

Conformational Flexibility Models for the Receptor in Structure Based Drug Design

M. L. Teodoro

mteodoro@rice.edu

tel: 713-348-3051

Department of Biochemistry and Cell Biology and Department of Computer Science

Rice University

6100 Main St, MS 140

Houston, TX 77005

USA

L. E. Kavraki

kavraki@rice.edu

tel: 713-348-5737

fax: 713-348-5930

Department of Computer Science and Department of Bioengineering,

Rice University

6100 Main St, MS 132

P.O. Box 1892

Houston, TX 77251-1892

USA

Abstract

The problem of incorporating protein flexibility in the routine *in silico* screening of large databases of small chemical compounds is still an unsolved and hard problem. The main reason behind this difficulty is the exponential explosion in computational complexity due to the inclusion of the large number of degrees of freedom that represent the receptor flexibility. In order to address this limitation several flexibility models for the receptor have been developed which try to limit the number of additional model parameters. These models can be roughly grouped divided into five different categories. These are the use of soft receptors which relax energetic penalties due to steric clashes, the selection of a few critical degrees of freedom in the receptor binding site, the use of multiple receptor structures either individually or by combining them using an averaging scheme, the use of modified molecular simulation methods, or the use of collective degrees of freedom as a new basis of representation for protein flexibility. All these flexible receptor models strive to balance an improvement in the accuracy of the docking predictions with an increase in computational cost. In addition, other challenges such as the development of accurate solvation models and scoring functions make the receptor flexibility problem even harder.

Introduction

The ability to predict the bound conformations and interaction energy between small organic molecules and biomacromolecules such as proteins and DNA is of extreme physiological and pharmacological importance. Hence, there has been a considerable effort from both academia and industry to develop computational methods that can be used to determine the affinity with which a ligand will bind a target receptor. These methods usually include docking algorithms that compute the three dimensional structure of the complex as would be determined experimentally using X-ray crystallography or Nuclear Magnetic Resonance (NMR) methods. This procedure entails not only determining the identity and three dimensional structure of the bound ligand, but also determine how the binding process affects the conformation of the receptor. In this paper we review the different receptor flexibility representations that have been explored to study receptor conformational changes in the context of structure based drug design.

A central paradigm which was used in the development of the first docking programs was the lock-and-key model first described by Fischer [1]. In this model the three dimensional structure of the ligand and the receptor complement each other in the same way that a lock complements a key. According to this model, one could find a good drug candidate by searching a database of small molecules for one that complemented the three dimensional structure of a given receptor. This rigid matching was supported by several studies of complexes of proteolytic enzymes with small protein inhibitors [2-4] and from the first example of an antibody-protein complex [5]. However, subsequent work has confirmed that the lock-and-key model is not the most correct description for ligand binding. A more accurate view of this process was first presented by Koshland [6]

in the induced fit model. In this model the three dimensional structure of the ligand and the receptor adapt to each other during the binding process. It is important to note that not only the structure of the ligand is changed but also the structure of the receptor changes during the binding process. This occurs because introduction of a ligand changes the chemical and structural environment of the receptor. As such, the unbound protein conformational substates, corresponding to the low energy regions of the protein energy landscape, are likely to change due to the introduction of the ligand in the environment of the receptor. The induced fit model is supported by multiple observations in different proteins such as streptavidin [7], HIV-1 protease [8], DHFR [9], aldose reductase [10] and many others. The qualitative and quantitative effects of ligand-induced changes in proteins have been described previously [11-15] and explain the ability of a protein to bind multiple drugs with considerably different three dimensional shapes [8, 16].

A more modern, but not contradictory, model for protein/ligand binding is to think of the binding process as a selection of a particular receptor conformation from an ensemble of metastable states [17-20]. The protein exists as a family of similar conformations in a hierarchical energy landscape [21]. Successful binding shifts the dynamic population equilibrium in favor of the bound receptor conformation. This model of ligand binding suggests that for the design of novel inhibitors we may need to explore receptor conformations located outside the narrow scope of the conformational ensemble presently determined using experimental methods. This is important for drug design because it clearly illustrates the need to consider protein flexibility and the existence of multiple receptor conformations. It also provides a justification for higher affinity inhibitors that do not mimic substrates at their transition state. Additionally, if a protein

exists in a population of states as discussed in [18, 22] then one could either design a moderate affinity ligand for a highly populated conformer (lower energy) or a high affinity ligand for a less populated conformer (higher energy).

Although it has been clearly established that a protein is able to undergo conformational changes during the binding process, most docking studies consider the protein as a rigid structure. The reason for this crude approximation is the extraordinary increase in computational complexity that is required to include the degrees of freedom of a protein in a modeling study. Pioneer efforts in the docking area [23, 24] were limited not only in methodology but mainly in computational capability. In the 1980s Kuntz and coworkers developed the program DOCK [25] which made structure-based drug design a staple of current pharmaceutical research methods. Currently, researchers have the ability to use not only improved versions of the original DOCK program but also other docking software such as FlexX [26] or Autodock [27], among many others, to computationally predict the spatial conformation and affinity of bound complexes between a flexible ligand and a rigid receptor. These programs use different search methods and scoring functions. A review of these is beyond the scope of this paper. For recent reviews in general structure based drug design, docking methods and scoring functions see [28-32].

The three dimensional conformation of a molecule can be represented by the values corresponding to its degrees of freedom. These are usually the Cartesian coordinates of its individual atoms or alternatively the values for its internal degrees of freedom. These are bond lengths, bond angles and dihedral angles (i.e., torsions around single bonds). A common approximation when modeling organic molecules is to consider that bond lengths and bond angles are constant and only dihedral angles are free to

change. Even using this approximation a protein can have thousands of degrees of freedom whereas a small organic molecule can be usually modeled using only five to twenty degrees of freedom. In the last decade, with the advent of improved computational capabilities, researchers have been trying to solve the high dimensional problem of modeling protein flexibility in docking applications. The effect of protein flexibility on structure based drug design has been reviewed by Carlson *et al.* [22, 33, 34].

There is currently no computationally efficient docking method that is able to screen a large database of potential ligands against a target receptor by considering the full flexibility of both ligand and receptor. In order for this process to become efficient it is necessary to find a representation for protein flexibility which avoids the direct search of a solution space comprised of thousands of degrees of freedom. Here we review the different representations that have been used to incorporate protein flexibility in the modeling of protein/ligand interactions. A common theme behind all these approaches is that the accuracy of the results is usually directly proportional to the computational complexity of the representation. We tried to group the different types of flexibility representations models into different categories that illustrate some of the key ideas that have been presented in the literature in recent years. However it is important to note that the boundaries between these categories are not rigid and in fact several of the publications referenced below could easily fall in more than one category.

