Annu. Rev. Pharmacol. Toxicol. 2003. 43:31–45
doi: 10.1146/annurev.pharmtox.43.100901.140216

First published online as a Review in Advance on July 24, 2002

PROTEIN FLEXIBILITY AND COMPUTER-AIDED DRUG DESIGN

Chung F. Wong[1,2] and J. Andrew McCammon[1,2,3]
[1]Howard Hughes Medical Institute, [2]Department of Pharmacology, and [3]Department of Chemistry and Biochemistry, University of California, San Diego, La Jolla, California 92093-0365; email: jmccammon@ucsd.edu

Key Words molecular dynamics simulations, free energy calculations, sensitivity analysis, dynamic pharmacophore, relaxed complex method

■ **Abstract** Although computational techniques are increasingly being used in computer-aided drug design, the effects due to protein flexibility are still ignored in many applications. This review revisits rigorous statistical mechanical methods for predicting binding affinity, discusses some recent developments for improving their speed and reliability, and examines faster approximate models for facilitating virtual screening and lead optimization.

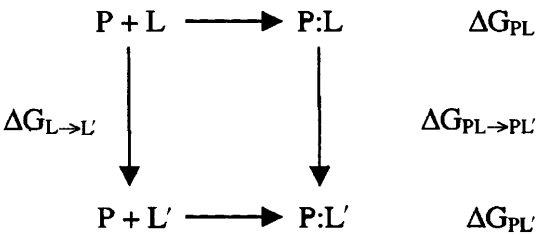
INTRODUCTION

The rapid advance of computer technology and the development of new modeling software have made computer-aided drug design an increasingly useful tool. This review focuses on addressing the role of protein flexibility in drug discovery, as this remains one of the most challenging problems in computer-assisted drug development and many modeling efforts still assume proteins to be rigid. The first molecular dynamics simulation of a protein, bovine pancreatic trypsin inhibitor (1), revealed in atomic details large structural fluctuations that were previously unexpected. The possible significance of these effects on determining protein functions and molecular recognition was quickly recognized and has triggered the development of methods for including receptor flexibility in modeling protein-ligand interactions. For example, rigorous simulation methods for calculating binding free energy were introduced in the early 1980s (2–8). Although these simulations were limited to several tens of picoseconds in the early stages, they moved a step beyond the fixed-conformation picture that dominated many scientists' thinking at the time. Despite this advance, an issue rapidly came up concerning the length of simulations required to generate representative ensembles for reliable predictions. Although the simulation length required to gain good statistics depends on the specific problem at hand, many applications of practical interest demand a large amount of computer time, preventing these modeling techniques from being used

routinely in day-to-day drug discovery. However, progress has been made in developing new algorithms to accelerate these calculations, in employing implicit rather than explicit solvent models to facilitate solute conformational samplings, and in introducing approximate methods to screen many derivatives of a drug lead more efficiently. Here, we briefly review some earlier work and then discuss various encouraging recent developments.

FREE ENERGY CALCULATIONS

In 1984 (2), the thermodynamic cycle-perturbation method was introduced to help compare the binding affinity of a group of similar inhibitors. This method is based on constructing a thermodynamic cycle relating two binding processes:



Because free energy is a state function,

$$\Delta\Delta G = \Delta G_{PL'} - \Delta G_{PL}, \qquad 1.$$

which compares the binding affinity of two ligands, L and L′, to a protein P, can also be obtained from

$$\Delta\Delta G = \Delta G_{PL\rightarrow PL'} - \Delta G_{L\rightarrow L'}. \qquad 2.$$

This trick converts the difficult problem of calculating $\Delta G_{PL'}$ and ΔG_{PL} into evaluating $\Delta G_{L\rightarrow L'}$ and $\Delta G_{PL\rightarrow PL'}$. $\Delta G_{PL'}$ and ΔG_{PL} are usually more difficult to calculate directly because they involve simulating the displacement of a large number of water molecules before the binding between P and L/L′ can occur. $\Delta G_{PL'}$ and ΔG_{PL} are often much easier to calculate, especially when the two ligands are similar, because they involve relatively small changes in chemical functional groups. One can use special simulation methods, such as Zwanzig's perturbation theory (9), to calculate the free energy differences, $\Delta G_{L\rightarrow L'}$ and $\Delta G_{PL\rightarrow PL'}$. For the Helmholtz free energy, Zwanzig's perturbation theory reads

$$\Delta A = -RT \ln \langle \exp(-\Delta H/RT) \rangle_r, \qquad 3.$$

where RT is the gas constant times the absolute temperature, $\Delta H = H_p - H_r$ in which H_p and H_r are the classical Hamiltonians of the perturbed and reference systems respectively, and $\langle \ldots \rangle_r$ represents an ensemble average over the reference

