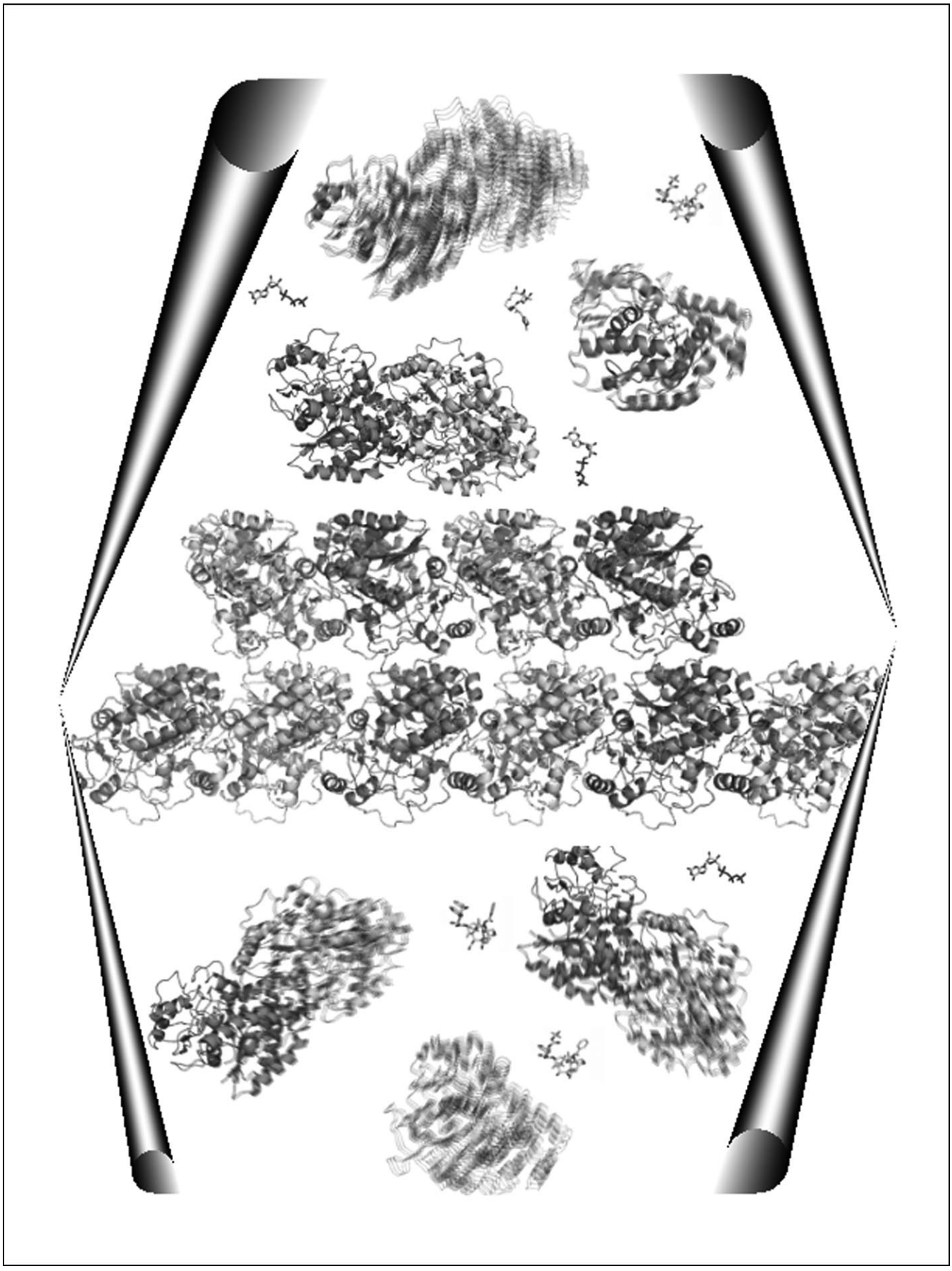

 DOI: 10.1002/cmdc.200700087 © 2007 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

DOI: 10.1002/cmdc.200700087

Overcoming the Inadequacies or Limitations of Experimental Structures as Drug Targets by Using Computational Modeling Tools and Molecular Dynamics Simulations**

Esther Marco*[a] and Federico Gago*[b]

X-ray crystallography, NMR spectroscopy, and cryoelectron microscopy stand out as powerful tools that enable us to obtain atomic detail about biomolecules that can be potentially targeted by drugs. This knowledge is essential if virtual screening or structure-based ligand-design methods are going to be used in drug discovery. However, the macromolecule of interest is not always amenable to these types of experiment or, as is often the case, the conformation found experimentally cannot be used directly for docking studies because of significant changes between apo and bound forms. Furthermore, sometimes the desired insight into the binding mechanism cannot be gained because the structure of the ligand–receptor complex, not having been time-resolved, represents the endpoint of the binding process and therefore retains little or no information about the intermediate stages that led to its creation. Molecular dynamics (MD) simulations are routinely applied these days to the study of biomolecular systems with the aims of sampling configuration space more efficiently and getting a better understanding of the factors that determine structural stability and relevant biophysical and biochemical processes such as protein folding, ligand binding, and enzymatic reactions. This field has matured significantly in recent years, and strategies have been devised (for example activated, steered, or targeted MD) that allow the calculated trajectories to be biased in attempts to properly shape a ligand binding pocket or simulate large-scale motions involving one or more protein domains. On the other hand, low-frequency motions can be simulated quite inexpensively by calculation of normal modes which allow the investigation of alternative receptor conformations. Selected examples in which these methods have been applied to several medicinal chemistry and in silico pharmacology endeavors are presented.

Introduction

Researchers' ideas about proteins and nucleic acids have changed dramatically in the last 50 years. Our perception of these important biological macromolecules has evolved, and we see them no more as formless chemical compounds but as precisely structured molecular machines with tailored biological purposes. The views of DNA as a rather monotonous repeating polymer and of proteins as colloidal micelles or globules with no specific structure have been replaced in the collective unconscious of most scientists with, respectively, the double helix that suggested a plausible mechanism by which the genetic material could be faithfully replicated, and exquisitely sculpted molecules that carry out an amazing variety of functions. This revolution in the way we now think about the macromolecules of life has come about largely driven by the huge success of powerful experimental techniques such as X-ray crystallography, cryoelectron microscopy and nuclear magnetic resonance (NMR) spectroscopy. In the 21st century, the study of the architecture and shape of biological macromolecules has come of age and has matured into a separate discipline, first known as structural biology[1] and now more often referred to as structural genomics,[2] the impact of which on the development of the medical sciences cannot be overemphasized. This expanding field is continuously benefiting from advances in protein expression and purification, sample preparation, microcrystallization,[3] use of synchrotron light[4] and automation,[5] as well as from the flexible linking and sharing of increasingly powerful computers and storage devices ("grid computing").[6]

As a direct consequence of this explosion in bio-macromolecular structure determination, the growth of the Protein Data

[a] *Dr. E. Marco*
Bioinformatics Unit, Centro de Biología Molecular "Severo Ochoa"
Universidad Autónoma de Madrid, Cantoblanco, 28049 Madrid (Spain)
Fax: (+34) 914 974 799
E-mail: emarco@cbm.uam.es

[b] *Prof. F. Gago*
Department of Pharmacology, University of Alcalá
28871 Alcalá de Henares, Madrid (Spain)
Fax: (+34) 918 854 591
E-mail: federico.gago@uah.es

[**] *This paper, which expands on an oral presentation delivered in September 2006 at the XIXth International Symposium on Medicinal Chemistry held in Istanbul, is dedicated to Professor W. Graham Richards, former supervisor and friend, on occasion of his retirement as Chairman of Chemistry at Oxford University, and to Dr. Peter Goodford, one of the pioneers of structure-based drug design, father of the GRID program, and a constant source of inspiration and encouragement to many workers in this field. Frontispiece: after M. C. Escher's lithograph* Liberation *(April 1955).*

 © 2007 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

Bank (PDB),[7] which is maintained by the nonprofit Research Collaboratory for Structural Bioinformatics, has been exponential. At the time of this writing, the PDB contained over 39800 protein structures in the public domain, both in the unbound form (that is, interacting only with solvent and buffer molecules or, in the case of those solved in the solid state, also with replicas of the same protein in the crystal lattice) and in complex with a variety of small ligands (organic and inorganic solutes, drugs, ions, co-solvent molecules, etc.), other proteins or nucleic acids. Importantly, knowledge of the three-dimensional (3D) structures of these proteins provides not only a wealth of information about the way they work but also a basis for the screening or design of novel ligands that can be eventually fine-tuned into useful drugs, hence the significance of solving the structure of therapeutically relevant drug targets in pharmaceutical research.[8] As a first step toward this end, it may be particularly instructive 1) to identify pockets, cavities, and clefts (for example, using CASTp)[9] that allow small molecules to bind with high affinity, and 2) to examine the affinity potentials of these putative binding sites for representative functional groups or probes (for example, using GRID).[10]

However, as more and more macromolecular structure representations are permeating into the covers and pages of many standard biochemistry and pharmacology textbooks, the risk exists that these enthralling but static images will keep conveying the same wrong impression of immobility and rigidity that was purported by the first brass models of DNA, sperm whale myoglobin, and hemoglobin.[11] The same is true of the majority of interactive views of macromolecules retrieved from the PDB that are generated using molecular graphics programs despite the fact that crystallographic files, in addition to the atomic Cartesian coordinates, also contain the informative[12] but often less appealing B-factors, which reflect the fluctuation of atoms about their average positions in the crystal.[13] Thus, although NMR specialists have undoubtedly helped to change this awareness by usually illustrating the macromolecule they solve by means of an ensemble of closely related structures, each with a similar likelihood of existing under the conditions of the experiment, limitations in our description and perception of these macromolecular entities still contribute to misguiding the viewer because they do not accentuate the fact that proteins are flexible molecules. Nonetheless, as early as 1959, and in view of results obtained from hydrogen-exchange experiments with α-helices, Linderstrom-Lang and Schellman concluded that *"a protein cannot be said to have 'a' secondary structure but exists mainly as a group of structures not too different from one another in free energy but frequently differing considerably in energy and entropy. In fact, the molecule must be conceived as trying every possible structure"*.[14]