Flexibility Representations

Soft Receptors

Perhaps the simplest solution to represent some degree of receptor flexibility in docking applications is the use of soft receptors. Soft receptors can be easily generated by relaxing the high energy penalty that the system incurs when an atom in the ligand overlaps an atom in the receptor structure. By reducing the van der Waals contributions to the total energy score the receptor is in practice made softer thus allowing, for example, a larger ligand to fit in the same binding site as determined experimentally for a smaller molecule (see Fig. (1)). The rationale behind this approach is that the receptor structure has some inherent flexibility which allows it to adapt to slightly differently shaped ligands by resorting to small variations in the orientation of binding site chains and backbone positions. If the change in the receptor conformation is small enough, it is assumed that the receptor is capable of such a conformational change, given its large number of degrees of freedom, even though the conformational change itself is not modeled explicitly. It is also assumed that the change in protein conformation does not incur a sufficiently high energetic penalty that it offsets the improved interaction energy between the ligand and the receptor. The main advantage of using soft receptors is ease of implementation (docking algorithms stay unchanged) and speed (the cost of evaluating the scoring function is the same as for the rigid case).

The first use of a soft docking approach was by Jiang *et al.* [35]. Their method consisted of constructing a three dimensional cube representation of the molecular volumes and surfaces. These were matched geometrically in a first phase. In a second

phase they were scored in accordance to the favorable energetic interactions in the buried surface areas. Schnecke *et al.* [36] also allowed for some tolerance when calculating van der Waals overlaps between atoms.

Another use of soft docking models is to improve convergence during energy minimization and avoid becoming trapped in local minima. Apostolakis *et al.* [37] developed a docking approach which is based on a combination of minimization with shifted nonbonded interactions and Monte Carlo minimization. In the initial stages of the conformational search the ligand is allowed to overlap with the receptor and nonbonded energy terms are modified to avoid high energy gradients. During the course of the minimization the interactions are then gradually restored to their original values simulating a ligand that is gradually exposed to the field of the receptor. This allows for initial ligand/receptor conformations, which due to steric clashes would result in a very high energy penalty, to slowly adapt to each other in a complementary conformation without overlaps. One potential pitfall of this approach is the possibility that the ligand may become interlocked with the protein leading to failure of the docking procedure.

Although the use of soft receptors presents a number of advantages such as ease of implementation and computation speed, it also makes use of conformational and energetic assumptions which are difficult to verify. This can easily result in errors especially if the soft region is made excessively large to account for larger conformational changes on the part of the receptor.

Selection of Specific Degrees of Freedom

In order to reduce the complexity of modeling the very large dimensional space representing the full flexibility of the protein, is it possible to obtain an approximate

solution by selecting only a few degrees of freedom to model explicitly. The degrees of freedom chosen usually correspond to rotations around single bonds (see Fig. (2)). The reason for this choice is that these degrees of freedom are usually considered the natural degrees of freedom in molecules. Rotations around bonds lead to deviations from ideal geometry that result in a small energy penalty when compared to deviations from ideality in bond lengths and bond angles. This assumption is in good agreement with current modeling force fields such as CHARMM [38] or AMBER [39]. Selection of which torsional degrees of freedom to model is usually the most difficult part of this method because it requires a considerable amount of a priori knowledge of alternative binding modes for a given receptor. This knowledge usually is a result of the availability of different experimental structures obtained under different conditions or using different ligands. If multiple experimental structures are not available some insight can be obtained from simulation methods such as Monte Carlo (MC) or molecular dynamics (MD). The torsions chosen are usually rotations of aminoacid side chains in the binding site of the receptor protein. It is also common to further reduce the search space by using rotamer libraries for the aminoacid side chains [40-42].

The first application of using select degrees of freedom to model receptor flexibility was carried out by Leach [43]. This work made use of the Dead End Elimination (DDE) [44] and the A* algorithm [45] to explore the conformational space for the degrees of freedom for both ligand and receptor. The DDE states that a rotamer r of residue i (i_r) is incompatible with the global energy minimum structure if it satisfies the following inequality:

$$E_{i_r,rigid} + \sum_j \min_s \varepsilon_{i_r,j_s} > E_{i_t,rigid} + \sum_j \max_s \varepsilon_{i_t,j_s} ,$$

where $E_{i_r, rigid}$ is the interaction energy between i_r and the rigid part of the protein, $\min_s \varepsilon_{i_r, j_s}$ is the minimum interaction energy between rotamer r of residue i with all permitted rotamers s of residue j , and $\max_s \varepsilon_{i_r, j_s}$ is the corresponding maximum value for rotamer i_r . The A* (pronounced "A star") algorithm is a well known and well studied best-first search algorithm that works by expansion of nodes, always expanding the current fringe node that looks like it is along the best path from the start node to the goal node. Besides using these two methods Leach also introduced an energy threshold to the global minimum and returned all structures under this threshold as potential binding candidates. The purpose of the threshold is to take into account the fact that the true global energy minimum of the bound complex does not necessarily correspond to that of the force field. This work was later extended by Leach and Lemon [46] to explore the conformational space of whole proteins. Schaffer *et al.* [47] also used DDE to perform flexible docking of two HIV-1 protease inhibitors with mutants of this protein. The DDE algorithm was applied using a rotamer library to perform discrete optimization of all possible side chain conformation combinations in the binding site. The best solutions were later optimized in conjunction with the ligand using a Monte Carlo simulated annealing technique. This two step method leads to a solution which is not restricted to the dihedral values present in the rotamer library and is also of lower energy. More recently Althaus *et al.* [48] used two alternative combinatorial optimization methods to also solve the side chain conformation problem. The first method consists of a heuristic multi-greedy approach, which is faster but does not necessarily produce an optimal solution. The second method is able to find the global minimum energy conformation and is based on a branch-and-cut algorithm and integer linear programming.

In the program GOLD Jones *et al.* [49] use a genetic algorithm (GA) to dock a flexible ligand to a semi-flexible protein. GAs are an optimization method which derives its behavior from a metaphor of the process of evolution. A solution to a problem is encoded in a chromosome and a fitness score is assigned to it based on the relative merit of the solution. A population of chromosomes then goes through a process of evolution in which only the fittest solutions “survive”. This program takes into account not only the position and conformation of the ligand but also the hydrogen bonding network in the binding site. This was achieved by encoding orientation information for donor hydrogen atoms and acceptors in the GA chromosome. This type of conformational information is very important because if the starting point for a docking study is a rigid crystallographic structure, the orientations of hydroxyl groups will be undetermined. Being able to model these orientations explicitly removes any bias that may result from positioning hydroxyl groups based upon a known ligand. One limitation of this work is that the binding site still remained essentially rigid because protein conformational changes are limited to a few terminal bonds. This program performed very well for hydrophilic ligands but encountered some difficulties when trying to dock hydrophobic ligands due to the reduced contribution of hydrogen bonding to the binding process.

In SPECITOPE Schnecke *et al.* [36] also make use of side chain rotations in the late stages of docking to remove steric overlaps between the protein side chains and the ligand. If an overlap clash is detected the program attempts to remove it by rotating the side chain through the minimal angle that resolves the clash. The single bond closest to the bumping atoms in the side chain is used first to resolve the overlap. If a bump free conformation cannot be generated with this rotation, the next rotatable bond closer to the

ligand backbone is rotated. This procedure will miss potential combinations of side chain conformations that do not overlap with the ligand and is not capable of finding the minimum energy conformation. Nevertheless, it will successfully resolve many cases of overlap.

Anderson *et al.* [50] introduced the algorithm SOFTSPOTS that addresses the problem of knowing which rotational degrees of freedom should be selected to represent receptor flexibility. Using a single protein structure, this algorithm is capable of identifying regions of high flexibility. The results were combined with a second algorithm named PLASTIC that provides a collection of possible conformations based on rotamer libraries effectively reducing the bias caused by structures of proteins co-crystallized with inhibitors. More recently, Kayrys *et al.* [51] have improved the Mining Minima optimizer method, first described by David *et al.* [52], to include select side chain degrees of freedom in the docking simulation of several proteins and ligands.

