FINAL

Protein Flexibility and Mobility in Structure-Based Drug Design

Aqeel Ahmed[§], Sina Kazemi[§] and Holger Gohlke[*]

Department of Biological Sciences, Molecular Bioinformatics Group, J.W. Goethe-University, Frankfurt, Germany

Abstract: Initially, structure-based drug design (SBDD) approaches relied on the validity of the "lock and key" model, although this assumption leads to clear limitations. Thus, there are considerable efforts nowadays to incorporate the influence of (changes of) protein flexibility and mobility into recent drug design approaches. These efforts are grounded on the "induced-fit" and "conformational selection" models of ligand binding to proteins. Here we will summarize computational approaches in SBDD that address these issues, with a focus on methods that account for receptor plasticity. In particular, we consider how protein plasticity can be incorporated into docking strategies. Two requirements need to be met for this: first, one needs to detect what can move and how; second, this knowledge needs to be transformed into a docking algorithm. With regard to the former, knowledge about moving protein parts can be gained from experimental information as well as established techniques such as molecular dynamics simulations, graph theoretical and geometry-based approaches, or harmonic analysis-based methods. With regard to the latter, a plethora of approaches has been presented recently that range from considering protein plasticity only implicitly to modeling sidechain movements to also including backbone changes. A hallmark of all these approaches is that they need to balance accuracy and efficiency. Case studies of SBDD for which the inclusion of protein plasticity was crucial to success are noted along these lines. This provides a picture of scope and limitations of the current approaches as well as guidelines for further developments.

Key Words: rigidity, harmonic modes, flexible docking, protein-ligand, conformational variability, plasticity.

1. INTRODUCTION

In a recent insightful essay about designing ligands that bind proteins Whitesides and coworker [1] concluded "We are certainly not close to the end of the problem, but we are at least at the beginning.". This cautiously optimistic statement very well reflects also the current status in a subfield of structure-based drug design (SBDD), that is, how to

*Corresponding Author: Tel: (+49) 69 798-29411; E-mail: gohlke@bioinformatik.uni-frankfurt.de
[§]Both authors contributed equally to this work.

Garry W. Caldwell / Atta-ur-Rahman / Michael R. D'Andrea / M. Iqbal Choudhary (Eds.)
All rights reserved – © 2007 Bentham Science Publishers.

deal with protein flexibility and mobility and their changing upon complex formation. It is very well received nowadays that our ability to successfully predict ligands crucially depends on our detailed understanding of molecular recognition processes [2]. Issues related to receptor flexibility and mobility are certainly essential here. In the following we will understand flexibility as a static property that only determines the *possibility* of motion, whereas nothing actually moves [3]; mobility (or motion) in turn describes *actual* movements in terms of directions and amplitudes. Flexibility is not necessarily a prerequisite for mobility, as rigid parts of a biomolecule (e.g., domains) can well move as a whole (e.g., when connected by hinges). However, mobility provides the origin for receptor plasticity, which enables binding partners to conformationally adapt to each other.

Plasticity is perhaps the foremost characteristic of flexibility and mobility of biomolecules. Pronounced plasticity upon ligand binding has been observed in several pharmacologically important proteins, e.g., HIV-1 protease [4], aldose reductase [5], FK506 binding protein [6], renin [7] and DHFR [6]. However, conformational adaptation based on inherent flexibility is also a hallmark of DNA [8] or RNA [9-11] targets. In the overwhelming number of cases reported in literature, hydrophobic interactions led to the observed conformational changes such that the notion of an "hydrophobic collapse" of a receptor around a ligand emerged [12]. In the extreme case, such a collapse can lead to disorder-order transitions of intrinsically unfolded proteins upon complex formation [13-15].

In view of the observed examples of receptor plasticity, the "rigid receptor hypothesis" [12], which is based on the "lock and key" model [16] of molecular recognition and has served as a foundation for the development of early docking approaches, is no longer tenable. However, recent studies suggest that biomolecular recognition processes and flexibility (or changes of flexibility) of the binding partners are more fundamentally interrelated than even acknowledged by the "induced fit" [17] model. As such, the "conformational selection" model [18,19] proposes that proper conformations are "picked" by the binding from the ensembles of rapidly interconverting conformational species of the unbound molecules. This is supported by experimental evidence for the presence of conformational variability of binding partners prior to their association [20] and yields an explanation as to why a single protein can bind multiple unrelated ligands at the same site [21]. The fact that high affinity ligands were identified in some cases that do not mimic the binding mode of the natural substrate then does not come as a surprise anymore. The model also provides the foundation for computational investigations of conformational fluctuations of the unbound protein state, which may reveal conformational states adopted by the bound proteins (see chapter 2). Despite the conceptual differences between "induced fit" and "conformational selection", one should note that both models at least agree with regard to the statement that in every receptor-ligand complex the conformation of both binding partners has to be a specific one for both to fit.

Receptor flexibility and mobility is not limited to influencing only steric complementarity (and, hence, direct interactions between the binding partners) in molecular recognition. Rather, pronounced additional (indirect) energetic and entropic contributions to the binding affinity arise. These are almost completely neglected in current SBDD approaches. First, *conformational variability* of the receptor leads to a disfavorable *reorganization energy* that can be large, even for relatively well preorganized binding sites. In binding free energy calculations of ligands binding to avidin and streptavidin, Laz-

aridis *et al.* obtained values between 9 and 36 kcal/mol [22]. These numbers are of the same order of magnitude than the binding free energies and, thus, should clearly be considered as correction terms when it comes to flexible receptor docking [23]. Yet, these values are in general difficult to compute because they correspond to the small difference between large conformational energies of the bound and unbound receptor. Further complications arise from the fact that the protein reorganization energy may not be the same even for slightly different ligands [22].

Second, *changes in the receptor flexibility* upon complex formation (i.e., changes in the internal degrees of freedom) lead to *configurational entropy* contributions. Usually recognition sites that are in direct contact with the binding partner become less flexible, although influences on the flexibility of residues distant from the epitope are also seen. This can be explained in that perturbations at the binding site can propagate to remote locations by altering the dynamic network of interactions in proteins [24-26]. Conversely, it is increasingly recognized that complex formation may also lead to an increase in configurational entropy (which is related to an increase in flexibility of the system), compensating for the loss of translational and rotational entropy upon association [27-35]. For some cases, transfer of flexibility to other protein parts (within one binding partner) has been described, leading to a redistribution of protein configurational entropy, also potentially reducing the total entropy loss [18,19,36-39]. Computing configurational entropies is notoriously difficult [40,41]. The methods applied range from empirical scales considering the restriction of sidechain motions [42] to harmonic analyses based on static structures [42,43] or covariance matrices obtained from MD simulations [44,45]. All but the first are inadequate in terms of computational requirements for applications in lead finding and virtual screening. Clearly, more work is required in this area.

Finally, protein flexibility and mobility provide a basis for *allosteric regulation*, which is considered to be advantageous compared to competitive regulation due to a generally higher selectivity of the modulator [46]. In the "classical view" [47,48], allostery is explained by the occurrence of conformational changes in the receptor. Complementary, the "modern view of allostery" [49] considers the allosteric modulator to change the population of conformational substates that are already accessible in the global ensemble of the system (see above) [18,19] and that differ in their binding or signal transduction properties [46,48]. Thus, this view emphasizes the role of (changes in) the protein flexibility as an entropic carrier of allosteric information [20,49-51]. From a SBDD point of view several opportunities exist to exploit these mechanisms, ranging from the identification of novel allosteric pockets [52] to actually identifying allosteric modulators. In all cases, a thorough understanding and accurate modeling of protein flexibility and mobility is necessary.

Despite the widespread influence of protein flexibility and mobility in molecular recognition, we will mainly focus in the remainder of this review on computational approaches in SBDD that address the issue of receptor *plasticity*. We feel that, from a computational point of view, this field has seen the most progress over recent years. In particular, we consider how protein plasticity can be incorporated into docking. Two requirements need to be met for this: first, one needs to detect what can move and how; second, this knowledge needs to be transformed into a docking algorithm. With regard to the former, knowledge about moving protein parts can be gained from experimental information as well as established techniques such as MD, graph-theoretical approaches,

or normal mode analysis. With regard to the latter, a plethora of approaches has been presented that range from considering protein plasticity only implicitly to modeling sidechain movements to also including backbone changes. In all these cases, making a tradeoff between accuracy and efficiency is required.

2. DETERMINING PROTEIN FLEXIBILITY AND MOBILITY

Protein plasticity comprises a range of possible movements, from single side-chains to drastic structural rearrangements as seen in calmodulin [53]. Depending on where the binding site is located, one or more of these movement types will be relevant for docking. What remains particularly challenging is to couple (and model) movements at different scales that occur in binding. Nevertheless, two recent studies indicate that with little effort much should be gained. First, Najmanovich *et al.* [54] reported that rotations in side-chains of up to three residues account for ≈ 85% of all the cases where there is a conformational change upon ligand binding. Thus, many conformational changes can be captured by sidechain motions only. Similarly, a recent study by Zavodszky and Kuhn [55] assessed for a large set of typical enzymatic targets to what extent side-chain rotations of amino acids contribute to the plasticity within the binding site. Although sidechain rotations were proven to be necessary for correctly docking about half of the protein-ligand complexes considered, 95% of the rotations were smaller than 45º.