The way of looking at molecules in general, and proteins in particular, should then be as an equilibrium distribution of all possible structures with probabilities depending on the free energy increase required for their formation.[15] On the other hand, protein motions, which can be categorized either by the magnitude of the displacement[16] or by their characteristic time scale, are often the essential link between structure and function (for example, catalysis, cellular locomotion, transport of ions and/or small molecules, regulation of activity, and formation of large assemblies). Bearing this in mind is particularly relevant if we aim to have in our hands ligands that interfere with protein function because the successful identification of new ligands through virtual screening may necessitate targeting *not only the narrow window of presently observable bound-receptor conformations but the complete ensemble of pre-existing receptor conformations*.[17]

Historical Perspective of Ligand Binding

The familiar "key–lock" paradigm was put forward by Emil Fischer to explain the manner in which the enzyme hexokinase exerted its specificity.[18] In essence, this theory said that the enzyme was a rather rigid negative of the substrate and that the substrate had to fit into this negative to react. However, numerous anomalies in enzyme chemistry, and especially the necessity for "activation" of an amino acid to allow it to form a peptide bond during protein synthesis, led Daniel Koshland to postulate that *reaction between enzyme and substrate can occur only after a change in protein structure induced by the substrate itself*.[19] While retaining the idea of the fit from the previous theory, it was appreciated that *the substrate may cause an appreciable change in the three-dimensional relationship of the amino acids at the active site* so as to *bring the catalytic groups into the proper orientation for reaction, whereas a non-substrate*

Federico Gago studied pharmacy at Complutense University in Madrid and completed his PhD at Alcalá University, near Madrid. After pursuing postdoctoral studies at Oxford University under the supervision of Professor W. Graham Richards, he returned to Alcalá, where he is currently associate professor of pharmacology. His research interests are in the areas of molecular modeling, drug design, structure–activity relationships, and computer simulations of biomolecular systems including DNA, enzymes, and membrane receptors.

Esther Marco studied pharmacy at Alcalá University, where she completed a PhD in Dr. Gago's research group. After enjoying a postdoctoral position in the Gastrointestinal Biology and Pathology Unit 531 of INSERM in Toulouse (France), under the supervision of Dr. Daniel Fourmy, she joined Dr. Angel R. Ortiz's team at the Centro de Biología Molecular "Severo Ochoa" in Madrid. Her research interests are in the areas of drug design, homology modeling, and computer simulations of biomolecular systems including DNA, DNA–protein complexes, and GPCRs.

 © 2007 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

will not. Over the years this "induced fit" theory was generalized and extensively used to explain nonenzymatic interactions as well, such as those taking place in drug–receptor, protein–protein, or protein–nucleic acid complexes. Ideally, to gain insight into this problem we would like time to be added as a fourth dimension to the three spatial dimensions of crystallography to literally *"watch" the protein as it executes its function*.[20] In practice, however, the number of examples of time-resolved crystallographic structures is only a very small percentage of the total,[21] although this may change in the future thanks to the use of last-generation synchrotron X-ray sources and major advances in data processing and analysis.[22]

A useful workaround to study these adaptive motions is to compare instances of the same macromolecule solved in different crystals. This exercise most often reveals interesting conformational changes that may range from very subtle movements of amino acid side chains to large motions involving entire protein domains. Fortunately, a comprehensive and highly valuable database already exists, MolMovDB,[16,23] that addresses this issue in a very systematic way. Furthermore, protein motions are classified hierarchically into a limited number of categories both on the basis of size (that is, concerning fragments, domains or subunits) and on whether or not the movement involves sliding over a tightly packed interface. To aid in the visualization of the path connecting two different conformers of the same protein in a rough fashion, a "morph server" is also available[24] that performs a restrained interpolation and produces a "morph movie" that gives the viewer an idea of whether the motion is a rigid-body displacement or entails significant internal deformations.[25] This same idea is implemented in the MovieMaker web server,[26] which creates colorful animations while allowing the user to have some control over rendering style and other image features. It is crucial to realize, however, that the purpose of morphing is to smooth the visual transition and highlight the structural relationships between two given conformations, which does not mean that the empirical interpolated intermediates are necessarily meaningful.

One of the first "laboratories" for the investigation of the connection between protein structure, dynamics, and function was the dioxygen-storage protein myoglobin, which in complex with the simple ligand, carbon monoxide, was found to exist in three major conformational substates, the enthalpies and entropies of which were shown to depend on environmental conditions such as temperature, solvent, pH, and pressure.[27] Probably less well known is the earlier work of Fred Karush, who, in view of the demonstrated high-affinity competitive binding of many small, hydrophobic, typically anionic molecules to serum albumin, concluded that the binding site in this macromolecule could assume a large number of configurations of similar energies that were in equilibrium with one another. Upon interacting with a given ligand, the best-fitting configuration would become selected from the whole structural ensemble. Karush called this phenomenon *configurational adaptability*.[28] Likewise, from observations of biphasic or triphasic reactions in stopped-flow kinetic experiments of hapten binding to antibodies, Foote and Milstein deduced the existence of an equilibrium between at least two antibody conformations, with a given ligand binding preferentially to one form over the other.[29] Besides challenging the common assumption at the time that a highly specific antibody folds into a single conformation at the combining site, these findings reinforced the view that ligand binding may involve some form of receptor isomerization.

These few examples should suffice to illustrate the fact that the concept of *conformational ensembles* has been around for some time, as well as the suggestion that ligands are usually faced with several receptor microstates rather than with a unique structure, as is usually assumed for simplification. As a consequence, the current view is that it is not a conformational change in the receptor that is brought about through the act of ligand binding, as previously surmised, but a shift of the conformational equilibrium in favor of those conformers in the dynamic ensemble that are most complementary to the ligand, at the expense of those for which the ligand has less affinity, as first proposed by Gregorio Weber in the early 1970s.[30] Then one should be aware, and this is especially true for structure-based drug-design efforts and virtual docking experiments, that the selected receptor conformation may not represent the actual target structure or the bound state for one or more of the ligands under consideration.

Similarities and Linkage between Protein Folding and Ligand Binding Processes

In a scenario of nearly isoenergetic conformational substates of a protein rather than a single conformation available for binding to other partners, it was just a matter of time that concepts already commonplace in the protein-folding field, such as "energy landscapes" and "funnels",[31] made an appearance in studies of ligand binding to proteins. In fact, a "binding landscape model" was formulated about 10 years ago in an attempt to establish a statistical mechanics framework connecting native and non-native protein structures to the thermodynamics of ligand binding.[32] It has been stated that *the sole difference between folding and binding is the presence, or absence, of chain connectivity*.[33] Indeed, the terms intermolecular and intramolecular recognition are often used when referring to binding and folding, highlighting the common ground between the two processes, and there is little doubt that knowledge from each process has been instrumental for the better understanding of the other.[34]

The realization and acceptance that protein folding progresses via multiple routes going downhill rather than through a single pathway led to the "folding funnel" concept, which has helped solve long-standing issues such as Levinthal's and the Blind Watchmaker's paradoxes.[31,35] A funnel-shaped energy landscape is implicitly referred to as a macrostate, that is, an ensemble as described in statistical mechanics.[31] At any point in time, the enumeration of the various numbers of conformations in any given state, summed over all possible receptor states, yields a Gaussian distribution. If the receptor is now challenged with a ligand displaying affinity for one of the microstates within the pre-existing ensemble, selective binding to this conformation will create a new Gaussian distribution of

 © 2007 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

receptor states. Therefore, just as a protein in the native state is thought to exist in a collection of microstates, so too will the resulting ligand–receptor complex. The extent to which ligand binding shifts the population distribution within the ensemble will depend on both the binding free energy of the ligand and the differences in free energy between the microstates (Figure 1).[36] The pharmacological characteristics of the

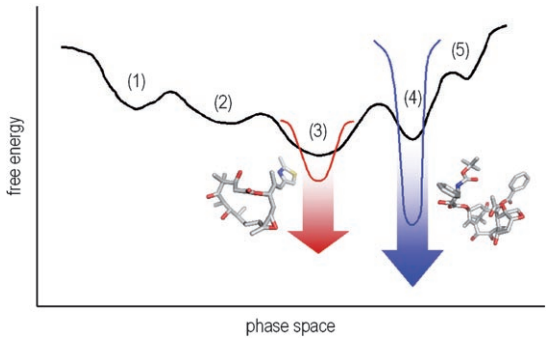

Figure 1. Schematic of binding of two ligands, A (red) and B (blue), to a protein that can adopt a number of conformational states in solution (1–5) that vary in free energy (black line). Ligand A binds preferentially, albeit weakly (small free energy change, red curve), to the most populated state (3), hence no large conformational change is observed. High-affinity binding of ligand B (more favorable free energy change, blue curve) to one of the minority states (4) shifts the population equilibrium within the ensemble toward this conformation ("induced fit"). Note that the binding free energy for ligand B needs to be larger than the difference in free energy between states 4 and 3 (after reference [36]).

ligand will then emerge from the intersection of the ligand-bound ensemble with the various ensembles defining pharmacological receptor behavior, including activation or inhibition, dimerization, internalization (in the case of membrane receptors), or coupling to other binding partners.[37]

Differences among microstates may simply consist of selected regions that are locally unfolded due to some stabilizing interactions being constantly formed and broken by changes in thermal energy. Of course, not all conformations are equally represented in solution, and the diversity within the ensemble is expected to be proportional to the flexibility of the molecule. Under this new light, protein molecules are viewed as dynamic distributions that can present a range of binding site shapes ("plasticity" or "malleability") to any incoming ligand, thus explaining the common observation that a "specificity pocket" can actually bind multiple dissimilar ligands and not just one.