A common theme among the work described in this section is that receptor side chain conformations are modeled using torsional degrees of freedom. In order to make the calculation of interactions energies more efficient it would be desirable to work with a force field which is also described in terms of internal coordinates to avoid repeated conversion between two coordinate systems. Use of internal coordinate force fields also leads to more efficient convergence of energy optimizations. Abagyan *et al.* described a method to carry out flexible protein-ligand docking by global energy optimization in internal coordinates [53] and more recently described a method to accurately "project" a Cartesian force field onto an internal coordinate molecular model with fixed-bond geometry [54].

Multiple Receptor Structures

One possible way of representing a flexible receptor for drug design applications is the use of multiple static receptor structures (see Fig. (3)). This concept is supported by the currently accepted model that proteins in solution do not exist in a single minimum energy static conformation but are in fact constantly jumping between low energy conformational substates [55-57]. In this way the best description for a protein structure is that of a conformational ensemble [20, 58] of slightly different protein structures coexisting in a low energy region of the potential energy surface. Moreover the binding process can be thought of as not exactly an induced fit model as described by Koshland in 1958 [6] but more like a selection of a particular substate from the conformational ensemble that best complements the shape of a specific ligand [17].

The use of multiple static conformations for docking poses two critical questions: “How can we obtain a representative subset of the conformational ensemble typical of a given receptor?” and “What is the best way of combining this large amount of structural information for a docking study?”. Currently there exist only a limited set of means to generate the three dimensional structure of macromolecules. The structures can be determined experimentally either from X-ray crystallography or NMR or generated via computational methods such as MC or MD simulations. Simulations typically use as a starting point a structure determined by one of the experimental methods. Ideally we would like to use the method that is able to provide the greatest conformational sampling of the structure space. Comparisons done between the experimental and computational techniques [59, 60] seem to indicate that X-ray crystallography and NMR structures seem to provide better coverage. However this balance can potentially change due to advances

in computational methods [61]. Another limitation in choosing data sources is availability. Although experimental data is preferable, the monetary and time cost of determining multiple structures experimentally is significantly higher than obtaining the same amount of data computationally. The second critical question also remains open and current approaches use diverse ways of combining multiple structures.

The first use of multiple structures for a drug design applications was by Pang and Kozikowski [62] to study the binding of huperzine A (HA) to acetylcholinesterase (AChE). In this study the authors ran a short MD simulation (40 ps) of AChE from which they extracted 69 conformations which were docked using rigid docking to HA. This study successfully predicted that HA binds to the bottom of the binding cavity of AChE (the gorge). More recently, other studies [63, 64] have exploited similar approaches but used a larger number of structures, longer MD sampling, and more accurate simulation conditions. Instead of resorting to computational methods to derive structural data Knegtel *et al.* [65] used a family of structures from an NMR structural determination or, as an alternative, several crystal structures of the same protein system. In that study the authors combined the different structures into a single interaction energy grid to be used for rigid receptor docking by the DOCK program. Interaction energy grids are calculated by placing a probe atom at discrete points in the space around a target protein and assigning to the grid point the value of the interaction energy between the probe and protein. This grid is then utilized as a fast lookup table for interaction energy calculations, effectively reducing the cost of computation from quadratic to linear. The averaged grids were constructed using energy-weighted and geometry-weighted averaging methods. The main limitation of these averaging approaches is that they can lead to loss of geometric

accuracy. The binding energies computed from the composite grid are also less favorable than for individual grids for each protein in the ensemble. In the case of geometry-weighted averaging the binding site can become too permissive in terms the size of the ligand that it can accommodate. Sudbeck *et al.* [66] superimposed the crystal structures of nine inhibitor complexes of HIV reverse transcriptase to generate a composite binding site which summarized its unique critical features. The overlaid coordinates of the nine different inhibitors were used to generate a combined molecular surface defining an enlarged binding pocket which represented the plasticity of the receptor. The combined binding pocket was used to verify the results obtained from the docking of small molecules to a single structure of reverse transcriptase through a conjugate gradient minimization method for the ligand and all residues within 5 Å. This study resulted in the development of two new inhibitors. Multiple crystal structures of HIV-1 protease were also utilized by Bouzida *et al.* [67] to account for receptor flexibility. Broughton *et al.* [68] used different conformation snapshots from a short MD simulation of dihydrofolate reductase to generate interactions grids which were also combined into a single grid by means of a weighted average method. Before calculation of the grids the structures were superimposed using the bound inhibitors as a reference.

As mentioned earlier one of the methods of combining multiple receptor structures is to create an average grid for the protein/ligand interaction potential [65]. Osterberg *et al.* [69] analyzed this problem in depth by using HIV-1 protease as a model system and comparing four different grid averaging methods. The first two naïve methods consisted of a mean grid that takes a simple point-by-point average across all the grids, and a minimum grid takes the minimum value across all the grids. These two methods

performed poorly. The third approach is similar to that described by Knegtel *et al.* [65] and consists of a weighted averaging scheme. In this case if one or more of the grids contain a favorable, negative value, their weights will dominate the average. On the other hand, if all the grids contain unfavorable positive values, all will have identical small weights resulting in an unfavorable region representing all grids. The fourth averaging scheme is similar to the previous one but uses a Boltzmann assumption to calculate the weight based on the interaction energy. The last two averaging schemes were able to efficiently represent multiple structures in a single grid and the docking results were satisfactory. However, as the authors point out this method for incorporation of conformational flexibility can introduce potential dangerous artifacts such as positive interaction regions for mutually exclusive solutions.

A rather different way of considering protein flexibility as represented in interaction grids for multiple static structures is GRID/CPCA (Consensus Principal Component Analysis). This method, introduced by Kastenholtz *et al.* [70] in the study of serine proteases, is an extension of GRID/PCA [71], and can be used to identify selectivity features for a receptor. One of the main advantages of this method versus its predecessor is that it allows the inclusion of more than two structures in the PCA calculation. Moreover, when several structures are used, it allows for some averaging of the individual structures reducing differences which might be due to experimental variations but which are not relevant to the specificity features of the receptors.

In the program FlexE, Claussen *et al.* [72] introduced a new method of combining multiple receptor structures to represent a flexible binding site. The algorithm starts by superimposing the set of conformations available for a given receptor and merging

similar parts of the structures. Dissimilar substructures are treated as independent alternatives and FlexE selects the combination of substructures which complements best a given ligand with respect to the scoring function. In practice, this results in the generation of alternative receptor conformations which were not present in the initial set but may constitute valid docking targets.

More recently Moreno and León [73] introduced a new receptor representation that allows the use of an ensemble of protein structures as input to DOCK instead of a single rigid structure. In this approach, an ensemble of protein/inhibitor complex structures is used to construct a set of templates of attached points (one for each type of amino acid) located in positions suitable for interactions with ligand atoms. The combination of templates gives a description of a flexible binding site. The authors propose the method of attached points as an alternative to SPHGEN [74, 75] or SURFSPH [76] to generate a binding site descriptor.