In this chapter, we focus on techniques for the prediction of protein flexibility and mobility. They have conceived renowned interest during the last decade due to their direct implication in rational ligand design on the one hand and limitations of experimental techniques to provide these insights in atomic detail on the other hand. We emphasize three distinct areas of research: force field-based, graph theoretical and geometry-based as well as harmonic analysis-based methods. These areas are characterized by either well established methods that are increasingly applied in SBDD or considerable developments recently.

2.1. Force Field-Based Methods

Here we consider techniques that model macromolecular movements by sampling the energy landscape of the molecule given by an empirical force field. In particular, molecular dynamics (MD) simulation is one of the most widely applied and accurate computational techniques currently being used in the field of macromolecular computation [56-58]. Therefore, in this section, we focus on MD simulations that were successfully applied in SBDD.

MD simulation is based on Newtonian dynamics where instantaneous forces present in the molecular structure are numerically integrated to generate a trajectory through phase space [59]. Important information related to SBDD can be derived by analyzing this trajectory. This includes the analysis of flexibility and mobility, generating multiple conformations of a protein, investigating allosteric mechanisms, and examining the influence of mutants.

MD has been utilized successfully for the investigation of receptor plasticity consequently enhancing SBDD. For example, a recent MD study [60] of HIV-1 integrase showed an intermittent opening of an unknown favorable binding trench adjacent to the catalytic site, which was experimentally validated later on [61]. Subsequent docking studies of novel ligands with the potential to bind to both regions showed greater selec-

tivity when interacting with the trench [60]. Similarly, Carlson *et al*. [62] developed a new method for generating dynamic pharmacophore models to compensate for the inherent plasticity of an active site. This method identifies conserved binding regions over multiple configurations of the active site and uses those regions to define a complementary model. A receptor based pharmacophore model of HIV-1 integrase derived from MD conformations was found to agree well with binding characteristics of known inhibitors of the system.

Improvements in MD simulation techniques and increased computational power have recently allowed performing MD simulations on *unbound* protein states that clearly show a potential for generating conformations that mimic *bound* states. These conformations may well be used subsequently in flexible docking approaches (see chapter 3). In a remarkable study, an artificially low solvent viscosity used in a MD simulation [63] of HIV-1 protease enabled a comprehensive sampling of the conformational space, which shaded light on the flap dynamics of the protein. The unliganded protease undergoes multiple conversions between the "closed" and "semiopen" forms observed in crystal structures of inhibitor bound and unliganded protein, respectively, including reversal of the flap "handedness". Although the overall dynamics of the unliganded protease was found to be predominantly populated by semiopen conformations, with closed and fully open structures being a minor component of the overall ensemble, these results strongly support the "conformational selection" model.

In another MD study [64] starting from the unbound form of aldose reductase a set of distinct conformational substates that may prove useful as alternative structural templates in virtual screening/docking for new inibitors was identified. Demonstrating that the ligand bound form can be predicted from an unbound structure, this finding again corroborates the "conformational selection" model. Along these lines, 41 proteins that form protein-protein complexes have been simulated in order to investigate the extent to which conformational fluctuations lead to novel conformational states [65]. Starting again from the unbound structures, it was found that fluctuations take some parts of the molecules into regions of conformational space closer to a bound state, although simulation times of 5 ns were not sufficient in any case to sample the complete bound state.

MD studies have also been applied for analyzing the influence of mutants and allosteric mechanisms on a structural basis. In the case of HIV-1 protease, a plausible mechanism for drug resistance due to the V82F/I84V double mutant was discerned, and the location of a potential new binding site for allosteric inhibitors was suggested [66]. This mechanism was also supported by a restrained MD simulation and docking study [67,68]. Another study [69] for the same protein elucidated the mechanism of multi-drug resistance due to non-active site mutations and, thus, provided hints for the design of novel drugs with reduced resistance potential. Regarding kinases, allosteric mechanisms were also studied by MD simulation. E.g., a dynamic coupling between the SH2 and SH3 domains of the inactive c-Src and Hck was observed, which impedes activation. Dephosphorylation of the protein tail reduces the coupling, as does the replacement of connector residues with glycine in the MD simulation [70].

2.2. Graph Theoretical and Geometry-Based Methods

Due to the computational burden needed by force field-based techniques to describe large scale motions and to analyze macromolecular flexibility, graph theoretical and ge-

ometry-based methods are getting popular. Graph theoretical methods can be utilized to analyze protein flexibility whereas geometric simulations, which aim at avoiding steric clashes thereby fulfilling bond length and angle constraints, can be used to efficiently generate macromolecular conformations.

Protein flexibility can be analyzed accurately and efficiently by a graph-theoretical technique, which allows identifying rigid and flexible regions from a single, static structure. Atoms and inter-atomic interactions in a protein structure are transformed into a bond-bending network by modeling atoms as nodes and covalent bonds and non-covalent interactions as edges (yielding distance and angle constraints). A fast combinatorial algorithm, the "pebble game", then identifies the rigid, over-rigid, and flexible regions by counting bond-rotational degrees of freedom in the network. This algorithm has been implemented into the FIRST (Floppy Inclusion and Rigid Substructure Topology) software [71]. FIRST analyses have been used to accurately identify rigid regions as well as collectively and independently moving regions in a series of proteins [72,73].

An interesting feature of the FIRST analysis is that changes in the flexibility of the binding partners due to complex formation can be investigated in detail e.g., in the case of a protein-protein complex formation [74], additional interactions across the interface led to a propagation of rigidity through the binding partners. This demonstrates the long-range aspect to rigidity percolation. Recently, FIRST was also applied in combination with MD for investigating the flexibility of prolyl oligopeptidase [75].

Flexibility information from FIRST, which leads to a natural coarse-graining of macromolecules based on rigid regions [76], was further exploited for simulating protein mobility using constrained geometric simulation (CGS) [77-79] or as input for an elastic network based approach (see chapter 2.3) [80]. CGS based methods such as ROCK (Rigidity Optimized Conformational Kinetics) [77] explore the rigidity-restricted conformational space by satisfying ring closure equations, whereas FRODA (Framework Rigidity Optimized Dynamic Algorithm) [79] makes use of a more efficient algorithm that moves flexible and rigid parts by ghost template rearrangements. The resulting structures from ROCK were used in flexible docking for the drug targets cyclophilin and estrogen receptor [81]. FRODA generated structures have not yet been applied for SBDD. However, docking studies of the multi-subunit protein complex photosystem I [82], which make use of FRODA conformations and aim at exploring alternative approaching pathways, have been reported recently.

CONCOORD (from CONstraint to COORDinates) by De Groot *et al.* [83,84] is another method that generates conformations by satisfying constraints. Starting from a random structure, conformational space is captured by fulfilling a set of upper and lower interatomic distance bounds that are derived from the experimental structure of the protein. The differences between upper and lower distance bounds depend on the strengths of interactions, with stronger interactions leading to smaller deviations. Frequently, CONCOORD derived structures are subsequently analyzed using essential dynamics (ED) [84], which simplifies the mobility study of a protein by decomposing its dynamics into essential modes.

CONCOORD can generate conformations very efficiently as compared to MD; therefore, it is well suited for larger systems. In the case of hyaluronate lyase [85], whose size precludes the application of MD to investigate biologically relevant time scales, flexibility (allosteric) information and functional implications were derived from

CONCOORD. Two ED modes of motion were identified: the first motion describes an opening and closing of a catalytic cleft, and the second motion demonstrates the mobility of a binding cleft, which may facilitate the binding of the negatively charged hyaluronan to the enzyme. Finally, Mustard and Ritchie [86] showed that docking to multiple eigen-structures (obtained by an ED study following a CONCOORD run) generates better docking predictions than docking only to unbound or model built structures.

Techniques from robotics motion planning [87] have so far been mainly applied to simulate the folding pathways of proteins [88-90] and RNAs [91,92]. However loop conformations of xylanase from *T. xylanilyticus* and amylosucrase from *N. polysaccharea* have been modeled [93,94] using an efficient geometric algorithm as a *a priori* filter for conformational search methods.

2.3. Harmonic Analysis-Based Methods

Large-amplitude molecular motions have been studied by normal mode analysis (NMA) for decades [95-97]. Alternatives in the form of elastic network models (ENM) have emerged recently, triggered by the development of simplified force-fields [98] and coarse-grained models of macromolecules: the Gaussian network model (GNM) [99,100] and the Anisotropic network model (ANM) [101-104] Since then ENM has been applied to a vast range of problems concerning flexibility/mobility of proteins and other large macro-molecules. Detailed reviews of these studies have been published recently [105-108].

The success of the ENM approach is based on a reduced protein representation and inherent coarse-graining. This allows predicting surprisingly accurately large conformational changes, which is difficult with force field based methods. In the original ENM approach, detailed atomic potentials were replaced with uniform residue-based harmonic potentials in a network of interacting residues (usually C-alpha positions). Later, further coarse-graining has been reported assuming structural rigidity based on secondary structure [109], rigidity of sequentially consecutive residues [109-111], or using a rigid cluster decomposition by FIRST [80]. Further coarse-graining can still be applied without losing atomic-level details [109,112,113]. That way, the method can be applied to even macromolecular assemblies.