state. (The quantities ΔG and ΔA are virtually the same in many cases.) This formula is exact, although it is hard to obtain a reliable estimate of ΔA when the difference between the reference and perturbed system is large. One way to alleviate this problem is to change the reference system into the perturbed one in N steps, rather than in one single step, and use the perturbation formula N times to estimate ΔA from

$$\Delta A = -RT \sum_{i=1}^{N} \ln \langle \exp(-\Delta H_i/RT) \rangle_i, \qquad 4.$$

where $i = 1$ refers to the reference state, $i = \mathrm{N} + 1$ corresponds to the final state, and $\Delta H_i = H_{i+1} - H_i$. These calculations can be very expensive to do when large modifications are made because many windows are required. Consequently, these calculations are usually only practical to use at the late-stage refinement of an already good lead. However, methods to further speed up free energy calculations are constantly being introduced and some recent developments are discussed here.

Locally Enhanced Sampling

Protein side chains or ligand functional groups can adopt multiple conformations, and the barriers separating these conformations may be rather high such that the transitions among them are difficult. Thus, it may require a long simulation to reliably estimate a free energy change if more than one conformation contributes significantly to binding. To improve the sampling of these conformations in free energy calculations, Verkhivker et al. (10) utilized the locally enhanced sampling (LES) method (11). The essence of this method is to replace a side chain or a functional group by N copies that do not interact with each other, and each copy only interacts with itself and its surrounding with $1/\mathrm{N}^{\mathrm{th}}$ of the original strength. The use of multiple copies and the reduction of the interaction potentials can significantly enhance the sampling of the conformations of the side chain or the functional group. This requires two additional perturbation calculations: one in changing from the single-copy representation of the reference state to the multiple-copy representation and the other in going from the multiple-copy representation of the perturbed state into the single-copy representation. But Verkhivker et al. (10) demonstrated that the benefits of adopting a multiple-copy representation outweighed the additional costs of introducing two more perturbation calculations. More recently, Simmerling et al. (12) applied this method to study the $\alpha \rightarrow \beta$ anomerization of glucose and found that the free energy calculations converged an order of magnitude faster than with the single-copy method.

λ Dynamics Method

The λ dynamics method (13, 14) is another technique introduced to speed up free energy calculations. In this method, multiple ligands are placed in the binding site of their receptor at once with the interaction potential of each ligand reduced from its full strength. The fraction, λ_i^2, of the interaction potential of each ligand

is determined dynamically during a simulation, with λ_t treated as a particle with a fictitious mass. Because the interaction potential of each ligand is reduced, the barriers for conformational transitions are lower. The reduced barriers can enable a ligand to explore different orientations and conformations more easily. Also, the ranking of the ligands can emerge quickly during the simulation as λ_t^2 can increase rapidly for the winners at the expense of the losers. The identification of the strong binders can therefore be much quicker than by doing many free energy perturbation calculations including only one ligand at a time. This method was able to quickly distinguish strong benzamidine inhibitors of trypsin from weaker ones (15).

Systematic Sensitivity Analysis

One does not always need to aim for highly accurate binding constants to be productive in drug design. It is already useful to generate rules or constraints from computational studies to guide the design of chemical libraries for high-throughput screening and to direct the optimization of a drug lead. To this end, one does not necessarily make physical modifications, but nonphysical ones that probe the relative significance of different features of functional groups in affecting binding. For example, by sequentially turning off the atomic partial charge or dipole moment of every relevant functional group in a lead compound, one can determine which charges or dipole moments are important to keep in optimizing a drug lead and which charges or dipole moments should be turned off to improve binding affinity. For this more modest goal, one can use mathematical tricks to carry out many free energy difference calculations simultaneously.

One way to do this is to adopt an approach that has been used by engineers for a long time (16, 17) and has recently been applied to study molecular and biomolecular systems (18–28). In order to identify the key model parameters determining system properties, engineers calculate the derivatives, $\frac{\partial O}{\partial \lambda_i}$s, of a property, O, of the system with respect to the model parameters, λ_is, to measure the sensitivity of the observable to parameter changes. (The dimensionless logarithmic derivatives $\frac{\partial \ln O}{\partial \ln \lambda_i} = \frac{\partial O}{\partial \lambda_i}\frac{\lambda_i}{O}$s are also often calculated to facilitate comparison among different types of parameters.) Parameters that do not affect system properties yield negligible $\frac{\partial O}{\partial \lambda_i}$s. On the other hand, important parameters would yield large $\frac{\partial O}{\partial \lambda_i}$s. Analytical expressions can be worked out for calculating these derivatives based on the dynamical behavior of a single reference system. Because the calculation of the derivatives of a number of observables with respect to all the parameters of a model only adds a small fraction to the costs of doing dynamical simulations on the reference system, it is not difficult to systematically compare the role of all the parameters in a model so that no important parameters are overlooked. This idea can be generalized to include higher-order derivatives so that a Taylor's series expansion,