Multiple protein structures can be used not only to generate flexible receptor representations for docking purposes, but also to generate pharmacophores. A pharmacophore is a template for the desired ligand. The pharmacophore is represented by a set of features that an effective ligand should possess and a set of spatial constraints among the features. The features can be specific atoms, positive or negative charges, hydrophobic or hydrophilic centers, hydrogen bond donors or acceptors, and others. The spatial arrangement of the features represents the relative 3D placements of these features in the docked conformation of the ligand. Carlson *et al.* introduced the concept of a dynamic pharmacophore by combining sets of structures derived by either X-ray crystallography [77] or snapshots of a MD simulation [78]. Potential sites of interest in

the receptor binding site are determined by running a multi-unit MC minimization using probe molecules for the different features of interest. The results of these simulations for each conformer are then overlaid. This procedure then reveals conserved binding regions that are highly occupied during the MD simulation despite the flexibility of the receptor. The conserved features define the dynamic pharmacophore. Studies similar to dynamic pharmacophore identification were performed by Stultz and Karplus [79] using a combination of the Multiple Copy Simultaneous Search (MCSS) and Locally Enhanced Sampling (LES) methods. This protocol uses quenched MD to identify energetically favorable positions and orientations of small functional groups in a flexible binding site. In this method multiple copies of the functional groups are distributed in the binding site and quenched to find energy minima. These functional groups can later be used as building blocks for larger ligands.

One of the main advantages of using multiple structures instead of using a selection of degrees of freedom to represent protein flexibility is that the flexible region is not limited to a specific small region of the protein. This kind of representation allows the protein to be fully flexible without the exponential blow up in terms of computational cost that would derive from including all the degrees of freedom of the protein. On the other hand flexibility is modeled implicitly and as such only a small fraction of the conformational space of the receptor is represented. In addition, the method by which the multiple receptor structures are combined has a drastic influence on the possible results of the docking computation.

Molecular Simulations

To simulate the binding process with as much detail as possible and avoid some of the limitations of previous flexibility models one can use force field based atomistic simulation methods such as MC or MD (see Fig. (4)). Whereas MD applies the laws of classical mechanics to compute the motion of the particles in a molecular system, MC methods are so called because they are based on a random sampling of the conformational space. The main advantage of these types of flexibility representations in docking studies is that they are very accurate and can model explicitly all degrees of freedom of the system including the solvent if necessary. Unfortunately, the high level of accuracy in the modeling process comes at the expense of a very high computational cost. For example, in the case of molecular dynamics, state of the art protein simulations can only simulate periods ranging from 10 to 100 ns, even when using large parallel computers or clusters. Given that diffusion and binding of ligands to proteins is a slower process it is clear that the use of these simulations techniques cannot be used as a general method to screen large databases of compounds at least in the near future. However, it is possible to carry out approximations that reduce the computational expense and lead to insight which would be impossible to gain using less flexible receptor representations. The cost of carrying out the computational approximations is usually a loss in accuracy.

In order to solve the time sampling limitations of traditional molecular dynamics Di Nola *et al.* [80] used a modified temperature coupling scheme to perform the docking of phosphocholine onto immunoglobulin McPC603. Instead of coupling the whole system to the same temperature bath, Di Nola used a regular coupling temperature to the internal degrees of freedom of the ligand and a very high temperature (1300-1700 K) for

the translational modes. In practice this allows the ligand to sample the surface of protein receptor much faster without disturbing internal motions. This method was extended later by Mangoni *et al.* [81] to also include the flexibility of the receptor which was also coupled to the lower temperature bath (300 K). In order to further reduce the computational cost of the simulation, the protein simulation was restricted to a sphere of 20 Å around the chain oxygen of the phosphocholine molecule in the crystallographic position. The remaining part of the protein was kept rigid. The same approach of restricting the full molecular simulation to the vicinity of the binding site was used by Luty *et al.* [82] to simulate the docking of benzamidine to trypsin and by Wasserman *et al.* [83] to simulate the docking of L-leucine hydroxamic acid to thermolysin. Given and Gilson [84] also restricted flexibility to the binding site area of HIV-1 protease within the context of a hierarchical docking protocol. In this method conformations are evolved in stages, with the lowest energy conformations from one stage serving as starting points for the next. The focus of this study was not to develop a computationally efficient method but rather generate a picture of the ligand-binding energy surface using different energy functions.

A different approach to enhance the sampling rate of force field based simulations methods is to smooth the potential energy surface in order to increase the rate of transition between metastable conformations. Nakajima *et al.* [85, 86] used the method of multicanonical molecular dynamics simulation based on the work of Berg *et al.* [87] to simulate the binding of a short proline-rich peptide to a Src homology 3 (SH3) domain. In this method the simulation is carried out in a deformed energy surface characterized by a flatter energy distribution and resulted in much faster sampling of the conformational

space of the ligand and the binding site of SH3. Pak and Wang [88] applied the Tsallis transformation to the non bonded interaction potential of the CHARMM force field and ran dynamics simulations with infrequent q -jumping and q -relaxation between the normal and the smooth energy surface. By combining potential smoothing and by restricting the flexibility of the receptor to aminoacid side chains in the binding site it was possible to successfully simulate the formation of streptavidin/biotin and protein kinase C/phorbol-13-acetate complexes. More recently Zhu *et al.* [89] introduced the program F-DycoBlock that performs the docking of a flexible ligand to a flexible receptor using multiple-copy stochastic molecular dynamics. In this method several copies of the ligand molecule are simulated simultaneously. These copies are constructed in a special way because they do not interact with each other. The protein moves in the mean field of all ligand copies. In this study the authors also used four different types of receptor flexibility: all-atom restrained, backbone restrained, intramolecular hydrogen-bond restrained and active-site flexible.

The alternative to the use of MD is the use of MC based methods. In [90] Caflisch *et al.* extended the MC minimization approach to take into account receptor flexibility by the use of a flexible enzyme binding site, whose side chains are submitted to random perturbations. This work used the Metropolis MC method as a global optimization method combined with a conjugate gradient minimization scheme for local optimization. Solute-solvent energies were calculated by solving the finite-difference linearized Poisson-Boltzmann equation. Trosset and Scheraga developed PRODOCK package for docking [91]. The global optimization method used in this tool is the scaled collective variables Monte Carlo method developed by Noguti and Go [92] with energy

minimization after each MC step. The minimization step was greatly improved by the use of a grid based energy evaluation technique using Bezier splines [93, 94] and the use of collective degrees of freedom. One of the main problems with conventional simulation methods is the propensity for the system to get trapped in local minima leading to a computationally inefficient sampling of the energy landscape. In order to minimize this problem, Verkhivker et al. [95] made use of parallel simulated tempering dynamics to investigate the specificity of binding and mechanisms of inhibitor resistance in HIV-1 protease. Parallel tempering is a replica-exchange Monte Carlo method that simulates several copies of the protein simultaneously using different temperatures and periodically exchanges conformations at neighboring temperatures. This process enhances conformational sampling by facilitating the escape from local minima.