In agreement with the "conformational selection" model, the conformational change captured by ligands is found for most of the proteins to occur along the lowest energy (frequency) modes calculated by normal mode analysis of the *unbound* protein. These modes usually involve hinge-bending, large-amplitude, and correlated motions [95,104,109,114]. Along these lines ENM has been mostly applied as a *a posteriori* analysis in combination with experimental studies, e.g., for examining functional dynamics in *E. coli* adenylate kinase, HIV-1 reverse transcriptase, and influenza virus hemagglutinin [115-117], cooperative and allosteric dynamics in tryptophan synthase [118], and binding effects in HIV-1 reverse transcriptase [119]. Moreover, several studies showed an efficient conformation and pathway generation by ENM-based techniques, which can be exploited for docking studies [120-123] (see chapter 3.4.3).

Apart from a large scale dynamical analysis, residue level analysis has also been successfully applied [124-127]. This is surprising, considering the simplicity and coarse-graining of the underlying model. For example, high frequency modes of GNM have been shown to be important for the identification of binding "hot spot" residues [126],

catalytic residues [125], and protein-binding sites [124]. Catalytic sites were found to be colocalized with global hinge centers predicted by GNM, whereas the ligand binding sites were found to be enjoying flexibility near the catalytic site [125]. In SBDD, these studies can be exploited for efficiently identifying binding hotspot and catalytic residues.

When harmonic analysis is combined with methods that provide atomic detail such as MD, one can utilize the benefit of both. In this respect MD/NMA hybrid methods have been proposed [128-130] that amplify collective motions along normal mode directions in a conventional MD. This method was successfully used for docking in the case of HIV-1 protease [129].

Loop motions are hard to predict but play an important role in accommodating ligands in binding pockets, e.g., in kinases, DOXP reductase, tyrosin phosphatase, and HIV-1 protease. Cavasotto *et al*. [123] introduced a measure of relevance of normal modes for important loop conformational changes upon ligand binding and found that only a few low-frequency modes (less than ten, but usually not the first low-frequency modes) are critical and sufficient to represent binding pocket mobility in protein kinases. Using these relevant modes an ensemble of alternative conformations for holo and apo structures of cAMP-dependent protein kinase was generated, which exhibit backbone rearrangements in two independent loop regions close to the binding pocket. Taking into account a limited set of relevant modes is particularly important for highly mobile receptors, for which the first low-frequency modes may not describe the movements of the region important for binding.

Recently, it was also shown that a FIRST analysis-based coarse-graining of ENM can lead to the accurate predictions of loop movements [80]. This can be explained by the fact that the appropriate coarse-graining removes irrelevant modes of the system (without losing the important functional modes), whereas the modes related to flexible regions become more emphasized. As a result, low-frequency modes representing functional motions in the flexible region are obtained. As an example, nearly the whole protein was found to be rigid by FIRST in the case of tyrosine phosphate, except the flexible loop region β7-α4. With a rigid cluster normal mode analysis (RCNMA), an accurate prediction of the loop movement by the lowest-frequency mode was then achieved (Fig. **1**).

3. INCORPORATING PROTEIN PLASTICITY INTO DOCKING ALGORITHMS

3.1. Efficiency Considerations

Once the dynamic properties of the protein are known, incorporating this information is the next challenge in algorithmic development. The main concern in this step is computational efficiency. Even with the simplest models of proteins where the bond lengths and angles are kept constant and only the backbone and sidechain dihedral angles are considered rotatable, there are still too many degrees of freedom to obtain a thorough sampling of the protein conformational space for available computational times. This is true more than ever for the short timeframe available for a docking run. Further difficulties arise when it comes to recognizing energetically accessible protein conformations among the set of decoys generated. Both findings led to the conclusion that docking including receptor plasticity is as difficult as the protein folding problem [131].

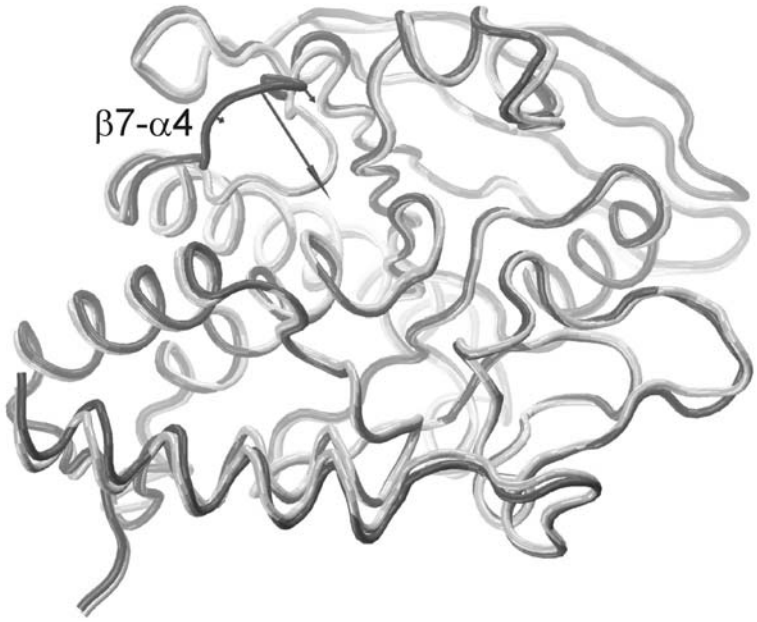

Fig. (1). The amplitudes and directions of motions as predicted by the mode most involved in the conformational change of the loop β7-α4 of tyrosine phosphatase are depicted as arrows. The mode was calculated using the rigid cluster normal mode analysis approach RCNMA [80] and the unbound (dark gray) protein conformation. The amplitudes of the motions were scaled for best graphical representation. In addition, the bound (light gray) conformation is superimposed, demonstrating a striking agreement between computed and experimentally determined protein plasticity. (The figure was reproduced from ref. [80]).

Several developments fostered the transition from early docking algorithms that described initially both the ligand and the receptor as rigid [132] and later considered ligand flexibility [131,133-137] to approaches where receptor plasticity is considered. First, the development of computer hardware and the resulting higher computational capability in terms of speed and memory as well as the common access to computer workstations for scientists had a big impact on the design of new docking algorithms. Since the very beginning, computational ligand design has also attracted the attention of computer scientist leading to better algorithms particularly in the field of search algorithms. Also, the availability of affordable computer clusters with high capacities led to a new generation of parallel applications able to overcome algorithmic shortcomings in terms of speed (see, e.g., FightAIDS@Home: http://fightaidsathome.scripps.edu). Finally, a better understanding of biophysical processes on the microscopic level and the development of new models for the description of movements in biomacromolecules (as described in previous chapters) has allowed considering receptor plasticity. Yet, despite these encouraging developments, fully flexibly docking still requires a tradeoff between prediction quality and calculation time.

The following chapters shall give an overview of current fully flexible docking methods, an exciting and steadily growing area (Fig. **2**) in the field of computational SBDD. The methods can be divided into three major groups according to the degree of the motions they consider: I) Plasticity is considered *implicitly*. In many cases this is achieved by modifying already existing methods; II) Sidechain conformational changes in the binding pocket are modeled; III) Large scale conformational changes including backbone motions are taken into account.

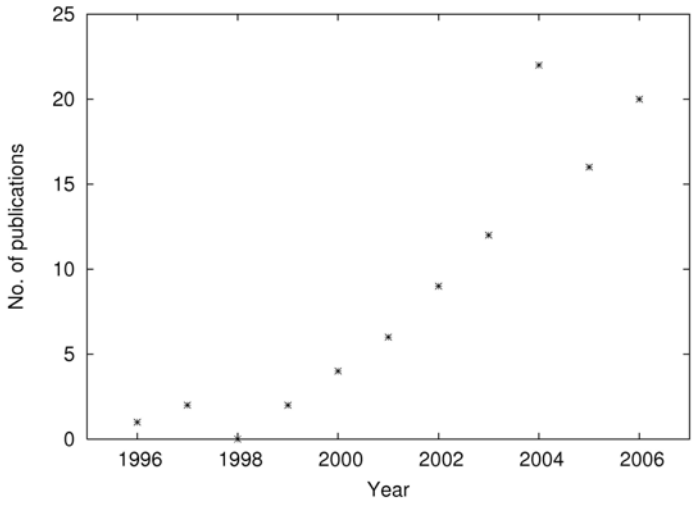

Fig. (2). The number of publications dealing with protein plasticity in docking over the past ten years. The data was derived from the ISI Web of Science database (updated November 17th, 2006) using the query "docking AND ("flexible protein" OR "protein flexibilit*" OR "protein plasticit*")".

3.2. Considering Plasticity Implicitly

3.2.1. Capped Van Der Waals Interactions and Reduced Radii for Soft Docking

Soft docking is the simplest approach for tackling the problem of receptor plasticity from an algorithmic point of view. Generally speaking, repulsive contributions to the energy function are reduced either by capping the van der Waals contributions or by reducing the van der Waals radii of the receptor atoms. In both cases, the interpenetration of the receptor surface by a ligand atom becomes possible to a certain degree, which allows for the binding of a larger ligand to a structure determined experimentally in complex with a smaller ligand. For a receptor in which small sidechain rearrangements in the binding pocket are sufficient to adapt to the bound conformation (e.g., sidechains moving away to make room for the ligand), this approach should yield the right binding pose for the ligand.

