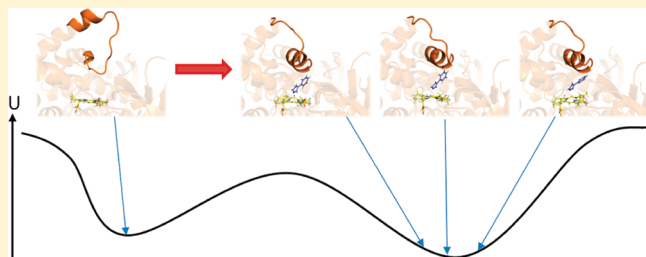


Efficient Incorporation of Protein Flexibility and Dynamics into Molecular Docking Simulations

Markus A. Lill*

Department of Medicinal Chemistry and Molecular Pharmacology, College of Pharmacy, Purdue University, 575 Stadium Mall Drive, West Lafayette, Indiana 47907, United States

ABSTRACT: Flexibility and dynamics are protein characteristics that are essential for the process of molecular recognition. Conformational changes in the protein that are coupled to ligand binding are described by the biophysical models of induced fit and conformational selection. Different concepts that incorporate protein flexibility into protein–ligand docking within the context of these two models are reviewed. Several computational studies that discuss the validity and possible limitations of such approaches will be presented. Finally, different approaches that incorporate protein dynamics, e.g., configurational entropy, and solvation effects into docking will be highlighted.



Molecular recognition between receptors and ligands plays a fundamental role in virtually all biochemical processes in living organisms. In many instances, ligands such as hormones or neurotransmitters are located outside the cell and noncovalently bind to receptors, e.g., G-protein-coupled receptors^{1–3} or ligand-gated ion channel receptors.⁴ This association process results in conformational changes that lead to subsequent signaling events inside the cell. Furthermore, environmental signals are also known to bind to receptors and cause downstream signaling events; for example, odors and volatile substances bind to olfactory and taste receptors and trigger downstream signaling events that are ultimately translated into smells and tastes.^{5,6}

In addition to molecular recognition leading to the downstream signaling effect, substrate recognition by enzymes is a process of central importance that allows the living organism's internal processes as well as interaction with the organisms' environment. For example, compounds that are absorbed from the environment, including drugs, might have potential adverse effects on the organism. Enzymes such as cytochrome P450 bind and chemically modify these compounds to more water-soluble products that subsequently are excreted from the body.^{7,8}

Analyzing examples for molecular recognition as shown in Figure 1 demonstrates that protein flexibility coupled to ligand binding often plays a key role in protein–ligand association.¹⁵ Individual residues in the binding site of the protein change their side-chain conformation to bind different ligands (Figure 1, top panel).^{9,10} Large-scale loop reorganization is observed upon ligand binding to cytochrome P450 119 (Figure 1, middle panel).^{16,17} In the estrogen receptor, the binding of agonists (Figure 1, bottom left) or selective estrogen receptor modulators (SERM) (Figure 1, bottom right) generates different conformations of helix 12 resulting in tissue-specific control of signals.¹⁸

In this review, the induced-fit and conformational-selection biophysical models will be discussed to elucidate the role of

protein flexibility in ligand binding. On the basis of this discussion, different approaches that incorporate protein flexibility into docking will be presented. Additional computational studies will be presented that discuss the validity of the different methods that incorporate protein flexibility during docking. Furthermore, three additional important issues will be discussed that need further consideration in the development of future flexible docking methods: the modeling of protein dynamics, i.e., configurational entropy, the modeling of solvation effects, and the precise quantification of the free energy associated with the protein's conformational change coupled to ligand binding.

MECHANISM OF PROTEIN FLEXIBILITY COUPLED TO LIGAND BINDING

The coupling between protein conformational change and ligand binding is commonly explained by one of two different biophysical models, either the induced-fit¹⁹ (or conformational-induction) or conformational-selection²⁰ (or population-shift) mechanism. In a simplified dynamic energy landscape model (Figure 2), the two mechanisms can be characterized as either a path from the ligand-unoccupied open (UO) state to the ligand-bound closed (BC) state either via the ligand-unoccupied closed (UC) state (in the case of the conformational-selection mechanism; red path in Figure 2) or via the ligand-bound open (BO) state to the ligand-bound closed (BC) state (induced-fit mechanism; green path in Figure 2). The two different mechanisms of protein flexibility coupled to ligand binding, induced fit and population shift,

Received: March 26, 2011

Revised: June 15, 2011

Published: June 16, 2011

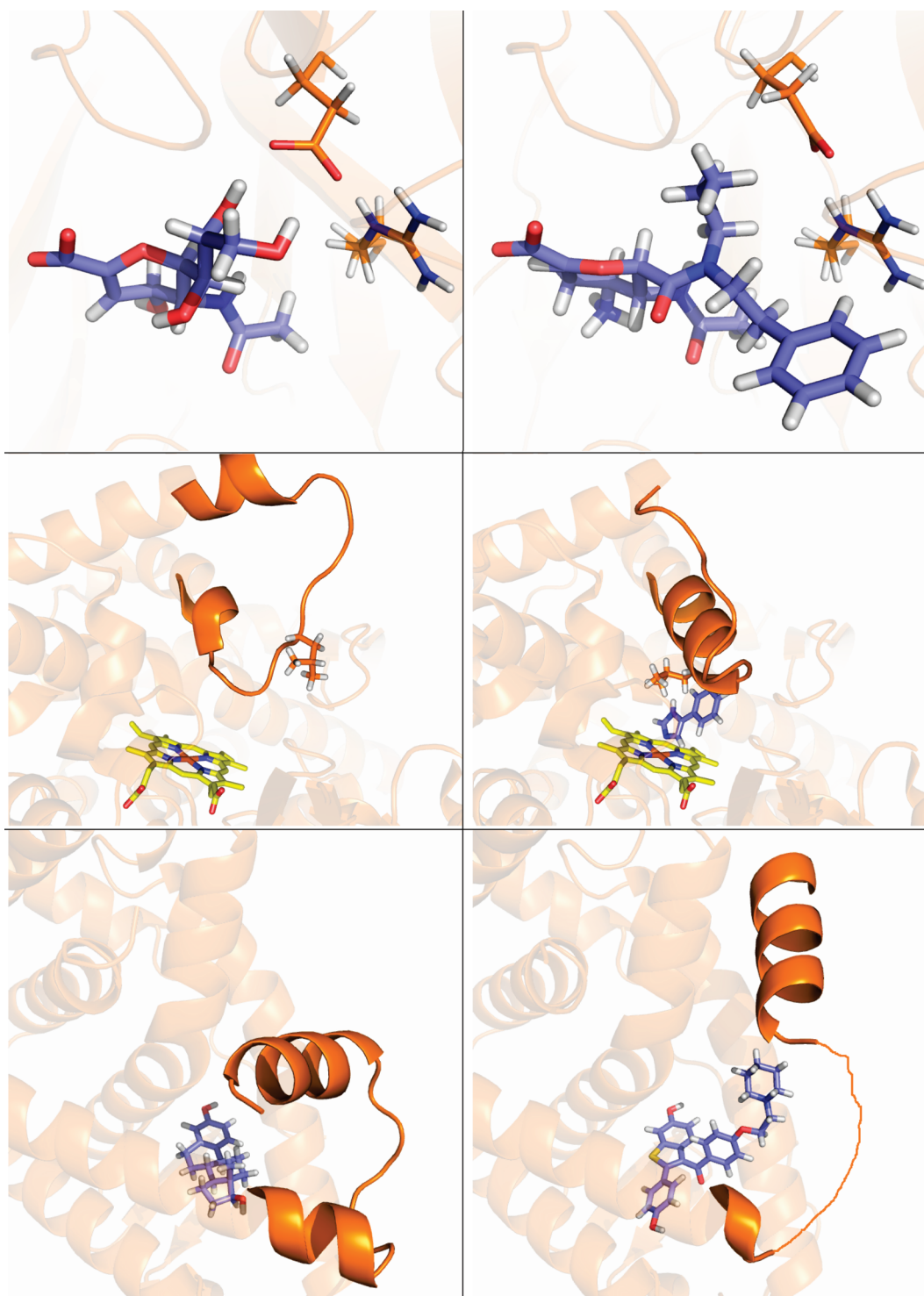


Figure 1. Three examples of different types of protein flexibility coupled to ligand binding. In the top panel, alternative conformations of the side chain of Glu276 are observed forming intermolecular hydrogen bonds between influenza A N9 neuraminidase and the inhibitor 2-deoxy-2,3-dehydro-*N*-acetylneuraminic acid (left, PDB entry 1nnb⁹) and forming intraprotein hydrogen bonds with Arg224 when 5-acetylamino-4-amino-6-(phenethyl-propylcarbomoyl)-5,6-dihydro-4*H*-pyran-2-carboxylic acid is bound (right, PDB entry 1bji¹⁰). In the middle panel, an extended loop in the apoprotein of cytochrome P450 119 (left, PDB entry 1io9¹¹) adapts a new fold upon binding of 4-phenylimidazole (right, PDB entry 1f4t¹²) stabilizing the ligand. In the bottom panel, binding of the agonist 17 β -estradiol (left, PDB entry 1ere¹³) to the ligand-binding domain of the estrogen receptor causes a different conformation of helix 12 compared to the selective estrogen receptor modulator raloxifene (right, PDB entry 1err¹⁴).

relate to the Koshland–Neméthy–Filmer (KNF)²¹ and Monod–Wyman–Changeux (MWC)²² models describing allostery, respectively.