$$\Delta O = \sum_i \frac{\partial O}{\partial \lambda_i}\Delta\lambda_i + \frac{1}{2}\sum_{i,j}\frac{\partial^2 O}{\partial \lambda_i \partial \lambda_j}\Delta\lambda_i\Delta\lambda_j + \ldots, \qquad 5.$$

can be used to predict the influence of larger parameter changes on system properties. The Taylor's series expansion also permits the effects of many different combinations of parameter changes to be examined. However, it is more difficult to calculate higher-order derivatives, and the convergence of the Taylor's series can be slow, if it converges at all, when parameter modifications are large. Different strategies have been introduced to deal with somewhat larger perturbations.

One-Step Application of Zwanzig's Perturbation Theory

The single-window Zwanzig perturbation theory (9) described above can provide quick estimates of free energy changes when parameter perturbations are sufficiently small. In fact, this strategy was used in earlier free energy calculations when computers were much less powerful. For example, an early free energy perturbation study focuses on examining the effects of making conservative modifications on free energy changes (3). In one case, benzamidine was modified into parafluorobenzamidine and the effects on trypsin binding were examined. In this calculation, only simulations on the reference systems, benzamidine and the trypsin-benzamidine complex, were performed, and the single-window Zwanzig perturbation formula was used to calculate $\Delta G_{L \to L'}$ and $\Delta G_{PL \to PL'}$ directly.

By focusing on small perturbations, a single-window Zwanzig formula (9) can be used to provide initial estimates of the effects of making many physical or nonphysical changes on binding affinity, without carrying out expensive molecular dynamics simulations for all the perturbed systems. Only simulations of the reference system are needed. This technique, or its close cousin in which the Helmholtz free energy change is obtained by expanding ΔA in terms of ΔH and keeping up to second-order term,

$$\Delta A = \langle \Delta H \rangle_r + \frac{1}{2RT} \left\langle (\Delta H - \langle \Delta H \rangle_r)^2 \right\rangle_r, \qquad 6.$$

has been used to examine the effects of adding or removing protons (29, 30) or of changing molecular charge distribution on free energy changes (31). In drug design applications, focusing on a small chemical subspace for which a single-window perturbation formula can be used to study many modifications should already be useful for finding better derivatives of a drug lead. One can then use a few of the promising derivatives for further single-window perturbation calculations to enlarge the chemical subspace for identifying other drug candidates. Although this full-blown molecular dynamics–based method has not yet been applied extensively to drug-design applications, an implicit solvent model has already been used in pilot studies on protein kinases to develop pharmacophore models for mining new drug leads from small-molecule libraries, to generate constraints for designing focused chemical libraries for specific targets, and to produce guiding principles for optimizing a drug lead (32, 33).

The range of application of this idea can be extended by using soft-core potentials in reference simulations (34). A single perturbation formula does not work well when larger atoms or atomic groups are added or deleted because the

reference simulation does not adequately sample the configuration states relevant to the modified systems. If a large atom or atomic group is going to be deleted, the reference simulation may not have sampled well the space, allowing the surrounding solute or solvent atoms to get closer. If a large atom or atomic group is going to be created, the reference simulations may have many configurations that create unfavorable steric clashes at the modified sites. To alleviate this problem, Liu et al. (34) utilized soft-core potentials. For example, a modified Lennard-Jones potential of the form

$$V(r_{ij}) = 4\varepsilon_{ij}\left[\frac{\sigma_{ij}^{12}}{\left(\alpha\sigma_{ij}^{6} + r_{ij}^{6}\right)^{2}} - \frac{\sigma_{ij}^{6}}{\alpha\sigma_{ij}^{6} + r_{ij}^{6}}\right] \qquad 7.$$

can be used at selected sites in a reference simulation so as to create more space where atoms are going to be added and to allow atoms to get closer to sites where atoms are going to be deleted. In the above equation, ε_{ij} and σ_{ij} are the Lennard-Jones parameters between atoms i and j, r_{ij} is the distance between the two atoms, and α is a softening parameter that prevents the potential from diverging as $r_{ij} \rightarrow 0$. This approach was able to predict well the free energy differences among a number of para-substituted phenols in water solvent (34). There is a limit to which this approach works well. Mordasini & McCammon (35) later examined the range of applicability of this single-reference approximation by introducing increasingly larger modifications. They found that this model could still yield reasonable qualitative scoring when functional groups involving up to three atoms were deleted.