Another important linkage between folding and binding is that as the protein progresses through the folding funnels, a potentially low-population time conformer can be specifically targeted by a ligand. Depletion of this conformer from the solution will shift the equilibrium in its favor, thus furthering the binding reaction and causing a redistribution of the whole ensemble. This can be interesting from a ligand-design perspective, particularly in the protein kinase field, as it has been found that phosphorylation of the critical activation loop tyrosine in some dual-specificity tyrosine-phosphorylation-regulated kinases (DYRKs)[38] or glycogen synthase kinase 3β (GSK3β)[39] is an intramolecular event that only takes place in the nascent kinase as it passes through a transitory chaperone-dependent or independent intermediate. Characterizing the features that make these transitional forms structurally different from the mature enzymes can provide interesting clues that can be exploited with the aim of blocking the crucial autophosphorylation reaction that is necessary for kinase activation.

Allostery as a Redistribution of the Population of the Substates

The coupling of conformational changes between two widely separated binding sites is usually referred to as allostery, and is part of the more universal phenomenon of cooperativity whereby the binding of one ligand can increase (or decrease) the affinity of the target towards a second ligand. The pioneering Monod–Wyman–Changeux "symmetry", "concerted", or "two-state" model[40] proposed that allosteric proteins are symmetric oligomers, with each subunit changing shape in a concerted manner to preserve symmetry of the entire molecule as it is transformed from a "weak-binding" (*tense*, T) to a "strong-binding" (*relaxed*, R) conformational state under the influence of the ligand. Shortly thereafter, Koshland, Nemethy, and Filmer postulated an alternative "induced fit" or "sequential" model[41] according to which each subunit changes shape as the ligand is bound, so that changes in one subunit lead to distortions in the shape or changes in the interactions involving the other protein subunits. This latter model allowed for "negative cooperativity" when binding of the first ligand makes it difficult for the second ligand to bind. In line with the preceding discussion, these earlier views have been replaced by a perception that allostery originates as a consequence of the redistribution of protein conformational ensembles. Ligand binding at one location of the macromolecule can indeed have an effect on the conformation at a site that is far away because it can promote an enrichment in the likelihood of a given conformer.[42] In light of this new paradigm of redistribution of the probabilities of the different conformations, allostery is not relegated to multidomain proteins but can be naturally extended to many single-domain proteins as well, because they have been observed to display allosteric transitions that are modulated not only by binding of small molecules but also by chemical modification or introduction of point mutations.[43]

A combinatorial algorithm (COREX), first described by Vincent Hilser and Ernesto Freire,[44] uses the crystallographic or NMR structure of a protein as a template to generate a large ensemble (typically 10^5) of conformational states that are characterized by having some regions unfolded and the remaining regions in the state specified by the native structure. Under this statistical thermodynamic formalism, the protein is partitioned into a number of folding units that are defined by placing a block of windows over the entire sequence. The thermodynamic quantities (ΔH, ΔS, ΔCp, and ΔG) for each state, as well as the partition function and probability of each state, are evaluated by using an empirical parameterization of the ener-

 © 2007 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

getics, and a stability constant is estimated for each residue. Despite the restrictive nature of its underlying assumptions, COREX has been successfully used to account for hydrogen exchange data obtained for several proteins. Following implementation of a random sampling technique, the updated COREX algorithm can now be applied to proteins of arbitrary size.[42] Potential users may benefit from a publicly available COREX/BEST (Biology through Ensemble-based Structural Thermodynamics) web interface[45] which allows 1) calculation of a conformational ensemble for a given protein, 2) assessment of relative differences in structural stability within the protein, and 3) determination of the temperature-dependent stability of the calculated conformational ensemble. Remarkably, application of this type of analysis to 16 non-structurally homologous proteins revealed a dual character for all the ligand binding sites studied, in the sense that they presented regions with both high and low structural stability. In many cases the latter regions were shown to correspond to loops that are stabilized upon binding a low-molecular-weight ligand, which becomes significantly buried as a result. On the contrary, catalytic residues in enzymes were found to be mostly located in regions with high structural stability, showing that active sites preferentially occupy cooperatively constrained regions. Subsequent work showed that, in allosteric enzymes, the low-stability regions in the regulatory site do play a crucial role in the transmission of information to the catalytic site.[46]

A wealth of information from both experimental and theoretical work supports the view that the energy of stabilization of the protein structure is not evenly distributed throughout the molecule.[47] Even under native conditions multiple independent local unfolding and refolding events occur that are thought to reflect specific functional requirements insofar as they are necessary for favoring (or precluding) interaction with other proteins or with small ligands. Very often, the effect of binding at one location is propagated to distal regions, triggering a structural response that can be translated into an intermolecular communication signal. Particularly noteworthy are those cases in which a bound ligand can modulate the binding affinity between two proteins by several orders of magnitude because the region containing the recognition determinants for formation of the interface is devoid of stable secondary or tertiary structural elements in its absence.[48] In this respect, it is worth recalling that some allosteric inhibitors of the human immunodeficiency virus type 1 (HIV-1) reverse transcriptase (RT) have been shown to enhance the association of p66 and p51 enzyme subunits dramatically and even induce heterodimerization of dimerization-defective mutants following tight binding to the p66 subunit.[49]

Allosteric modulators are receiving a great deal of attention nowadays, especially in the field of G-protein-coupled receptors (GPCRs),[50] due to a number of theoretical advantages they may possess over classic orthosteric ligands, such as a ceiling level to the maximum effect they can produce and a potential for greater subtype selectivity. This renewed interest is not truly surprising because allosteric sites in ligand-gated ion channels have been known for many years, as probably best exemplified by the regulatory site present in a subset of $GABA_A$ receptors to which the exceedingly successful and widely prescribed benzodiazepines bind.[51]

Computational Approaches to Conformational Ensembles

In view of the growing evidence relating pre-existing conformations of the receptor-unbound state in equilibrium and observed structural changes upon ligand binding, it becomes essential to understand these movements and interactions if accurate and useful models of receptor–ligand complexes are to be constructed. Having been established experimentally (e.g. X-ray crystallography, hydrogen/deuterium exchange, or NMR spectroscopic techniques) that there is a range of conformational isomers in equilibrium with each other available for ligand binding, the question arises of how to gain insight into these population redistributions using only computational tools when only one or very few structures of a given protein are known and relatively large conformational changes are at play.

At least two published methods facilitate the investigation of domain motions in proteins when two conformations of the same protein are available: Hingefind[52] and DynDom.[53] Both programs allow 1) the identification of the reference domain ("rigid core"), 2) the determination of the effective rotation axes that characterize the domain movements by effective rotation axes ("hinges"), and 3) the visualization of the hinge-bending residues. When only one structure is available, however, dynamic information and alternate conformations have to be obtained through alternative means, as illustrated by the next four strategies.

1. Normal mode analysis

Atomic fluctuations in a macromolecular system can be separated into local oscillations (with a sub-picosecond time scale) superposed on motions with a more collective character that are brought about by concerted variations in dihedral angles and give rise to vibrational modes with low frequencies (below 200 cm^{-1}), that is, large amplitudes.[54] Normal mode analysis (NMA) is a fast and relatively simple time-independent approach to the calculation of these vibrational modes that is based on the harmonic approximation of the potential energy function around a minimum energy conformation: it finds the analytical solution of the equations of motion by numerical diagonalization of the Hessian matrix (the mass-weighted second derivatives of the potential energy matrix). The eigenvectors of this matrix are the normal modes, and the eigenvalues are the squares of the associated frequencies.

Initially, the stationary point on the potential energy surface for starting an NMA calculation was most often computed by subjecting to energy minimization an all-atom description of the macromolecule employing a conventional force-field empirical potential. Subsequent work demonstrated the validity (without loss of accuracy) of using simplified approaches in terms of both the potential and the chain representation. Thus, amino acids can be represented by point masses or "par-

 © 2007 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

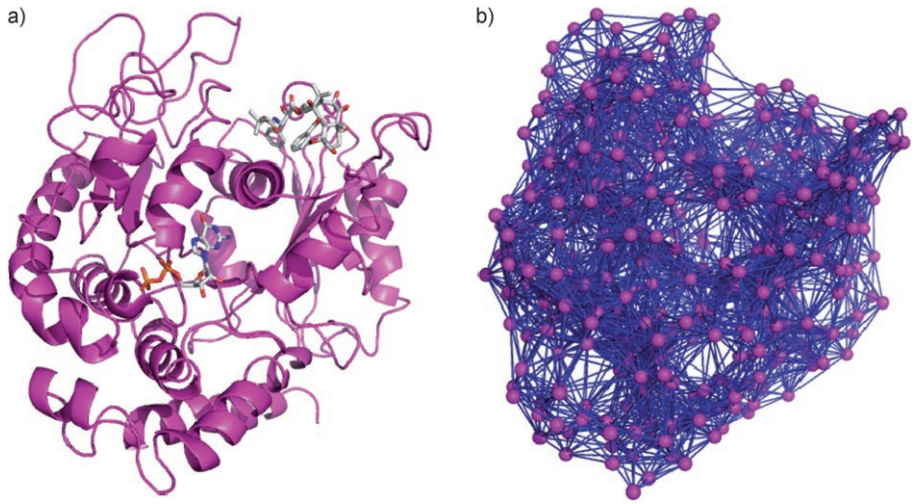

Figure 2. Two representations of the microtubule-forming protein β-tubulin in complex with GDP and the antitumor drug docetaxel (PDB entry 1tub): a) a standard ribbon structure and atom-type-colored sticks for the ligands, and b) an elastic network model (protein Cα atoms as spheres joined by blue springs).

ticles" (usually the Cα atoms) which are connected by springs of equal strength resulting in a 3D elastic network (Figure 2). The pairwise Hookean potential is controlled by a single parameter, the cut-off distance used to define the structural neighbors, as first described by Monique Tirion[55] and then further developed by others.[56,57]

The purpose of normal mode perturbation is to capture much larger amplitude motions than the atomic displacements reflected by the B-factors, which are usually calculated at room temperature. By analyzing the lowest-frequency modes of a protein one after the other, one can identify the conformational changes that are stabilized by ligand binding or are most relevant for a particular function. Test cases for validating the usefulness of this type of analysis are provided by the increasing number of protein structures present in the PDB in both "open" and "closed" forms. In this respect, a number of surveys have already demonstrated that a handful of the lowest-frequency modes is generally sufficient to explain the observed movements.[58,59] This good agreement is as pleasing as it is striking given that standard NMA is performed in vacuo and is usually restricted to a single minimum, therefore neglecting anharmonic effects due to transitions from one local minimum to another.