To be able to categorize protein–ligand systems by one of these mechanisms, Okazaki et al.²³ developed a double-basin Hamiltonian quantitatively representing the dynamic energy

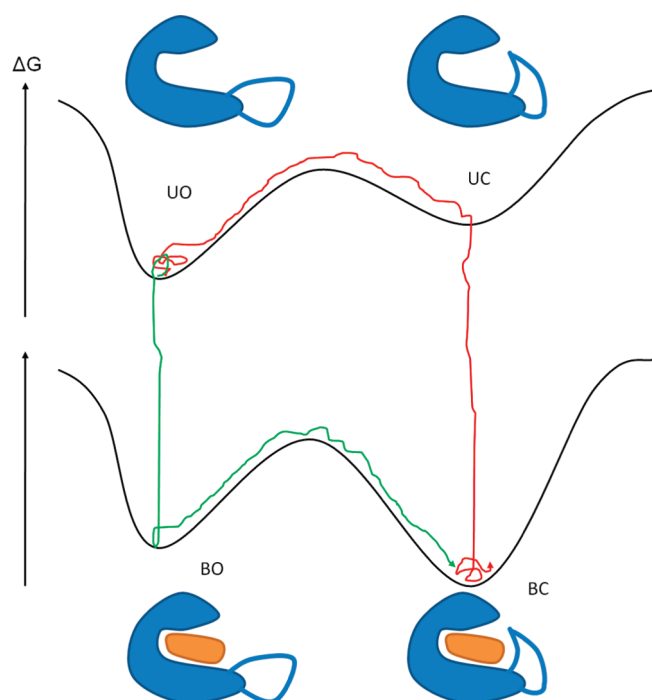


Figure 2. Scheme representing the two mechanisms of protein conformational change coupled to ligand binding. In the conformational-selection mechanism (red line), the apo form of the protein (unoccupied open, UO) adapts to a certain probability a closed conformation (UC) to which the ligand binds and stabilizes the conformation (bound closed, BC). In the induced-fit mechanism (green line), the ligand binds to the apo conformation of the protein (BO). This binding event triggers the conformational change to the final BC state. The top curve shows a model of the free energy of the unbound form of the protein as a function of an arbitrary reaction coordinate describing the conformational change between the open and closed form; the bottom curve displays the associated model of the free energy of the bound form.

landscape model of Figure 2 in which two basins of the Hamiltonian describe the open and closed protein state. A ligand–protein interaction term is added to the Hamiltonian that characterizes the influence of the binding of the ligand to the overall Hamiltonian of the system. From simulations using this Hamiltonian with varying strength and range of protein–ligand interactions, the authors deduced that strong and long-range interactions favor the induced-fit mechanism whereas weak and short-range interactions favor the conformational-selection mechanism. The trend between the ligand–protein interaction strength and the mechanism of conformational change is intuitive and can be interpreted utilizing the free energy landscape model (Figures 2 and 3). If the free energy difference between the apo and holo conformations (ΔG_{OC}) of the ligand-free protein is large (Figure 3, top panel), the unbound protein visits the holo form infrequently. For protein systems in which this situation occurs, strong protein–ligand interactions (ΔG_{PL}) are required to induce and stabilize the holo conformation and drive the overall free energy of the protein–ligand complex to a more negative value versus that of the unbound apo conformation. If the energy difference between the apo- and holoprotein form is small (Figure 3, bottom panel), the holo conformation is frequently visited by the unbound protein and the ligand can bind to this protein conformation. In this case, a weaker protein–ligand interaction is required to stabilize the holo form

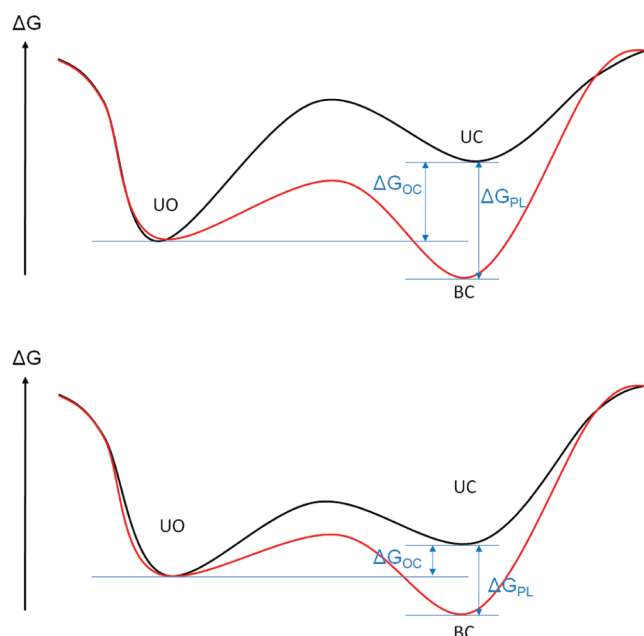


Figure 3. Two different free energy models that characterize protein–ligand systems, likely representing the induced-fit (top) and conformational-selection (bottom) mechanisms of binding, are shown. The black curve models the free energy of the unbound form of the protein as a function of an arbitrary reaction coordinate that describes the conformational change between the open and closed form of the protein. The red curve displays the associated model of the free energy of the bound form of the protein. ΔG_{OC} is the free energy difference between the open and closed state of the ligand-free protein, and ΔG_{PL} is the stabilizing free energy of the protein upon ligand binding. The free energy ΔG_{PL} contains many contributions, including the direct protein–ligand interactions typically affected by water screening, hydrophobic association, and the conformational and vibrational entropy of protein and ligand.

of the protein. However, direct interactions between protein and ligand atoms, such as hydrogen bonds or electrostatic interactions, are not always the dominant contributors to ΔG_{PL} . The hydrophobic association between protein and ligand moieties often dominates the free energy of binding.^{24–27} Until recently, hydrophobic interactions were considered to be always entropically driven.²⁸ In this model, water molecules form ordered water shells²⁹ around hydrophobic ligand moieties or amino acids creating a highly ordered water structure that is entropically unfavorable compared to the structure of bulk water with its highly dynamic hydrogen bond network. As the ligand binds to the protein, the water molecules in the shell surrounding the hydrophobic moieties of the ligand and the binding site of the protein are released into the bulk solvent and gain translational, rotational, and vibrational entropy, thereby lowering the free energy of the protein–ligand complex. However, recent studies^{26,27,30,31} demonstrated that hydrophobic interactions can also be enthalpically driven. Using a model system for hydrophobic receptor–ligand association, McCammon and co-workers^{26,27} found that disordered water molecules with a density lower than the bulk density bind to small but solvent-accessible hydrophobic cavities. When the ligand binds, these disordered water molecules are released into the bulk solvent and actually lose entropy. The gain of water–water interactions is the major contributor to the net change in enthalpy that overcompensates for the loss of entropy.

To separate the two different mechanisms of protein flexibility coupled to ligand binding, experimental techniques such as X-ray crystallography, NMR experiments, and kinetic measurements are utilized to identify holo-like protein conformations preexisting in the ligand-free state of the protein,^{32–34} in agreement with the population-shift mechanism, or to identify intermediate states³⁵ that potentially point toward an induced-fit mechanism. Experiments also suggested the coexistence of both the population-shift and induced-fit mechanisms for ligand binding to a protein.^{36–39} For example, X-ray crystallography and kinetic measurements of ligand binding to antibody SPE7 showed that ligands bind to preexisting conformations of SPE7 followed by an induced fit generating the final high-affinity protein–ligand complex.⁴⁰ Also, single-molecule AFM measurements showed that protein–protein binding between a nuclear transport effector, RanBP1, and a nuclear import receptor, importin β , can follow the population-shift or induced-fit mechanism, depending on which additional substrate binds to importin β .³⁶

■ COMPUTATIONAL METHODS FOR SIMULATING PROTEIN FLEXIBILITY COUPLED TO LIGAND BINDING

Over the past few decades, many computational concepts that incorporate different levels of protein flexibility into the modeling of protein–ligand association have emerged. Many methods are used to study individual protein–ligand complexes, such as molecular-dynamics (MD) or Monte Carlo (MC) simulations. Such methods can be used to locally sample energetically accessible substates of the protein–ligand complex's free energy landscape. Algorithms such as free energy perturbation (FEP)⁴¹ or thermodynamic integration (TI)⁴² in combination with MD or MC simulations can calculate the relative free energy of compounds binding to the same drug target within 1–2 kcal/mol of the experimental value. Unfortunately, the procedures associated with FEP and TI are computationally demanding and are typically limited to the comparison of structurally similar compounds. In addition to being computationally expensive, the application of these techniques requires a priori knowledge of the configuration of the ligand and protein in their bound form.

Recently, standard FEP and TI protocols have been extended to efficiently incorporate conformational changes associated with ligand binding.^{43,44} Using umbrella sampling in conjunction with the “confine-and-release” framework, Mobley et al.⁴³ accurately predicted the free energy of a ligand binding to a designed binding site in T4 lysozyme that involves a conformational change of a Val residue. In a separate study, Lawrenz et al.⁴⁴ developed an independent-trajectory thermodynamics-integration (IT-TI) method that computes absolute and relative binding free energies using multiple independent trajectories to improve configurational sampling. In their study on peramivir binding to H5N1 avian influenza virus neuraminidase, IT-TI was utilized to extensively sample the phase space accessible to the flexible loop regions that interact with the ligand. IT-TI was designed for use with distributed computing and consequently allows a significantly more rapid computation of absolute free energies of binding compared to standard TI or FEP calculations for a series of ligands.

An alternative method designed to predict the configuration and binding affinity of a ligand for a specific receptor is receptor–ligand docking. Docking methods are less computationally demanding than simulation-based free energy methods and are used to virtually screen large libraries of existing or

hypothetical chemicals with the goal of identifying new potentially active chemotypes that bind to the target receptor. Throughout the docking process, many different ligand orientations and conformations (binding poses) are generated in the binding site of the protein using a search algorithm. The free energy of binding of each pose is subsequently estimated using a scoring function.

For some target proteins, the search algorithm of a docking program may generate bioactive conformations (rmsd < 2 Å) throughout the search process for up to 90% of all ligands; for other proteins, this percentage can be as low as 40%.⁴⁵ However, existing scoring functions used in docking are not sufficiently accurate to reliably estimate the free energy of binding of a protein–ligand complex. Thus, the bioactive pose is not always energetically ranked top among all poses generated throughout the search process.⁴⁵ The correlation between experimental and predicted binding affinities using scoring functions for a series of compounds binding to the same protein is usually weak and often influenced by considerations of ligand size rather than by the correct modeling of the underlying physicochemical contributions to binding affinity.^{46,47}

The limitations of docking programs described above originate from the necessity of finding a balance between accuracy and speed to screen large ligand libraries. Consequently, scoring functions quantify a simplified representation of the full protein–ligand interaction by typically including only elements such as hydrogen bonds and hydrophobic contacts and typically neglecting effects such as solvation, polarization, and entropy. The individual terms of the scoring function are weighted by prefactors that are optimized using a training set of protein–ligand complexes. As a consequence of the simplified representation of the physics of protein–ligand interactions and the optimization procedure, the success of scoring functions for binding-pose prediction and virtual screening is known to be dependent on the protein system.⁴⁵ In addition to using a simplified scoring function to reduce the computational time required to screen ligand libraries, only critical degrees of freedom of the ligand are considered during the search algorithm to limit the conformational space that must be sampled (i.e., translation, rotation, and torsional rotations of ligand). On the basis of the need to keep the computational costs low, protein flexibility is commonly only partially incorporated into docking, e.g., side-chain flexibility. Considering only side-chain flexibility in the binding site, typically containing between 10 and 30 residues, there are approximately 20–70 rotatable bonds that represent the protein's degrees of flexibility. The computational cost significantly increases if additional degrees of freedom, representing the backbone motion, of the protein are incorporated.