Combining Explicit and Implicit Solvent Models

Recently, explicit and implicit solvent models have been combined to facilitate free energy calculations (36–40). This approach uses explicit solvent molecular dynamics simulations to relax crystal structures to solution ones, and then uses the simulated solution structures in implicit-solvent calculations to obtain free energy. Using implicit-solvent models eliminates the extensive simulation time required for sampling solvent configurations. This approach assumes that the free energy of a system can be obtained by averaging the potential of mean force, obtained from an implicit solvent model, of dynamics snapshots generated from explicit-solvent models. Entropy contributions of the solute can be estimated from the harmonic or quasiharmonic model. These approximations appear to work well. Vorobjev et al. (36) used one such approach successfully in distinguishing correctly folded protein conformations from misfolded ones. In their study, Vorobjev et al. (36) ran a quick molecular dynamics simulation of a protein for approximately 50–100 ps and used an implicit-solvent model to calculate the free energy of the correctly and incorrectly folded protein. The free energy was found to be lower for the correctly folded protein. The implicit solvent model included contributions from the gas-phase energy of the solute, the energy of cavity formation, the solute-solvent interaction energy, and the solvent electrostatic polarization energy. The energy of

cavity formation was assumed to be proportional to the solvent accessible surface area of the solute. The solvent electrostatic polarization energy was obtained by solving the Poisson-Boltzmann equation and the other terms were obtained from a molecular mechanics force field.

A similar MM/PBSA (molecular mechanics/Poisson-Boltzmann-surface area) approach was used to study protein-ligand interactions. For example, Kuhn & Kollman (38) obtained encouraging results by applying this method to predict the binding affinity of seven ligands to avidin and streptavidin. They obtained a correlation coefficient of 0.92 between the calculated and the experimental binding affinities. The root-mean-square difference between calculated and experimental results, which covered a range of approximately 16 kcal/mol, was on the order of 1.7 kcal/mol. In these calculations, the length of the molecular dynamics simulation used for the averaging was 300 ps. No simulations on the separated protein and ligands were done. Instead, the protein and ligands were assumed to adopt the same conformation as that in the molecular dynamics simulations of the complexes. They also used the harmonic approximation to calculate the entropy change upon binding using six quenched dynamics snapshots. This approximation introduced a relatively large uncertainty in calculating entropy changes; the discrepancy among results from the six snapshots could amount to 5 kcal/mol in the worst case that they studied. So the solute entropy contributions remain a challenge to calculate.

To further speed up this approach, one can replace the expensive explicit-solvent simulations with implicit ones. Statistical mechanical theory gives the Helmholtz free energy A, apart from the scaling constant of the classical partition function that cancels out in binding energy calculations, as

$$A = -RT \ln \iint \exp(-\beta H(u, v)) du\, dv, \qquad 8.$$

where R is the gas constant, T is the absolute temperature, $\beta = {}^{1}/_{RT}$, and $H(u, v)$ is the classical Hamiltonian expressed in terms of the solute coordinates u and the solvent coordinates v. Integrating over the solvent coordinates gives

$$A = -RT \ln \int \exp(-\beta W(u)) du, \qquad 9.$$

where $W(u)$ is the potential of mean force of the solute with conformation defined by u and the kinetic energy term for the solute is ignored because it cancels out in binding energy calculations. $W(u)$ can be estimated by a continuum solvent model as in the previous examples. If one can calculate the atomic forces resulting from $W(u)$, one can carry out a molecular dynamics simulation to generate an ensemble of structures for calculating the Helmholtz free energy according to Equation 9, although it is known to be difficult to obtain a free energy from this equation directly. If one adopts similar approximations as in previous MD/PBSA calculations, one can calculate the internal energy E of a system via

$$E = \langle W(u) \rangle, \qquad 10.$$

where $\langle \ldots \rangle$ represents an ensemble average over snapshots obtained from an implicit solvent simulation and the Helmholtz free energy can be obtained from

$$A = E - TS. \qquad 11.$$

Again, one can use the harmonic or quasiharmonic approximation to estimate the entropy term, although a more realistic implicit solvent model can be used here. Most of the previous MM/PBSA calculations employed more approximate distance-dependent dielectric models.