For the non-experts, a good starting point for carrying out NMA is one of the excellent and user-friendly interactive WWW servers available in the public domain, such as NOMAD-Ref[60] or El Némo,[61] which also allow the retrieval of protein coordinates and 3D animated views of the generated modes using either Molscript[62] (El Némo) or PyMOL[63] (NOMAD-Ref). These calculations are quite fast and impose virtually no limit on the size of the macromolecule under study because they take advantage of the elastic network model (ENM) of proteins and a so-called rotation–translation block method.[64] NMA has been successfully applied, for instance, to the homopentameric α7 nicotinic acetylcholine receptor (nAChR), which is a ligand-operated ion channel, and to the mechano-sensitive channel (MscL). Remarkably, the results revealed that the first step of the gating mechanisms can be explained by just the lowest (nAChR)[65] or the three lowest-frequency modes (MscL).[66] The structural reorganization leading to opening of the channel pore in both cases was shown to be caused by concerted tilt and twist motions of the protein with opposing rotations of the extracellular and transmembrane domains.

NMA has been advocated as a computationally inexpensive technique to generate multiple receptor conformations in attempts to incorporate binding pocket flexibility in protein–protein[67] and ligand–protein[68] docking. Despite its recognized value, a possible caveat with this method is that highly localized low-amplitude motions such as those giving rise to loop rearrangements are unlikely to be associated with the lowest-frequency modes making it necessary to calculate a large number of them (up to several hundred). Nonetheless, a *measure of relevance* can be calculated that expresses the contribution of the relevant mode(s) to the motion in the region of interest.[68] The receptor structure can then be perturbed along the single mode identified or along a linear combination of several modes to generate a diversity of backbone conformations. Prior to the intended docking procedure the orientation of the amino acid side chains in the binding pocket can be optimized either in the absence or in the presence of known ligands.

2. Gaussian network model

Conceptually similar to NMA, and built on statistical mechanical theory developed for describing the fluctuation dynamics of polymer networks, the Gaussian network model (GNM) is also entirely based on inter-residue contact topology in the folded state, requires no a priori knowledge of empirical energy parameters, and provides a unique, closed mathematical solution for each structure.[69] An important feature of the GNM is the possibility of dissecting the observed motion into a collection of normal modes. The GNM mode analysis is similar to that in conventional NMA, with the slowest modes also providing information on collective motions that are usually relevant to biological function. Despite its simplicity, the GNM has proven to yield results in good quantitative and qualitative agreement with experimental data and molecular dynamics (MD) simulations. Importantly, catalytic residues in enzymes have been shown to occupy preferentially cooperatively constrained regions (minima in slow modes), in good agreement with the COREX results discussed above. GNM-generated dynamic data for over 20000 PDB structures have been collected

 © 2007 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

in a publicly available database, iGNM,[69] and structural dynamics computations using the GNM can also be performed online.[70]

3. Floppy inclusion and rigid substructure topography

The intrinsic flexibility within a protein can also be evaluated by means of a graph-theory algorithm, floppy inclusion and rigid substructure topography (FIRST), that is based on a more elaborate bond network definition consisting of covalent bonds, hydrogen bonds, and hydrophobic interactions.[71] By identifying which bonds are constrained and which bonds remain free to rotate in a single, static 3D structure, the relative rigidity or flexibility index of a given region can be rapidly calculated. More recently, an approach that combines concepts from the ENM of proteins and rigidity theory has been shown to be useful in reliably predicting the directions and magnitudes of the motions in a test set of 10 proteins that show conformational changes on ligand binding.[72]

4. Molecular dynamics simulations

This time-dependent method samples phase space and provides structural variation in atomic detail by numerically solving Newton's equations of motion. The last 30 years have witnessed a spectacular evolution in the theoretical treatment and application of MD simulations to both the refinement of crystallographic structures and the study of protein and nucleic acid structure.[73] The progress has been astonishing: from the first MD study on bovine pancreatic trypsin inhibitor (BPTI) in vacuum for only 9.2 ps using a crude molecular mechanics potential,[74] or the first theoretical demonstration that the potential energy surface of myoglobin is characterized by a large number of thermally accessible minima in the neighborhood of the native structure,[75] to the systematic analysis of literally thousands of protein structures using state-of-the-art potentials and explicit solvent and counterions.[76] These achievements bring us a little step closer to the long-standing goal of describing living systems in terms of chemistry and physics.[73]

The huge amount of information contained in the multiple sets of atomic coordinates that are stored during a MD run ("trajectory") can be distilled in the form of a visual animation on a graphics screen and also as a variety of plots representing the evolution of the properties of interest (geometries, root-mean-square fluctuations, energy values, etc.) as a function of time. Therefore these trajectories provide a means to identify and study motions that are crucial for function and constitute a valuable means of interpreting diverse experimental data on macromolecular structure in solution, both in the presence and absence of experimentally derived restraints. However, it may be necessary for proper analysis to separate functionally important large-scale correlated motions from random thermal local fluctuations. To this end, a principal components analysis (PCA) can be carried out on a large number of configurations of MD trajectory data (which involves the diagonalization of the covariance matrix of atomic fluctuations upon removal of the overall translation and rotation) to yield collective variables that are sorted according to their contribution to the total mean-square fluctuation. In the so-called quasi-harmonic analysis it is the mass-weighted coordinates of all atoms that are employed, thus relating it to NMA, whereas in essential dynamics (ED) analysis[77] only the backbone or Cα atoms are usually selected for the PCA. Interestingly, the deformation patterns determined by ED extracted from atomistic simulations in aqueous solution of a set of 30 proteins from the PDB (representative of all protein metafolds)[76] were very recently shown to be similar to those obtained from coarse-grained NMA.[78] Furthermore, the trace flexibility of the proteins simulated in water was accurately represented by the first, most relevant NMA eigenvectors.

The standard implementation of MD preferentially samples regions of configuration space in the vicinity of potential energy minima (hence its importance for studying local conformational changes in the binding site),[79] whereas those at the top of high reaction barriers are rarely visited on the finite time scale at room or physiological temperatures. Nonetheless, larger conformational transitions can be "accelerated" through the implementation of various strategies,[80] a couple of which taken from work in our lab are briefly presented below.

Energetics of Complex Formation

The reliable computation of binding free energies is a primary objective in virtual high-throughput screening and computer-aided ligand design.[81] From the discussion so far, it is apparent that a correct evaluation of the binding energy cannot be performed by examining just a single conformation of the molecular complex under study. Besides, the educated although subjective interpretation of the strength of certain interactions that stabilize a given ligand in its binding site can be biased by misconceptions or incorrect assumptions about the relative importance, among others, of 1) ionic interactions in the presence of competing water molecules,[82] 2) the hydrophobic effect, and 3) entropic considerations, which not only are often neglected or overlooked but are also difficult to foresee and calculate, especially in the context of enthalpy–entropy compensations. Furthermore, extrapolation of the conclusions to other ligands, either congeneric or structurally unrelated, is often subjective and problematic.

Accurate binding free energy *differences* between pairs of very similar ligands (or receptors) could, in principle, be obtained in the context of MD or Monte Carlo simulations making use of free energy perturbation[83] or thermodynamic integration,[83] which use conformational sampling to generate thermodynamic averages and compute the sum of small changes along a multistep pathway defining the conversion of ligand A into ligand B both free in solution and within the receptor binding site ("computational alchemy").[84] However, despite the potential of these rigorous methods to provide great physical insight about the reasons for the differences in binding affinities between pairs of ligands,[85] the necessity to allocate huge computational resources for their implementation makes them impractical for the comparative study of even small series of receptor–ligand complexes. To circumvent this

 © 2007 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

problem, two approaches that generate conformations of only the free and bound species (and compute the absolute binding free energy by taking a difference) have attracted a lot of interest in recent years: the linear interaction energy approach[86,87] and the so-called molecular mechanics Poisson–Boltzmann surface area (MM-PBSA) approach. The latter makes use of a continuum representation of the solvent to replace the computer-intensive treatment of explicit water molecules.[85,88]

In a much simpler way, when a high-resolution ligand–receptor complex is available and an experimental measure of affinity has been obtained for a series of related ligands, it is relatively straightforward to build a whole set of complexes using standard molecular modeling tools and then compute desolvation-corrected ligand–receptor interaction energies in order to derive quantitative structure–activity relationships (QSAR).[89] Although this comparatively unsophisticated method is not devoid of hurdles because of uncertainties in the modeling procedure (for example, insufficient sampling of internal degrees of freedom, wrong choice of ionization states for ligands and/or side chains of titratable amino acids, systematic errors in the refinement of the complexes, inadequate evaluation of changes in solvation, and neglect of polarization effects) and the fact that each complex is represented by only one structure, the resulting QSAR equation can be objectively interpreted and also challenged for predictive ability. Furthermore, we and others have shown that a useful alternative to considering just one global L–R interaction energy for each complex is to decompose this quantity into a set of van der Waals (vdW) and electrostatic (Ele) contributions ("descriptors") emanating from individual amino acids making up the receptor, and to project the resulting matrix of energy terms onto a small number of orthogonal "latent variables" (or principal components) using partial least squares (PLS) regression.[90] At the end of the procedure, which is formally similar to that used in comparative molecular field analysis (CoMFA),[91] those pairwise interactions between the ligands and individual protein residues that are predictive of activity or binding free energy are selected and assigned weights according to their importance in the model. Since its inception, this chemometric method, which demands only modest computational requirements and is known as comparative binding energy (COMBINE) analysis (Figure 3), has been successfully applied to a variety of biologically relevant targets.[92] Interestingly, in one application to a series of non-nucleoside inhibitors of HIV-1 RT, the PLS coefficients derived from the analysis not only yielded a robust QSAR model but also correctly identified the effect of mutations on relevant protein residues.[93] In recent times, the potential of this method has been productively extended by using it as a scoring function in virtual screening against the serine protease, factor Xa,[94] and also for the purpose of addressing the issue of ligand selectivity across a family of related receptors.[95]