■ CONCEPTS THAT INCORPORATE PROTEIN CONFORMATIONAL CHANGES INTO RECEPTOR–LIGAND DOCKING

As several reviews that discuss existing methods and software used to model conformational changes during docking have been published,^{48–54} it is not the aim of this review to produce a detailed description of each method. Instead, general concepts that include protein flexibility during docking are highlighted in the context of their relation to the two biophysical mechanisms of protein flexibility coupled to ligand binding, the induced-fit versus conformational-selection mechanism.

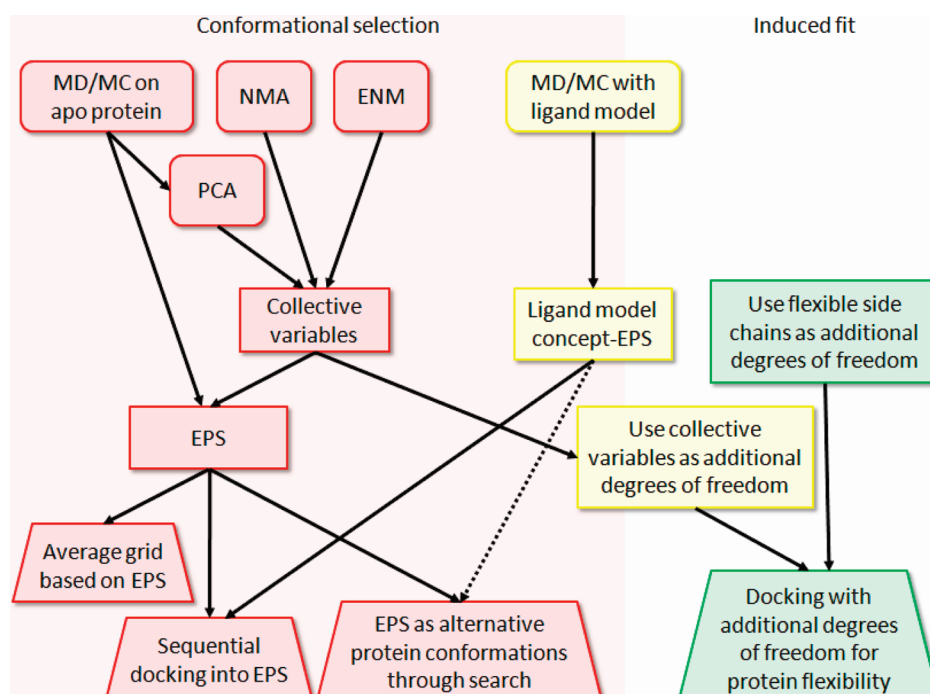


Figure 4. Overview of methods discussed in this review that incorporate protein flexibility in docking simulations. Methods that can be assigned to the conformational-selection mechanism of protein flexibility are colored red, methods utilizing the induced-fit mechanism green, and methods that have elements from both mechanisms yellow. The different sources of protein flexibility for docking are shown as rectangular shapes, the methods for generating these sources as rounded rectangles, and the different concepts for using these sources in the docking algorithm as trapezoids. Abbreviations: MD, molecular-dynamics simulations; MC, Monte Carlo simulations; NMA, normal-mode analysis; ENM, elastic-network model; PCA, principal-component analysis; EPS, ensemble of protein structures.

In general, algorithms can be divided into two categories (Figure 4). The first class of methods generates an ensemble of protein structures (EPS) prior to docking. On the basis of the ideas of conformational selection, each ligand in the library is then subsequently docked to each member of the EPS. The second category of methods generates alternative protein conformations on the fly in parallel to the pose generation of the ligand during docking. In this concept, additional degrees of freedom are selected that describe important contributions of protein flexibility. Variables that characterize the magnitude of the conformational change along the selected degrees of protein flexibility are optimized in parallel to the ligand degrees of freedom, such as translational, rotational, and torsional changes. As the accessible conformational space to the protein–ligand system scales, in the worst case, as polynomial with the number of degrees of freedom, the degrees of protein flexibility must be selected very carefully. This approach can be interpreted as an implementation of the induced-fit idea into docking.

Docking with Conformational Change Represented by Additional Degrees of Freedom. Many docking programs allow side chains of selected binding-site residues to be flexible throughout the search process. Alternative side-chain conformations are generated either by an exhaustive search,^{55,56} by selecting states from a pregenerated rotamer library,^{57–59} or by a minimal change in conformations to resolve steric clashes between protein and the current ligand binding pose.⁶⁰ The estimation of free energy changes between different protein conformations is an important issue when including protein flexibility. The energetic cost associated with torsional changes of side chains is typically estimated in one of two ways. The

interaction between the flexible residues and the rest of the protein is computed using the same scoring function used to estimate the protein–ligand interactions,^{61,62} or an empirical energy function is used that is proportional to the size of torsional changes.⁶⁰ The latter estimation assumes that only small changes in the torsion angles are observed.

The question of whether the inclusion of selected side chains is sufficient to accurately model protein flexibility in docking arises. For protein systems such as neuraminidase (shown in Figure 1, top panel) this may be the case, but for other systems [e.g., HIV protease, cAMP-dependent protein kinase, CYP450 (Figure 1, middle panel) or estrogen receptor (Figure 1, bottom panel)], protein flexibility beyond the side-chain level plays a significant role during ligand binding.

To incorporate protein-backbone flexibility into docking, it is possible to include conformational fluctuations of the protein structure in terms of collective motions, or “degrees of freedom”. These collective degrees of freedom are then included as additional searchable variables in the binding-pose optimization procedure. Zacharias and co-workers derived collective degrees of freedom by performing normal-mode analysis (NMA) originally using a full atomistic representation of the protein.⁶³ NMA is based on the assumption that protein motion can be approximated by a combination of harmonic vibrations. In NMA, the Hessian matrix containing the partial second derivatives of the potential energy with respect to the atom coordinates is diagonalized. The resulting eigenvectors describing the lowest-frequency modes characterize the slow collective or large-scale motions of the protein. To increase the computational efficiency, Zacharias and co-workers later used the collective variables

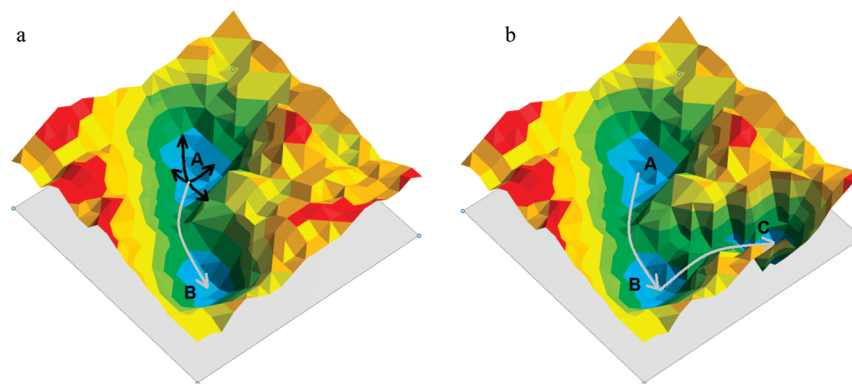


Figure 5. (a) Two-dimensional model of a free energy landscape of a protein system with the apo structure represented by energy minimum (A) and the holo structure by energy minimum (B). On the basis of the encoding hypothesis, protein motion can be reduced to collective variables (black arrows). A linear combination of these collective variables allows the transition over low-energy barriers (gray arrow) to the holo structure of the protein. (b) If the holo structure (C) and apo structure (A) are separated by another stable conformation (B), the apo structure may only encode the transition to state B and ligand binding may be necessary to stabilize state B and induce a conformational change to holo structure C.

derived from a NMA of an elastic-network model (ENM).⁶⁴ In an ENM, the residues of the protein are represented by single three-dimensional points that are connected by mechanical springs and the fluctuations of the protein are governed by the difference in local particle density. An alternative approach used to reduce protein dynamics to a few collective degrees of freedom is principle-component analysis (PCA) of a protein trajectory obtained from a MD simulation.⁶⁵

The previously described procedures utilize the potential function of the ligand-free form of the protein to determine the collective degrees of freedom of the protein. Such methods assume that the collective degrees of freedom encode the protein fluctuations necessary for ligand binding to occur. This encoding hypothesis^{66–68} is an implication of the conformational-selection model assuming that the fluctuations in the apoprotein structure trigger frequent transitions to the holo forms of the protein (Figure 5a). To test this hypothesis, Ikeguchi et al.⁶⁹ used linear-response theory where ligand binding was modeled as an external perturbation to the atomic fluctuations of the apo form of the protein. They demonstrated that upon ligand binding, the overall large-scale change (up to 15 Å C_α displacement between apo and holo conformations) of three different protein systems (ferric-binding protein, citrate synthase, and F₁-ATPase) could be predicted on the basis of five collective modes. However, differences on the order of 2–5 Å between the predicted and experimental holo structure were observed in some portions of each protein system.

In another study to test the encoding hypothesis, Cukier applied PCA analysis to an apo structure of an adenylate kinase known to undergo large-scale conformational changes in the LID and AMP binding domain to form a closed form of the binding site when the substrates AMP and Mg²⁺-ATP are bound.⁶⁶ The authors observed that the conformational change of the LID domain was encoded in 12 collective modes of the apo form, whereas the additional AMP-induced conformational change in the AMP binding domain was not encoded in the collective modes. These examples demonstrate that holo-like protein conformations can be generated using collective degrees of freedom from the apoprotein for many systems, but that additional induced fit may be necessary to predict a holo-like protein conformation in some protein systems (cf. Figure 5b).