An innovative approach to predicting the binding conformation of a flexible ligand in a flexible binding pocket was recently introduced by Ota and Agard [96] by combining the simulated annealing and the crystallographic refinement search methods. This scheme starts by using a shrunken ligand for which the bond lengths and the non bonded interactions have been greatly reduced. The ligand is then grown in the binding site using a simulated annealing protocol to search for a bound conformation. This procedure is repeated several times and an averaged pseudo electron density map is calculated by averaging amplitudes and phases calculated from each structure. The final bound conformation is determined by conventional crystallographic refinement using the calculated structure factors. This method has the advantages of being able to model individual water molecules relevant to the binding configuration and providing a series of crystallographic measures, such as B-factors, that facilitate the comparison with X-ray

crystallographic data. Unfortunately, due to the high computational cost, this technique is not suitable for large scale database screening but could be useful in the late stages of a docking study.

Collective Degrees of Freedom

An alternative representation for protein flexibility is the use of collective degrees of freedom. This approach enables the representation of full protein flexibility, including loops and domains, without a dramatic increase in computational cost. Collective degrees of freedom are not native degrees of freedom of molecules. Instead they consist of global protein motions that result from a simultaneous change of all or part of the native degrees of freedom of the receptor.

Collective degrees of freedom can be determined using different methods. One method is the calculation of normal modes for the receptor [97-99]. Normal modes are simple harmonic oscillations about a *local* energy minimum which depends on the structure of the receptor and the energy function. For a purely harmonic energy function, any motion can be exactly expressed as a superposition of normal modes. In proteins, the lowest frequency modes correspond to delocalized motions, in which a large number of atoms oscillate with considerable amplitude. The highest frequency motions are more localized such as the stretching of bonds. By assuming that the protein is at an energy minimum we can represent its flexibility by using the low frequency normal modes as degrees of freedom for the system. Zacharias and Sklenar [100] applied a method similar to normal mode analysis to derive a series of harmonic modes that were used to account for receptor flexibility in the binding of a small ligand to DNA. This in practice reduced the number of degrees of freedom of the DNA molecule from 822 ($3 \times 276 \text{ atoms} - 6$) to

approximately 5 to 40. Keserû and Kolossváry [101] also used a normal mode based model [102, 103] to study inhibitor binding to HIV integrase.

An alternative method of calculating collective degrees of freedom for macromolecules is the use of dimensional reduction methods. The most commonly used dimensional reduction method for the study of protein motions is principal component analysis (PCA). This method was first applied by Garcia [104] in order to identify high-amplitude modes of fluctuations in macromolecular dynamics simulations. It has also been used to identify and study protein conformational substates [105-107], as a possible method to extend the timescale of molecular dynamics simulations [108-110] and as a method to perform conformational sampling [110-112]. The most significant principal components have a direct physical interpretation. They correspond to a concerted motion of the protein where all the atoms move in specific spatial directions and with fixed ratios in overall displacement. The directions and ratios are indicated by the direction and size of the arrows in Fig. (5). Recently, we have presented a protocol [113] to derive a reduced basis representation of protein flexibility which can be used to reduce the complexity of modeling protein/ligand interactions. By considering only the most significant principal components as the valuable degrees of freedom of the system, it is possible to reduce an initial search space of thousands of degrees of freedom to less than fifty. This is achievable because the fifty most significant principal components usually account for 80-90% of the overall conformational variance of the system. We applied this protocol to two different model systems: HIV-1 protease and aldose reductase. Although there is inevitably some loss in accuracy, we show that we can obtain good approximations to experimentally determined conformations while working in a

drastically reduced search space. Moreover, in the case of HIV-1 protease we were able to identify biological relevant motions such as the flap opening for the binding site. The PCA approach avoids some of the limitations of normal modes such as deficient solvent modeling and existence of multiple energy minima during a large motion. The last limitation contradicts the initial assumption of a single well energy potential.

An alternative representation using a concept similar to collective degrees of freedom is based on the concept of molecular hinges [114-116]. This research is based on methods used in the fields of computer vision and robotics. The hinge-bending approach was originally used to model flexibility for the ligand, but the roles of the ligand and the protein can be swapped since the mathematical problem is symmetrical. Hinges are articulation points placed at specific places in the protein which allow for relative movement of domains or substructural parts. Several simultaneous hinges can be modeled. These hinge points do not correspond to single degrees of freedom of the original model but are instead hinges which are allowed to rotate in three dimensions, representing implicitly rotations about consecutive or nearby bonds. The ligand is also considered flexible and the search for a docking conformation is done simultaneously mimicking the induced fit process. Like pliers closing on a screw, the receptor adapts its shape to that of the ligand. This method does not model conformational changes for sidechains explicitly. However, it models large conformational changes efficiently and can be easily combined with some of the methods described above in order to model conformational changes for specific areas of the receptor at an atomistic level. One of the main problems of the molecular hinges approach is determining the location of the hinges. Recently, Jacobs *et al.* [117] introduced a flexibility prediction algorithm based

on graph theory which can help solve this problem. The algorithm computes a constraint network for the protein defined by the bonds (covalent and hydrogen) and salt bridges and identifies all the rigid and flexible substructures in the protein, including overconstrained regions and underconstrained or flexible regions.

Using collective degrees of freedom as a flexibility representation has a number of advantages and disadvantages. One advantage is that protein flexibility is not limited to a specific small region of the protein as was the case when using only select degrees of freedom (such as binding site torsional angles). Furthermore, since only a few independent degrees of freedom are used in the optimization procedure the computational cost is similar to using only select degrees of freedom and is much less than the cost of techniques which consider all degrees of freedom such as traditional MD or MC. On the other hand, the degrees of freedom that are searched during the drug design procedure are not the native degrees of freedom of the protein, but collective modes of motion that try to account for most of the variance observed during protein motion. This may result in a loss of accuracy and difficulty in obtaining exact solutions. For example, there may not exist a combination of values for the reduced basis formed by the most significant collective degrees of freedom that results in the exact placement of all binding site sidechains as observed in an experimentally determined structure. However, this is probably a minor problem since exact solutions are rarely obtained using other methods, and as shown in [113] it is possible to obtain very good approximations using only a small number of collective degrees of freedom. Furthermore, in order to avoid high energy penalties that might result from van der Waals clashes it is possible to combine

collective modes of motion either with a soft receptor representation or with a post processing minimization procedure.

Other challenges

The docking problem poses many other challenges. In addition to the problem of modeling protein flexibility in current docking methodologies, there are other critical problems which still need to be addressed. Two of the most important are developing better solvation models and improving scoring functions. If solvation is not handled correctly, the results of a docking study usually include candidate ligands which are too highly charged or too large. The second problem is even more critical and the best search procedure can be rendered useless if matched to a bad scoring function.

The problem of modeling the solvent can be further subdivided into two different components. The first issue is the placement of discrete water molecules during the docking procedure. This issue is analogous to the protein flexibility problem because the consequence of including water molecules in the simulation procedure is also an increase in the number of degrees of freedom. The number of degrees of freedom introduced in the computation depends on the level of detail at which the water is modeled. It can range from three degrees of freedom, if the water is modeled as a sphere, to nine, if it is modeled using the Cartesian coordinates of its constituent atoms. The increase is also directly proportional to the number of water molecules explicitly modeled. This explains the high cost of including even a reduced number of waters in the modeling process. Raymer *et al.* [118] used a K-nearest-neighbors genetic algorithm to predict conserved waters in protein/ligand interactions. This method is useful because it can help to identify sites favored to be occupied by a mediating water molecule or a polar ligand atom, as

well as water molecules likely to be displaced by the ligand. Rarey *et al.* [119] also extended the program FlexX to include the placement of single water molecules during the incremental construction approach. The results were mixed and showed an improvement in 27.5% of test cases and worsening in 23%. The results obtained are illustrative of the difficulty of this problem. The second issue critical to solvent modeling is calculating the contribution of the solvation energy to the overall affinity of binding. Current results [120-122] indicate that the use of continuum electrostatic models can be efficiently used to address this problem.