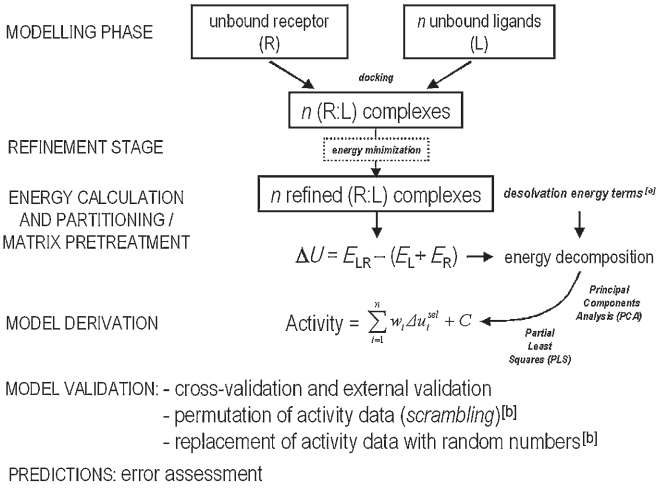

Figure 3. Flowchart of the COMBINE procedure. [a] Optional but strongly recommended. [b] Optional to rule out the feasibility of getting spurious models in these two cases.

Alternative Scenarios in the Drug-Discovery Process

Many current medicinal chemistry research projects are typically guided by the structure of the target macromolecule. Nonetheless, for the reasons outlined above, the available structure may not represent the ideal target conformation that is needed for docking or ligand-design purposes. In the simplest but far from trivial case, local conformational changes in the protein binding site may preclude or hamper the proper positioning of the small molecule. In relatively more complex cases, the binding process additionally entails significant changes in the conformation of the peptide backbone. To illustrate some of these typical scenarios we have selected some practical cases that originated in our own research, and we show the workaround that we used employing some of the methods discussed above. Of course, interested readers can benefit from many other excellent and far more detailed recent reviews on the subject.[36,96,97]

The 3D structure of the target protein has been solved experimentally and is available in both apo and bound forms

The usual objective in this case is to understand the motions and conformational changes involved in ligand binding as well as the energetics of the binding process. An additional common aim is to use the available structure or ensemble of structures[98] directly for ligand docking and virtual screening.

If the protein consists of two hinge-connected domains that behave virtually as rigid bodies,[99] the path connecting the "open" and "closed" conformations can be studied by restraining the Cα atoms of the hinge region in one structure to the positions of equivalent atoms in the alternative structure by means of a harmonic potential that is applied during the course of an MD simulation. In contrast to a mere linear interpolation, this "template forcing" or "targeted" MD (tMD) simulation[100] provides a more realistic interconversion pathway

 © 2007 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

that takes full account of backbone and side chain flexibility. The target distance is a measure of the structural root-mean-square distance between selected atoms or residues in the two molecular conformations. As no prior knowledge of the barrier height is usually available, the value of the force constant can be progressively increased during a series of consecutive MD runs. It is advisable to use different combinations of time spans and force constant values to check for possible protocol dependencies. Importantly, any difference in the trajectories in the presence or absence of a bound ligand can be attributed to stabilization of the closed form by that particular ligand. Moreover, if ligand interactions are asymmetric with respect to the two domains being separated, the distinct behavior can provide clues about which domain is more likely to establish the primary interactions with the ligand and thus serve as the docking platform. Once the ligand is docked onto the open form the closure motion can be simulated using the closed form as the target. We have used this procedure to simulate, for example, the opening and closing events taking place in the ligand binding core of the ionotropic glutamate AMPA receptor GluR2, both in the apo form[101] and in complex with agonists (kainate and glutamate) and antagonists (6,7-dinitroquinoxaline-2,3-dione and 5′-guanosine monophosphate).[102]

The tMD approach is also useful when the protein rearrangement upon ligand binding is multifarious, as exemplified by the structural differences observed when the apo or DNA-bound forms of HIV-1 RT are compared with the enzyme in complex with one of several non-nucleoside RT inhibitor (NNRTI) analogues bound in the allosteric site (Figure 4). Thus, tMD simulations supported the inference that the energy barrier associated with the process of creation of the cavity where this type of inhibitor binds is higher in a common mutant RT form (Lys 103 Asn) than in the wild-type enzyme. This is so because of the extra stabilization brought about by a hydrogen bond between the side chains of Asn 103 and Tyr 188 for which no equivalent exists in the wild-type RT structure.[103] Furthermore, by docking three different NNRTIs at the proposed entrance to the pocket in the unbound form and gradually forcing them into this previously inexistent cavity, the processes of inhibitor entry and pocket creation could be monitored in detail. As a result, a rationale for the resilience of the Lys 103 Asn mutant enzyme to the inhibitory effect of etravirine could be gained, namely, the ability of this second-generation NNRTI to disrupt the extra hydrogen bonding interaction that hampers entrance of first-generation NNRTIs.[104] Of note, it is highly unlikely that this explanation could have been derived solely from QSAR studies on the complexes between a series of NNRTIs and both wild-type and mutant enzymes.

The 3D structure of the target protein has been solved experimentally and is available in a given conformation not suitable for docking studies

The primary objective in this case is to model multiple conformations in silico by considering either side chain rearrangements alone[105] or in combination with essential backbone movements[68] to perform ligand docking and scoring on the discrete set of selected receptor conformations. This can then be followed by a merging and shrinking step, where the results of the multiple virtual screenings are condensed to improve the enrichment factor.[68]

MD simulations can also be particularly useful to generate alternative receptor conformations, especially in cases that will benefit from incorporation into the sampling procedure of some experimentally substantiated information, for example in the form of distance restraints. A case in point is the eukaryotic elongation factor 1A (eEF-1A), which has been solved in complex with the guanine nucleotide exchange factor domain of eEF-1Bα but is not available in the GTP conformation to which the potent antitumor agent didemnin B binds with the highest affinity.[106] However, this conformation was easily achieved by implementing a steered MD (sMD) protocol consisting of 1) pairwise forcing the distances between Cα atoms of selected residues to adopt the same values that are measured between Cα atoms of equivalent residues in the crystallographic GTP form of its prokaryotic counterpart, EF-1A (formerly EF-Tu) and 2) optimizing the coordination environment of the crucial magnesium ion. The desired structural rearrangement was achieved by progressively reducing each distance in a zipper-like fashion (Figure 5) using a harmonic potential with a force constant that was progressively increased during several consecutive sMD runs.[107] As in the tMD

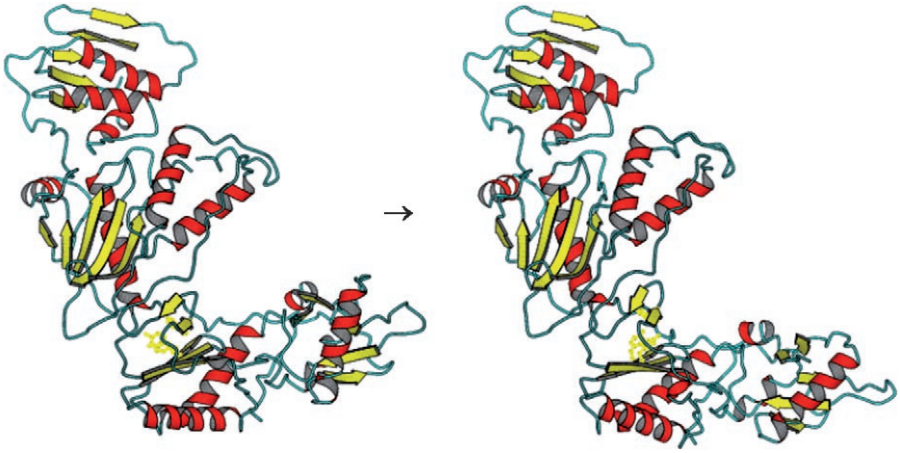

Figure 4. Conformational changes in HIV-1 RT subunit p66 in going from the apo form in a covalently trapped catalytic complex (PDB entry 1rtd) to a non-nucleoside-bound inhibited form (PDB entry 1fk9). Aromatic residues making up the non-nucleoside binding site are displayed as yellow sticks. The figure was created from standard output produced by MovieMaker, which also provides Cartesian coordinates in PDB format for a user-selected number of interpolated intermediates.[26]

 © 2007 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

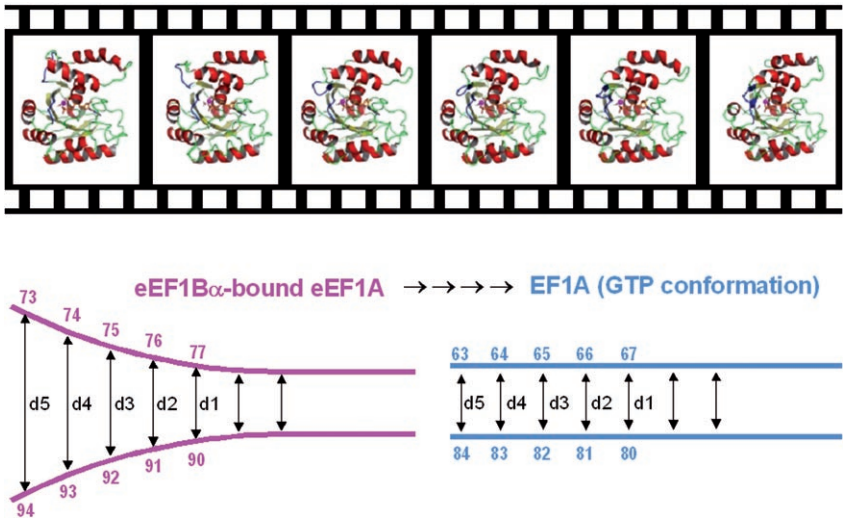

Figure 5. Schematic of the steered molecular dynamics protocol implemented to drive the conformation of the catalytic domain of eEF-1A (residue numbers in pink) progressively from that found in the crystallographic complex with the guanine nucleotide exchange factor eEF-1Bα (left) to that adopted when it binds GTP (right), which is equivalent to the "GTP conformation" experimentally observed for its prokaryotic counterpart, EF-1A, also known as EF-Tu (residue numbers in cyan).[107]

cases reported above, the incremental use of this term prevents artifactually overcoming larger energy barriers that could deprive the simulation of any physical sense.