One issue with many of the studies used to validate the ability of the collective-mode approach to generate holoprotein structures was that the experimental end state of the protein is known a priori (open and closed forms of the protein were obtained beforehand from experimental studies). Without this structural information, we are faced with the question of how the different collective modes should be weighted and what magnitude of conformational change should be expected.

The weighting of the collective modes and the magnitude of conformational change to be expected from them directly relate to the energy difference required to undergo the transition between the apo and holo forms of the protein (represented as ΔG_{OC} in Figure 3). In an attempt to quantify this energy, May et al.⁶⁴ estimated the energy needed to move the protein between states as an empirical fourth-order function of the magnitude of deformation along each collective mode *m*:

$$V_{\text{deform}} = \sum_{m=1}^M \kappa_m |R_m - R_m^0|^4 \quad (1)$$

where $|R_m - R_m^0|$ is the magnitude of conformational change along the *m*th collective mode and κ_m is the square of the eigenvalue of mode *m*. This empirical quantification approach was successfully applied to cyclin-dependent kinase 2 cross-docking studies. The experimental holo structures for the ligands used in the cross-docking studies differ from the apo form by rather small shifts of secondary structural elements near the binding site (backbone rmsd of 0.7–2.2 Å). However, it is not obvious how this quantification of ΔG_{OC} can be applied to protein systems known to undergo large-scale conformational changes between the apo and holo forms and are characterized by two local minima separated by an energy barrier (see Figure 5). In this situation, the energetic penalty used in the study of May et al.⁶⁴ would constantly increase with the deviation from the apo form in contrast to the free energy profile. In another attempt to quantify the free energy difference associated with conformational shifts in the loop regions, Cavaotto et al.⁷⁰ used a sophisticated scoring procedure comprised of force-field energies and an implicit solvation model for the protein region of interest, i.e., the residues in the proximity of the binding site. Even more difficult than estimating the free energy associated with conformational

shifts of secondary-structure elements is accurately quantifying the relative free energies of large-scale conformations that involve partial refolding of the protein structure, such as unstructured protein regions in the absence of the ligand or significant refolding of loop conformations. As the difficulties of this problem are related to those of *ab initio* protein structure prediction, an accurate estimation (1–3 kcal/mol) of the free energy of such protein conformational changes is currently unlikely to be achieved with current scoring procedures.

An important question concerning the use of collective variables is the number of degrees of freedom required to generate a holo-like protein structure. The answer to this question certainly depends on the required level of accuracy. Pande and co-workers⁷¹ applied NMA on four different protein systems (myosin, calmodulin, NtrC, and hemoglobin) that undergo significant conformational changes (2.0–15.1 Å rmsd between experimental apo and holo forms) during ligand binding. Using the 20 most significant normal modes, less than 50% of the observed conformational change (measured by the rmsd between the projected and experimental target structure) was reproduced and deviations between the predicted and experimental target structure ranged from 2 to 8 Å. To achieve an rmsd of 2 Å between the predicted and experimental structures, hundreds of normal modes were required. In an attempt to limit the number of additional degrees of freedom required to describe backbone motion, new approaches were recently devised to project the collective variables on important regions of the protein, or individual atoms.^{66,70}

Docking to a Pregenerated Ensemble of Protein Conformations. NMA, PCA, or ENM can also be used to generate an ensemble of protein structures (EPS) of alternative templates for docking (Figure 4). The simplest way to include the EPS is to perform sequential docking to all members of the ensemble, linearly increasing the required computational time with the size of the EPS. To accelerate the docking process, the EPS alternatively can be integrated into the docking algorithm as alternative states accessible by the protein–ligand system during the pose generation phase^{72–74} or can be combined in an average interaction energy grid used throughout docking.⁷⁵ A detailed discussion of different means of handling the EPS throughout docking can be found in a review by Totrov and Abagyan.⁵⁰

Whereas the previously discussed methods focus on medium- to large-scale conformational changes of the protein, it has been recognized that small-scale changes (1–2 Å rmsd) are often critical for successful docking.^{76–78} As an alternative to using collective variables to generate EPS, MD simulations can be conducted on the protein system. Whereas large-scale conformational changes are typically missed using a standard MD simulation, this approach has the ability to include small-scale backbone and side-chain fluctuations in the EPS.

In ensemble docking studies,⁷⁹ it has been recognized that using an EPS with only a few protein conformations can increase the success rate for correctly predicting binding poses and enrichment in virtual screening. Using a very large EPS, however, leads to a poorer docking performance because of the generation of a large number of false positives, thereby reducing the enrichment rate in virtual screening. Thus, an important consideration when using EPS for docking is how many and which protein conformations should be utilized.

Amaro et al.⁸⁰ utilized different approaches, e.g., hierarchical clustering based on pairwise rmsd between MD snapshots, to reduce the initial EPS generated by MD simulations to a

structurally diverse subset of protein structures representing the configurational space sampled by MD. This procedure allowed the authors to reduce the initial size of the EPS by 90–99% without any observable loss of accuracy in virtual screening experiments. Bolstad et al.⁸¹ developed a method for selecting protein conformations from the EPS on the basis of their ability to conserve the relative orientations of a core of amino acids critical for binding. Using dihydrofolate reductase (DHFR) as a test case, atoms of binding-site residues that have a conserved relative position in various DHFR cocomplexes with ligands were manually identified. Distances between these protein atoms were computed for all MD snapshots and compared with the conserved distances in the X-ray structures. Only MD snapshots that approximately preserve the interatomic distances between the selected protein atoms were chosen as members of the EPS used for docking. This procedure allowed the reduction of the initial EPS by 50–75%, increasing the efficiency and ranking accuracy of ensemble docking.

Armen et al.⁸² studied the influence of different levels of modeled protein flexibility on cross-docking accuracy using p38 α MAP kinase that displays significant side-chain and loop flexibility in the binding site. The level of protein flexibility was defined by the number and size of segments of nonrestrained atoms or torsions in the MD simulations utilized to generate alternative template structures, ranging from a rigid protein structure over flexible side chains and flexible loop regions to the fully flexible protein for which no restraints on protein atoms or torsions are applied. They found that limiting the modeled protein flexibility to the fewest degrees of freedom necessary to adequately represent the experimentally observed flexibility (i.e., binding-site side chains and two flexible loops in the studied test case) showed superior cross-docking performance compared to using rigid or fully flexible proteins as template structures for docking. The authors reasoned that fully flexible protein models exhibited decreased docking performance because incorporating unnecessary degrees of protein flexibility or alternative protein structures into docking increased the potential for generating many non-holo-like protein conformations or protein–decoy complex structures that were highly ranked because of insufficient scoring functions.

Independent of the approach used to generate the EPS, each method is based on the conformational-selection model. However, it must be noted that existing computational methods and infrastructure typically allow the protein to sample only a portion of the configurational space accessible to the protein system. In other words, even if the population-shift mechanism is the dominant mechanism of protein flexibility associated with ligand binding, can we actually identify the holo conformation by computational means? Xu et al.⁷⁸ recently performed docking to an EPS generated by long MD simulations on two different protein systems that involve only small-scale conformational changes between the apo and holo structures (rmsd < 1 Å). Despite the long simulation time, docking to an EPS generated by short MD simulations with bound ligands clearly outperformed docking to the EPS generated by long MD simulations on the apo form of the proteins. Two possible conclusions were drawn from the results. Either there are too many alternative protein conformations generated by the long MD simulations, making the identification of the relevant holo-like structures difficult and ultimately causing many false positive poses, or ligand binding was required to induce the holo conformation of the protein.

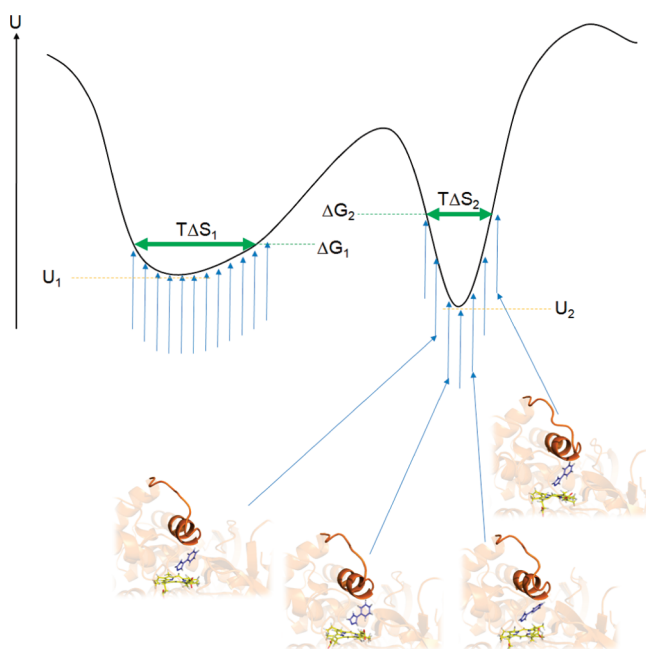


Figure 6. Model of the potential energy U of a protein–ligand system of two different binding modes corresponding to the two energy minima. For each binding mode, the protein–ligand system samples slightly different configurations (blue vertical arrows and four selected protein–ligand snapshots of one binding mode). The left local minimum with a shallow shape (minimum potential energy U_1) allows for more flexibility of the protein–ligand system and a smaller loss of configurational entropy compared to the steeper right minimum (minimum potential energy U_2) and consequently has a lower free energy ΔG_1 than the right minimum (ΔG_2). Neglecting configurational entropy would cause the binding mode corresponding to the right minimum to be predicted as the top-ranked binding pose in disagreement with the real free energy values. The vertical blue arrows represent possible protein–ligand conformations at a given temperature T . The horizontal green arrows represent the remaining configurational entropy of the system in the two different binding modes, $T\Delta S_1$ and $T\Delta S_2$.