Besides the representation of receptor flexibility and the subsequent conformational search for the docked conformation, another critical aspect of molecular docking is the ability to correctly score potential binding candidates. Ideally these would be scored according to their binding free energy. Unfortunately the process of accurately deriving the interaction energy between a receptor and a ligand is complicated by the need to add large energy contributions of opposite signs to compute a small energy value. Furthermore, some of the steps for accurate energy calculation such as entropy estimation and complete conformational sampling are still problematic. Methods such as free energy perturbation [123] are too computationally expensive to use routinely to screen a large number of drug candidates. As an alternative to an accurate calculation of the free energy of binding current docking methods use scoring functions. For reviews on this topic see [30, 32, 124-128].

Current scoring methods can be roughly divided into four categories. These are shape and chemical complementary, force field based, empirical, and knowledge based. All these scoring schemes have advantages and disadvantages. The main difficulty in

designing good scoring functions for docking purposes is achieving a balance between accuracy and computational cost. Some of the current methods available can be very accurate, but they take more time than what is computationally feasible. One solution that addresses some of the limitations of specific scoring schemes is to combine the scores using different methods and use this information as a new score. This is the underlying principle for consensus scoring [129]. Presently, the most practical solution is the use of a fast scoring function for the early stages to eliminate potential drug candidates which are clearly inappropriate and a more accurate method for estimating the free energy of binding which is able to correctly rank different ligand candidates.

Conclusions

The problem of incorporating protein flexibility in routine screening of chemical databases is still an unsolved and hard problem. One of the main reasons for this is the exponential explosion in complexity due to the inclusion of the flexibility parameters for the receptor. In order to address this limitation several flexibility models for the receptor have been developed which try to limit the number of additional model parameters. These models can be roughly grouped divided into five different categories. These are the use of soft receptors, the selection of a few critical degrees of freedom in the receptor binding site, the use of multiple receptor structures either individually or by combining them using an averaging scheme, the use of traditional molecular simulations, or the use of collective degrees of freedom as a new basis of representation for protein flexibility. All the flexible receptor models presented in this paper strive to balance an improvement in the accuracy of the docking predictions with an increase in computational cost.

Currently the most promising solution to include protein flexibility in a drug design effort is to use a combination of methods. Docking should proceed in a hierarchical fashion. In an initial phase, a large number of possible drug candidates are eliminated using simpler and faster schemes not only in terms of the conformational search procedure but also in regard to the scoring function. In later stages protein flexibility can be taken into account using more elaborate and computationally expensive methods. These types of hierarchical schemes will also benefit greatly in the future from advances in the field of parallel computation. The increasing availability of parallel systems will also result in the development of a new generation of docking algorithms to make the best use of these machines and clusters.

Acknowledgements

M. Teodoro has been partially supported by a PRAXIS XXI Pre-doctoral Fellowship from the Portuguese Ministry of Science, a Whitaker Biomedical Engineering Grant, an Autrey Fellowship Award from Rice University and a Pre-doctoral Fellowship from the Keck Center for Computational Biology. Work on this paper by L. Kavraki has been supported in part by NSF IRI-9702281, NSF 0114796 Grant, a Whitaker Biomedical Engineering Grant, a Sloan Fellowship and a Texas ATP Award.

References

- [1] Fischer, E. *Ber. Dtsch. Chem. Ges.*, **1894**, 27, 2985.
- [2] Blow, D. M. *Acc.Chem.Res.*, **1976**, 9, 145.
- [3] Huber, R.; Bode, W. *Acc. Chem. Res.*, **1978**, 11, 114.
- [4] Hubbard, S. J.; Campbell, S. F.; Thornton, J. M. *J. Mol. Biol.*, **1991**, 220, 507.
- [5] Amit, A. G.; Mariuzza, R. A.; Phillips, S. E.; Poljak, R. J. *Science*, **1986**, 233, 747.
- [6] Koshland, D. E. *P. Natl. Acad. Sci. USA*, **1958**, 44, 98.
- [7] Weber, P. C.; Ohlendorf, D. H.; Wendoloski, J. J.; Salemme, F. R. *Science*, **1989**, 243, 85.
- [8] Wlodawer, A.; Vondrasek, J. *Annu. Rev. Bioph. Biom.*, **1998**, 27, 249.
- [9] Bystroff, C.; Kraut, J. *Biochemistry-US*, **1991**, 30, 2227.
- [10] Wilson, D. K.; Tarle, I.; Petrash, J. M.; Quioco, F. A. *P. Natl. Acad. Sci. USA*, **1993**, 90, 9847.
- [11] Betts, M. J.; Sternberg, M. J. *Protein Eng.*, **1999**, 12, 271.
- [12] Murray, C. W.; Baxter, C. A.; Frenkel, A. D. *J. Comput. Aid. Mol. Des.*, **1999**, 13, 547.
- [13] Najmanovich, R.; Kuttner, J.; Sobolev, V.; Edelman, M. *Proteins*, **2000**, 39, 261.
- [14] Zhao, S.; Goodsell, D. S.; Olson, A. J. *Proteins*, **2001**, 43, 271.
- [15] Fradera, X.; Cruz, X.; Silva, C. H. T. P.; Gelpi, J. L.; Luque, F. J.; Orozco, M. *Bioinformatics*, **2002**, 18, 939.
- [16] Vazquez-Laslop, N.; Zheleznova, E. E.; Markham, P. N.; Brennan, R. G.; Neyfakh, A. A. *Biochem. Soc. T.*, **2000**, 28, 517.
- [17] Ma, B.; Kumar, S.; Tsai, C. J.; Nussinov, R. *Protein Eng.*, **1999**, 12, 713.