In another representative example, we focused on *E. coli* thymidine phosphorylase (TPase), which appears in the X-ray crystal structure as a protein containing two distinct domains separated by a cleft,[108] with a thymine base bound in the thymidine binding site present in the α-domain, and a sulfate ion bound in the phosphate binding site in the α/β domain. When the sulfate anion is replaced with a phosphate in an attempt to figure out how the transition state (TS) of the reaction is stabilized, the distance between any anion oxygen and the C1′ of thymidine is ~8 Å, strongly suggesting that domain closure is necessary to generate a catalytically competent active site in the enzyme. Apart from the academic interest in studying the motions leading to this alternative "closed" conformation, characterization of the TS and identification of the residues involved in its stabilization are mandatory if our inhibitor design strategy is based, as is often the case, in mimicking this TS.[109] Therefore, in a first step, we obtained accurate geometries for the thymidine substrate, the corresponding TS, and the final product, thymine (in both enol and keto forms) from high-level quantum mechanical calculations.[110] Then we simulated the reaction within a reduced representation of the active site using the TRITON software,[111] which also allowed us to analyze essential interactions between the substrate and surrounding protein residues. On the basis of the results obtained we assigned a dual role to the imidazole ring of the highly conserved His85: deprotonation of the incoming phosphate and donation of the proton to O2 of thymidine. The resulting enol form of thymine would then spontaneously tautomerize to the more stable keto form as one of the final products. This S_N1 mechanism proposed for the phosphorolysis reaction led us to expect a higher catalytic rate for 2-thiothymidine (because S2 protonation by His85 would be facilitated relative to protonation of O2 in thymidine), and this prediction was confirmed experimentally. Also as anticipated, domain separation starting from the previously closed form required more energy when the TS was bound in the active site than in the presence of either substrates or products.

The 3D structure of the target protein has not been solved experimentally or is not publicly available

The major focus in this case is on obtaining an approximate but reliable 3D structure of the target protein starting just from its primary sequence. Major advances in secondary structure prediction, fold recognition, identification of potentially remote homologues that can be used as templates, improved sequence[112] and structural[113] alignments, and advances in model refinement[114] have largely contributed to making comparative modeling[115] a successful and widely used technique, particularly in cases where the sequence identity between target and template is greater than 30%.[116] In fact, publicly available repositories of high-quality models of whole proteins or protein fragments or domains already exist, such as MODBASE,[117] which can provide at least a first approximation to the structure of the protein of interest.

It is important to realize, however, that there can be instances in which the structural homology between two proteins can be quite high despite a notoriously low sequence identity, practically on the border of questionable biological significance. For example, both metabotropic and ionotropic receptors for the major excitatory neurotransmitter glutamate have large extracellular N-terminal domains structurally similar to bacterial periplasmic amino acid binding proteins despite the low (~20%) sequence identity.[118] Similarly, a search for proteins structurally homologous to the previously discussed *E. coli* TPase using the structural alignment program Dali[119] pointed to the pyrimidine nucleoside phosphorylase (PyNPase) from *Bacillus stearothermophilus* (42% sequence identity) as a structural neighbor: in their open forms, the root-mean-square deviation of Cα atoms is barely 2.0 Å, with just 5 out of 442 of these atoms being found in nonequivalent positions. The happy coincidence that this PyNPase had also been solved in the closed state provided us with a suitable template for carrying out the tMD protocol described above.

As a last example, we recently modeled the structure of human mitochondrial thymidine kinase (TK-2) using a number of putative structural neighbors[120] belonging to the same homologous superfamily of P-loop-containing nucleotide triphos-

 © 2007 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

phate hydrolases characterized by a three-layer ($\alpha\beta\alpha$) sandwich architecture and a Rossmann fold. Examination of the normal modes for these enzymes and exploration of the resulting structures with the CAVER tool[121] revealed that some internal motions coupled closure of the P-loop in the ATP binding site to the creation of a hydrophobic channel in the helical region that connected the deoxythymidine binding site with the bulk solvent. We thought this observation was interesting because TK-2 inhibition by some nucleoside analogues is dependent on the presence of bound ATP. Indeed, when a protein conformation selected from the NMA was used as target in an automated docking procedure, the best solution placed part of the inhibitor inside this tunnel. Taken together, these findings strongly suggest that this methodology could be useful for other targets in which "cryptic" binding sites[122] are exposed following a local rearrangement of the protein backbone that can be promptly studied using NMA.

Summary

A current bottleneck in many structure-based ligand-design projects is that the conformation of the target protein as found in a crystal structure, be it unbound (apo form) or in complex with a binding partner (such as a small-molecule ligand or another macromolecule) may not be truly representative of the dominant conformation in solution or the bound conformation in complex with the desired ligand(s). This is so because different ligands usually display affinity for different microstates that already exist within the ensemble of free receptor conformations in solution, although this diversity is not immediately apparent from inspection of a "static" structure due to the stabilization brought about by crystal packing constraints and limitations in our visual perception of macromolecular structure using standard molecular graphics programs. Useful workarounds to overcome these limitations are to color the protein atoms according to the B-factor column in the PDB entry, to visualize the crystal lattice beyond the asymmetric unit by carrying out the corresponding symmetry operations, to compare structures of the same protein in different crystal forms and/or trapped in different conformations, and to try and understand the motions involved in the conformational changes observed. NMA and MD simulations, either independently or in combination, are arguably the tools of choice for studying many of the motions that are important for ligand binding and molecular recognition.

Outlook

For researchers involved in understanding drug effects and the development of new medicines, the successful merge of structural biology with functional genomics is generating a wealth of data that is yielding, with unprecedented atomic detail, important clues about the mechanism of action of numerous drugs. With fast and widespread access to curated databases containing information on DNA and protein sequences, gene expression, cellular roles, protein families, and taxonomy for microbes, plants and humans,[123] pharmacology and medicinal chemistry are increasingly taking genomic and computational approaches in their stride because these avenues hold extraordinary promise of a rich harvest of new information that can help explain how target enzymes are inhibited or how drug receptors are activated or blocked.

Now that cooperative motions in proteins or protein assemblies have been efficiently characterized through sampling of the conformational space using conventional MD simulations,[76,78] and the output of these efforts has been observed to compare very well with that of the much more inexpensive method of NMA or GNM, it will probably be worth exploring whether the joint use of these two methodologies can be successfully exploited to generate the ensemble that best represents the actual repertoire of receptor conformations that can be targeted by one or more ligands.

Universities have a large share of the responsibility to ensure that biomedical students, before leaving the classrooms, are properly trained to understand that the tantalizing molecular organization and operation of cells, tissues, and whole organisms relies on interactions among the molecules of life, most importantly proteins and nucleic acids. The fact that these interactions, as well as those between macromolecules and drugs, can nowadays be visualized in atomic detail and normally understood in terms of the general laws of physical chemistry represents the present-day culmination of the pioneering aspirations of Alfred Joseph Clark (1885–1941), one of the originators of receptor theory and the father of quantitative pharmacology.[124] We hope that the ideas and hints outlined in this review will be helpful to others in pursuing their research efforts towards the goal of designing and developing safer and more effective drugs.

Acknowledgements

We thank past and present students and collaborators for their enthusiasm and hard work. We are grateful to Dr. M. Jesús Pérez-Pérez and Dr. Pablo Chacón for careful reading and useful suggestions, and we apologize to all colleagues whose key contributions could not be cited owing to space restrictions and focus. The authors' research at the University of Alcalá described herein was supported by the European Commission (Grant no. QLRT-2000-30291), the Spanish Comisión Interministerial de Ciencia y Tecnología (Grant nos. SAF 2000-0153-C02, SAF2003-7219-C02, and SAF2006-12713-C02), Comunidad de Madrid (Grant nos. 08.1/0039/2000 and BIPEDD-CM/2006), PharmaMar S.A., and the National Foundation for Cancer Research.

Keywords: ligand docking · molecular dynamics · normal mode analysis · quantitative structure–activity relationships · virtual screening

[1] R. E. Dickerson, *Present at the flood: How structural biology came about*, Sinauer Associates, Sunderland, MA, **2005**.
[2] J. M. Chandonia, S. Brenner, *Science* **2006**, *311*, 347 – 351.
[3] E. Abola, P. Kuhn, T. Earnest, R. C. Stevens, *Nat. Struct. Biol.* **2000**, *Suppl. 7*, 973 – 977.