The docking results for the apo MD ensemble led Xu et al.⁷⁸ to the development of the ligand-model concept that is capable of sampling protein conformations that are relevant for binding structurally diverse ligands. In this method, MD simulations are performed with a dynamically changing set of restrained functional groups in the binding site of the protein, essentially representing a large hypothetical ensemble of different chemical species binding to the same target protein. Beginning from an apo structure, the ligand-model approach was used to derive an EPS used for docking, and the results outperformed docking to an EPS generated from an apo-MD simulation. Furthermore, the method was only slightly less successful than docking to the experimentally known holo form of each individual protein–ligand complex.

All four previously discussed studies demonstrated that utilizing a large EPS generated from an apo MD simulation significantly increased the required computing time without necessarily improving the docking performance. A reduction of the EPS to a small set of protein structures relevant for ligand binding or the reduction of protein flexibility to the smallest relevant number of degrees of freedom has been shown to increase not only the efficiency but also the accuracy of ensemble docking.

CONCEPTS THAT INCORPORATE PROTEIN DYNAMICS INTO RECEPTOR–LIGAND DOCKING

As mentioned previously, a significant problem in current docking protocols is the failure to accurately score the different docking poses and different ligands. Even if the docking study incorporates all inherent conformational changes coupled to ligand binding, for example by retrospective docking using the experimentally determined holoprotein structure for each ligand, often no correlation between experimental and predicted binding affinity is observed.⁴⁶ An important factor contributing to the failure of docking to accurately predict binding affinities is the lack of inclusion of dynamic information about the protein–ligand complex. Even if conformational changes of the protein are included, the docking-predicted free energy of binding is generally based on a single protein–ligand structure. However, in reality, the protein–ligand complex samples local substates in the conformational vicinity of a given binding mode (Figure 6).

From statistical mechanics, the free energy of binding ΔG° at standard concentration C° can be determined by integrating over all configurations accessible to the protein–ligand complex, and the protein and ligand in their unbound state:^{83,84}

$$\Delta G^\circ = -RT \ln \left(\frac{C^\circ}{8\pi^2} \frac{\int \exp\{-\beta[U(r_{PL})]\} dr_{PL}}{\int \exp\{-\beta[U(r_P)]\} dr_P \times \int \exp\{-\beta[U(r_L)]\} dr_L} \right) \quad (2)$$

where U_{PL} , U_P , and U_L are the potential energies for the protein–ligand complex, protein, and ligand, respectively, as a function of their internal coordinates. The potential energy includes the solvation of the individual entities. Via reformulation of eq 2, the free energy of binding can be calculated, for example, using the molecular-mechanics Poisson–Boltzmann or generalized Born surface area (MMPBSA/GBSA)^{85,86} method to determine the average potential energy of the protein–ligand complex, the free protein, and the unbound ligand and the difference in configurational entropy:

$$\Delta G^\circ = \langle U_{PL} \rangle - \langle U_P \rangle - \langle U_L \rangle - T\Delta S_{\text{config}} \quad (3)$$

The average potential energy terms given in eq 3 are determined using an implicit solvation model, either PBSA or GBSA, for example, and thus include a simplified representation of entropic contributions of solvation typically represented by a cavity term ΔG_{cav} that is proportional to the difference in the solvent-accessible surface between the bound and unbound form of the protein–ligand complex. ΔG_{cav} represents the entropy penalty associated with the reorganization of water molecules around the solutes.⁸⁶ The ensemble of protein–ligand complex, free protein, and unbound ligand configurations used in the averaging process are typically sampled by means of MD or MC simulations.

In an attempt to rapidly predict native binding poses and rank libraries of compounds, scoring functions used in docking neglect the averaging process over multiple protein–ligand conformations. Entropic contributions to ligand binding are incorporated into scoring functions in simplified forms. For example, terms describing hydrophobic contacts between the protein and the ligand are used to model changes in solvation

entropy upon protein–ligand binding;^{87,88} a count of the number of rotatable bonds is utilized as a measure of the conformational restriction of the bound ligand.^{61,89} Changes in vibrational, translational, and rotational entropy upon ligand binding, however, are typically neglected. Furthermore, changes in the configurational entropy of the protein upon ligand binding are typically neglected in standard scoring functions.^{90,91} Nevertheless, the entropic contributions associated with protein–ligand binding commonly neglected in today's scoring functions are important contributions to the binding affinity, and current research tries to develop extensions to current scoring schemes that try to address such factors.

Postprocessing of Docking Poses. One category of methods, named postprocessing, incorporates the dynamic information of the protein–ligand system after the docking process has been completed. The top-scored binding pose or several low-scored poses are used as input for subsequent MD or MC simulations. In combination with free energy methods such as FEP, TI, MMPBSA/GBSA, or linear interaction energy analysis (LIE),⁹² a more accurate estimation of the free energy of binding is possible.⁹³ This post-processing step can significantly improve the successful prediction of binding affinities.^{93,94} However, this process is relatively time-consuming and therefore requires that the bioactive binding mode be within the top-ranked binding poses for each ligand.

Modeling of Configurational Entropy in Docking. Methods in a second category try to directly estimate the configurational entropy or averages over the predicted free energy of similar protein–ligand configurations generated throughout docking. Chang et al.⁹⁵ utilized a simple cluster size method to estimate the configurational entropy of each binding mode. For each pose generated by the docking program, all conformationally related poses (rmsd < 2 Å) are identified. The number of poses within this rmsd range is used as a measure for accessible configurational entropy of the protein–ligand system at a particular local minimum. The underlying assumption is that the search algorithm of the docking software (AutoDock4) more easily identifies binding poses in wide energy wells compared to narrow wells. On the basis of this assumption, the probability of identifying similar binding poses correlates with the width of the energy well and the accessible configurational entropy of the ligand in the binding site of the protein. A similar estimate of configurational entropy has also been implemented in a knowledge-based scoring function.⁹⁶

Ruvinsky et al.^{97,98} clustered the ensemble of binding poses on the basis of their pairwise rmsd values to estimate the configurational entropy of each unique binding mode. The entropy of each unique binding mode was calculated by measuring the variance in translational (measured by relative translational coordinate r), rotational (measured by Euler angles Φ_1 , Φ_2 , and Φ_3), and torsion (torsion angles Ω_k) values among the members of a cluster:

$$\Delta S = \Delta S_0 - k_B T \ln \left\{ \prod_{i=1}^3 [\max(r_i) - \min(r_i)] \times \prod_{j=1}^3 [\max(\Phi_j) - \min(\Phi_j)] \times \prod_{k=1}^{tor} [\max(\Omega_k) - \min(\Omega_k)] \right\} \quad (4)$$

The underlying concept of this method is that binding modes that are restrained within the binding site (i.e., smaller variation

in translational, rotational, and torsion variables) yield a larger loss of entropy upon binding to the protein than less restrained binding modes (Figure 6) and are energetically less favorable.

Applying the previously described methods to estimate the configurational entropy of 22 ligands binding to APS reductase⁹⁵ produced a significant improvement in identifying native binding poses as the highest ranked for all utilized entropy measures. Whereas the simple cluster-size method relies on the details of the search algorithm throughout docking, the method of Ruvinsky et al.^{97,98} was conceptually designed to represent characteristics of the underlying free energy landscape of a given binding mode. However, analysis revealed⁹⁵ that the results of the latter method were strongly dependent on the number of poses generated in a binding-mode cluster and therefore primarily correlate with cluster size rather than the underlying free energy landscape.

Unwalla and co-workers⁹⁹ derived a scoring method, titled the partition function-based scoring (PFS) method, that includes an estimate for the change in ligand entropy upon binding:

$$PFS = -k_B T \ln \left[\sum_{i=1}^{N_{\text{bound poses}}} \exp(-\beta \times f \times \text{score}_i) \right] + k_B T \ln N_{\text{unbound conformations}} \quad (5)$$

The first term includes estimates for the remaining conformational, positional, and rotational entropy of the bound ligand, and the second term estimates the conformational entropy of the unbound ligand; $N_{\text{bound poses}}$ is the number of distinct binding poses identified by the docking program, and $N_{\text{unbound conformations}}$ is the number of ligand conformations within 5 kcal/mol of the lowest identified ligand conformation. An initial study using PFS⁹⁹ suggested that this method can improve ranking of compounds with similar number of rotatable bonds, a benefit that is inaccessible to the frequently used approximation that adds a constant entropy penalty per rotatable bond. In addition, the study also suggested that the PFS method would be more accurate if the scoring function accurately mimics the form of the free energy landscape. However, a simple scaling of existing scoring functions (factor f in eq 5) is unlikely to provide such a mimic.

Compared to the previously discussed method for incorporating protein dynamics into docking, the Mining Minima approach^{100–103} is computationally more expensive. Mining Minima aims to directly calculate the configurational integrals in eq 2; e.g., $Z = \int \exp\{-\beta[U(r_{\text{PL}})]\} dr_{\text{PL}}$ for the protein–ligand complex. Local minima corresponding to different binding modes are identified throughout the conformational search procedure. The eigenvectors corresponding to the largest eigenvalues of the Hessian matrix in bond–angle–torsion coordinates are determined. Protein–ligand conformations are generated along the eigenvectors and used to compute the local configurational integral Z_i for local minimum i . The overall free energy of a compound is then computed from the sum of local configurational integrals: $\Delta G = -k_B T \ln \sum_i Z_i$. This method was successfully applied to binding affinity predictions of ligands binding to HIV-1 protease and phosphodiesterase.¹⁰³

Another successful approach that incorporates protein dynamics into docking is the relaxed complex scheme (RCS).^{104–106} In this scheme, MD simulations of the apo form of the protein are used to generate an EPS. Sequential docking to the EPS is then performed, and for each ligand, similar binding poses are clustered

across all EPS protein templates and the score of all members of a cluster is averaged. Using the average score of each unique binding mode was shown to provide a more accurate estimation of binding free energies than using individual scores.^{106,107} Whereas the original RCS scheme is based on the conformational-selection model, Xu et al.⁷⁸ extended the RCS approach to allow the incorporation of the contribution of induced fit to protein flexibility. In combination with the previously discussed ligand-model concept, the RCS scheme was successfully applied to predict binding modes and affinities of structurally diverse compounds.

Incorporating Solvation Effects into Docking. As discussed previously in this review, the change in solvation entropy plays a fundamental role in the strength of protein–ligand binding. In empirical scoring functions, desolvation effects are typically represented by a term characterizing hydrophobic contacts between the protein and the ligand.^{87,88} In scoring functions that are based on physical terms, desolvation of nonpolar atoms is often approximated by a function that is proportional to the solvent-accessible surface.⁶¹ A similar term has also been utilized for knowledge-based scoring functions.⁹⁶ The electrostatic component of desolvation that is dominant for polar groups can be approximated using an implicit solvation model by means of the Poisson–Boltzmann (PB) or Generalized-Born (GB) methodology. These methods, however, are too time-consuming for use in docking methods applied to virtual screening. To improve the efficiency of implicit solvation models in docking, Mysinger et al.¹⁰⁸ recently developed a concept that computes the fractional desolvation of a ligand atom inside of the protein based on the protein's environment surrounding the atom. The energy associated with the atom's transfer from a high- to low-dielectric medium (i.e., water to protein) is then weighted by the fractional desolvation of the atom; this energy estimates the desolvation free energy of each ligand atom. The overall desolvation energy of the ligand is computed by summing over all atomic desolvation contributions.

The previously discussed approaches are based on a continuum approximation of the solvent and do not include the effects associated with the directionality of water-mediated hydrogen bonds; the entropic contribution of desolvation is approximated by a solvent accessibility term. Directional hydrogen bonds, however, can play an important role in mediating polar interactions between solute and solvent.¹⁰⁹ Docking programs, e.g., GOLD,¹⁰⁹ have been modified to incorporate water-mediated interactions into docking. Water molecules are switched on and off and are allowed to rotate throughout the docking process. To more accurately estimate desolvation energies, Abel et al.¹¹⁰ have developed the WaterMap concept. Explicit protein–water simulations are performed; water sites are identified on the basis of conserved water positions throughout the MD simulation, and the enthalpy and entropy of each water site relative to bulk water are computed using the MD trajectory and inhomogeneous solvation theory.^{111,112} The contribution of desolvation to the binding free energy has been successfully predicted using WaterMap by considering the excess enthalpy and entropy of the water molecules in the binding site that are replaced by ligand binding.^{110,113,114}

CONCLUDING REMARKS

Many methods have been developed in recent years to incorporate conformational changes in the protein into receptor–ligand

docking. Most of the methods derive an EPS or collective modes of protein motion from the apo structure of the protein, thus assuming the conformational-selection model of protein flexibility coupled to ligand binding is valid. However, experimental and computational studies have demonstrated that for some systems the induced-fit mechanism of ligand binding plays an important role during protein–ligand recognition. Furthermore, if the conformational-selection model is the dominant mechanism of protein flexibility coupled to ligand binding, elucidating or determining the holo-like conformation on the time scale sampled *in silico* is still challenging. In particular, sampling partial folding or unfolding of segments of the protein associated with ligand binding may not be accessible with current computational methods. For such systems, novel approaches that combine protein folding and ligand binding may be necessary to accurately predict and quantify ligand binding, and this is a current area of research.^{115,116}

Another area of future research is the precise quantification of protein–ligand binding in the context of protein flexibility. Improvement is needed in scoring functions used to quantify the direct interaction between ligand and protein. Improvement is also necessary for the scoring procedure used to estimate the free energy associated with the different conformational states accessible to the protein structure. In general, the inclusion of additional degrees of freedom to simulate protein flexibility adds to the difficulty of accurately predicting the free energy of binding. This difficulty is due to the fact that more contributions to the free energy must be considered that vary between different poses and different ligands, i.e., the interaction between flexible residues and the core of the protein, and typically, these additional contributions also introduce additional inaccuracies into the calculated binding affinity. The inaccuracy of scoring protein–ligand complexes in combination with the generation of additional possible binding poses when including protein flexibility contributes to the failure of some virtual screening experiments that incorporate protein flexibility.^{50,79} To improve the scoring process in docking, it may be necessary to include protein dynamics, for example, by sampling conformational substates of a given binding mode or by including an estimate of configurational entropy. This area of research may contribute to a more precise estimation of binding affinities in the future.

AUTHOR INFORMATION

Corresponding Author

*E-mail: mlill@purdue.edu. Phone: (765) 496-9375. Fax: (765) 494-1414.

Funding Sources

This work has in part been supported by the National Institutes of Health (Grants GM085604 and GM092855).

ACKNOWLEDGMENT

I thank Matthew Danielson and Jared Thompson for critical reading of the manuscript and the reviewers for valuable suggestions to improve the manuscript.

ABBREVIATIONS

MD, molecular dynamics; MC, Monte Carlo; NMA, normal-mode analysis; ENM, elastic-network model; PCA, principal-component analysis; FEP, free energy perturbation; TI, thermodynamic integration; IT-TI, independent-trajectory thermodynamics

integration; LIE, linear interaction energy analysis; MMPBSA, molecular mechanics Poisson–Boltzmann surface area method; MMGBSA, molecular mechanics generalized Born surface area; rmsd, root-mean-square deviation; EPS, ensemble of protein structures; PFS, partition function-based scoring; KNF, Koshland–Nemethy–Filmer; MWC, Monod–Wyman–Changeux; SERM, selective estrogen receptor modulators; DHFR, dihydrofolate reductase; PDB, Protein Data Bank.

REFERENCES

- (1) Wess, J., Han, S. J., Kim, S. K., Jacobson, K. A., and Li, J. H. (2008) Conformational changes involved in G-protein-coupled-receptor activation. *Trends Pharmacol. Sci.* 29, 616–625.
- (2) Kobilka, B. K., and Deupi, X. (2007) Conformational complexity of G-protein-coupled receptors. *Trends Pharmacol. Sci.* 28, 397–406.
- (3) Kobilka, B. K. (2007) G protein coupled receptor structure and activation. *Biochim. Biophys. Acta* 1768, 794–807.
- (4) Hucho, F., and Weise, C. (2001) Ligand-gated ion channels. *Angew. Chem., Int. Ed.* 40, 3101–3116.
- (5) Shepherd, G. M. (1994) Discrimination of Molecular Signals by the Olfactory Receptor Neuron. *Neuron* 13, 771–790.
- (6) Rozengurt, E., and Sternini, C. (2007) Taste receptor signaling in the mammalian gut. *Curr. Opin. Pharmacol.* 7, 557–562.
- (7) Denisov, I. G., Makris, T. M., Sligar, S. G., and Schlichting, I. (2005) Structure and chemistry of cytochrome P450. *Chem. Rev.* 105, 2253–2277.
- (8) Guengerich, F. P. (2006) Cytochrome P450s and other enzymes in drug metabolism and toxicity. *AAPS J.* 8, E101–E111.
- (9) Bossart-Whitaker, P., Carson, M., Babu, Y. S., Smith, C. D., Laver, W. G., and Air, G. M. (1993) Three-dimensional Structure of Influenza A N9 Neuraminidase and Its Complex with the Inhibitor 2-Deoxy 2,3-Dehydro-N-Acetyl Neuraminic Acid. *J. Mol. Biol.* 232, 1069–1083.
- (10) Taylor, N. R., Cleasby, A., Singh, O., Skarzynski, T., Wonacott, A. J., Smith, P. W., Sollis, S. L., Howes, P. D., Cherry, P. C., Bethell, R., Colman, P., and Varghese, J. (1998) Dihydropyranocarboxamides Related to Zanamivir: A New Series of Inhibitors of Influenza Virus Sialidases. 2. Crystallographic and Molecular Modeling Study of Complexes of 4-Amino-4H-pyran-6-carboxamides and Sialidase from Influenza Virus Types A and B. *J. Med. Chem.* 41, 798–807.
- (11) Park, S. Y., Yamane, K., Adachi, S. i., Shiro, Y., Weiss, K. E., Maves, S. A., and Sligar, S. G. (2002) Thermophilic cytochrome P450 (CYP119) from *Sulfolobus solfataricus*: High resolution structure and functional properties. *J. Inorg. Biochem.* 91, 491–501.
- (12) Yano, J. K., Koo, L. S., Schuller, D. J., Li, H., Ortiz de Montellano, P. R., and Poulos, T. L. (2000) Crystal Structure of a Thermophilic Cytochrome P450 from the Archaeon *Sulfolobus solfataricus*. *J. Biol. Chem.* 275, 31086–31092.
- (13) Brzozowski, A. M., Pike, A. C., Dauter, Z., Hubbard, R. E., Bonn, T., Engstrom, O., Ohman, L., Greene, G. L., Gustafsson, J. A., and Carlquist, M. (1997) Molecular basis of agonism and antagonism in the oestrogen receptor. *Nature* 389, 753–758.
- (14) Brzozowski, A. M., Pike, A. C., Dauter, Z., Hubbard, R. E., Bonn, T., Engstrom, O., Ohman, L., Greene, G. L., Gustafsson, J. A., and Carlquist, M. (1997) Molecular basis of agonism and antagonism in the oestrogen receptor. *Nature* 389, 753–758.
- (15) Teague, S. J. (2003) Implications of protein flexibility for drug discovery. *Nat. Rev. Drug Discovery* 2, 527–541.
- (16) Park, S. Y., Yamane, K., Adachi, S. i., Shiro, Y., Weiss, K. E., Maves, S. A., and Sligar, S. G. (2002) Thermophilic cytochrome P450 (CYP119) from *Sulfolobus solfataricus*: High resolution structure and functional properties. *J. Inorg. Biochem.* 91, 491–501.
- (17) Yano, J. K., Koo, L. S., Schuller, D. J., Li, H., Ortiz de Montellano, P. R., and Poulos, T. L. (2000) Crystal Structure of a Thermophilic Cytochrome P450 from the Archaeon *Sulfolobus solfataricus*. *J. Biol. Chem.* 275, 31086–31092.
- (18) Brzozowski, A. M., Pike, A. C., Dauter, Z., Hubbard, R. E., Bonn, T., Engstrom, O., Ohman, L., Greene, G. L., Gustafsson, J. A., and Carlquist, M. (1997) Molecular basis of agonism and antagonism in the oestrogen receptor. *Nature* 389, 753–758.
- (19) Koshland, D. E. (1958) Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proc. Natl. Acad. Sci. U.S.A.* 44, 98–104.
- (20) Ma, B. Y., Kumar, S., Tsai, C. J., and Nussinov, R. (1999) Folding funnels and binding mechanisms. *Protein Eng.* 12, 713–720.
- (21) Koshland, D. E., Jr., Nemethy, G., and Filmer, D. (1966) Comparison of experimental binding data and theoretical models in proteins containing subunits. *Biochemistry* 5, 365–385.
- (22) Monod, J., Wyman, J., and Changeux, J. P. (1965) On the nature of allosteric transitions: A plausible model. *J. Mol. Biol.* 12, 88–118.
- (23) Okazaki, K. I., and Takada, S. (2008) Dynamic energy landscape view of coupled binding and protein conformational change: Induced-fit versus population-shift mechanisms. *Proc. Natl. Acad. Sci. U.S.A.* 105, 11182–11187.
- (24) Tanford, C. (1978) The hydrophobic effect and the organization of living matter. *Science* 200, 1012–1018.
- (25) Tanford, C. (1980) *The Hydrophobic Effect*, Wiley, New York.
- (26) Baron, R., Setny, P., and McCammon, J. A. (2010) Water in cavity-ligand recognition. *J. Am. Chem. Soc.* 132, 12091–12097.
- (27) Setny, P., Baron, R., and McCammon, J. A. (2010) How Can Hydrophobic Association Be Enthalpy Driven? *J. Chem. Theory Comput.* 6, 2866–2871.
- (28) Dunitz, J. D. (1994) The entropic cost of bound water in crystals and biomolecules. *Science* 264, 670.
- (29) Gao, S., House, W., and Chapman, W. G. (2005) NMR/MRI study of clathrate hydrate mechanisms. *J. Phys. Chem. B* 109, 19090–19093.
- (30) Barratt, E., Bingham, R. J., Warner, D. J., Laughton, C. A., Phillips, S. E., and Homans, S. W. (2005) Van der Waals interactions dominate ligand-protein association in a protein binding site occluded from solvent water. *J. Am. Chem. Soc.* 127, 11827–11834.
- (31) Bingham, R. J., Findlay, J. B., Hsieh, S. Y., Kalverda, A. P., Kjellberg, A., Perazzolo, C., Phillips, S. E., Seshadri, K., Trinh, C. H., Turnbull, W. B., Bodenhausen, G., and Homans, S. W. (2004) Thermodynamics of binding of 2-methoxy-3-isopropylpyrazine and 2-methoxy-3-isobutylpyrazine to the major urinary protein. *J. Am. Chem. Soc.* 126, 1675–1681.
- (32) Volkman, B. F., Lipson, D., Wemmer, D. E., and Kern, D. (2001) Two-state allosteric behavior in a single-domain signaling protein. *Science* 291, 2429–2433.
- (33) Boehr, D. D., Nussinov, R., and Wright, P. E. (2009) The role of dynamic conformational ensembles in biomolecular recognition. *Nat. Chem. Biol.* 5, 789–796.
- (34) Fraser, J. S., Clarkson, M. W., Degnan, S. C., Erion, R., Kern, D., and Alber, T. (2009) Hidden alternative structures of proline isomerase essential for catalysis. *Nature* 462, 669–673.
- (35) Dyer, C. M., and Dahlquist, F. W. (2006) Switched or not? The structure of unphosphorylated CheY bound to the N terminus of FlhM. *J. Bacteriol.* 188, 7354–7363.
- (36) Nevo, R., Brumfeld, V., Elbaum, M., Hinterdorfer, P., and Reich, Z. (2004) Direct discrimination between models of protein activation by single-molecule force measurements. *Biophys. J.* 87, 2630–2634.
- (37) Hammes, G. G., Chang, Y. C., and Oas, T. G. (2009) Conformational selection or induced fit: A flux description of reaction mechanism. *Proc. Natl. Acad. Sci. U.S.A.* 106, 13737–13741.
- (38) Weikl, T. R., and von Deuster, C. (2009) Selected-fit versus induced-fit protein binding: Kinetic differences and mutational analysis. *Proteins* 75, 104–110.
- (39) Ma, B. Y., and Nussinov, R. (2010) Enzyme dynamics point to stepwise conformational selection in catalysis. *Curr. Opin. Chem. Biol.* 14, 652–659.
- (40) James, L. C., and Tawfik, D. S. (2005) Structure and kinetics of a transient antibody binding intermediate reveal a kinetic discrimination

mechanism in antigen recognition. *Proc. Natl. Acad. Sci. U.S.A.* 102, 12730–12735.

(41) Zwanzig, R. W. (1954) High-Temperature Equation of State by A Perturbation Method. 1. Nonpolar Gases. *J. Chem. Phys.* 22, 1420–1426.

(42) Kirkwood, J. G. (1935) Statistical mechanics of fluid mixtures. *J. Chem. Phys.* 3, 300–313.

(43) Mobley, D. L., Chodera, J. D., and Dill, K. A. (2007) The Confine-and-Release Method: Obtaining Correct Binding Free Energies in the Presence of Protein Conformational Change. *J. Chem. Theory Comput.* 3, 1231–1235.

(44) Lawrenz, M., Baron, R., and McCammon, J. A. (2009) Independent-Trajectories Thermodynamic-Integration Free-Energy Changes for Biomolecular Systems: Determinants of H5N1 Avian Influenza Virus Neuraminidase Inhibition by Peramivir. *J. Chem. Theory Comput.* 5, 1106–1116.

(45) Warren, G. L., Andrews, C. W., Capelli, A. M., Clarke, B., LaLonde, J., Lambert, M. H., Lindvall, M., Nevins, N., Semus, S. F., Senger, S., Tedesco, G., Wall, I. D., Woolven, J. M., Peishoff, C. E., and Head, M. S. (2006) A critical assessment of docking programs and scoring functions. *J. Med. Chem.* 49, 5912–5931.

(46) Ferrara, P., Gohlke, H., Price, D. J., Klebe, G., and Brooks, C. L., III (2004) Assessing scoring functions for protein-ligand interactions. *J. Med. Chem.* 47, 3032–3047.

(47) Kuntz, I. D., Chen, K., Sharp, K. A., and Kollman, P. A. (1999) The maximal affinity of ligands. *Proc. Natl. Acad. Sci. U.S.A.* 96, 9997–10002.

(48) Carlson, H. A. (2002) Protein flexibility and drug design: How to hit a moving target. *Curr. Opin. Chem. Biol.* 6, 447–452.

(49) Teodoro, M. L., and Kavraki, L. E. (2003) Conformational flexibility models for the receptor in structure based drug design. *Curr. Pharm. Des.* 9, 1635–1648.

(50) Totrov, M., and Abagyan, R. (2008) Flexible ligand docking to multiple receptor conformations: A practical alternative. *Curr. Opin. Struct. Biol.* 18, 178–184.

(51) Beier, C., and Zacharias, M. (2010) Tackling the challenges posed by target flexibility in drug design. *Expert Opin. Drug Discovery* 5, 347–359.

(52) Rao, C., Subramanian, J., and Sharma, S. D. (2009) Managing protein flexibility in docking and its applications. *Drug Discovery Today* 14, 394–400.

(53) Sotriffer, C. A. (2011) Accounting for Induced-Fit Effects in Docking: What is Possible and What is Not? *Curr. Top. Med. Chem.* 11, 179–191.

(54) Lin, J. H. (2011) Accommodating Protein Flexibility for Structure-Based Drug Design. *Curr. Top. Med. Chem.* 11, 171–178.

(55) Yang, A. Y. C., Kallblad, P., and Mancera, R. L. (2004) Molecular modelling prediction of ligand binding site flexibility. *J. Comput.-Aided Mol. Des.* 18, 235–250.

(56) Meiler, J., and Baker, D. (2006) ROSETTALIGAND: Protein-small molecule docking with full side-chain flexibility. *Proteins* 65, 538–548.

(57) Leach, A. R. (1994) Ligand Docking to Proteins with Discrete Side-Chain Flexibility. *J. Mol. Biol.* 235, 345–356.

(58) Nabuurs, S. B., Wagener, M., and De Vlieg, J. (2007) A flexible approach to induced fit docking. *J. Med. Chem.* 50, 6507–6518.

(59) Hartmann, C., Antes, I., and Lengauer, T. (2009) Docking and scoring with alternative side-chain conformations. *Proteins* 74, 712–726.

(60) Zavodszky, M. I., and Kuhn, L. A. (2005) Side-chain flexibility in protein-ligand binding: The minimal rotation hypothesis. *Protein Sci.* 14, 1104–1114.

(61) Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K., and Olson, A. J. (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* 19, 1639–1662.

(62) Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., and Olson, A. J. (2009) AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* 30, 2785–2791.

(63) Zacharias, M., and Sklenar, H. (1999) Harmonic modes as variables to approximately account for receptor flexibility in ligand-receptor docking simulations: Application to DNA minor groove ligand complex. *J. Comput. Chem.* 20, 287–300.

(64) May, A., and Zacharias, M. (2008) Protein-ligand docking accounting for receptor side chain and global flexibility in normal modes: Evaluation on kinase inhibitor cross docking. *J. Med. Chem.* 51, 3499–3506.

(65) Amadei, A., Linssen, A. B., and Berendsen, H. J. (1993) Essential dynamics of proteins. *Proteins* 17, 412–425.

(66) Cukier, R. I. (2009) Apo adenylate kinase encodes its holo form: A principal component and varimax analysis. *J. Phys. Chem. B* 113, 1662–1672.

(67) Lou, H., and Cukier, R. I. (2006) Molecular dynamics of apo-adenylate kinase: A distance replica exchange method for the free energy of conformational fluctuations. *J. Phys. Chem. B* 110, 24121–24137.

(68) Lou, H., and Cukier, R. I. (2006) Molecular dynamics of apo-adenylate kinase: A principal component analysis. *J. Phys. Chem. B* 110, 12796–12808.

(69) Ikeguchi, M., Ueno, J., Sato, M., and Kidera, A. (2005) Protein structural change upon ligand binding: Linear response theory. *Phys. Rev. Lett.* 94, 078102.

(70) Cavasotto, C. N., Kovacs, J. A., and Abagyan, R. A. (2005) Representing receptor flexibility in ligand docking through relevant normal modes. *J. Am. Chem. Soc.* 127, 9632–9640.

(71) Petrone, P., and Pande, V. S. (2006) Can conformational change be described by only a few normal modes? *Biophys. J.* 90, 1583–1593.

(72) Claussen, H., Buning, C., Rarey, M., and Lengauer, T. (2001) FlexE: Efficient molecular docking considering protein structure variations. *J. Mol. Biol.* 308, 377–395.

(73) Zhao, Y., and Sanner, M. F. (2007) FLIPDock: Docking flexible ligands into flexible receptors. *Proteins* 68, 726–737.

(74) Ferrari, A. M., Wei, B. Q., Costantino, L., and Shoichet, B. K. (2004) Soft docking and multiple receptor conformations in virtual screening. *J. Med. Chem.* 47, 5076–5084.

(75) Knegtel, R. M., Kuntz, I. D., and Oshiro, C. M. (1997) Molecular docking to ensembles of protein structures. *J. Mol. Biol.* 266, 424–440.

(76) Erickson, J. A., Jalaie, M., Robertson, D. H., Lewis, R. A., and Vieth, M. (2004) Lessons in molecular recognition: The effects of ligand and protein flexibility on molecular docking accuracy. *J. Med. Chem.* 47, 45–55.

(77) Kitchen, D. B., Decornez, H., Furr, J. R., and Bajorath, J. (2004) Docking and scoring in virtual screening for drug discovery: Methods and applications. *Nat. Rev. Drug Discovery* 3, 935–949.

(78) Xu, M., and Lill, M. A. (2011) Significant Enhancement of Docking Sensitivity Using Implicit Ligand Sampling. *J. Chem. Inf. Model.* 51, 693–706.

(79) Barril, X., and Morley, S. D. (2005) Unveiling the full potential of flexible receptor docking using multiple crystallographic structures. *J. Med. Chem.* 48, 4432–4443.

(80) Amaro, R. E., Baron, R., and McCammon, J. A. (2008) An improved relaxed complex scheme for receptor flexibility in computer-aided drug design. *J. Comput.-Aided Mol. Des.* 22, 693–705.

(81) Bolstad, E. S., and Anderson, A. C. (2009) In pursuit of virtual lead optimization: Pruning ensembles of receptor structures for increased efficiency and accuracy during docking. *Proteins* 75, 62–74.

(82) Armen, R. S., Chen, J., and Brooks, C. L. (2009) An Evaluation of Explicit Receptor Flexibility in Molecular Docking Using Molecular Dynamics and Torsion Angle Molecular Dynamics. *J. Chem. Theory Comput.* 5, 2909–2923.

(83) Gilson, M. K., and Zhou, H. X. (2007) Calculation of protein-ligand binding affinities. *Annu. Rev. Biophys. Biomol. Struct.* 36, 21–42.

(84) Gilson, M. K., Given, J. A., Bush, B. L., and McCammon, J. A. (1997) The statistical-thermodynamic basis for computation of binding affinities: A critical review. *Biophys. J.* 72, 1047–1069.

(85) Kollman, P. A., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, L., Lee, M., Lee, T., Duan, Y., Wang, W., Donini, O., Cieplak, P.,

Srinivasan, J., Case, D. A., and Cheatham, T. E., III (2000) Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Acc. Chem. Res.* 33, 889–897.

(86) Srinivasan, J., Miller, J., Kollman, P. A., and Case, D. A. (1998) Continuum solvent studies of the stability of RNA hairpin loops and helices. *J. Biomol. Struct. Dyn.* 16, 671–682.

(87) Bohm, H. J. (1994) The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. *J. Comput.-Aided Mol. Des.* 8, 243–256.

(88) Eldridge, M. D., Murray, C. W., Auton, T. R., Paolini, G. V., and Mee, R. P. (1997) Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J. Comput.-Aided Mol. Des.* 11, 425–445.

(89) Huey, R., Morris, G. M., Olson, A. J., and Goodsell, D. S. (2007) A semiempirical free energy force field with charge-based desolvation. *J. Comput. Chem.* 28, 1145–1152.

(90) Moy, F. J., Chanda, P. K., Chen, J., Cosmi, S., Edris, W., Levin, J. I., Rush, T. S., Wilhelm, J., and Powers, R. (2002) Impact of mobility on structure-based drug design for the MMPs. *J. Am. Chem. Soc.* 124, 12658–12659.

(91) Thielges, M. C., Chung, J. K., and Fayer, M. D. (2011) Protein dynamics in cytochrome P450 molecular recognition and substrate specificity using 2D IR vibrational echo spectroscopy. *J. Am. Chem. Soc.* 133, 3995–4004.

(92) Aqvist, J., Medina, C., and Samuelsson, J. E. (1994) A new method for predicting binding affinity in computer-aided drug design. *Protein Eng.* 7, 385–391.

(93) Alonso, H., Bliznyuk, A. A., and Gready, J. E. (2006) Combining docking and molecular dynamic simulations in drug design. *Med. Res. Rev.* 26, 531–568.

(94) Naim, M., Bhat, S., Rankin, K. N., Dennis, S., Chowdhury, S. F., Siddiqi, I., Drabik, P., Sulea, T., Bayly, C. I., Jakalian, A., and Purisima, E. O. (2007) Solvated interaction energy (SIE) for scoring protein-ligand binding affinities. 1. Exploring the parameter space. *J. Chem. Inf. Model.* 47, 122–133.

(95) Chang, M. W., Belew, R. K., Carroll, K. S., Olson, A. J., and Goodsell, D. S. (2008) Empirical entropic contributions in computational docking: Evaluation in APS reductase complexes. *J. Comput. Chem.* 29, 1753–1761.

(96) Huang, S. Y., and Zou, X. (2010) Inclusion of solvation and entropy in the knowledge-based scoring function for protein-ligand interactions. *J. Chem. Inf. Model.* 50, 262–273.

(97) Ruvinsky, A. M., and Kozintsev, A. V. (2005) New and fast statistical-thermodynamic method for computation of protein-ligand binding entropy substantially improves docking accuracy. *J. Comput. Chem.* 26, 1089–1095.

(98) Ruvinsky, A. M. (2007) Calculations of protein-ligand binding entropy of relative and overall molecular motions. *J. Comput.-Aided Mol. Des.* 21, 361–370.

(99) Salaniwal, S., Manas, E. S., Alvarez, J. C., and Unwalla, R. J. (2007) Critical evaluation of methods to incorporate entropy loss upon binding in high-throughput docking. *Proteins* 66, 422–435.

(100) Chang, C. E., and Gilson, M. K. (2004) Free energy, entropy, and induced fit in host-guest recognition: Calculations with the second-generation mining minima algorithm. *J. Am. Chem. Soc.* 126, 13156–13164.

(101) David, L., Luo, R., and Gilson, M. K. (2001) Ligand-receptor docking with the Mining Minima optimizer. *J. Comput.-Aided Mol. Des.* 15, 157–171.

(102) Kairys, V., and Gilson, M. K. (2002) Enhanced docking with the mining minima optimizer: Acceleration and side-chain flexibility. *J. Comput. Chem.* 23, 1656–1670.

(103) Chen, W., Gilson, M. K., Webb, S. P., and Potter, M. J. (2010) Modeling Protein-Ligand Binding by Mining Minima. *J. Chem. Theory Comput.* 6, 3540–3557.

(104) Kua, J., Zhang, Y., and McCammon, J. A. (2002) Studying enzyme binding specificity in acetylcholinesterase using a combined

molecular dynamics and multiple docking approach. *J. Am. Chem. Soc.* 124, 8260–8267.

(105) Lin, J. H., Perryman, A. L., Schames, J. R., and McCammon, J. A. (2003) The relaxed complex method: Accommodating receptor flexibility for drug design with an improved scoring scheme. *Biopolymers* 68, 47–62.

(106) Lin, J. H., Perryman, A. L., Schames, J. R., and McCammon, J. A. (2002) Computational drug design accommodating receptor flexibility: The relaxed complex scheme. *J. Am. Chem. Soc.* 124, 5632–5633.

(107) Kortvelyesi, T., Dennis, S., Silberstein, M., Brown, L., III, and Vajda, S. (2003) Algorithms for computational solvent mapping of proteins. *Proteins* 51, 340–351.

(108) Mysinger, M. M., and Shoichet, B. K. (2010) Rapid context-dependent ligand desolvation in molecular docking. *J. Chem. Inf. Model.* 50, 1561–1573.

(109) Verdonk, M. L., Chessari, G., Cole, J. C., Hartshorn, M. J., Murray, C. W., Nissink, J. W., Taylor, R. D., and Taylor, R. (2005) Modeling water molecules in protein-ligand docking using GOLD. *J. Med. Chem.* 48, 6504–6515.

(110) Abel, R., Young, T., Farid, R., Berne, B. J., and Friesner, R. A. (2008) Role of the active-site solvent in the thermodynamics of factor Xa ligand binding. *J. Am. Chem. Soc.* 130, 2817–2831.

(111) Lazaridis, T. (1998) Inhomogeneous fluid approach to solvation thermodynamics. 1. Theory. *J. Phys. Chem. B* 102, 3531–3541.

(112) Lazaridis, T. (1998) Inhomogeneous fluid approach to solvation thermodynamics. 2. Applications to simple fluids. *J. Phys. Chem. B* 102, 3542–3550.

(113) Abel, R., Salam, N. K., Shelley, J., Farid, R., Friesner, R. A., and Sherman, W. (2011) Contribution of Explicit Solvent Effects to the Binding Affinity of Small-Molecule Inhibitors in Blood Coagulation Factor Serine Proteases. *ChemMedChem* 6, 1049–1066.

(114) Higgs, C., Beuming, T., and Sherman, W. (2010) Hydration Site Thermodynamics Explain SARs for Triazolylpurines Analogues Binding to the A2A Receptor. *ACS Med. Chem. Lett.* 1, 160–164.

(115) Das, R., Andre, I., Shen, Y., Wu, Y., Lemak, A., Bansal, S., Arrowsmith, C. H., Szyperki, T., and Baker, D. (2009) Simultaneous prediction of protein folding and docking at high resolution. *Proc. Natl. Acad. Sci. U.S.A.* 106, 18978–18983.

(116) Wong, S., and Jacobson, M. P. (2008) Conformational selection in silico: Loop latching motions and ligand binding in enzymes. *Proteins* 71, 153–164.