Several methods have already been introduced to use the relatively sophisticated Poisson-Boltzmann model to calculate electrostatic forces during molecular dynamics simulations (41, 42). A method for calculating forces resulting from solvent accessible surface area–dependent hydrophobic term have also been developed (43). However, it is still expensive to evaluate Poisson-Boltzmann forces on the fly during molecular dynamics simulations. An alternative is to use the significantly cheaper generalized Born model. Dominy & Brooks (44) have parametrized the generalized Born model by Qui et al. (45) for the CHARMM force field (45a) and found that the model performed quite well in reproducing molecular solvation energy and conformational free energy. And using this model in molecular dynamics simulations of a 56-residue protein yielded results agreeing well with corresponding explicit-solvent simulations. These results are encouraging as they demonstrate that this type of model can make it much easier to include protein flexibility, via molecular simulations, in modeling protein-drug interactions. This model has also been incorporated into the UHBD program (46, 47) to use together with constrained Brownian dynamics simulations (48). The constrained Brownian dynamics simulation algorithm can use a larger time step than molecular dynamics simulation algorithms to speed up the conformational sampling of small molecules. This is useful for improving the calculation of the free energy of floppy ligands in solution. As mentioned before, many earlier MM/PBSA simulations assumed the ligand conformational distribution in solution to be the same as that in a protein-ligand complex. This approximation may not serve well for floppy ligands and for ligands that adopt very different conformations in the bound and unbound states.

Chemical-Scanning Computational Experiments

The explicit/implicit solvent approach just described requires doing at least one simulation for each protein-ligand complex. Therefore, it is still difficult to examine the binding of a large number of compounds to a receptor. However, if one focuses on a small subset of chemical space around a lead compound, one can adopt the same approximations as described earlier in free energy calculations so that simulations on the reference systems alone can be used to predict the effects of making many modifications on a lead compound. In these calculations, no molecular dynamics simulation needs to be performed on the derivatives of a lead compound. Instead, snapshots of the reference simulations are modified

to change different functional groups of the lead compound into new ones. For example, Kuhn & Kollman (49) were able to predict a derivative that binds stronger than biotin to avidin by changing different C–H groups of biotin into C–F groups. This approach has also been applied to cases where larger modifications are made. For example, Massova & Kollman (50) performed a computational alanine scanning experiment and were still able to obtain qualitative agreement with experiments when studying protein-protein interactions even though the chemical modifications were rather large (they involved changing non-alanine amino acids into alanines).

Semi-Empirical Linear Response Theory

Another way to facilitate the comparison of the binding affinity among a number of rather different ligands is the semi-empirical linear response approach (51–56). This method assumes the binding affinity, ΔG, between a ligand-receptor pair to be approximated by the relation

$$\Delta G = \alpha\left(\left\langle U_{electrostatic}^{Bound}\right\rangle - \left\langle U_{electrostatic}^{Unbound}\right\rangle\right) + \beta\left(\left\langle U_{Lennard\text{-}Jones}^{Bound}\right\rangle - \left\langle U_{Lennard\text{-}Jones}^{Unbound}\right\rangle\right) + \gamma\left(\left\langle \Omega^{Bound}\right\rangle - \left\langle \Omega^{Unbound}\right\rangle\right), \quad 12.$$

where $\langle U_{electrostatic}^{Bound,\,Unbound}\rangle$ is the averaged ligand-surrounding electrostatic interaction energy, $\langle U_{Lennard\text{-}Jones}^{Bound,\,Unbound}\rangle$ is the averaged ligand-surrounding Lennard-Jones interaction energy, and $\langle \Omega^{Bound,\,Unbound}\rangle$ is the averaged solvent-accessible surface area of the complex or the uncomplexed molecule(s). α, β, and γ are empirical parameters determined by a least-square fit of the experimental binding free energy of a number of inhibitors to the ensemble averaged quantities of Equation 12 obtained from molecular simulations. Once α, β, and γ are determined, they can be used to predict the binding affinity of inhibitors whose binding affinity has not been measured. Encouraging results have appeared in a number of applications. For example, a recent application of this approach to study the binding of the tetrahydroimidazo[4,5,1-jk][1,4]benzodiazepine-2(1H)-thione and -one class of compounds to HIV reverse transcriptase (55) yielded a root-mean-square deviations of less than 1 kcal/mol from experimental results when the observed range of binding affinity was ∼4 kcal/mol. The Åqvist group did not use the solvent-accessible surface area term in their semi-empirical model but still obtained good correlation with experimental data in a number of applications (57).