 © 2007 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

[4] T. L. Sorensen, K. E. McAuley, R. Flaig, E. M. Duke, *Trends Biotechnol.* **2006**, *24*, 500 – 508.
[5] T. L. Blundell, H. Jhoti, C. Abell, *Nat. Rev. Drug Discovery* **2002**, *1*, 45 – 54.
[6] A. Chien, I. Foster, D. Goddette, *Drug Discovery Today* **2002**, *7*, S176 – S180.
[7] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, *Nucleic Acids Res.* **2000**, *28*, 235 – 242.
[8] D. Bailey, E. Zanders, P. Dean, *Nat. Biotechnol.* **2001**, *19*, 207 – 209.
[9] a) J. Dundas, Z. Ouyang, J. Tseng, A. Binkowski, Y. Turpaz, J. Liang, *Nucleic Acids Res.* **2006**, *34 (Web Server issue)*, W116 – 118; b) URL: http://cast.engr.uic.edu.
[10] a) P. Goodford, *J. Med. Chem.* **1985**, *28*, 849 – 857; b) URL: http://www.moldiscovery.com.
[11] D. C. Phillips in *Biomolecular Stereodynamics II* (Ed.: R. H. Sarma), Adenine Press, Guilderland, NY, **1981**, p. 497.
[12] B. Halle, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 1274 – 1279.
[13] Z. Yuan, T. L. Bailey, R. D. Teasdale, *Proteins* **2005**, *58*, 905 – 912.
[14] K. U. Linderstrom-Lang, J. A. Schellman in *The Enzymes, Vol. 1* (Eds.: P. D. Boyer, A. Lardy, K. Myrbick), Academic Press, New York, **1959**, p. 443.
[15] J. A. Schellman, C. G. Schellman, *Protein Sci.* **1997**, *6*, 1092 – 1100.
[16] M. Gerstein, W. A. Krebs, *Nucleic Acids Res.* **1998**, *26*, 4280 – 4290.
[17] M. G. Bursavich, D. H. Rich, *J. Med. Chem.* **2002**, *45*, 541 – 558.
[18] E. Fischer, *Ber. Dtsch. Chem. Ges.* **1894**, *27*, 2985 – 2993.
[19] D. E. Koshland, *Proc. Natl. Acad. Sci. USA* **1958**, *44*, 98 – 104.
[20] F. Schotte, M. Lim, T. A. Jackson, A. V. Smirnov, J. Soman, J. S. Olson, G. N. Phillips, Jr., M. Wulff, P. A. Anfinrud, *Science* **2003**, *300*, 1944 – 1947.
[21] K. Moffat, *Acta Crystallogr. Sect. A* **1998**, *54*, 833 – 841.
[22] M. Schmidt, H. Ihee, R. Pahl, V. Srajer, *Methods Mol. Biol.* **2005**, *305*, 115 – 154.
[23] S. Flores, N. Echols, D. Milburn, B. Hespenheide, K. Keating, J. Lu, S. Wells, E. Z. Yu, M. Thorpe, M. Gerstein, *Nucleic Acids Res.* **2006**, *34(Database issue)*, D296 – D301.
[24] W. G. Krebs, M. Gerstein, *Nucleic Acids Res.* **2000**, *28*, 1665 – 1675.
[25] N. Echols, D. Milburn, M. Gerstein, *Nucleic Acids Res.* **2003**, *31*, 478 – 482.
[26] a) R. Maiti, G. H. van Domselaar, D. S. Wishart, *Nucleic Acids Res.* **2005**, *33(Web Server issue)*, W358-W362; b) URL: http://wishart.biology.ualberta.ca/moviemaker.
[27] a) H. Frauenfelder, S. G. Sligar, P. G. Wolynes, *Science* **1991**, *254*, 1598 – 1603; b) M. K. Hong, D. Braunstein, B. R. Cowen, H. Frauenfelder, I. E. Iben, J. R. Mourant, P. Ormos, R. Scholl, A. Schulte, P. J. Steinbach, *Biophys. J.* **1990**, *58*, 429 – 436.
[28] F. Karush, *J. Am. Chem. Soc.* **1950**, *72*, 2705 – 2713.
[29] J. Foote, C. Milstein, *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 10370 – 10374.
[30] G. Weber, *Biochemistry* **1972**, *11*, 864 – 878.
[31] K. A. Dill, *Protein Sci.* **1999**, *8*, 1166 – 1180.
[32] D. W. Miller, K. A. Dill, *Protein Sci.* **1997**, *6*, 2166 – 2179.
[33] C. J. Tsai, S. Kumar, B. Ma, R. Nussinov, *Protein Sci.* **1999**, *8*, 1181 – 1190.
[34] C. J. Tsai, D. Xu, R. Nussinov, *Folding Des.* **1998**, *3*, R71 – R80.
[35] K. A. Dill, H. S. Chan, *Nat. Struct. Biol.* **1997**, *4*, 10 – 19.
[36] H. A. Carlson, *Curr. Opin. Chem. Biol.* **2002**, *6*, 447 – 452.
[37] T. Kenakin, *Annu. Rev. Pharmacol. Toxicol.* **2002**, *42*, 349 – 379.
[38] P. A. Lochhead, G. Sibbet, N. Morrice, V. Cleghon, *Cell* **2005**, *121*, 925 – 936.
[39] P. A. Lochhead, R. Kinstrie, G. Sibbet, T. Rawjee, N. Morrice, V. Cleghon, *Mol. Cell* **2006**, *24*, 627 – 633.
[40] J. Monod, J. Wyman, J. P. Changeux, *J. Mol. Biol.* **1965**, *12*, 88 – 118.
[41] D. E. Koshland, Jr., G. Nemethy, D. Filmer, *Biochemistry* **1966**, *5*, 365 – 385.
[42] E. Freire, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 10118 – 10122.
[43] K. Gunasekaran, B. Ma, R. Nussinov, *Proteins Struct. Funct. Bioinf.* **2004**, *57*, 433 – 443.
[44] V. J. Hilser, E. Freire, *J. Mol. Biol.* **1996**, *262*, 756 – 772.
[45] a) J. Vertrees, P. Barritt, S. Whitten, V. J. Hilser, *Bioinformatics* **2005**, *21*, 3318 – 3319; b) URL: http://best.utmb.edu/BEST/.
[46] I. Luque, E. Freire, *Proteins* **2000**, *41*, 63 – 71.
[47] For a review, see: I. Luque, S. A. Leavitt, E. Freire, *Annu. Rev. Biophys. Biomol. Struct.* **2002**, *31*, 235 – 256.
[48] S. A. Leavitt, A. Schon, J. C. Klein, U. Manjappara, I. M. Chaiken, E. Freire, *Curr. Protein Pept. Sci.* **2004**, *5*, 1 – 8.
[49] G. Tachedjian, M. Orlova, S. G. Sarafianos, E. Arnold, S. P. Goff, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 7188 – 7193.
[50] a) A. Christopoulos, T. Kenakin, *Pharmacol. Rev.* **2002**, *54*, 323 – 374; b) C. J. Langmead, A. Christopoulos, *Trends Pharmacol. Sci.* **2006**, *27*, 475 – 481.
[51] M. Chebib, G. A. R. Johnston, *J. Med. Chem.* **2000**, *43*, 1427 – 1447.
[52] a) W. Wriggers, K. Schulten, *Proteins Struct. Funct. Bioinf.* **1997**, *29*, 1 – 14; b) URL: http://biomachina.org/disseminate/hingefind/hingefind.html.
[53] a) V. Hayward, V. , H. J. C. Berendsen, *Proteins Struct. Funct. Bioinf.* **1998**, *30*, 144 – 154; b) S. Hayward, R. Lee, *J. Mol. Graph. Model.* **2002**, *21*, 181 – 183; c) URL: http://www.cmp.uea.ac.uk/dyndom.
[54] a) B. Brooks, M. Karplus, *Proc. Natl. Acad. Sci. USA* **1983**, *80*, 6571 – 6575; b) N. Go, T. Noguti, T. Nishikawa, *Proc. Natl. Acad. Sci. USA* **1983**, *80*, 3696 – 3700.
[55] M. M. Tirion, *Phys. Rev. Lett.* **1996**, *77*, 1905 – 1908.
[56] K. Hinsen, *Proteins Struct. Funct. Bioinf.* **1998**, *33*, 417 – 429.
[57] I. Bahar, A. R. Atligan, B. Erman, *Folding Des.* **1997**, *2*, 173 – 181.
[58] a) F. Tama, Y. H. Sanejouand, *Protein Eng.* **2001**, *14*, 1 – 6; b) M. Delarue, Y.-H. Sanejouand, *J. Mol. Biol.* **2002**, *320*, 1011 – 1024.
[59] W. G. Krebs, V. Alexandrov, C. A. Wilson, N. Echols, H. Yu, M. Gerstein, *Proteins Struct. Funct. Bioinf.* **2002**, *48*, 682 – 695.
[60] a) E. Lindahl, C. Azuara, P. Koehl, M. Delarue, *Nucleic Acids Res.* **2006**, *34(Web Server issue)*, W52 – W56; b) URL: http://lorentz.immstr.pasteur.fr/nma/index.php.
[61] a) K. Suhre, Y. H. Sanejouand, *Nucleic Acids Res.* **2004**, *32(Web Server issue)*, W610 – W614; b) URL: http://www.igs.cnrs-mrs.fr/elnemo/.
[62] a) P. J. Kraulis, *J. Appl. Crystallogr.* **1991**, *24*, 946 – 950; b) URL: http://www.avatar.se/molscript/.
[63] W. L. DeLano, The PyMOL Molecular Graphics System, **2002**; URL: http://pymol.sourceforge.net/.
[64] F. Tama, F. X. Gadea, O. Marques, Y. H. Sanejouand, *Proteins Struct. Funct. Bioinf.* **2000**, *41*, 1 – 7.
[65] A. Taly, M. Delarue, T. Grutter, M. Nilges, N. Le Novere, P. J. Corringer, J. P. Changeux, *Biophys. J.* **2005**, *88*, 3954 – 3965.
[66] H. Valadie, J. J. Lacapcre, Y. H. Sanejouand, C. Etchebest, *J. Mol. Biol.* **2003**, *332*, 657 – 674.
[67] G. M. Keseru, I. Kolossvary, *J. Am. Chem. Soc.* **2001**, *123*, 12708 – 12709.
[68] C. N. Cavasotto, J. A. Kovacs, R. A. Abagyan, *J. Am. Chem. Soc.* **2005**, *127*, 9632 – 9640.
[69] L. W. Yang, X. Liu, C. J. Jursa, M. Holliman, A. J. Rader, H. A. Karimi, I. Bahar, *Bioinformatics* **2005**, *21*, 2978 – 2987.
[70] L. W. Yang, A. J. Rader, X. Liu, C. J. Jursa, S. C. Chen, H. A. Karimi, I. Bahar, *Nucleic Acids Res.* **2006**, *34(Web Server issue)*, W24 – W31.
[71] a) D. J. Jacobs, A. J. Rader, L. A. Kuhn, M. F. Thorpe, *Proteins* **2001**, *44*, 150 – 165; b) M. I. Zavodszky, M. Lei, M. F. Thorpe, A. R. Day, L. A. Kuhn, *Proteins* **2004**, *57*, 243 – 261.
[72] A. Ahmed, H. Gohlke, *Proteins* **2006**, *63*, 1038 – 1051.
[73] a) J. A. McCammon, M. Karplus, *Nat. Struct. Biol.* **2002**, *9*, 646 – 652; b) T. E. Cheatham III, P. A. Kollman, *Annu. Rev. Phys. Chem.* **2000**, *51*, 435 – 471.
[74] J. A. McCammon, B. R. Gelin, M. Karplus, *Nature* **1977**, *267*, 585 – 590.
[75] R. Elber, M. Karplus, *Science* **1987**, *235*, 318 – 321.
[76] M. Rueda, C. Ferrer-Costa, T. Meyer, A. Pérez, J. Camps, A. Hospital, J. L. Gelpí, M. Orozco, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 796 – 801.
[77] For reviews, see: a) A. Kitao, N. Go, *Curr. Opin. Struct. Biol.* **1999**, *9*, 164 – 169; b) H. J. C. Berendsen, S. Hayward, *Curr. Opin. Struct. Biol.* **2000**, *10*, 165 – 169.
[78] M. Rueda, P. Chacón, M. Orozco, *Structure* **2007**, *15*, 565 – 575.
[79] H. Alonso, A. A. Bliznyuk, J. E. Gready, *Med. Res. Rev.* **2006**, *26*, 531 – 568.
[80] a) B. J. Berne, J. E. Straub, *Curr. Opin. Struct. Biol.* **1997**, *7*, 181 – 189; b) A. Mitsutake, Y. Sugita, Y. Okamoto, *Biopolymers* **2001**, *60*, 96 – 123; c) D. Hamelberg, J. Mongan, J. A. McCammon, *J. Chem. Phys.* **2004**, *120*, 11919 – 11929.
[81] Ajay, M. A. Murcko, *J. Med. Chem.* **1995**, *38*, 4953 – 4967.
[82] B. Honig, A. Nicholls, *Science* **1995**, *268*, 1144 – 1149.
[83] T. Rodinger, R. Pomès, *Curr. Opin. Struct. Biol.* **2005**, *15*, 164 – 170.