The main advantage of this approach is that it can be implemented very easily. Established docking algorithms that already consider ligand flexibility only require small changes [138]. The second advantage of soft docking is its speed. The approach is as fast

as the algorithms considering only ligand flexibility because no additional degrees of freedom are added to the search problem. Two other approaches in which van der Waals softening was applied were reported by Schnecke *et al.* [139] and Mizutani *et al.* [140] (see below). In both cases, very good docking results were obtained if only local protein movements occurred.

3.2.2. Soft Knowledge-Based Potentials

The reduced steepness (i.e., increased softness) of knowledge-based potentials compared to force field-based or empirical scoring functions has been recognized as an advantage [141], as the former functions are more robust to small changes in a receptor conformation. Knowledge-based potentials are derived by transforming pair distribution functions obtained for receptor-ligand atom pair interactions observed in experimental complex structures [142]. The potential softness can be amplified during the derivation process by "smearing" observed interactions over a larger distance range. In the case of the DrugScore scoring function [143], this has been accomplished by a moving equilateral triangle function of 0.4 Å basis length. This idea was recently extended to more explicitly consider atom mobilities of protein atoms based on crystallographic B factors. Surprisingly, however, with the resulting M-score function [144] (which is based on the DrugScore formalism otherwise) only a slight improvement in the correlation between experimental and calculated binding affinities was found compared to when protein atomic mobility was not considered.

3.2.3. Combining Information from Multiple Structures

Receptor plasticity can also be represented based on multiple structures. The different receptor conformations can be derived from X-ray structures complexed with different ligands, NMR structure ensembles, MD (see chapter 2.1) and geometrical simulation trajectories (see chapter 2.2), or harmonic modes (see chapter 2.3). The advantage is that conformational variability is not limited to a small region of the protein such as the binding pocket in this case. Here, we do not consider in more detail the trivial solution to run a parallel docking against every conformation [145] as the main potential drawback is that the dynamic picture might not be complete, and some relevant conformations might not be present. Instead, we will focus on approaches where, following a mean-field approach, properties of several conformations are combined into one protein representation.

Knegtel *et al.* [146] used such an approach with NMR structure ensembles and also with several crystal conformations. Interaction energy contributions of different conformations of the target protein were merged into one "average" grid, and a rigid ligand-rigid receptor docking was then performed with DOCK [132]. The approach was tested on a database of about 150 small-molecules to which known HIV-1 protease ligands were added. All active ligands could be ranked within the top 17% of the database. In contrast, in the case of single structure grids at most three inhibitors could be identified within the top 25%. Similarly, Broughton *et al.* [147] used combinations of several conformations to derive interaction grids, which afterwards were combined into merged grids. The conformations were obtained by applying short MD simulations to cocrystallized structures of COX-2 and dihydrofolate-reductase (DHFR). In this case the authors made use of ligands contained in the MD snapshots to superimpose the structures before they combined the grids.

The way the grids are combined and which method is used for weighting the single contributions in the combination process has a dramatic influence on the outcome of the docking result, as found by Oesterberg *et al.* [148] for the test case HIV-1 protease. The authors compared several weighting schemes and analyzed the results with respect to the predictive power in docking. They could show that naïve mean, minimum, and maximum grids performed badly because they were dominated either by repulsive or attractive contributions of the combined grids. Alternatively, they also tested a weighted averaging scheme similar to that of Knegtel *et al.* as well as one where the weight factors were determined by a Boltzmann term. Both approaches succeeded in carrying over important binding determinants of multiple structures into a single grid representation, resulting in good dockings. Furthermore, for a dataset of 21 docked HIV-1 protease inhibitors, a rmsd between experimental and predicted binding free energies of about 1.4 kcal/mol was obtained. The authors also pointed out a cyclic urea inhibitor, which was challenging: only in the corresponding protein conformation the water 301, which generates a repulsive region unfavorable for the right pose of the urea inhibitor, was missing. Nevertheless, the inhibitor was also docked correctly, because the combined maps softened the repulsion of the water molecules present in the other 20 structures. We note, however, that such a mean-field representation also bears a risk. If, e.g., a moving sidechain leads to two distinct interaction regions in two different receptor conformations, the combined grid may (non-physically) represent both these regions.

Sudbeck *et al.* [149] pursued a different way of combining multiple structures. From nine HIV reverse transcriptase crystal structures complexed with different inhibitors they derived a summarized representation of a composite binding site surface. Afterwards the authors used a linear gradient minimization technique to optimize the structure of a ligand and the binding pocket residues within a radius of 5 Å of one reference structure. The results from this docking run were ranked by the constructed combined binding pocket. It is notable that with this approach Sudbeck *et al.* affected the development of two new inhibitors.

3.3. Avoiding Collisions with Sidechains

According to the "minimal rotation hypothesis" [55] developed on 20 different proteins by Zavodszky and Kuhn protein sidechains move as little as necessary in order to achieve a collision-free complex formation. This holds true for proteins that undergo small conformational changes upon protein binding. The hypothesis provides the basis for many docking approaches that accommodate protein plasticity by avoiding collisions of sidechains with the ligand in the binding pocket. An implication of this hypothesis is that small deviations from ideal geometries only result in small energy penalties and, hence, should lead to accessible protein conformations.

3.3.1. Optimizing Local Hydrogen Bond Networks

One of the early docking programs taking partial receptor plasticity into account was GOLD from Jones *et al.* [150,151], which is nowadays very broadly used [152]. It uses a genetic algorithm (GA) to not only optimize ligand placement, but also the conformation of some terminal bonds of sidechains in the receptor binding pocket. Thereby it fosters the construction of a hydrogen bond network. The authors consider this an important point especially in those cases where X-ray receptor structures are used for docking because this information is not directly given by the X-ray structure as a result of the lack of hydrogen positions. GOLD was evaluated on a test set of 100 protein complexes of

which it could predict "good" complex configurations in 73 % of the cases. Unfortunately, the results for hydrophobic ligands are rather modest, which is attributed to the fact that I) the chromosome encoding favors the sampling of hydrogen-bonding motifs and II) the fitness function does not have a term for desolvation and, hence, underestimates hydrophobic contributions. Also, systematic problems in ranking very polar ligands and in ranking general ligands in large cavities have been reported [153]. Finally, the amount of conformational change that can be accommodated by GOLD is limited.

3.3.2. Moving a Pre-Defined Set of Sidechains

Since 1990, when Goodsell and Olson [154] reported the first version of AutoDock, several important improvements both with respect to the search problem and the energetic description of the protein-ligand complexes have been added [134,155]. The version AutoDock 3.0 is the most widely used docking program currently [152]. In the latest version 4.0, AutoDock now also supports sidechain mobility. The user can chose sidechains that should be considered explicitly during the docking run. Similar to GOLD, the conformations of these sidechains are encoded into the chromosome representation of the docking solutions, as is the ligand configuration, and both are simultaneously optimized during the docking run. The drawback of this method is that considering sidechains that way requires to restrict the number of sidechains set flexible to assure suitable calculation times.

3.3.3. Sidechain Rotations to Avoid Clashes

In the program SPECITOPE, a tool for screening large structural databases for potential ligands for a protein [139], Schnecke *et al.* also use sidechain rotations to optimize the complex structure. These are introduced at a late stage of the docking to remove clashes between the ligand and the receptor. The sidechain rotations are done in an iterative manner: If a clash is detected the bond of the sidechain next to the clashing atoms is rotated to remove the clash. If a clash free conformation cannot be obtained that way, another bond next to the ligand will be used. The method will finally identify complex structures without steric overlaps. However, as sidechains that are not in direct contact with the ligand are not optimized, the obtained solution is not a global minimum energy conformation. The approach was used to screen a library of 140,000 peptide fragments for the targets serine protease, a DNA repair enzyme, an aspartic proteinase, and a glycosyltransferase. This could be achieved in about an hour by the algorithm. SPECITOPE was able to narrow down the library of ≈140,000 peptides to 23 potential ligands for subtilisin, 14 for uracil-DNA glycosylase, 35 for aspartic proteinase, and 13 for cyclodextrin glycosyltransferase.

Schnecke and Kuhn have also proposed the new docking algorithm SLIDE [156] that incorporates sidechain mobility. The algorithm places an anchor fragment of the ligand in a first step. Afterwards the remaining fragments are added according to their database conformation, and clashes of the protein sidechains and the ligand are corrected by rotation around the sidechain and ligand bonds in the non-anchor regions. One disadvantage of the algorithm is that only a restricted flexibility of the ligand and receptor is considered, another one is that the placement of the anchor fragment is critical for the docking success.

Mizutani *et al.* [140] recently proposed an approach based on ADAM [157] (a docking tool originally developed for rigid docking) that allows the consideration of side-

chain movements. It combines both the "van der Waals offset method" (see above) and a post-docking optimization. The van der Waals energy curve is shifted so that the interatomic contact distances become smaller and the protein cavity is widened. After a docking with ADAM into such a cavity, clashes with the protein atoms are detected. With the energy minimization program BLUTO, which was developed during this study, the complex structure is then corrected to become collision free by considering local motions of protein atoms. BLUTO considers intermolecular van der Waals and electrostatic energies and intramolecular energies of the ligand and the protein. The authors showed for a test case of 18 complexes that this approach performed better than the docking algorithms FlexX [133], Glide [158], and GOLD [151] (for which only 12 of the 18 cases were tested and reported in the literature) with respect to ligand rmsd of the top ranked solutions.

3.3.4. Using Rotamer Libraries

One way to account for sidechain mobility without sampling the full conformational space of the sidechain is to use rotamer libraries that contain information about favorable discrete conformations [159-161]. These libraries are derived from crystallographic data. The approach proposed by Leach [161] along these lines was among the first to predict the conformations of biomolecular complexes taking into account receptor plasticity. The main problem of this early approach was that pre-determined conformations of both the ligand and the sidechains were used.

In another approach Frimurer *et al*. [159] used a rotamer library to generate an ensemble of tyrosine phosphatase B1 structures by using three selected active site residues. The conformers were then used for flexible-ligand docking with FlexX [133]. In a similar approach Kallblad and Dean [160] used a rotamer library to generate numerous conformers from a single MMP-1 structure. Subsequently, a core ensemble of conformers was extracted by statistical methods. A synthetic inhibitor was then rigidly docked to this core ensemble using GOLD. The result was compared to docking to the experimental conformation and also to an energy minimized version. The rmsd of the ligand to the native structure was 7.63 Å for the docking to the start conformation and 5.97 Å for the minimized structure. With the structures of the core ensemble, an rmsd of 1.25 Å was obtained instead.

3.4. Considering Backbone Mobility

The consideration of receptor backbone degrees of freedom results in a very large search space even if only φ and ψ angles are considered and no changes in bond lengths and angles. Therefore a number of approaches were published in which the search space has been further narrowed down. In addition, when large scale macromolecular motions are modeled the issue of scoring the generated receptor conformation becomes more and more important: disfavorable receptor conformations need to be recognized as such, as they will contribute adversely to the overall binding affinity.

3.4.1. Parallel Docking into Multiple Conformations

The docking algorithm IFREDA (ICM-flexible receptor docking algorithm) by Cavasotto and Abagyan [162] uses a combination of different conformers to account for receptor plasticity. The IFREDA algorithm consists of three main steps. In the first step, an ensemble of receptor conformations is generated by placing the ligand at arbitrary

positions and orientations into the binding pocket of the receptor followed by an energy minimization of these randomly produced conformations. In this step side-chain and essential backbone movements are taken into account. Thus, an "induced-fit" of the binding pocket is generated, making this procedure especially appropriate for cases where only one bound crystal structure or an apo form is available. However, it is required that a number of binders are already known. In the second step, the conformations generated are used for a fast rigid receptor, flexible ligand docking based on energy grids. This is finally followed by merging the screening scores and keeping the best rank for each compound. In the end there is only one score for every compound although the compounds were docked against several conformations. IFREDA was tested for seven protein kinase complexes from four different subfamilies for which it was able to generate correct conformations (rmsd < 2 Å) for 70% of the ligands in a cross-docking experiment. The dataset also included structures with a backbone motion of up to 2 Å rmsd. The method was also tested in a small virtual screening scenario where the known active ligands were mixed into a library of 1000 randomly chosen compounds. This resulted in about 80% of the native binders sorted in the top 1.5% of the screened database.

3.4.2. Structurally Combining Multiple Conformations

The underlying idea of the FlexE approach [163] is to use experimentally determined or generated conformations as a "basis set" of receptor plasticity. By structural combination, novel adopted conformations of the receptor are generated. The search for receptor conformations is based on a so-called united protein description. In this description similar parts of superposed structures from a given structure ensemble are merged whereas dissimilar parts of the protein are treated as alternatives during docking. The united protein description prunes the search space but at the same time allows combining the alternatives to new structures that were not contained in the ensemble. That way FlexE can account for mobility in terms of loop movements and sidechain movements. The authors also propose to combine FlexE with a rotamer library approach to achieve even better sidechain predictions.

The downside of the purely geometry-directed receptor conformer generation from the united protein representation is that the conformational energy of the solutions is not considered. In virtual screening runs, this can easily lead to the generation of rather large binding pockets that are able to accommodate particularly large ligands. Likewise, FlexE's capability to consider large main-chain variations, such as domain movements, is limited.

3.4.3. Modeling Protein Motions in Reduced Coordinates

Harmonic modes obtained from normal mode analysis, modes resulting from ENM, or principal components derived from MD trajectories all describe receptor movements in terms of collective vibrational motions. Modes that correspond to low frequencies thereby characterize movements with large amplitudes, such as loop or even domain movements, which are usually considered to be biologically relevant. Using modes as independent variables to describe receptor motions considerably reduces the complexity of the conformational search space.

One of the earliest approaches along these lines was presented by Zacharias and Sklenar [164]. In docking ligands to DNA, harmonic modes derived from the energy

minimized free DNA were combined to model receptor plasticity. Between 5 to 40 of the first eigenvectors (sorted by their eigenfrequencies) were sufficient to successfully deform the DNA from the unbound to the ligand-bound structure. The authors point out that the sterical fit between ligand and receptor improved significantly in that case. The relaxation protocol resulted in structures with a significantly reduced rmsd relative to the bound receptor structure (up to 0.5 Å to the ligand-bound form).

As a drawback, only those motions can be modeled that are represented by the subset of harmonic modes applied. High frequency movements such as rotations of binding pocket sidechains are not considered, which can have a big impact on the success of docking. Similarly, the assumption of harmonic modes as independent variables may not be valid for every receptor movement. Finally, the use of modes may not be successful when it comes to modeling curvilinear receptor motions. In turn, it is important to point out that with the use of harmonic modes it is possible to calculate approximate energies for the protein deformation during the docking run. Thus, the energy contribution due to the protein conformational change can be considered in the binding free energy estimation. This is an advantage over approaches that are purely geometry-based such as FlexE.

In another approach Cavasotto *et al.* [165] used relevant low-frequency modes to describe loop mobility in cAMP-dependent protein kinase (see chapter 2.3). Thus generated representative receptor backbone conformations were afterwards optimized by placing a non-native binder into the binding pockets. The optimized conformations are then used for virtual screening.

3.4.4. Pose Optimization of Docking Solutions with MD Simulations

A number of docking approaches have been published recently that are based on a two-step protocol. Different levels of detail are pursued in both steps. In the first step, methods with low accuracy are used, which therefore are very fast and can be applied to large compound databases. The second step then consists of a more accurate and time-consuming method. The overall goal here is to screen many compounds without compromising the accuracy. This is in most of the cases accomplished by docking in combination with MD simulations. The docking is used for two purposes: First, as a filter to identify compounds that can actually bind to the considered target and, second, to provide decoy complex structures as starting points for MD because the latter method is not able to sample the right orientation of a ligand in a feasible time. In the subsequent MD step, induced fit effects of the protein will be considered.

Although the primary objective for the trajectory information is to obtain better insight into the binding process, the knowledge can also be used in later steps of ligand optimization with respect to ADMET [166] properties. During the MD step, it is also possible to incorporate explicit water molecules into the complex structure prediction, which is not provided by most of the approaches solely relying on docking.

An intriguing example of such an application is given by the insight gained into the preferential binding of cyclin-dependent kinase inhibitors, which helped to understand at a molecular level the differential inhibition observed in experimental data. Here Park *et al.* [167] studied the inhibition of CDK2 and CDK4 by three different selective inhibitors. The inhibitors reduced the mobility of a disordered loop of CDK4 but did not seriously affect CDK2. It was also discovered that the tighter binding of the inhibitors to

CDK4 was an effect of a smaller number of water molecules moving into the binding site upon ligand binding. Such insights would not have been possible without the MD step.

Cavalli *et al.* [168] pointed out that the combination of docking with subsequent MD simulations allows identifying those starting structures as being in good agreement with the experimental binding mode that lead to stable trajectories. Thus, in a study with human acetylcholinesterase they could not only account for induced fit effects but were also able to discriminate among conformations of different stability. Similar conclusions were also reported by Zoete *et al.* [169] for the binding of D-glucose to insulin. The authors observed binding at different surface sites of insulin for D-glucose. With the help of MD simulations, the validity of binding sites was then judged based on the dynamical stability of the bound glucose. Finally, the MD simulations revealed a small number and an unstable character of hydrogen bonds between glucose and insulin, which reflects the low binding free energy known from experiment.

For the approaches presented it has to be considered that even short MD runs require considerable computational time and, hence, are limited to only selected cases. Second, the accuracy of the docking tool used in the first step is critical for the stability of the MD trajectories and, consequently, for the success of the complex prediction. Finally, force field parameters for ligands are usually not provided with biomolecular force fields, and parameterization of the missing ones may be cumbersome if it comes to database screening. Recent developments such as the general Amber force field (GAFF) [170], which is designed to be compatible with existing Amber force fields and has parameters for most organic and pharmaceutical small-molecules, together with a set of auxiliary programs [171] to (semi-)automatically setup small-molecule parameters are expected to alleviate this problem.

4. SUMMARY AND OUTLOOK

Computational approaches in SBDD that address issues of protein flexibility and mobility have been reviewed, with a focus on methods that account for receptor plasticity in docking. These methods are grounded on the "induced fit" and "conformational selection" models that superseed the long-standing "lock and key" paradigm. Particularly, the "conformational selection" model provides an encouraging viewpoint, because it implies that conformational states adopted by the bound protein may be found already by investigating conformational fluctuations of the unbound structure. With respect to docking considering receptor plasticity this then leads to the suggestion to pursue a two step approach: initially, one detects what can move and how; subsequently, this knowledge is included into a docking strategy.

Methods of the first step can be grouped into three classes: force field-based, graph theoretical and geometry-based, or harmonic analysis-based. Common to most methods is that they describe protein motions at a certain scale: MD simulations usually provide a very detailed picture of the dynamics, but fail to account for large-scale motions; in turn, harmonic analysis-based methods allow modeling biologically relevant collective motions, but are less suitable for, e.g., sidechain motions. Yet, notable exceptions have been reported that are able to describe motions on multiple scales. These approaches usually benefit from combining two or more of the established "single scale" techniques. We anticipate that developments along these lines will be very fruitful in the future, although coupling and modeling movements at different scales will remain challenging.

Regarding the incorporation of receptor plasticity into docking, three main routes of algorithmic development have been identified: I) Implicit consideration of plasticity; II) modeling of sidechain motions; III) accounting for large scale conformational changes including backbone motions. Only in the latter class, receptor motions at different scales can be accounted for, although at the expense of utilizing a predefined ensemble of receptor structures as a "basis set". Clearly, there is still room for further improvement.

Whereas the impact of receptor flexibility and mobility on structure is by now well received, energetic and entropic consequences mentioned in the Introduction section are – with a few exceptions – largely neglected in SBDD. Progress in this area is highly necessary. However, this requires a deepening of our basic understanding first, as Nature seems to be particularly capable to provide surprises in this area [172].

Not least because of progress in various experimental areas, we are confident that we are awaiting exciting times for understanding and modeling receptor flexibility and mobility. After all, there is a beginning!

ACKNOWLEDGEMENTS

This work was supported by startup funds from J.W. Goethe-University, Frankfurt and funds from Merck KGaA, Darmstadt, and Novartis Pharma AG, Basel.

ABBREVIATIONS

rmsd = Root mean-square deviation

SBDD = Structure-based drug design

REFERENCES

[1] Whitesides, G. M.; Krishnamurthy, V. M. *Q. Rev. Biophys.*, **2005**, *38*, 385.
[2] Gohlke, H.; Klebe, G. *Angew. Chem.*, **2002**, *41*, 2644.
[3] Gohlke, H.; Thorpe, M. F. *Biophys. J.*, **2006**, *91*, 2115.
[4] Wlodawer, A.; Vondrasek, J. *Annu. Rev. Biophys. Biomol. Struct.*, **1998**, *27*, 249.
[5] Wilson, D. K.; Tarle, I.; Petrash, J. M.; Quiocho, F. A. *Proc. Natl. Acad. Sci. USA*, **1993**, *90*, 9847.
[6] Lin, J. H.; Perryman, A. L.; Schames, J. R.; McCammon, J. A. *J. Am. Chem. Soc.*, **2002**, *124*, 5632.
[7] Vieira, E.; Binggeli, A.; Breu, V.; Bur, D.; Fischli, W.; Guller, R.; Hirth, G.; Marki, H. P.; Muller, M.; Oefner, C.; Scalone, M.; Stadler, H.; Wilhelm, M.; Wostl, W. *Bioorg. Med. Chem. Lett.*, **1999**, *9*, 1397.
[8] Allemann, R. K.; Egli, M. *Chem. Biol.*, **1997**, *4*, 643.
[9] Puglisi, J. D.; Chen, L.; Frankel, A. D.; Williamson, J. R. *Proc. Natl. Acad. Sci. USA*, **1993**, *90*, 3680.
[10] Puglisi, J. D.; Tan, R.; Calnan, B. J.; Frankel, A. D.; Williamson, J. R. *Science*, **1992**, *257*, 76.
[11] Fourmy, D.; Yoshizawa, S.; Puglisi, J. D. *J. Mol. Biol.*, **1998**, *277*, 333.
[12] Davis, A. M.; Teague, S. J. *Angew. Chem. Int. Ed. Engl.*, **1999**, *38*, 736.
[13] Wright, P. E.; Dyson, H. J. *J. Mol. Biol.*, **1999**, *293*, 321.
[14] Levy, Y.; Cho, S. S.; Onuchic, J. N.; Wolynes, P. G. *J. Mol. Biol.*, **2005**, *346*, 1121.
[15] Liu, J.; Tan, H.; Rost, B. *J. Mol. Biol.*, **2002**, *322*, 53.
[16] Fischer, E. *Ber. Dtsch. Chem. Ges.*, **1894**, *27*, 2985.
[17] Koshland, D. E. *Science*, **1967**, *156*, 540.
[18] Tsai, C.-J.; Ma, B.; Nussinov, R. *Proc. Natl. Acad. Sci. USA*, **1999**, *96*, 9970.
[19] Tsai, C.-J.; Kumar, S.; Ma, B.; Nussinov, R. *Protein Sci.*, **1999**, *8*, 1181.
[20] Volkman, B. F.; Lipson, D.; Wemmer, D. E.; Kern, D. *Science*, **2001**, *291*, 2429.
[21] Ma, B.; Shatsky, M.; Wolfson, H. J.; Nussinov, R. *Protein Sci.*, **2002**, *11*, 184.
[22] Lazaridis, T.; Masunov, A.; Gandolfo, F. *Proteins*, **2002**, *47*, 194.

[23] Wei, B. Q.; Weaver, L. H.; Ferrari, A. M.; Matthews, B. W.; Shoichet, B. K. *J. Mol. Biol.*, **2004**, *337*, 1161.
[24] Schymkowitz, J. W.; Rousseau, F.; Wilkinson, H. R.; Friedler, A.; Itzhaki, L. S. *Nat. Struct. Biol.*, **2001**, *8*, 888.
[25] Freire, E. *Proc. Natl. Acad. Sci. USA*, **2000**, *97*, 11680.
[26] Pan, H.; Lee, J. C.; Hilser, V. J. *Proc. Natl. Acad. Sci. USA*, **2000**, *97*, 12020.
[27] Yu, L.; Zhu, C. X.; Tse-Dinh, Y. C.; Fesik, S. W. *Biochemistry*, **1996**, *35*, 9661.
[28] Olejniczak, E. T.; Zhou, M. M.; Fesik, S. W. *Biochemistry*, **1997**, *36*, 4118.
[29] Zidek, L.; Novotny, M. V.; Stone, M. J. *Nat. Struct. Biol.*, **1999**, *6*, 1118.
[30] Loh, A. P.; Pawley, N.; Nicholson, L. K.; Oswald, R. E. *Biochemistry*, **2001**, *40*, 4590.
[31] Arumugam, S.; Gao, G.; Patton, B. L.; Semenchenko, V.; Brew, K.; Van Doren, S. R. *J. Mol. Biol.*, **2003**, *327*, 719.
[32] Forman-Kay, J. D. *Nat. Struct. Biol.*, **1999**, *6*, 1086.
[33] Kay, L. E.; Muhandiram, D. R.; Wolf, G.; Shoelson, S. E.; Forman-Kay, J. D. *Nat. Struct. Biol.*, **1998**, *5*, 156.
[34] Kay, L. E.; Muhandiram, D. R.; Farrow, N. A.; Aubin, Y.; Forman-Kay, J. D. *Biochemistry*, **1996**, *35*, 361.
[35] Steinberg, I. Z.; Scheraga, H. A. *J. Biol. Chem.*, **1963**, *238*, 172.
[36] Stivers, J. T.; Abeygunawardana, C.; Mildvan, A. S.; Whitman, C. P. *Biochemistry*, **1996**, *35*, 16036.
[37] Hodsdon, M. E.; Cistola, D. P. *Biochemistry*, **1997**, *36*, 2278.
[38] Lee, A. L.; Kinnear, S. A.; Wand, A. J. *Nat. Struct. Biol.*, **2000**, *7*, 72.
[39] Canino, L. S.; Shen, T.; McCammon, J. A. *J. Chem. Phys.*, **2002**, *117*, 9927.
[40] Gohlke, H.; Case, D. A. *J. Comput. Chem.*, **2004**, *25*, 238.
[41] van Gunsteren, W. F.; Bakowies, D.; Baron, R.; Chandrasekhar, I.; Christen, M.; Daura, X.; Gee, P.; Geerke, D. P.; Glattli, A.; Hunenberger, P. H.; Kastenholz, M. A.; Oostenbrink, C.; Schenk, M.; Trzesniak, D.; van der Vegt, N. F.; Yu, H. B. *Angew. Chem. Int. Ed. Engl.*, **2006**, *45*, 4064.
[42] Pickett, S. D.; Sternberg, M. J. *J. Mol. Biol.*, **1993**, *231*, 825.
[43] Case, D. A. *Curr. Opin. Struct. Biol.*, **1994**, *4*, 285.
[44] Levy, R. M.; Karplus, M.; Kushick, J.; Perahia, D. *Macromolecules*, **1984**, *1984*, 1370.
[45] Schlitter, J. *Chem. Phys. Lett.*, **1993**, *215*, 617.
[46] Christopoulos, A. *Nat. Rev. Drug Discov.*, **2002**, *1*, 198.
[47] Monod, J.; Wyman, J.; Changeux, J.-P. *J. Mol. Biol.*, **1965**, *12*, 88.
[48] Changeux, J.-P.; Edelstein, S. J. *Science*, **2005**, *308*, 1424.
[49] Kern, D.; Zuiderweg, E. R. *Curr. Opin. Struct. Biol.*, **2003**, *13*, 748.
[50] Rousseau, F.; Schymkowitz, J. *Curr. Opin. Struct. Biol.*, **2005**, *15*, 23.
[51] Gunasekaran, K.; Ma, B.; Nussinov, R. *Proteins*, **2004**, *57*, 433.
[52] Hardy, J. A.; Wells, J. A. *Curr. Opin. Struct. Biol.*, **2004**, *14*, 706.
[53] Barnard, D.; Diaz, B.; Hettich, L.; Chuang, E.; Zhang, X. F.; Avruch, J.; Marshall, M. *Oncogene*, **1995**, *10*, 1283.
[54] Najmanovich, R.; Kuttner, J.; Sobolev, V.; Edelman, M. *Proteins*, **2000**, *39*, 261.
[55] Zavodszky, M. I.; Kuhn, L. A. *Protein Sci.*, **2005**, *14*, 1104.
[56] Hansson, T.; Oostenbrink, C.; van Gunsteren, W. *Curr. Opin. Struct. Biol.*, **2002**, *12*, 190.
[57] Karplus, M.; McCammon, J. A. *Nat. Struct. Biol.*, **2002**, *9*, 646.
[58] Adcock, S. A.; McCammon, J. A. *Chem. Rev.*, **2006**, *106*, 1589.
[59] Moraitakis, G.; Purkiss, A.; Goodfellow, J. *Rep. Progr. Phys.*, **2003**, *66*, 383.
[60] Schames, J. R.; Henchman, R. H.; Siegel, J. S.; Sotriffer, C. A.; Ni, H.; McCammon, J. A. *J. Med. Chem.*, **2004**, *47*, 1879.
[61] Hazuda, D. J.; Anthony, N. J.; Gomez, R. P.; Jolly, S. M.; Wai, J. S.; Zhuang, L.; Fisher, T. E.; Embrey, M.; Guare, J. P., Jr.; Egbertson, M. S.; Vacca, J. P.; Huff, J. R.; Felock, P. J.; Witmer, M. V.; Stillmock, K. A.; Danovich, R.; Grobler, J.; Miller, M. D.; Espeseth, A. S.; Jin, L.; Chen, I. W.; Lin, J. H.; Kassahun, K.; Ellis, J. D.; Wong, B. K.; Xu, W.; Pearson, P. G.; Schleif, W. A.; Cortese, R.; Emini, E.; Summa, V.; Holloway, M. K.; Young, S. D. *Proc. Natl. Acad. Sci. USA*, **2004**, *101*, 11233.
[62] Carlson, H. A.; Masukawa, K. M.; Rubins, K.; Bushman, F. D.; Jorgensen, W. L.; Lins, R. D.; Briggs, J. M.; McCammon, J. A. *J. Med. Chem.*, **2000**, *43*, 2100.
[63] Hornak, V.; Okur, A.; Rizzo, R. C.; Simmerling, C. *Proc. Natl. Acad. Sci. USA*, **2006**, *103*, 915.

[64] Sotriffer, C.; Krämer, O.; Klebe, G. *Proteins*, **2004**, *56*, 52.
[65] Smith, G.; Sternberg, M.; Bates, P. *J. Mol. Biol.*, **2005**, *347*, 1077.
[66] Perryman, A.; Lin, J.-H.; McCammon, A. *Protein Sci.*, **2004**, *13*, 1108.
[67] Perryman, A.; Lin, J.-H.; McCammon, A. *Chem. Biol. Drug Des.*, **2006**, *67*, 336.
[68] Perryman, A.; Lin, J.-H.; McCammon, A. *Biopolymers*, **2006**, *82*, 272.
[69] Ode, H.; Neya, S.; Hata, M.; Sugiura, W.; Hoshino, T. *J. Am. Chem. Soc.*, **2006**, *128*, 7887.
[70] Young, M. A.; Gonfloni, S.; Superti-Furga, G.; Roux, B.; Kuriyan, J. *Cell*, **2001**, *105*, 115.
[71] Jacobs, D.; Rader, A. J.; Kuhn, L.; Thorpe, M. F. *Proteins*, **2001**, *44*, 150.
[72] Jacobs, D. J.; Kuhn, L. A.; Thorpe, M. F. Rigidity theory and applications. New York: Kluwer Academic/Plenum, **1999**, 7, 961.
[73] Jacobs, D. J.; Rader, A. J.; Kuhn, L. A.; Thorpe, M. F. *Proteins*, **2001**, *44*, 150.
[74] Gohlke, H.; Kuhn, L. A.; Case, D. A. *Proteins*, **2004**, *56*, 322.
[75] Fuxreiter, M.; Magyar, C.; Juhasz, T.; Szeltner, Z.; Polgar, L.; Simon, I. *Proteins*, **2005**, *60*, 504.
[76] Gohlke, H.; Thorpe, M. *Biophys. J.*, **2006**.
[77] Lei, M.; Zavodszky, M.; Kuhn, L.; Thorpe, M. F. *J. Comput. Chem.*, **2004**, *25*, 1133.
[78] Thorpe, M. F.; Lei, M.; Rader, A. J.; Jacobs, D. J.; Kuhn, L. A. *J. Mol. Graph. Model.*, **2001**, *19*, 60.
[79] Wells, S.; Menor, S.; Hespenheide, B.; Thorpe, M. F. *Phys. Biol.*, **2005**, *2*, S127-36.
[80] Ahmed, A.; Gohlke, H. *Proteins*, **2006**, *63*, 1038.
[81] Zavodszky, M.; Lei, M.; Thorpe, M. F.; Day, A.; Kuhn, L. *Proteins*, **2004**, *57*, 243.
[82] Jolley, C. C.; Wells, S. A.; Hespenheide, B. M.; Thorpe, M. F.; Fromme, P. *J. Am. Chem. Soc.*, **2006**, *128*, 8803.
[83] de Groot, B. L.; Vriend, G.; Berendsen, H. J. *J. Mol. Biol.*, **1999**, *286*, 1241.
[84] de Groot, B. L.; van Aalten, D. M. F.; Scheek, R. M.; Amadei, A.; Vriend, G.; Berendsen, H. J. C. *Proteins*, **1998**, *29*, 240.
[85] Jedrzejas, M.; Mello, L.; de Groot, B.; Li, S. *J. Biol. Chem.*, **2002**, *277*, 28287.
[86] Mustard, D.; Ritchie, D. *Proteins*, **2005**, *60*, 269.
[87] Kavraki, L. E.; Svestka, P.; Latombe, J. C.; Overmars, M. H. *IEEE Robot. Automat.*, **1996**, *12*, 566.
[88] Song, G.; Thomas, S.; Dill, K. A.; Scholtz, J. M.; Amato, N. M. *Pac. Symp. Biocomput.*, **2003**, 240.
[89] Apaydin, M.; Brutlag, D.; Guesttin, C.; Hsu, D.; Latombe, J. 2002.
[90] Amato, N. M.; Song, G. *J. Comput. Biol.*, **2002**, *9*, 149.
[91] Amato, N. M.; Dill, K. A.; Song, G. *J. Comput. Biol.*, **2003**, *10*, 239.
[92] Tang, X.; Kirkpatrick, B.; Thomas, S.; Song, G.; Amato, N. M. *J. Comput. Biol.*, **2005**, *12*, 862.
[93] Cortes, J.; Simeon, T.; Remaud-Simeon, M.; Tran, V. *J. Comput. Chem.*, **2004**, *25*, 956.
[94] Cortes, J.; Simeon, T.; Ruiz de Angulo, V.; Guieysse, D.; Remaud-Simeon, M.; Tran, V. *Bioinformatics*, **2005**, *21*(Suppl 1), i116-25.
[95] Go, N.; Noguti, T.; Nishikawa, T. *Proc. Natl. Acad. Sci. USA*, **1983**, *80*, 3696.
[96] Brooks, B.; Karplus, M. *Proc. Natl. Acad. Sci. USA*, **1983**, *80*, 6571.
[97] Case, D. *Curr. Opin. Struct. Biol.*, **1994**, *4*, 285.
[98] Tirion, M. *Phys. Rev. Lett.*, **1996**, *77*, 1905.
[99] Bahar, I.; Atilgan, A. R.; Erman, B. *Fold. Des.*, **1997**, *2*, 173.
[100] Haliloglu, T.; Bahar, I.; Erman, B. *Phys. Rev. Lett.*, **1997**, *79*, 3090.
[101] Hinsen, K. *Proteins*, **1999**, *33*, 417.
[102] Hinsen, K.; Thomas, A.; Field, M. J. *Proteins*, **1999**, *34*, 369.
[103] Atilgan, A. R.; Durell, S. R.; Jernigan, R. L.; Demirel, M. C.; Keskin, O.; Bahar, I. *Biophys. J.*, **2001**, *80*, 505.
[104] Tama, F.; Sanejouand, Y. H. *Protein Eng.*, **2001**, *14*, 1.
[105] Tozzini, V. *Curr. Opin. Struct. Biol.*, **2005**, *15*, 144.
[106] Ma, J. *Structure*, **2005**, *13*, 373.
[107] Ma, J. *Curr. Prot. Pept. Sci.*, **2004**, *5*, 119.
[108] Bahar, I.; Rader, A. *Curr. Opin. Struct. Biol.*, **2005**, *15*, 586.
[109] Tama, F.; Gadea, F. X.; Marques, O.; Sanejouand, Y. H. *Proteins*, **2000**, *41*, 1.
[110] Kurkcuoglu, O.; Jernigan, R.; Doruker, P. *Polymer*, **2004**, *45*, 649.
[111] Doruker, P.; Jernigan, R. L.; Bahar, I. *J. Comput. Chem.*, **2002**, *23*, 119.
[112] Li, G.; Cui, Q. *Biophys. J.*, **2002**, *83*, 2457.

[113] Durand, P.; Trinquier, G.; Sanejouand, Y.-H. *Biopolymers*, **2004**, *34*, 759.
[114] Brooks, B.; Karplus, M. *Proc. Natl. Acad. Sci. USA*, **1985**, *82*, 4995.
[115] Bahar, I.; Erman, B.; Jernigan, R.; Atilgan, A.; Covell, D. *J. Mol. Biol.*, **1999**, *285*, 1023.
[116] Temiz, N. A.; Meirovitch, E.; Bahar, I. *Proteins*, **2004**, *57*, 468.
[117] Isin, B.; Doruker, P.; Bahar, I. *Biophys. J.*, **2002**, *82*, 569.
[118] Bahar, I.; Jernigan, R. L. *Biochemistry*, **1999**, *38*, 3478.
[119] Temiz, N. A.; Bahar, I. *Proteins*, **2002**, *49*, 61.
[120] Kim, M. K.; Jernigan, R. L.; Chirikjian, G. S. *Biophys. J.*, **2002**, *83*, 1620.
[121] Kim, M. K.; Chirikjian, G. S.; Jernigan, R. L. *J. Mol. Graph. Model.*, **2002**, *21*, 151.
[122] Miyashita, O.; Onuchic, J. N.; Wolynes, P. G. *Proc. Natl. Acad. Sci. USA*, **2003**, *100*, 12570.
[123] Cavasotto, C. N.; Kovacs, J. A.; Abagyan, R. A. *J. Am. Chem. Soc.*, **2005**, *127*, 9632.
[124] Ertekin, A.; Nussinov, R.; Haliloglu, T. *Protein Sci.*, **2006**, *15*, 2265.
[125] Yang, L. W.; Bahar, I. *Structure*, **2005**, *13*, 893.
[126] Haliloglu, T.; Keskin, O.; Ma, B.; Nussinov, R. *Biophys. J.*, **2005**, *88*, 1552.
[127] Demirel, M. C.; Atilgan, A. R.; Jernigan, R. L.; Erman, B.; Bahar, I. *Protein Sci.*, **1998**, *7*, 2522.
[128] He, J.; Zhang, Z.; Shi, Y.; Liu, H. *J. Chem. Phys.*, **2003**, *119*, 4005.
[129] Tatsumi, R.; Fukunishi, Y.; Nakamura, H. *J. Comput. Chem.*, **2004**, *25*, 1995.
[130] Zhang, Z.; Shi, Y.; Liu, H. *Biophys. J.*, **2003**, *84*, 3583.
[131] Halperin, I.; Ma, B.; Wolfson, H.; Nussinov, R. *Proteins*, **2002**, *47*, 409.
[132] Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. *J. Mol. Biol.*, **1982**, *161*, 269.
[133] Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. *J. Mol. Biol.*, **1996**, *261*, 470.
[134] Morris, G. M.; Goosell, D. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. *J. Comput. Chem.*, **1998**, *19*, 1693.
[135] Gane, P. G.; Dean, P. M. *Curr. Opin. Struct. Biol.*, **2000**, *10*, 401.
[136] Muegge, I.; Rarey, M. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; Wiley-VCH: New York, 2001; Vol. 17, pp. 1-60.
[137] Shoichet, B. K.; McGovern, S. L.; Wei, B.; Irwin, J. J. *Curr. Opin. Chem. Biol.*, **2002**, *6*, 439.
[138] Jiang, F.; Kim, S. H. *J. Mol. Biol.*, **1991**, *219*, 79.
[139] Schnecke, V.; Swanson, C.; Getzoff, E.; Tainer, J.; Kuhn, L. A. *Proteins*, **1998**, *33*, 74.
[140] Mizutani, M. Y.; Takamatsu, Y.; Ichinose, T.; Nakamura, K.; Itai, A. *Proteins*, **2006**, *63*, 878.
[141] Ferrara, P.; Gohlke, H.; Price, D. J.; Klebe, G.; Brooks, C. L. *J. Med. Chem.*, **2004**, *47*, 3032.
[142] Gohlke, H.; Klebe, G. *Curr. Opin. Struct. Biol.*, **2001**, *11*, 231.
[143] Gohlke, H.; Hendlich, M.; Klebe, G. *J. Mol. Biol.*, **2000**, *295*, 337.
[144] Yang, C. Y.; Wang, R.; Wang, S. *J. Med. Chem.*, **2006**, *49*, 5903.
[145] Carlson, H. A. *Curr. Opin. Chem. Biol.*, **2002**, *6*, 447.
[146] Knegtel, R. M.; Kuntz, I. D.; Oshiro, C. M. *J. Mol. Biol.*, **1997**, *266*, 424.
[147] Broughton, H. B. *J. Mol. Graph. Model.*, **2000**, *18*, 247.
[148] Österberg, F.; Morris, G. M.; Sanner, M. F.; Olson, A. J.; Goodsell, D. S. *Proteins*, **2002**, *46*, 34.
[149] Sudbeck, E. A.; Mao, C.; Vig, R.; Venkatachalam, T. K.; Tuel-Ahlgren, L.; Uckun, F. M. *Antimicrob. Agents Chemother.*, **1998**, *42*, 3225.
[150] Jones, G.; Willett, P.; Glen, R. C. *J. Mol. Biol.*, **1995**, *245*, 43.
[151] Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. *J. Mol. Biol.*, **1997**, *267*, 727.
[152] Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. *Proteins*, **2006**, *65*, 15.
[153] Kellenberger, E.; Rodrigo, J.; Muller, P.; Rognan, D. *Proteins*, **2004**, *57*, 225.
[154] Goodsell, D. S.; Olson, A. J. *Proteins*, **1990**, *8*, 195.
[155] Goodsell, D. S.; Morris, G. M.; Olson, A. J. *J. Mol. Recognit.*, **1996**, *9*, 1.
[156] Schnecke, V.; Kuhn, L. A. *Persp. Drug Discov. Des.*, **2000**, *20*, 171.
[157] Yamada, M.; Itai, A. *Chem. Pharm. Bull.*, **1993**, *41*, 1200.
[158] Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. *J. Med. Chem.*, **2004**, *47*, 1739.
[159] Frimurer, T. M.; Peters, G. H.; Iversen, L. F.; Andersen, H. S.; Moller, N. P.; Olsen, O. H. *Biophys. J.*, **2003**, *84*, 2273.
[160] Kallblad, P.; Dean, P. M. *J. Mol. Biol.*, **2003**, *326*, 1651.

[161] Leach, A. *J. Mol. Biol.*, **1994**, *235*, 345.
[162] Cavasotto, C. N.; Abagyan, R. A. *J. Mol. Biol.*, **2004**, *337*, 209.
[163] Claußen, H.; Buning, C.; Rarey, M.; Lengauer, T. *J. Mol. Biol*., **2001**, *308*, 377.
[164] Zacharias, M.; Sklenar, H. *J. Comput. Chem*., **1999**, *20*, 287.
[165] Cavasotto, C. N.; Kovacs, J. A.; Abagyan, R. A. *J. Am. Chem. Soc*., **2005**, *127*, 9632.
[166] Clark, D. E.; Grootenhuis, P. D. *Curr. Opin. Drug Discov. Devel*., **2002**, *5*, 382.
[167] Park, H.; Yeom, M. S.; Lee, S. *Chembiochem*, **2004**, *5*, 1662.
[168] Cavalli, A. *J. Med. Chem.*, **2004**, *47*, 3991.
[169] Zoete, V.; Meuwly, M.; Karplus, M. *Proteins*, **2004**, *55*, 568.
[170] Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.*, **2004**, *25*, 1157.
[171] Case, D. A.; Cheatham, T. E., 3rd; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *J. Comput. Chem.*, **2005**, *26*, 1668.
[172] Krishnamurthy, V. M.; Bohall, B. R.; Semetey, V.; Whitesides, G. M. *J. Am. Chem. Soc*., **2006**, *128*, 5802.