- [18] Ma, B.; Shatsky, M.; Wolfson, H. J.; Nussinov, R. *Protein Sci.*, **2002**, *11*, 184.
- [19] Ma, B.; Wolfson, H. J.; Nussinov, R. *Curr. Opin. Struct. Biol.*, **2001**, *11*, 364.
- [20] Bursavich, M. G.; Rich, D. H. *J. Med. Chem.*, **2002**, *45*, 541.
- [21] Verkhivker, G. M.; Bouzida, D.; Gehlhaar, D. K.; Rejto, P. A.; Freer, S. T.; Rose, P. W. *Curr. Opin. Struct. Biol.*, **2002**, *12*, 197.
- [22] Carlson, H. A.; McCammon, J. A. *Mol. Pharmacol.*, **2000**, *57*, 213.
- [23] Holtje, H. D.; Kier, L. B. *J Pharm Sci*, **1974**, *63*, 1722.
- [24] Kier, L. B.; Aldrich, H. S. *J Theor Biol*, **1974**, *46*, 529.
- [25] Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. *J. Mol. Biol.*, **1982**, *161*, 269.
- [26] Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. *J. Mol. Biol.*, **1996**, *261*, 470.
- [27] Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. *J. Comput. Chem.*, **1998**, *19*, 1639.
- [28] Gane, P. J.; Dean, P. M. *Curr. Opin. Struct. Biol.*, **2000**, *10*, 401.
- [29] Klebe, G. *J. Mol. Med.*, **2000**, *78*, 269.
- [30] Muegge, I.; Rarey, M. *Rev. Comp. Chem.*, **2001**, *17*, 1.
- [31] Shoichet, B. K.; McGovern, S. L.; Wei, B.; Irwin, J. J. *Curr. Opin. Chem. Biol.*, **2002**, *6*, 439.
- [32] Halperin, I.; Ma, B.; Wolfson, H.; Nussinov, R. *Proteins*, **2002**, *47*, 409.
- [33] Carlson, H. A. *Curr. Opin. Chem. Biol.*, **2002**, *6*, 447.
- [34] Carlson, H. A. *Curr. Pharm. Des.*, **2002**, *8*, 1571.
- [35] Jiang, F.; Kim, S. H. *J. Mol. Biol.*, **1991**, *219*, 79.

- [36] Schnecke, V.; Swanson, C. A.; Getzoff, E. D.; Tainer, J. A.; Kuhn, L. A. *Proteins*, **1998**, *33*, 74.
- [37] Apostolakis, J.; Pluckthun, A.; Caflisch, A. *J. Comput. Chem.*, **1998**, *19*, 21.
- [38] MacKerell, A. D.; Bashford, D.; Bellot, M.; Karplus, M. *J. Phys. Chem. B*, **1998**, *102*, 3586.
- [39] Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.*, **1995**, *117*, 5179.
- [40] Tuffery, P.; Etchebest, C.; Hazout, S.; Lavery, R. *J. Biomol. Struct. Dyn.*, **1991**, *8*, 1267.
- [41] Lovell, S. C.; Word, J. M.; Richardson, J. S.; Richardson, D. C. *Proteins*, **2000**, *40*, 389.
- [42] Dunbrack, R. *Curr. Opin. Struct. Biol.*, **2002**, *12*, 431.
- [43] Leach, A. R. *J. Mol. Biol.*, **1994**, *235*, 345.
- [44] Desmet, J.; DeMaeyer, M.; Hazes, B.; Lasters, I. *Nature*, **1992**, *356*, 539.
- [45] Hart, P. E.; N.J., N.; Raphael, B. *IEEE T. Syst. Sci. Cyb.*, **1968**, *4*, 100–114.
- [46] Leach, A. R.; Lemon, A. P. *Proteins*, **1998**, *33*, 227.
- [47] Schaffer, L.; Verkhivker, G. M. *Proteins*, **1998**, *33*, 295.
- [48] Althaus, E.; Kohlbacher, O.; Lenhof, H. P.; Muller, P. *J. Comput. Biol.*, **2002**, *9*, 597.
- [49] Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. *J. Mol. Biol.*, **1997**, *267*, 727.

- [50] Anderson, A. C.; O'Neil, R. H.; Surti, T. S.; Stroud, R. M. *Chem. Biol.*, **2001**, *8*, 445.
- [51] Kairys, V.; Gilson, M. K. *J. Comput. Chem.*, **2002**, *23*, 1656.
- [52] David, L.; Luo, R.; Gilson, M. K. *J. Comput. Aid. Mol. Des.*, **2001**, *15*, 157.
- [53] Totrov, M.; Abagyan, R. *Proteins*, **1997**, *Suppl*, 215.
- [54] Katritch, V.; Totrov, M.; Abagyan, R. *J. Comput. Chem.*, **2003**, *24*, 254.
- [55] Noguti, T.; Go, N. *Proteins*, **1989**, *5*, 97.
- [56] Frauenfelder, H.; Sligar, S. G.; Wolynes, P. G. *Science*, **1991**, *254*, 1598.
- [57] Kitao, A.; Hayward, S.; Go, N. *Proteins*, **1998**, *33*, 496.
- [58] Rich, D. H.; Bursavich, M. G.; Estiarte, M. A. *Biopolymers*, **2002**, *66*, 115.
- [59] Clarage, J. B.; Romo, T.; Andrews, B. K.; Pettitt, B. M.; Phillips, G. N., Jr. *P. Natl. Acad. Sci. USA*, **1995**, *92*, 3288.
- [60] Philippopoulos, M.; Lim, C. *Proteins*, **1999**, *36*, 87.
- [61] Karplus, M.; McCammon, J. A. *Nat. Struct. Biol.*, **2002**, *9*, 646.
- [62] Pang, Y. P.; Kozikowski, A. P. *J. Comput. Aid. Mol. Des.*, **1994**, *8*, 669.
- [63] Lin, J. H.; Perryman, A. L.; Schames, J. R.; McCammon, J. A. *J. Am. Chem. Soc.*, **2002**, *124*, 5632.
- [64] Kua, J.; Zhang, Y.; McCammon, J. A. *J. Am. Chem. Soc.*, **2002**, *124*, 8260.
- [65] Knegtel, R. M.; Kuntz, I. D.; Oshiro, C. M. *J. Mol. Biol.*, **1997**, *266*, 424.
- [66] Sudbeck, E. A.; Mao, C.; Vig, R.; Venkatachalam, T. K.; Tuel-Ahlgren, L.; Uckun, F. M. *Antimicrob. Agents Ch.*, **1998**, *42*, 3225.

- [67] Bouzida, D.; Rejto, P. A.; Arthurs, S.; Colson, A. B.; Freer, S. T.; Gehlhaar, D. K.; Larson, V.; Luty, B. A.; Rose, P. W.; Verkhivker, G. M. *Int. J. Quantum Chem.*, **1999**, 72, 73.
- [68] Broughton, H. B. *J. Mol. Graph. Model.*, **2000**, 18, 247.
- [69] Osterberg, F.; Morris, G. M.; Sanner, M. F.; Olson, A. J.; Goodsell, D. S. *Proteins*, **2002**, 46, 34.
- [70] Kastenholz, M. A.; Pastor, M.; Cruciani, G.; Haaksma, E. E.; Fox, T. *J. Med. Chem.*, **2000**, 43, 3033.
- [71] Pastor, M.; Cruciani, G. *J. Med. Chem.*, **1995**, 38, 4637.
- [72] Claussen, H.; Buning, C.; Rarey, M.; Lengauer, T. *J. Mol. Biol.*, **2001**, 308, 377.
- [73] Moreno, E.; Leon, K. *Proteins*, **2002**, 47, 1.
- [74] Bolin, J. T.; Filman, D. J.; Matthews, D. A.; Hamlin, R. C.; Kraut, J. *J. Biol. Chem.*, **1982**, 257, 13650.
- [75] DesJarlais, R. L.; Sheridan, R. P.; Seibel, G. L.; Dixon, J. S.; Kuntz, I. D.; Venkataraghavan, R. *J. Med. Chem.*, **1988**, 31, 722.
- [76] Oshiro, C. M.; Kuntz, I. D. *Proteins*, **1998**, 30, 321.
- [77] Carlson, H. A.; Masukawa, K. M.; McCammon, J. A. *J. Chem. Inf. Comput. Sci.*, **1999**, 103, 10213.
- [78] 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.
- [79] Stultz, C. M.; Karplus, M. *Proteins*, **1999**, 37, 512.
- [80] Di Nola, A.; Roccatano, D.; Berendsen, H. J. *Proteins*, **1994**, 19, 174.
- [81] Mangoni, M.; Roccatano, D.; Di Nola, A. *Proteins*, **1999**, 35, 153.

- [82] Luty, B. A.; Wasserman, R.; Stouten, P.; Hodge, C. N.; Zacharias, M.; McCammon, J. A. *J. Comput. Chem.*, **1995**, *16*, 454.
- [83] Wasserman, Z. R.; Hodge, C. N. *Proteins*, **1996**, *24*, 227.
- [84] Given, J. A.; Gilson, M. K. *Proteins*, **1998**, *33*, 475.
- [85] Nakajima, N.; Higo, J.; Kiderab, A.; Nakamura, H. *Chem. Phys. Lett.*, **1997**, *278*, 297.
- [86] Nakajima, N.; Nakamura, H.; Kidera, A. *J. Phys. Chem. B*, **1997**, *101*, 817.
- [87] Berg, B. A.; Neuhaus, T. *Phys. Rev. Lett.*, **1992**, *68*, 9.
- [88] Pak, Y.; Wang, C. *J. Phys. Chem. B*, **2000**, *104*, 354.
- [89] Zhu, J.; Fan, H.; Liu, H.; Shi, Y. *J. Comput. Aid. Mol. Des.*, **2001**, *15*, 979.
- [90] Caflisch, A.; Fischer, S.; Karplus, M. *J. Comput. Chem.*, **1997**, *18*, 723.
- [91] Trosset, J. Y.; Scheraga, H. A. *J. Comput. Chem.*, **1999**, *20*, 244.
- [92] Noguti, T.; Go, N. *Biopolymers*, **1985**, *24*, 527.
- [93] Trosset, J. Y.; Scheraga, H. A. *P. Natl. Acad. Sci. USA*, **1998**, *95*, 8011.
- [94] Trosset, J. Y.; Scheraga, H. A. *J. Comput. Chem.*, **1999**, *20*, 412.
- [95] Verkhivker, G. M.; Rejto, P. A.; Bouzida, D.; Arthurs, S.; Colson, A. B.; Freer, S. T.; Gehlhaar, D. K.; Larson, V.; Luty, B. A.; Marrone, T.; Rose, P. W. *Chem. Phys. Lett.*, **2001**, *337*, 181.
- [96] Ota, N.; Agard, D. A. *J. Mol. Biol.*, **2001**, *314*, 607.
- [97] Levy, R. M.; Karplus, M. *Biopolymers*, **1979**, *18*, 2465.
- [98] Go, N.; Noguti, T.; Nishikawa, T. *P. Natl. Acad. Sci. USA*, **1983**, *80*, 3696.
- [99] Levitt, M.; Sander, C.; Stern, P. S. *J. Mol. Biol.*, **1985**, *181*, 423.
- [100] Zacharias, M.; Sklenar, H. *J. Comput. Chem.*, **1999**, *20*, 287.

- [101] Keseru, G. M.; Kolossvary, I. *J. Am. Chem. Soc.*, **2001**, *123*, 12708.
- [102] Kolossvary, I.; Guida, W. C. *J. Comput. Chem.*, **1999**, *20*, 1671.
- [103] Kolossvary, I.; Keseru, G. M. *J. Comput. Chem.*, **2001**, *22*, 21.
- [104] Garcia, A. E. *Phys. Rev. Lett.*, **1992**, *68*, 2696.
- [105] Romo, T. D.; Clarage, J. B.; Sorensen, D. C.; Phillips, G. N., Jr. *Proteins*, **1995**, *22*, 311.
- [106] Caves, L. S.; Evanseck, J. D.; Karplus, M. *Protein Sci.*, **1998**, *7*, 649.
- [107] Kitao, A.; Go, N. *Curr. Opin. Struct. Biol.*, **1999**, *9*, 164.
- [108] Amadei, A.; Linssen, A. B.; Berendsen, H. J. *Proteins*, **1993**, *17*, 412.
- [109] Amadei, A.; Linssen, A. B.; de Groot, B. L.; van Aalten, D. M.; Berendsen, H. J. *J. Biomol. Struct. Dyn.*, **1996**, *13*, 615.
- [110] Abseher, R.; Nilges, M. *Proteins*, **2000**, *39*, 82.
- [111] de Groot, B. L.; Amadei, A.; van Aalten, D. M.; Berendsen, H. J. *J. Biomol. Struct. Dyn.*, **1996**, *13*, 741.
- [112] de Groot, B. L.; Amadei, A.; Scheek, R. M.; van Nuland, N. A.; Berendsen, H. J. *Proteins*, **1996**, *26*, 314.
- [113] Teodoro, M. L.; Phillips, G. N., Jr.; Kavraki, L. E. *J. Comput. Biol.*, **2003**, to appear.
- [114] Sandak, B.; Nussinov, R.; Wolfson, H. J. *Comput. Appl. Biosci.*, **1995**, *11*, 87.
- [115] Sandak, B.; Nussinov, R.; Wolfson, H. J. *J. Comput. Biol.*, **1998**, *5*, 631.
- [116] Sandak, B.; Wolfson, H. J.; Nussinov, R. *Proteins*, **1998**, *32*, 159.
- [117] Jacobs, D. J.; Rader, A. J.; Kuhn, L. A.; Thorpe, M. F. *Proteins*, **2001**, *44*, 150.

- [118] Raymer, M. L.; Sanschagrin, P. C.; Punch, W. F.; Venkataraman, S.; Goodman, E. D.; Kuhn, L. A. *J. Mol. Biol.*, **1997**, 265, 445.
- [119] Rarey, M.; Kramer, B.; Lengauer, T. *Proteins*, **1999**, 34, 17.
- [120] Majeux, N.; Scarsi, M.; Apostolakis, J.; Ehrhardt, C.; Caflisch, A. *Proteins*, **1999**, 37, 88.
- [121] Shoichet, B. K.; Leach, A. R.; Kuntz, I. D. *Proteins*, **1999**, 34, 4.
- [122] Zhang, L. Y.; Gallicchio, E.; Friesner, R. A.; Levy, R. M. *J. Comput. Chem.*, **2001**, 22, 591.
- [123] Kollman, P. A. *Chem. Rev.*, **1993**, 93, 2395
- [124] Ajay; Murcko, M. A. *J. Med. Chem.*, **1995**, 38, 4953.
- [125] Oprea, T. I.; Marshall, G. R. *Perpect. Drug Discov.*, **1998**, 9-11, 35.
- [126] Holloway, M. K. *Perpect. Drug Discov.*, **1998**, 9-11, 63.
- [127] Vieth, M.; Hirst, J. D.; Kolinski, A.; Brooks, C. L. I. *J. Comput. Chem.*, **1998**, 19, 1612.
- [128] Tame, J. R. *J. Comput. Aid. Mol. Des.*, **1999**, 13, 99.
- [129] Charifson, P. S.; Corkery, J. J.; Murcko, M. A.; Walters, W. P. *J. Med. Chem.*, **1999**, 42, 5100.