 © 2007 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

[84] T. P. Straatsma, J. A. McCammon, *Annu. Rev. Phys. Chem.* **1992**, *43*, 407 – 435.
[85] M. K. Gilson, H. X. Zhou, *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 21 – 42.
[86] J. Åqvist, V. B. Luzhkov, B. O. Brandsdal, *Acc. Chem. Res.* **2002**, *35*, 358 – 365.
[87] N. Foloppe, R. Hubbard, *Curr. Med. Chem.* **2006**, *13*, 3583 – 3608.
[88] P. A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D. A. Case, T. E. Cheatham III, *Acc. Chem. Res.* **2000**, *33*, 889 – 897.
[89] A. Checa, A. R. Ortiz, B. de Pascual-Teresa, F. Gago, *J. Med. Chem.* **1997**, *40*, 4136 – 4145.
[90] S. Wold, E. Johansson, M. Cocchi in *3D QSAR in Drug Design. Theory, Methods and Applications* (Ed.: H. Kubinyi), ESCOM Science Publishers B.V., Leiden, **1993**, pp. 523 – 550.
[91] R. D. Cramer III, D. E. Patterson, J. D. Bunce, *J. Am. Chem. Soc.* **1988**, *110*, 5959 – 5967.
[92] For reviews, see: a) R. C. Wade, A. R. Ortiz, F. Gago, *Perspect. Drug Discovery Des.* **1998**, *9/10/11*, 19 – 34; b) R. C. Wade, A. R. Ortiz, F. Gago in *3D QSAR in Drug Design, Vol. 2* (Eds.: H. Kubinyi, G. Folkers, Y. C. Martin), Kluwer Academic, Dordrecht, **1998**, pp. 19 – 34; c) J. Damborský, J. Kmuníèek, T. Jedlièka, S. Luengo, F. Gago, A. R. Ortiz, R. C. Wade, *Enzyme Functionality: Design, Engineering and Screening* (Ed.: A. Svendsen), Marcel Dekker, New York · Basel, **2004**, pp. 79 – 96.
[93] F. Rodríguez-Barrios, F. Gago, *J. Am. Chem. Soc.* **2004**, *126*, 2718 – 2719.
[94] M. Murcia, A. R. Ortiz, *J. Med. Chem.* **2004**, *47*, 805 – 820.
[95] M. Murcia, A. Morreale, A. R. Ortiz, *J. Med. Chem.* **2006**, *49*, 6241 – 6253.
[96] D. González-Ruiz, H. Gohlke, *Curr. Med. Chem.* **2006**, *13*, 2607 – 2625.
[97] S. F. Sousa, P. A. Fernandes, M. J. Ramos, *Proteins* **2006**, *65*, 15 – 26.
[98] R. M. Knegtel, I. D. Kuntz, C. M. Oshiro, *J. Mol. Biol.* **1997**, *266*, 424 – 440.
[99] M. Gerstein, A. M. Lesk, C. Chothia, *Biochemistry* **1994**, *33*, 6739 – 6749.
[100] J. Schlitter, M. Engels, P. Krüger, *J. Mol. Graph.* **1994**, *12*, 84 – 89.
[101] J. Mendieta, G. Ramírez, F. Gago, *Proteins Struct. Funct. Bioinf.* **2001**, *44*, 460 – 469.
[102] J. Mendieta, F. Gago, G. Ramírez, *Biochemistry* **2005**, *44*, 14470 – 14476.
[103] F. Rodríguez-Barrios, J. Balzarini, F. Gago, *J. Am. Chem. Soc.* **2004**, *126*, 15386 – 15387.
[104] F. Rodríguez-Barrios, J. Balzarini, F. Gago, *J. Am. Chem. Soc.* **2005**, *127*, 7570 – 7578.
[105] D. M. Lorber, M. K. Udo, B. K. Shoichet, *Protein Sci.* **2002**, *11*, 1393 – 1408.
[106] C. M. Crews, J. L. Collins, W. S. Lane, M. L. Snapper, S. L. Schreiber, *J. Biol. Chem.* **1994**, *269*, 15411 – 15414.
[107] E. Marco, S. Martín-Santamaría, C. Cuevas, F. Gago, *J. Med. Chem.* **2004**, *47*, 4439 – 4452.
[108] M. R. Walter, W. J. Cook, L. B. Cole, S. A. Short, G. W. Koszalka, T. A. Krenitsky, S. Ealick, *J. Biol. Chem.* **1990**, *265*, 14016 – 14022.
[109] V. L. Schramm, *Curr. Opin. Struct. Biol.* **2005**, *15*, 604 – 613.
[110] J. Mendieta, S. Martín-Santamaría, E. M. Priego, J. Balzarini, M. J. Camarasa, M. J. Pérez-Pérez, F. Gago, *Biochemistry* **2004**, *43*, 405 – 414.
[111] a) J. Damborsky, M. Prokop, J. Koca, *Trends Biochem. Sci.* **2001**, *26*, 71 – 73; b) URL: http://ncbr.chemi.muni.cz/triton/triton.html.
[112] a) O. Poirot, K. Suhre, C. Abergel, E. O'Toole, C. Notredame, *Nucleic Acids Res.* **2004**, *32(Web Server issue)*, W37 – W40; b) URL: http://www.igs.cnrs-mrs.fr/Tcoffee/.
[113] a) D. Lupyan, A. Leo-Macias, A. R. Ortiz, *Bioinformatics* **2005**, *21*, 3255 – 3263; b) URL: http://ub.cbm.uam.es/mammoth/mult/.
[114] J. Chen, C. L. Brooks III, *Proteins Struct. Funct. Bioinf.* **2007**, *67*, 922 – 930.
[115] R. Sánchez, U. Pieper, F. Melo, N. Eswar, M. A. Marti-Renom, M. S. Madhusudhan, N. Mirkovic, A. Sali, *Nat. Struct. Biol.* **2000**, *Suppl. 7*, 986 – 990.
[116] K. Ginalski, *Curr. Opin. Struct. Biol.* **2006**, *16*, 172 – 177.
[117] a) U. Pieper, N. Eswar, F. P. Davis, H. Braberg, M. S. Madhusudhan, A. Rossi, M. Marti-Renom, R. Karchin, B. M. Webb, D. Eramian, M. Y. Shen, L. Kelly, F. Melo, A. Sali, *Nucleic Acids Res.* **2006**, *34(Database issue)*, D291 – D295; b) URL: http://salilab.org/modbase/.
[118] P. J. O'Hara, P. O. Sheppard, H. Thogersen, D. Venezia, B. A. Haldeman, V. McGrane, K. M. Houamed, C. Thomsen, T. L. Gilbert, E. R. Mulvihill, *Neuron* **1993**, *11*, 41 – 52.
[119] a) L. Holm, C. Sander, *Nucleic Acids Res.* **1997**, *25*, 231 – 234; b) URL: http://ekhidna.biocenter.helsinki.fi/dali/.
[120] A. I. Hernández, O. Familiar, A. Negri, F. Rodríguez-Barrios, F. Gago, A. Karlsson, M. J. Camarasa, J. Balzarini, M. J. Pérez-Pérez, *J. Med. Chem.* **2006**, *49*, 7766 – 7773.
[121] a) M. Petrek, M. Otyepka, P. Banas, P. Kosinova, J. Koca, J. Damborsky, *BMC Bioinf.* **2006**, *7*, 316; b) URL: http://loschmidt.chemi.muni.cz/caver.
[122] S. J. Teague, *Nat. Rev. Drug Discovery* **2003**, *2*, 527 – 541.
[123] The Institute for Genomic Research; URL: http://www.tigr.org/tdb/.
[124] T. Kenakin, *Trends Pharmacol. Sci.* **2004**, *25*, 186 – 192.

Received: April 16, 2007
Revised: July 4, 2007
Published online on September 5, 2007

 © 2007 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim