky MI, Kuhn LA. Side-chain flexibility in protein-ligand binding: the minimal rotation hypothesis. Protein Sci 2005;14:1104–1114. [PubMed: 15772311]
41. Gunasekaran K, Nussinov R. How different are structurally flexible and rigid binding sites? Sequence and structural features discriminating proteins that do and do not undergo conformational change upon ligand binding. J Mol Biol 2007;365:257–273. [PubMed: 17059826]
42. Gutteridge A, Thornton J. Conformational changes observed in enzyme crystal structures upon substrate binding. J Mol Biol 2005;346:21–28. [PubMed: 15663924]
43. Jensen L. Refinement and reliability of macromolecular models based on X-ray diffraction data. Methods Enzymol 1997;B277:353–366. [PubMed: 18488317]
44. Dunitz J, Shomaker V, Trueblood K. Interpretation of atomic displacement parameters from diffraction studies of crystals. J Phys Chem 1988;92:856–867.
45. Merritt EA. Expanding the model: anisotropic displacement parameters in protein structure refinement. Acta Crystallogr D Biol Crystallogr 1999;55:1109–1117. [PubMed: 10329772]
46. Vitkup D, Ringe D, Karplus M, Petsko GA. Why protein R-factors are so large: a self-consistent analysis. Proteins 2002;46:345–354. [PubMed: 11835510]
47. Schmidt A, Lamzin VS. From atoms to proteins. Cell Mol Life Sci 2007;64:1959–1969. [PubMed: 17497239]
48. Bossi RT, Aliverti A, Raimondi D, Fischer F, Zanetti G, et al. A covalent modification of NADP+ revealed by the atomic resolution structure of FprA, a Mycobacterium tuberculosis oxidoreductase. Biochemistry 2002;41:8807–8818. [PubMed: 12102623]
49. Ringe D, Petsko GA. Mapping protein dynamics by X-ray diffraction. Prog Biophys Mol Biol 1985;45:197–235. [PubMed: 3892584]
50. Bourgeois D, Royant A. Advances in kinetic protein crystallography. Curr Opin Struct Biol 2005;15:538–547. [PubMed: 16129597]
51. Schmidt M, Ihee H, Pahl R, Srajer V. Protein-ligand interaction probed by time-resolved crystallography. Methods Mol Biol 2005;305:115–154. [PubMed: 15939996]
52. Katona G, Carpentier P, Niviere V, Amara P, Adam V, et al. Raman-assisted crystallography reveals end-on peroxide intermediates in a nonheme iron enzyme. Science 2007;316:449–453. [PubMed: 17446401]
53. http://www.sciencemag.org/content/vol316/issue5823/images/data/449/DC1/1138885s1.movhttp://www.sciencemag.org/content/vol316/issue5823/images/data/449/DC1/1138885s1.mov
54. Horst R, Wider G, Fiaux J, Bertelsen EB, Horwich AL, et al. Proton-proton Overhauser NMR spectroscopy with polypeptide chains in large structures. Proc Natl Acad Sci U S A 2006;103:15445–15450. [PubMed: 17032756]
55. Grishaev A, Tugarinov V, Kay LE, Trewhella J, Bax A. Refined solution structure of the 82-kDa enzyme malate synthase G from joint NMR and synchrotron SAXS restraints. J Biomol NMR 2008;40:95–106. [PubMed: 18008171]
56. Liang B, Tamm LK. Structure of outer membrane protein G by solution NMR spectroscopy. Proc Natl Acad Sci U S A 2007;104:16140–16145. [PubMed: 17911261]
57. Cioffi M, Hunter CA, Packer MJ, Spitaleri A. Determination of protein-ligand binding modes using complexation-induced changes in (1)h NMR chemical shift. J Med Chem 2008;51:2512–2517. [PubMed: 18366177]
58. Pertinhez TA, Sforca ML, Alves AC, Ramos CR, Ho PL, et al. 1H, 15N and 13C resonance assignments of the apo Sm14-M20(C62V) protein, a mutant of Schistosoma mansoni Sm14. J Biomol NMR 2004;29:553–554. [PubMed: 15243195]
59. Ramos CR, Figueredo RC, Pertinhez TA, Vilar MM, do Nascimento AL, et al. Gene structure and M20T polymorphism of the Schistosoma mansoni Sm14 fatty acid-binding protein. Molecular,

functioanl, and immunoprotection analysis. J Biol Chem 2003;278:12745–12751. [PubMed: 12551912]

60. Ramos C, Spisni A, Oyama SJ, Sforca ML, Ramos H, et al. Stability Improvement of the Fatty Acid Binding Protein Sm14 from S. mansoni by Cys replacement: structural and functional characterization of a vaccine candidate. 2008, submitted.
61. Ishima R, Torchia DA. Protein dynamics from NMR. Nat Struct Biol 2000;7:740–743. [PubMed: 10966641]
62. Eisenmesser EZ, Millet O, Labeikovsky W, Korzhnev DM, Wolf-Watz M, et al. Intrinsic dynamics of an enzyme underlies catalysis. Nature 2005;438:117–121. [PubMed: 16267559]
63. Mulder FA, Mittermaier A, Hon B, Dahlquist FW, Kay LE. Studying excited states of proteins by NMR spectroscopy. Nat Struct Biol 2001;8:932–935. [PubMed: 11685237]
64. Wuthrich K. NMR studies of structure and function of biological macromolecules (Nobel Lecture). J Biomol NMR 2003;27:13–39. [PubMed: 15143746]
65. Damm KL, Carlson HA. Exploring experimental sources of multiple protein conformations in structure-based drug design. J Am Chem Soc 2007;129:8225–8235. [PubMed: 17555316]
66. Fanelli F, De Benedetti PG. Computational modeling approaches to structure-function analysis of G protein-coupled receptors. Chem Rev 2005;105:3297–3351. [PubMed: 16159154]
67. Meagher KL, Carlson HA. Incorporating protein flexibility in structure-based drug discovery: using HIV-1 protease as a test case. J Am Chem Soc 2004;126:13276–13281. [PubMed: 15479081]
68. McCammon JA. Target flexibility in molecular recognition. Biochim Biophys Acta 2005;1754:221–224. [PubMed: 16181817]
69. Lin JH, Perryman AL, Schames JR, McCammon JA. The relaxed complex method: Accommodating receptor flexibility for drug design with an improved scoring scheme. Biopolymers 2003;68:47–62. [PubMed: 12579579]
70. Lin JH, Perryman AL, Schames JR, McCammon JA. Computational drug design accommodating receptor flexibility: the relaxed complex scheme. J Am Chem Soc 2002;124:5632–5633. [PubMed: 12010024]
71. Schames JR, Henchman RH, Siegel JS, Sotriffer CA, Ni H, et al. Discovery of a novel binding trench in HIV integrase. J Med Chem 2004;47:1879–1881. [PubMed: 15055986]
72. McCammon JA, Gelin BR, Karplus M. Dynamics of folded proteins. Nature 1977;267:585–590. [PubMed: 301613]
73. Freddolino PL, Arkhipov AS, Larson SB, McPherson A, Schulten K. Molecular dynamics simulations of the complete satellite tobacco mosaic virus. Structure 2006;14:437–449. [PubMed: 16531228]
74. Onaran, H.; Scheer, A.; Cotecchia, S.; Costa, T. Handbook of Experimental Pharmacology. Springer; Heidelberg: 2000. A look into receptor efficacy. From the signalling network of the cell to the intramolecular motion of the receptor; p. 217-280.
75. Kenakin T. Efficacy at G-protein-coupled receptors. Nat Rev Drug Discov 2002;1:103–110. [PubMed: 12120091]
76. Cremades N, Sancho J, Freire E. The native-state ensemble of proteins provides clues for folding, misfolding and function. Trends Biochem Sci 2006;31:494–496. [PubMed: 16870449]
77. Barnett-Norris J, Lynch D, Reggio PH. Lipids, lipid rafts and caveolae: their importance for GPCR signaling and their centrality to the endocannabinoid system. Life Sci 2005;77:1625–1639. [PubMed: 15993425]
78. Palczewski K. G protein-coupled receptor rhodopsin. Annu Rev Biochem 2006;75:743–767. [PubMed: 16756510]
79. Cherezov V, Rosenbaum DM, Hanson MA, Rasmussen SG, Thian FS, et al. High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. Science 2007;318:1258–1265. [PubMed: 17962520]
80. Rasmussen SG, Choi HJ, Rosenbaum DM, Kobilka TS, Thian FS, et al. Crystal structure of the human beta2 adrenergic G-protein-coupled receptor. Nature 2007;450:383–387. [PubMed: 17952055]
81. Salom D, Lodowski DT, Stenkamp RE, Le Trong I, Golczak M, et al. Crystal structure of a photoactivated deprotonated intermediate of rhodopsin. Proc Natl Acad Sci U S A 2006;103:16123–16128. [PubMed: 17060607]

82. Fanelli F, De Benedetti PG. Inactive and active states and supramolecular organization of GPCRs: insights from computational modeling. J Comput Aided Mol Des 2006;20:449–461. [PubMed: 17009093]
83. Feng X, Muller T, Mizrachi D, Fanelli F, Segaloff DL. An intracellular loop (IL2) residue confers different basal constitutive activities to the human lutropin receptor and human thyrotropin receptor through structural communication between IL2 and helix 6, via helix 3. Endocrinology 2008;149:1705–1717. [PubMed: 18162522]
84. Angelova K, Fanelli F, Puett D. Contributions of intracellular loops 2 and 3 of the lutropin receptor in Gs coupling. Mol Endocrinol 2008;22:126–138. [PubMed: 17872379]
85. Seeber M, De Benedetti PG, Fanelli F. Molecular dynamics simulations of the ligand-induced chemical information transfer in the 5-HT(1A) receptor. J Chem Inf Comput Sci 2003;43:1520–1531. [PubMed: 14502486]
86. Vitale RM, Pedone C, De Benedetti PG, Fanelli F. Structural features of the inactive and active states of the melanin-concentrating hormone receptors: insights from molecular simulations. Proteins 2004;56:430–448. [PubMed: 15229878]
87. Raimondi F, Seeber M, Benedetti PG, Fanelli F. Mechanisms of inter- and intramolecular communication in GPCRs and G proteins. J Am Chem Soc 2008;130:4310–4325. [PubMed: 18335928]
88. Yan F, Mosier PD, Westkaemper RB, Roth BL. Galpha-subunits differentially alter the conformation and agonist affinity of kappa-opioid receptors. Biochemistry 2008;47:1567–1578. [PubMed: 18205395]
89. http://www.chim.unipr.it/modellistica/http://www.chim.unipr.it/modellistica/
90. Brooks B, Bruccoleri R, Olafson B, States D, Swaminathan S, et al. A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. J Comput Chem 1983;4:187–217.
91. Berendsen, HJ.; Postma, J.; van Gusteren, W.; Hermans, J. Intermolecular forces. D. Reidel Publishing Company; Dordrecht: 1981. Interaction model for water in relation to protein hydration; p. 331-342.
92. Jorgensen W, Maxwell D, Tirado-Rives J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. J Am Chem Soc 1996;118:11225–11236.
93. Case DA, Cheatham TE 3rd, Darden T, Gohlke H, Luo R, et al. The Amber biomolecular simulation programs. J Comput Chem 2005;26:1668–1688. [PubMed: 16200636]
94. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, et al. GROMACS: fast, flexible, and free. J Comput Chem 2005;26:1701–1718. [PubMed: 16211538]
95. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, et al. Scalable molecular dynamics with NAMD. J Comput Chem 2005;26:1781–1802. [PubMed: 16222654]
96. Spyrakis, F.; Kellogg, GE.; Amadasi, A.; Cozzini, P. Frontiers in Drug Design and Discovery. Bentham Science Publisher LTD; 2007. Scoring functions for virtual screening; p. 317-379.
97. Warren GL, Andrews CW, Capelli AM, Clarke B, LaLonde J, et al. A critical assessment of docking programs and scoring functions. J Med Chem 2006;49:5912–5931. [PubMed: 17004707]
98. Kitchen DB, Decornez H, Furr JR, Bajorath J. Docking and scoring in virtual screening for drug discovery: methods and applications. Nat Rev Drug Discov 2004;3:935–949. [PubMed: 15520816]
99. Leach AR, Shoichet BK, Peishoff CE. Prediction of protein-ligand interactions. Docking and scoring: successes and gaps. J Med Chem 2006;49:5851–5855. [PubMed: 17004700]
100. Feher M. Consensus scoring for protein-ligand interactions. Drug Discov Today 2006;11:421–428. [PubMed: 16635804]
101. Teramoto R, Fukunishi H. Structure-based virtual screening with supervised consensus scoring: evaluation of pose prediction and enrichment factors. J Chem Inf Model 2008;48:747–754. [PubMed: 18318474]
102. Clark RD, Strizhev A, Leonard JM, Blake JF, Matthew JB. Consensus scoring for ligand/protein interactions. J Mol Graph Model 2002;20:281–295. [PubMed: 11858637]
103. Li G, Cui Q. pKa Calculations with QM/MM Free Energy Perturbations. J. Phys. Chem. B 2003;107:14521–14528.

104. Riccardi D, Schaefer P, Yang Y, Yu H, Ghosh N, et al. Development of Effective Quantum Mechanical/Molecular Mechanical (QM/MM) Methods for Complex Biological Processes. J. Phys. Chem. B 2006;110:6458–6469. [PubMed: 16570942]
105. Rosta E, Klahn M, Warshel A. Towards Accurate Ab Initio QM/MM Calculations of Free-Energy Profiles of Enzymatic Reactions. J. Phys. Chem. B 2006;110:2934–2941. [PubMed: 16471904]
106. Reddy MR, Singh UC, Erion MD. Development of a Quantum Mechanics-Based Free-Energy Perturbation Method: Use in the Calculation of Relative Solvation Free Energies. J. Am. Chem. Soc 2004;126:6224–6225. [PubMed: 15149207]
107. Reddy MR, Erion MD. Relative Binding Affinities of Fructose-1,6-Bisphosphatase Inhibitors Calculated Using a Quantum Mechanics-Based Free Energy Perturbation Method. J. Am. Chem. Soc 2007;129:9296–9297. [PubMed: 17616196]
108. Bohm HJ. The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. J Comput Aided Mol Des 1994;8:243–256. [PubMed: 7964925]
109. Gohlke H, Klebe G. Statistical potentials and scoring functions applied to protein-ligand binding. Curr Opin Struct Biol 2001;11:231–235. [PubMed: 11297933]
110. Spyrakis F, Amadasi A, Fornabaio M, Abraham DJ, Mozzarelli A, et al. The consequences of scoring docked ligand conformations using free energy correlations. Eur J Med Chem 2007;42:921–933. [PubMed: 17346861]
111. Schulz-Gasch T, Stahl M. Scoring functions for protein-ligand interactions: a critical perspective. Drug Discov. Today: Technol 2004;1:231–239.
112. Rosenfeld RJ, Goodsell DS, Musah RA, Morris GM, Goodin DB, et al. Automated docking of ligands to an artificial active site: augmenting crystallographic analysis with computer modeling. J Comput Aided Mol Des 2003;17:525–536. [PubMed: 14703123]
113. Zavodszky MI, Lei M, Thorpe MF, Day AR, Kuhn LA. Modeling correlated main-chain motions in proteins for flexible molecular recognition. Proteins 2004;57:243–261. [PubMed: 15340912]
114. Zacharias M, Sklenar H. Harmonic modes as variables to approximately account for receptor flexibility in ligand-receptor docking simulations: applications to DNA minor groove ligand complex. J Comput Chem 1999;20:287–300.
115. Cavasotto CN, Kovacs JA, Abagyan RA. Representing receptor flexibility in ligand docking through relevant normal modes. J Am Chem Soc 2005;127:9632–9640. [PubMed: 15984891]
116. Zentgraf M, Fokkens J, Sotriffer CA. Addressing protein flexibility and ligand selectivity by “in situ cross-docking”. Chem Med Chem 2006;1:1355–1359. [PubMed: 17024701]
117. Totrov M, Abagyan R. Flexible ligand docking to multiple receptor conformations: a practical alternative. Curr Opin Struct Biol 2008;18:178–184. [PubMed: 18302984]
118. Nabuurs SB, Wagener M, de Vlieg J. A flexible approach to induced fit docking. J Med Chem 2007;50:6507–6518. [PubMed: 18031000]
119. Kallblad P, Dean PM. Efficient conformational sampling of local side-chain flexibility. J Mol Biol 2003;326:1651–1665. [PubMed: 12595271]
120. Taylor RD, Jewsbury PJ, Essex JW. FDS: flexible ligand and receptor docking with a continuum solvent model and soft-core energy function. J Comput Chem 2003;24:1637–1656. [PubMed: 12926007]
121. Tietze S, Apostolakis J. GlamDock: development and validation of a new docking tool on several thousand protein-ligand complexes. J Chem Inf Model 2007;47:1657–1672. [PubMed: 17585857]
122. Sherman W, Day T, Jacobson MP, Friesner RA, Farid R. Novel procedure for modeling ligand/receptor induced fit effects. J Med Chem 2006;49:534–553. [PubMed: 16420040]
123. Schlessinger A, Rost B. Protein flexibility and rigidity predicted from sequence. Proteins 2005;61:115–126. [PubMed: 16080156]
124. Bottegoni G, Kufareva I, Totrov M, Abagyan R. A new method for ligand docking to flexible receptors by dual alanine scanning and refinement (SCARE). J Comput Aided Mol Des 2008;22:311–325. [PubMed: 18273556]
125. Krol M, Tournier AL, Bates PA. Flexible relaxation of rigid-body docking solutions. Proteins 2007;68:159–169. [PubMed: 17397060]

126. Zhao Y, Sanner MF. FLIPDock: docking flexible ligands into flexible receptors. Proteins 2007;68:726–737. [PubMed: 17523154]
127. Zavodszky MI, Sanschagrin PC, Korde RS, Kuhn LA. Distilling the essential features of a protein surface for improving protein-ligand docking, scoring, and virtual screening. J Comput Aided Mol Des 2002;16:883–902. [PubMed: 12825621]
128. Heringa J, Argos P. Strain in protein structures as viewed through nonrotameric side chains: II. effects upon ligand binding. Proteins 1999;37:44–55. [PubMed: 10451549]
129. Schnecke V, Kuhn LA. Virtual screening with solvation and ligand-induced complementarity. Perspectives in Drug Design and Discovery 2000;20:171–190.
130. Kuhn, LA. Computational and Structural Approaches to Drug Discovery: Ligand-Protein Interactions. RSC Publishing; Cambridge: 2008. Strength in flexibility: modeling side-chain conformational change in docking and screening; p. 181-191.
131. Arora, S. Optimizing side-chain interactions in protein-ligand interfaces. Michigan State University; East Lansing: 2005.
132. Sun Y, MacRae TH. Small heat shock proteins: molecular structure and chaperone function. Cell Mol Life Sci 2005;62:2460–2476. [PubMed: 16143830]
133. Kurachi Y, North A. Ion channels: their structure, function and control - an overview. J Physiol 2004;554:245–247.
134. Mattevi A, Rizzi M, Bolognesi M. New structures of allosteric proteins revealing remarkable conformational changes. Curr Opin Struct Biol 1996;6:824–829. [PubMed: 8994883]
135. Ahmed A, Gohlke H. Multiscale modeling of macromolecular conformational changes combining concepts from rigidity and elastic network theory. Proteins 2006;63:1038–1051. [PubMed: 16493629]
136. Jacobs DJ, Rader AJ, Kuhn LA, Thorpe MF. Protein flexibility predictions using graph theory. Proteins 2001;44:150–165. [PubMed: 11391777]
137. Tama F, Gadea FX, Marques O, Sanejouand YH. Building-block approach for determining low-frequency normal modes of macromolecules. Proteins 2000;41:1–7. [PubMed: 10944387]
138. Atilgan AR, Durell SR, Jernigan RL, Demirel MC, Keskin O, et al. Anisotropy of fluctuation dynamics of proteins with an elastic network model. Biophys J 2001;80:505–515. [PubMed: 11159421]
139. Wells S, Menor S, Hespenheide B, Thorpe MF. Constrained geometric simulation of diffusive motion in proteins. Phys Biol 2005;2:S127–136. [PubMed: 16280618]
140. Arkin MR, Randal M, DeLano WL, Hyde J, Luong TN, et al. Binding of small molecules to an adaptive protein-protein interface. Proc Natl Acad Sci U S A 2003;100:1603–1608. [PubMed: 12582206]
141. Knegtel RM, Kuntz ID, Oshiro CM. Molecular docking to ensembles of protein structures. J Mol Biol 1997;266:424–440. [PubMed: 9047373]
142. Sotriffer CA, Dramburg I. "In situ cross-docking" to simultaneously address multiple targets. J Med Chem 2005;48:3122–3125. [PubMed: 15857117]
143. Macchiarulo A, Nobeli I, Thornton JM. Ligand selectivity and competition between enzymes in silico. Nat Biotechnol 2004;22:1039–1045. [PubMed: 15286657]
144. Freire E. Statistical thermodynamic linkage between conformational and binding equilibria. Adv Protein Chem 1998;51:255–279. [PubMed: 9615172]
145. Ma B, Kumar S, Tsai CJ, Nussinov R. Folding funnels and binding mechanisms. Protein Eng 1999;12:713–720. [PubMed: 10506280]
146. Egbertson MS. HIV integrase inhibitors: from diketoacids to heterocyclic templates: a history of HIV integrase medicinal chemistry at Merck West Point and Merck Rome (IRBM). Curr Top Med Chem 2007;7:1251–1272. [PubMed: 17627556]
147. Klebe G, Kramer O, Sotriffer C. Strategies for the design of inhibitors of aldose reductase, an enzyme showing pronounced induced-fit adaptations. Cell Mol Life Sci 2004;61:783–793. [PubMed: 15095003]

148. Sotriffer CA, Kramer O, Klebe G. Probing flexibility and "induced-fit" phenomena in aldose reductase by comparative crystal structure analysis and molecular dynamics simulations. Proteins 2004;56:52–66. [PubMed: 15162486]

149. Steuber H, Zentgraf M, Gerlach C, Sotriffer CA, Heine A, et al. Expect the unexpected or caveat for drug designers: multiple structure determinations using aldose reductase crystals treated under varying soaking and co-crystallisation conditions. J Mol Biol 2006;363:174–187. [PubMed: 16952371]

150. Steuber H, Zentgraf M, Podjarny A, Heine A, Klebe G. High-resolution crystal structure of aldose reductase complexed with the novel sulfonylpyridazinone inhibitor exhibiting an alternative active site anchoring group. J Mol Biol 2006;356:45–56. [PubMed: 16337231]

151. Da Settimo F, Primofiore G, La Motta C, Sartini S, Taliani S, et al. Naphtho[1,2-d]isothiazole acetic acid derivatives as a novel class of selective aldose reductase inhibitors. J Med Chem 2005;48:6897–6907. [PubMed: 16250648]

152. Zentgraf M, Steuber H, Koch C, La Motta C, Sartini S, et al. How reliable are current docking approaches for structure-based drug design? Lessons from aldose reductase. Angew Chem Int Ed Engl 2007;46:3575–3578. [PubMed: 17394265]

153. Soliva R, Gelpi JL, Almansa C, Virgili M, Orozco M. Dissection of the recognition properties of p38 MAP kinase. Determination of the binding mode of a new pyridinyl-heterocycle inhibitor family. J Med Chem 2007;50:283–293. [PubMed: 17228870]

154. Almansa C, Virgili M. Pyrazolopyridine Derivatives.

155. Van Drie J. Computer-aided drug design: the next 20 years. J Comput Aided Mol Des 2007;21:591–601. [PubMed: 17989929]

156. Richards FM. Whatever happened to the fun? An autobiographical investigation. Annu Rev Biophys Biomol Struct 1997;26:1–25. [PubMed: 9241411]

157. Allen, G. Life Science in the Twentieth Century. Cambridge University Press; Cambridge: 1978.

158. Mitsutake A, Sugita Y, Okamoto Y. Generalized-ensemble algorithms for molecular simulations of biopolymers. Biopolymers 2001;60:96–123. [PubMed: 11455545]

159. Fornabaio M, Cozzini P, Mozzarelli A, Abraham DJ, Kellogg GE. Simple, intuitive calculations of free energy of binding for protein-ligand complexes. 2. Computational titration and pH effects in molecular models of neuraminidase-inhibitor complexes. J Med Chem 2003;46:4487–4500. [PubMed: 14521411]

160. Spyrakis F, Fornabaio M, Cozzini P, Mozzarelli A, Abraham DJ, et al. Computational titration analysis of a multiprotic HIV-1 protease-ligand complex. J Am Chem Soc 2004;126:11764–11765. [PubMed: 15382890]

161. Huang SY, Zou X. Ensemble docking of multiple protein structures: considering protein structural variations in molecular docking. Proteins 2007;66:399–421. [PubMed: 17096427]

162. Limongelli V, Marinelli L, Cosconati S, Braun HA, Schmidt B, et al. Ensemble-Docking Approach on BACE-1: Pharmacophore Perception and Guidelines for Drug Design. Chem Med Chem 2007;2:667–678. [PubMed: 17407105]

163. Barril X, Fradera X. Incorporating protein flexibility into docking and structure-based drug design. Expert Opin Drug Discov 2006;1:335–349.

164. Chang CE, Gilson MK. Free energy, entropy, and induced fit in host-guest recognition: calculations with the second-generation mining minima algorithm. J Am Chem Soc 2004;126:13156–13164. [PubMed: 15469315]

165. Chen W, Chang CE, Gilson MK. Calculation of cyclodextrin binding affinities: energy, entropy, and implications for drug design. Biophys J 2004;87:3035–3049. [PubMed: 15339804]

166. Chang CE, Chen W, Gilson MK. Ligand configurational entropy and protein binding. Proc Natl Acad Sci U S A 2007;104:1534–1539. [PubMed: 17242351]

167. Rao S, Sanschagrin PC, Greenwood JR, Repasky MP, Sherman W, et al. Improving database enrichment through ensemble docking. J Comput Aided Mol Des. 2008

168. DeLano, WL. The PyMOL Molecular Graphics System. DeLano Scientific; Palo Alto, CA, USA:

169. Hespenheide BM, Rader AJ, Thorpe MF, Kuhn LA. Identifying protein folding cores from the evolution of flexible regions during unfolding. J Mol Graph Model 2002;21:195–207. [PubMed: 12463638]

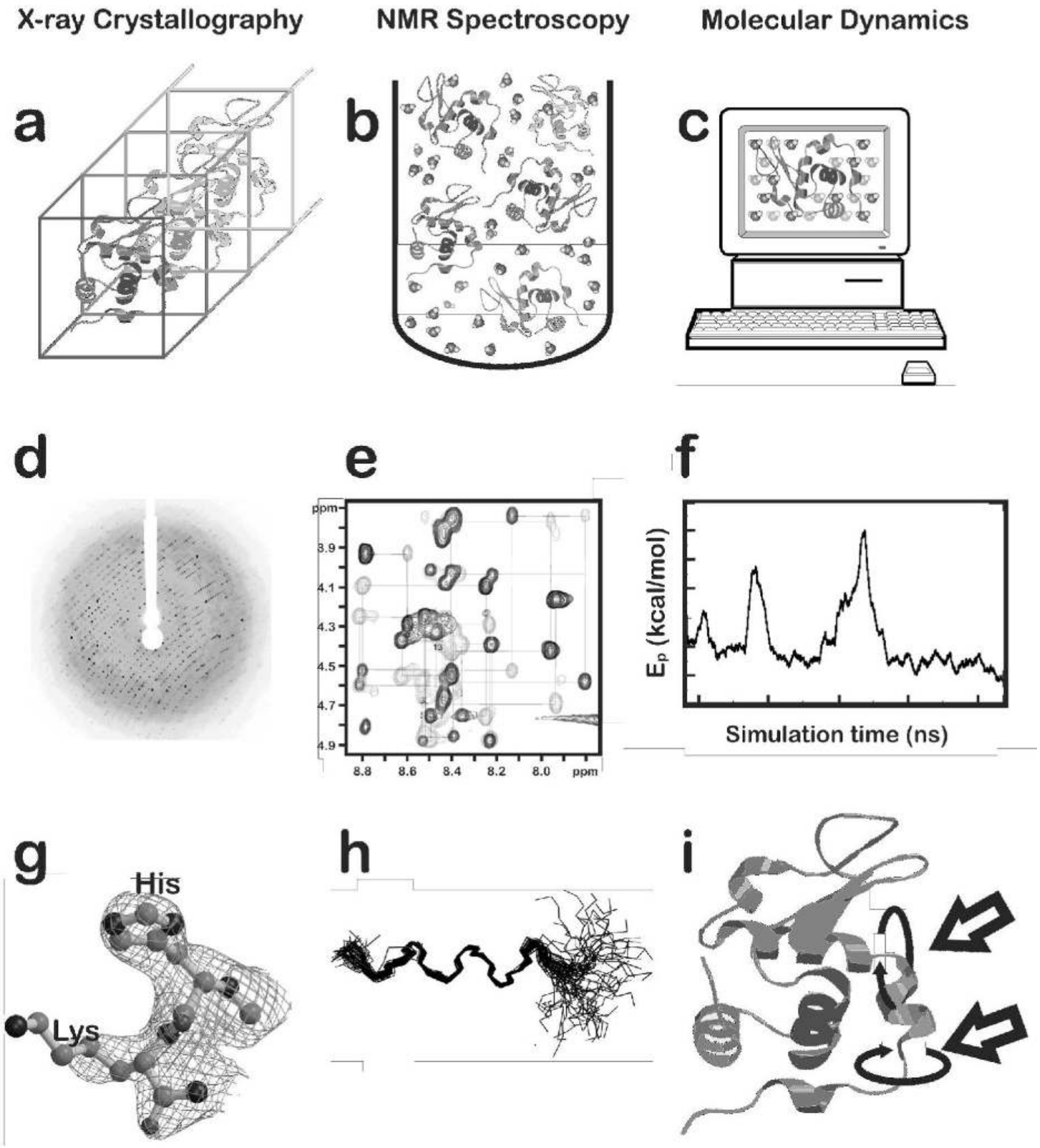


**Figure 1.**

X-ray crystallography, NMR spectroscopy and computational Molecular Dynamics. a) A crystal illustrating the uniformity of the protein molecular structures within their unit cells necessary for X-ray diffraction. Ordered water molecules (not shown) partially hydrate the structure; b) protein molecule in solution for NMR experiment. Ordered water and other (not shown) ions solvate the structure, but many other solvent molecules are not structurally ordered; c) protein immersed in virtual (water) solvent for computational Molecular Dynamics with periodic boundary conditions; d) diffraction pattern from an X-ray data collection; e) typical 2D NOESY spectrum (http://www.sanger.ac.uk/Users/sgj/thesis/html/node86.html>) from protein NMR; f) Molecular Dynamics potential energy as a function of simulation time for coarse-grained motions; g) electron density maps for histidine and lysine residues. The density around more labile (high B-factor) atoms is either diffuse or non-existent (e.g., the NZ atom of lysine); h) typical backbone traces for structures meeting constraints determined by NMR NOESY experiments. The high degree of flexibility at either end of the structure is evident; and i) protein structure illustrating (see arrows) typical large scale motions that may be observed with computational Molecular Dynamics.

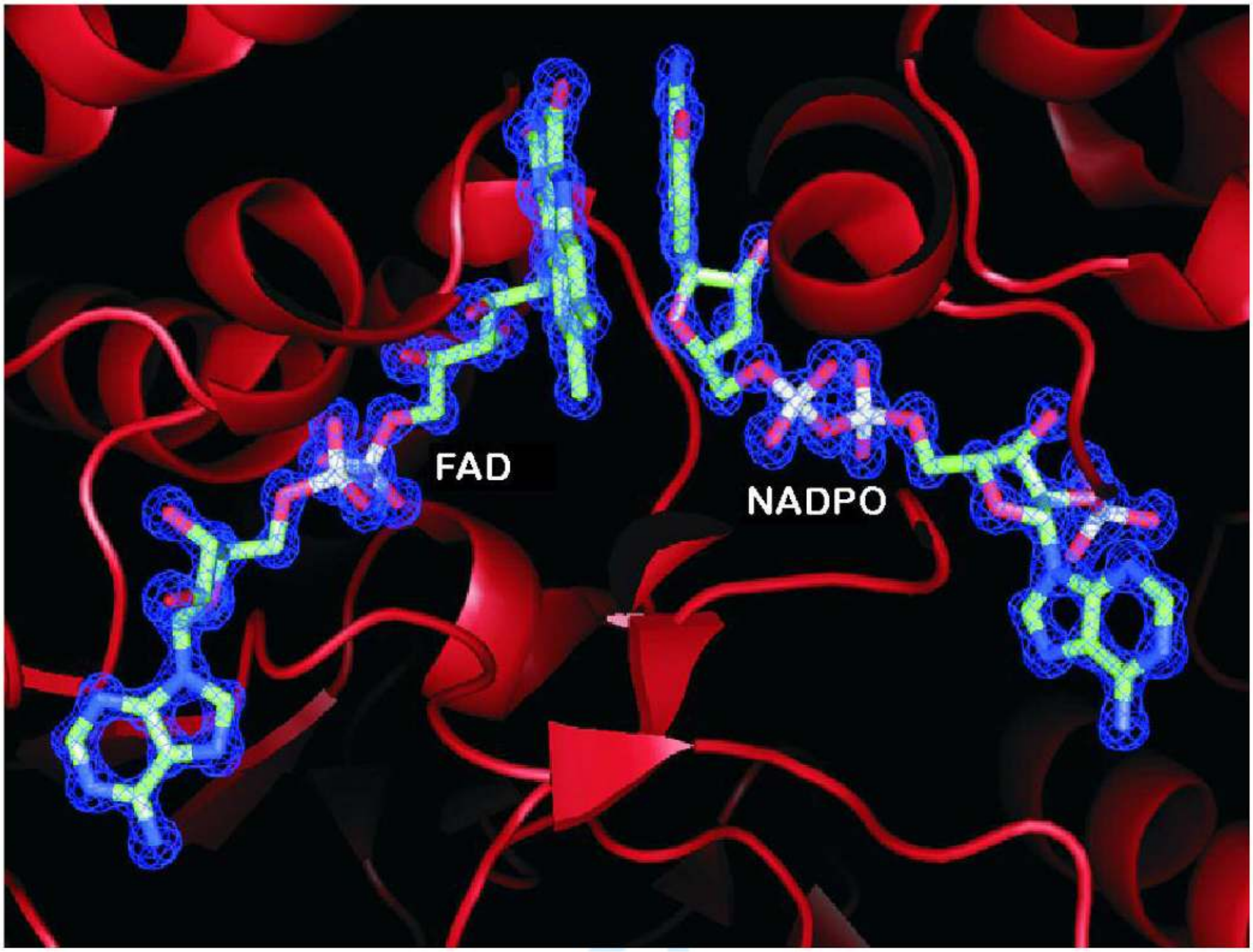


**Figure 2.**
View of the active site of Mycobacterium tuberculosis FprA. The FAD cofactor and a covalently modified NADP+ (labeled NADPO) identified in the 1.05 Å resolution crystal structure (PDB code, 1LQT) are highlighted.[48] Shown in blue is the $2F_o$-$F_c$ map contoured at the level of 2 σ above its mean. (Prepared using PyMOL.)[168]

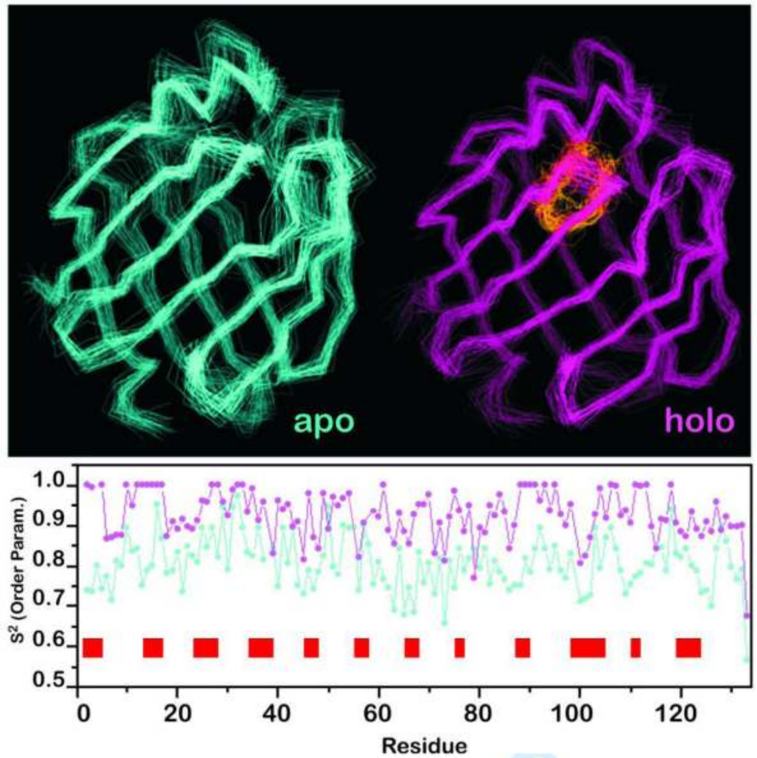


**Figure 3.**

Backbone dynamics data derived from $^{15}N$-$^{1}H$ NMR experiments on mutant Sm14-M20 (C62V). The NMR-derived backbone traces for both apo and holo protein forms are shown on top, with the bound fatty acid shown in yellow (in the holo form). The order parameters $S^2$ as functions of residue number are shown in the bottom graph for apo (cyan) and holo (magenta). The red bars along the residue axis indicate the loop regions of the protein, where a larger difference in $S^2$ between the two forms is observed.

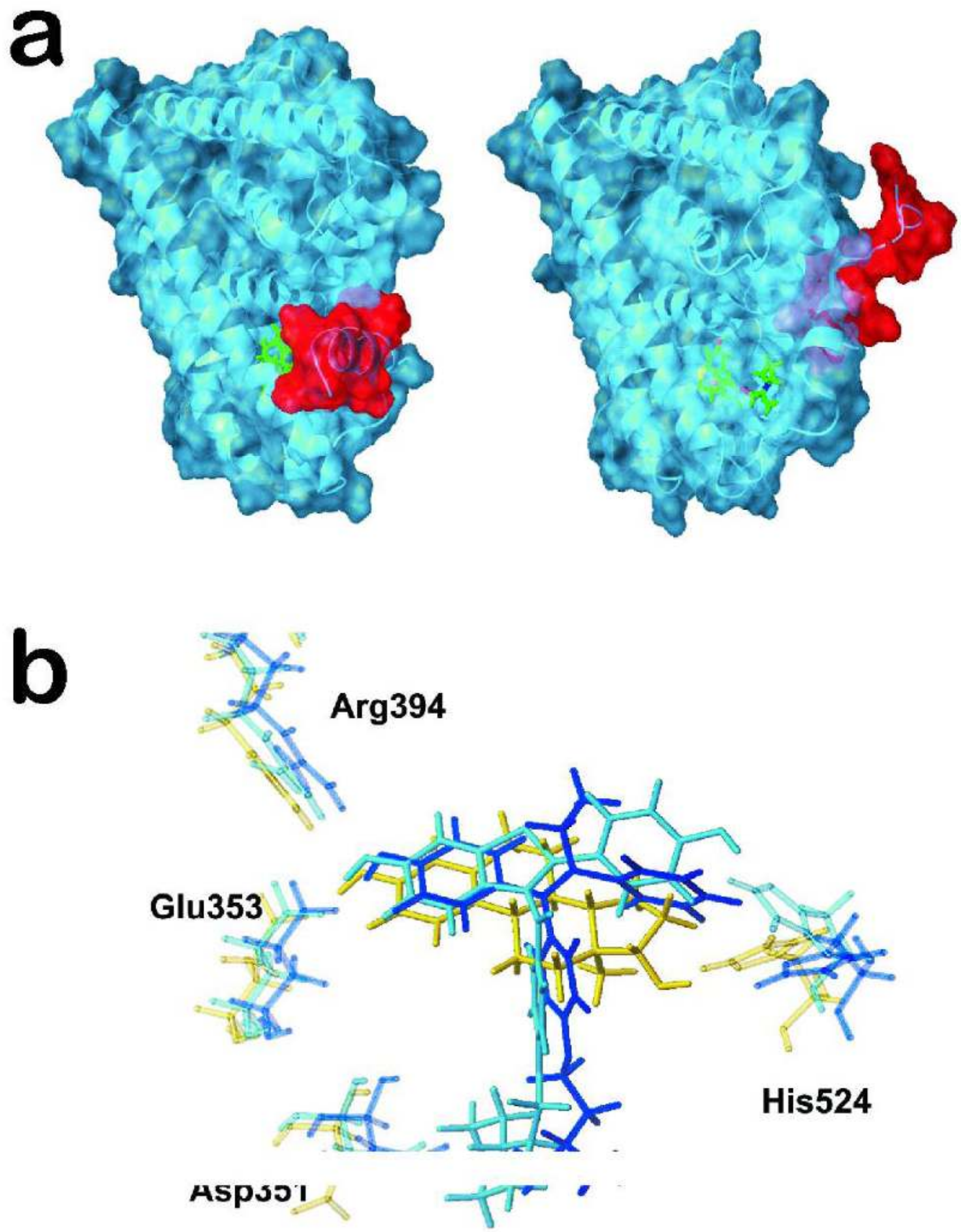


**Figure 4.**
Examples of large motions (coarse-grained) and small motions in the estrogen receptor α protein. a) The different positions of helix 12 (red) depending of the type of bound ligand (green): agonists such as estradiol induce and stabilize the closed conformation (left) while antagonists such as tamoxifen prevent helix 12 from adopting the agonist-induced conformation (right); b) small adjustments of the His524 residue within the ERα binding site depending on the ligand. Yellow bonds indicate the positions of His524 when the natural ligand estradiol (also yellow) is bound; light blue bonds illustrate the antitumor drug raloxifen and its effect on His524; dark blue bonds represents the drug tamoxifen and its effect on His524.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

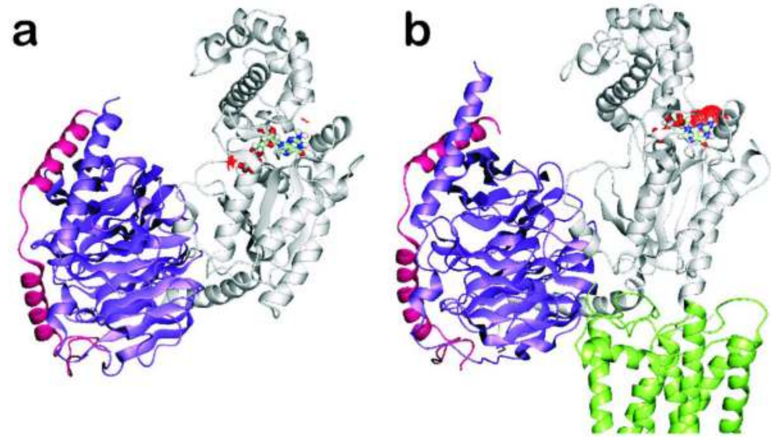


**Figure 5.**

Average minimized structures of a) free and b) TXA2-bound heterotrimeric $G_q$.[80] TXA2 and the G protein α-, β- and γ-subunits are, respectively, colored in green, gray, violet and magenta. The GDP molecule is colored by atom type. Red dots indicate the solvent accessible surface of GDP, which is exposed to solvent upon receptor binding.[80] Only the intracellular half of the receptor is shown.

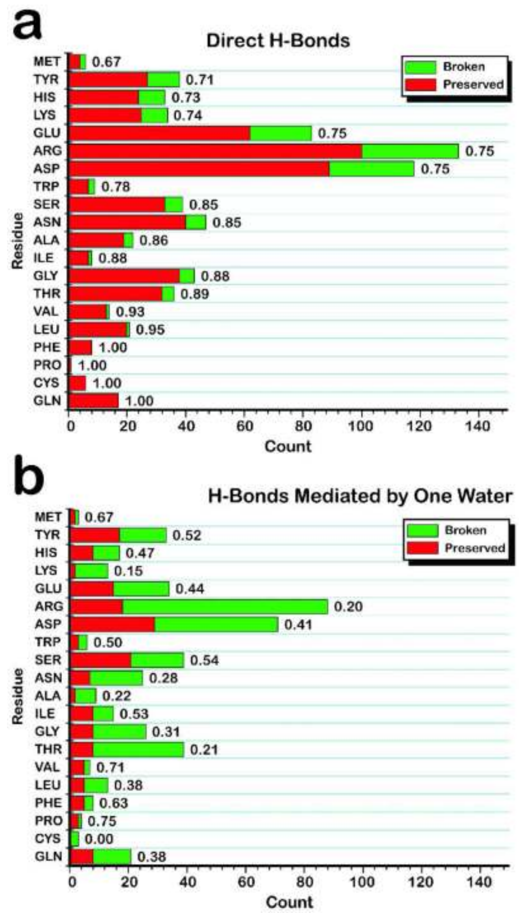


**Figure 6.**

a) Direct intra-protein hydrogen bonds tend to be preserved upon ligand binding. Ligand-free and ligand-bound crystal structures for 32 proteins with low pairwise sequence identity (<30%) were analyzed. The percentage of direct, intra-protein hydrogen bonds preserved upon ligand binding (red) is compared with the percentage broken (green) for each residue type, including main-chain and side-chain hydrogen bonds. Typically 75% or more of direct hydrogen bonds are preserved. b) Intra-protein water-mediated hydrogen bonds tend to break upon ligand binding. Analysis was performed for intra-protein hydrogen bonds mediated by one water molecule. The trend is opposite to that found for direct hydrogen bonds: 50-80% of water-mediated hydrogen bonds are broken upon ligand binding. Details: All residues containing an atom ≤4 Å from the ligand (in the ligand-bound structure, or ≤ 4 Å from the ligand superimposed into the ligand-free structure) or within 4 Å of a water molecule bridging between the protein and ligand, were kept for analysis. Intra-protein hydrogen bonds were initially identified as having a donor-acceptor distance ≤ 3.6 Å, hydrogen-acceptor distance ≤ 2.6 Å, and donor-H-acceptor angle of 90-180°. This set was screened by a hydrogen bond energy function[169] evaluating detailed atom chemistry-dependent features, to ensure that very weak hydrogen bonds were excluded. 60 of the 64 structures had resolution of 2.2 Å or better, and the remaining 4 had resolution ≤ 2.6 Å.

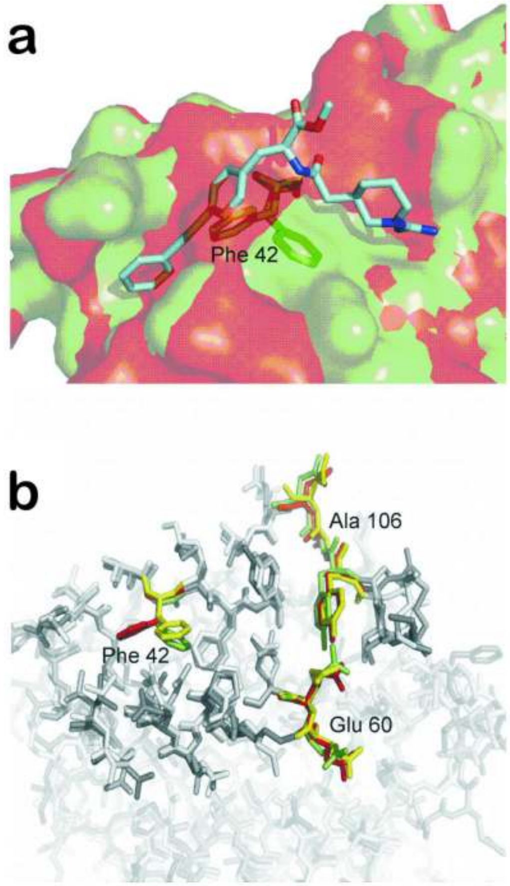


**Figure 7.**

Adaptive nature of the protein-protein interface of the cytokine interleukin-2. a) Overlay of the unbound (red) (PDB code: 1M47) and bound (green) (PDB code: 1M48) protein structure of IL-2 together with a small molecule (cyan) that buries into a groove not seen in the free structure of IL-2. Residue Phe42, whose resultant movement after small molecule binding opens the groove, is depicted in stick representation. b) Overlay of the protein-protein interface region (opaque sticks) of IL-2 in unbound (red) and bound (green) form. In addition, a snapshot from a FRODA simulation started from the unbound state is shown (yellow), which demonstrates that the movement of Phe42 can even be observed in the absence of the ligand, leading to a transient pocket opening. Interestingly, regions for which no movement was observed by experiment (around Glu60 and Ala106) also remain immobile during the simulation. (Pfleger, Metz, Kopitz and Gohlke, unpublished results.)

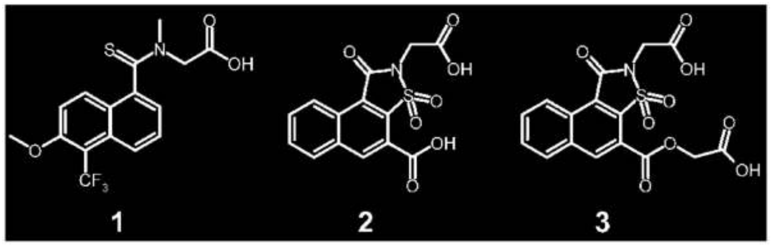


**Figure 8.**
Significantly different binding-site conformations are induced in aldose reductase upon binding of the inhibitors tolrestat (**1**) and analogs (**2,3**).

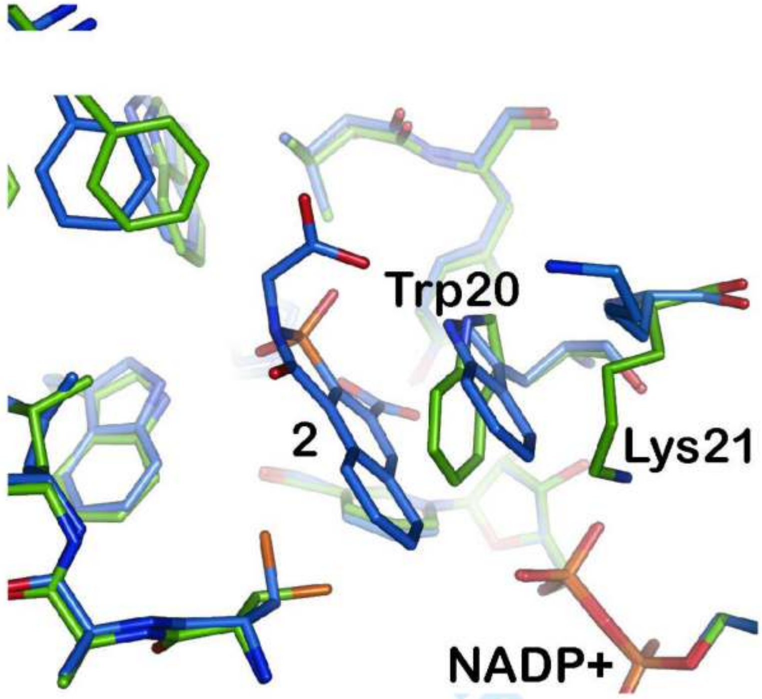


**Figure 9.**

The complex of aldose reductase with **2** (blue) shows an unexpected conformational change in the binding site compared to the standard conformation previously observed (green). The side chain of Trp20 is rotated by 35° and, more importantly, the salt bridge between the side chain of Lys21 and the $NADP^+$ cofactor is broken.

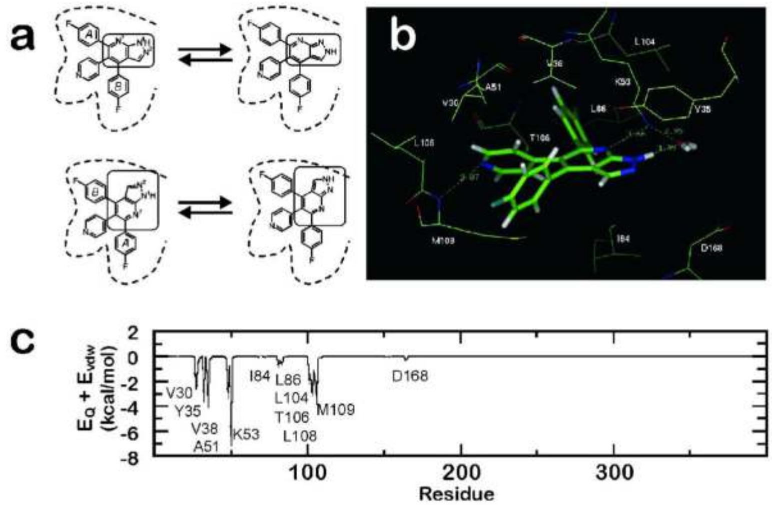


**Figure 10.**

Determination of the binding mode of a pyridinyl-heterocycle inhibitor binding to p38 MAP kinase. a) The four possible binding modes of compound **4** corresponding to the two possible tautomers (N1-H (upper left) and N2-H (upper right)) and their respective pseudosymmetric arrangements arising from a 180° rotation around the pyridine-pyrazolopyridine axis (bottom). For clarity, the central core of the molecule is marked in all cases and key residues defining the binding pocket are displayed as reference; b) detail of the optimum binding mode for compound **4** derived from MD simulations; significant interactions, including a water bridge between **4** and Lys53, are shown; and c) interaction profile in kcal/mol for the sum of electrostatic and van der Waals energy for the residues of p38α MAP kinase and compound **4**. Key residues for binding are noted. (Adapted from ref. [153].)