Dynamic Pharmacophore Method

Most receptor-based pharmacophore models have been developed by using one crystal, NMR, or model structure. Pharmacophore models based on a single receptor structure could fail to identify inhibitors that bind to structures that are somewhat different from the experimental or model structure but that are still readily accessible at physiological temperatures. To address this issue, Carlson et al. (58–60) developed a dynamic pharmacophore model in which a number of

snapshots from molecular dynamics simulations were used to construct the model. For each snapshot, they determined components of a pharmacophore model by identifying favorable binding sites of chemical functional groups using the multiunit search for interacting conformers (MUSIC) program available in the BOSS program (61). The MUSIC procedure identifies favorable binding sites of probe molecules by simultaneously energy refining a large number of probe molecules, which do not interact with other, in the potential field of a drug target. Strong binding sites tend to cluster many probe molecules in well-defined orientations and locations. By carrying out MUSIC calculations on a number of dynamic snapshots, one can identify strong binding sites consistently appearing in many rather than only one or a few snapshots. These sites can form important components in a pharmacophore model. This approach can also uncover useful binding sites that are not presented by the initial starting structure. By using methanol as probe molecules, Carlson et al. (58–60) developed dynamic pharmacophore models that perform better than the single conformation model in identifying potent inhibitors of HIV-1 integrase. Unfortunately, the dynamic model also increased the number of false positives.

Relaxed Complex Methods

Recently, another computational approach has been described to discover ligands that may bind with "induced fit" of their target molecules (62, 62a). The new methods, which are called "relaxed complex methods," are inspired by two successful experimental methods for rapid discovery of ligands that bind strongly to a receptor, namely the "SAR by NMR" method (63) and the "tether method" (64). These methods recognize that ligands may bind to conformations that occur only rarely in the dynamics of the receptor, and that strong binding often reflects multivalent attachment of the ligand to the receptor. The new computational approach includes a single ligand method and a double ligand method.

The basic element of these new methods is the automated docking of small libraries of compounds to a diverse selection of target conformations. The first phase of the approach involves generating the target conformations. This might make use of a long molecular dynamics simulation of the unliganded target molecule, an ensemble of short molecular dynamics simulations, or some other way of generating target conformations. The second phase involves the rapid docking of mini-libraries of candidate inhibitors to the conformational snapshots of the target. In this phase, a relatively simple scoring algorithm is used to allow fast docking. The third phase attempts to improve the scoring of the best complexes found in the docking calculations by use of a slower but more accurate algorithm for estimating the standard free energies of binding.

The scheme described above represents the single-ligand method. The double-ligand variant recognizes that two ligands with relatively low binding affinities to the target can be linked to form a high-affinity ligand. Because the binding of the first ligand could introduce unfavorable interactions for the binding of the second

ligand, the combination of the best-ranked ligands for respective binding sites does not necessarily produce the best composite compound. Continuing from the previous single-ligand studies, the first ligand is therefore treated as part of target, and the docking simulations of the second ligand are repeated in a limited search space, based on the allowable lengths of linkers. Again, the binding of the second ligand is subsequently rescored by other more accurate approaches.

The first applications of the relaxed complex methods have focused on an experimentally well-characterized system, FKBP (FK506 binding protein) (62, 62a). A long molecular dynamics calculation was used to sample the FKBP conformations and the AutoDock software (65–67) was used for the initial docking. The rescoring was done using the MM/PBSA routines from the Amber software (68) and APBS evaluation of the electrostatic energies (69). The results to date encourage the further development and application of these methods.

CONCLUSIONS

Although protein flexibility can play an important role in determining molecular recognition and drug design, most modeling efforts ignore these effects because they are costly to include. However, with the rapid advance of computer technology and algorithmic development, it should become feasible to take these effects into account more frequently in practical drug discovery applications. More expensive but rigorous free energy calculations can be used in the later stage of a lead optimization process. Approximate but faster methods relying on single-reference states, such as sensitivity analysis and chemical scanning, can be used to quickly identify productive and nonproductive features of a lead compound in molecular recognition. The identification of these features can help decide how a lead compound should be modified to improve binding affinity, by pointing out features that are profitably kept and those that should be modified. These features can also help to construct pharmacophore models for mining new drug leads from small-molecule libraries and generate constraints for designing combinatorial chemical libraries targeted towards the desired receptors. Intermediate between these two extremes are methods such as MM/PBSA and semi-empirical linear response approach that can further screen out less promising compounds suggested by the single-reference-state models before more rigorous free energy calculations are carried out. Sophisticated implicit solvent models may also play an important role in speeding up conformational sampling so that the effects of protein flexibility can be accounted for in the earlier stages of a drug design process.

ACKNOWLEDGMENTS

The authors acknowledge support from the Howard Hughes Medical Institute, the National Institutes of Health, the National Science Foundation, the Keck Foundation, the National Biomedical Computing Resource, the San Diego Supercomputer Center, and Accelrys, Inc.

The *Annual Review of Pharmacology and Toxicology* is online at http://pharmtox.annualreviews.org

LITERATURE CITED

1. McCammon JA, Gelin BR, Karplus M. 1977. Dynamics of folded proteins. *Nature* 267:585–90
2. Tembe BL, McCammon JA. 1984. Ligand-receptor interactions. *Comput. Chem.* 8:281–83
3. Wong CF, McCammon JA. 1986. Dynamics and design of enzymes and inhibitors. *J. Am. Chem. Soc.* 108:3830–32
4. Wong CF, McCammon JA. 1986. Computer simulation and the design of new biological molecules. *Isr. J. Chem.* 27:211–15
5. Mezei M, Beveridge DL. 1986. Free energy simulations. *Ann. Acad. Sci. NY* 482: 1–23
6. Bash PA, Singh UC, Langridge R, Kollman PA. 1987. Free energy calculation by computer simulation. *Science* 236:564–68
7. Jorgensen WL. 1989. Free energy calculations: a breakthrough for modeling organic chemistry in solution. *Acc. Chem. Res.* 22:184–89
8. Beveridge DL, DiCapua FM. 1989. Free energy via molecular simulation: applications to chemical and biomolecular systems. *Annu. Rev. Biophys. Biophys. Chem.* 18:431–92
9. Zwanzig RW. 1954. High-temperature equation of state by perturbation method. I. Nonpolar gases. *J. Chem. Phys.* 22: 1420–26
10. Verkhivker G, Elber R, Nowak W. 1992. Locally enhanced sampling in free energy calculations—application of mean field approximation to accurate calculation of free energy differences. *J. Chem. Phys.* 97:7838
11. Elber R, Karplus M. 1990. Enhanced sampling in molecular dynamics: use of the time-dependent Hartree approximation for a simulation of carbon monoxide diffusion through myoglobin. *J. Am. Chem. Soc.* 112:9161–75
12. Simmerling C, Fox T, Kollman PA. 1998. Use of locally enhanced sampling in free energy calculations: testing and application to the alpha → beta anomerization of glucose. *J. Am. Chem. Soc.* 120:5771–82
13. Banba S, Guo ZY, Brooks CL. 2000. Efficient sampling of ligand orientations and conformations in free energy calculations using the λ-dynamics method. *J. Phys. Chem. B* 104:6903–10
14. Damodaran KV, Banba S, Brooks CL. 2001. Application of multiple topology λ-dynamics to a host-guest system: beta-cyclodextrin with substituted benzenes. *J. Phys. Chem. B* 105:9316–22
15. Guo ZY, Brooks CL. 1998. Rapid screening of binding affinities: application of the λ-dynamics method to a trypsin-inhibitor system. *J. Am. Chem. Soc.* 120:1920–21
16. Franck P. 1987. *Introduction to System Sensitivity Theory*. New York: Academic
17. Tomovick R, Vukobratovic M. 1972. *General Sensitivity Theory*. New York: Am. Elsevier
18. Thacher TS, Hagler AT, Rabitz H. 1991. Application of sensitivity analysis to the establishment of intermolecular potential function. *J. Am. Chem. Soc.* 113:2020
19. Susnow R, Nachbar RB Jr, Schutt C, Rabitz H. 1991. Sensitivity of molecular structure to intramolecular potentials. *J. Phys. Chem.* 95:8585–97
20. Wong CF. 1991. Systematic sensitivity analyses in free energy perturbation calculations. *J. Am. Chem. Soc.* 113:3208–9
21. Wong CF, Rabitz H. 1991. Sensitivity analysis and principal component analysis in free energy calculations. *J. Phys. Chem.* 95:9628–30
22. Zhu S-B, Wong CF. 1993. Sensitivity

analysis of distribution functions of liquid water. *J. Chem. Phys.* 99:9047–53
23. Zhu S-B, Wong CF. 1993. Sensitivity analysis of water thermodynamics. *J. Chem. Phys.* 98:8892–99
24. Susnow R, Schutt C, Rabitz H, Subramaniam S. 1994. Conformational study of dipeptides: a sensitivity analysis approach. *J. Comp. Chem.* 15:947–62
25. Zhu S-B, Wong CF. 1994. Sensitivity analysis of a polarizable water model. *J. Phys. Chem.* 98:4695–701
26. Bleil RE, Wong CF, Rabitz H. 1995. Sensitivity analysis of a 2-dimensional lattice model of protein folding. *J. Phys. Chem.* 99:3379–86
27. Zhang H, Wong CF, Thacher T, Rabitz H. 1995. Parametric sensitivity analysis of avian pancreatic polypeptide (APP). *Proteins Struct. Funct. Genet.* 23:218
28. Wong CF, Thacher T, Rabitz H. 1998. Sensitivity analysis in biomolecular simulation. In *Reviews in Computational Chemistry*, ed. KB Lipkowitz, DB Boyd, pp. 281–326. New York: Wiley-VCH
29. Del Buono GS, Freire E, Levy RM. 1994. Intrinsic pK_as of ionizable residues in proteins: an explicit solvent calculation for lysozyme. *Proteins Struct. Funct. Genet.* 20:85–97
30. Levy RM, Belhadj M, Kitchen DB. 1991. Gaussian fluctuation formula for electrostatic free-energy changes in solution. *J. Chem. Phys.* 95:3627–33
31. Simonson T, Wong CF, Brünger AT. 1997. Classical and quantum simulations of tryptophan in solution. *J. Phys. Chem. A* 101:1935–45
32. Wong CF, Hünenberger PH, Akamine P, Narayana N, Diller T, et al. 2001. Computational analysis of PKA-balanol interactions. *J. Med. Chem.* 44:1530–39
33. Gould C, Wong CF. 2002. Designing specific protein kinase inhibitors: insights from computer simulations and comparative sequence/structure analysis. *Pharmacol. Ther.* In press
34. Liu HY, Mark AE, van Gunsteren WF. 1996. Estimating the relative free energy of different molecular states with respect to a single reference state. *J. Phys. Chem.* 100:9485–94
35. Mordasini TZ, McCammon JA. 2000. Calculations of relative hydration free energies: a comparative study using thermodynamic integration and an extrapolation method based on a single reference state. *J. Phys. Chem. B* 104:360–67
36. Vorobjev YN, Almagro JC, Hermans J. 1998. Discrimination between native and intentionally misfolded conformations of proteins: ES/IS, a new method for calculating conformational free energy that uses both dynamics simulations with an explicit solvent and an implicit solvent continuum model. *Proteins Struct. Funct. Genet.* 32:399–413
37. Vorobjev YN, Hermans J. 1999. ES/IS: estimation of conformational free energy by combining dynamics simulations with explicit solvent with an implicit solvent continuum model. *Biophys. Chem.* 78:195–205
38. Kuhn B, Kollman PA. 2000. Binding of a diverse set of ligands to avidin and streptavidin: an accurate quantitative prediction of their relative affinities by a combination of molecular mechanics and continuum solvent models. *J. Med. Chem.* 43:3786–91
39. Lee MR, Duan Y, Kollman PA. 2000. Use of MM-PB/SA in estimating the free energies of proteins: application to native, intermediates, and unfolded villin headpiece. *Proteins Struct. Funct. Genet.* 39:309–16
40. Massova I, Kollman PA. 2000. Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding. *Perspect. Drug Discov. Des.* 18:113–135
41. Gilson MK, Davis ME, Luty BA, McCammon JA. 1993. Computation of electrostatic forces on solvated molecules using the Poisson-Boltzmann equation. *J. Phys. Chem.* 97:3591–600

42. Im W, Beglov D, Roux B. 1998. Continuum solvation model: computation of electrostatic forces from numerical solutions to the Poisson-Boltzmann equation. *Comput. Phys. Commun.* 111:59–75
43. Sridharan S, Nicholls A, Sharp KA. 1995. A rapid method for calculating derivatives of solvent accessible surface areas of molecules. *J. Comput. Chem.* 16:1038–44
44. Dominy BN, Brooks CL. 1999. Development of a generalized Born model parametrization for proteins and nucleic acids. *J. Phys. Chem. B* 103:3765–73
45. Qiu D, Shenkin PS, Hollinger FP, Still WC. 1997. The GB/SA continuum model for solvation. A fast analytical method for the calculation of approximate Born radii. *J. Phys. Chem. A* 101:3005–14
45a. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. 1983. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* 4:187–217
46. Madura JD, Briggs JM, Wade RC, Davis ME, Luty BA, et al. 1995. Electrostatics and diffusion of molecules in solution—simulations with the University of Houston Brownian Dynamics Program. *Comput. Phys. Commun.* 91:57–95
47. Davis ME, Madura JD, Luty BA, McCammon JA. 1991. Electrostatics and diffusion of molecules in solution—simulations with the University-of-Houston-Brownian Dynamics Program. *Comput. Phys. Commun.* 62:187–97
48. Shen TY, Wong CF, McCammon JA. 2001. Atomistic Brownian dynamics simulation of peptide phosphorylation. *J. Am. Chem. Soc.* 123:9107–11
49. Kuhn B, Kollman PA. 2000. A ligand that is predicted to bind better to avidin than biotin: insights from computational fluorine scanning. *J. Am. Chem. Soc.* 122:3909–16
50. Massova I, Kollman PA. 1999. Computational alanine scanning to probe protein-protein interactions: a novel approach to evaluate binding free energies. *J. Am. Chem. Soc.* 121:8133–43
51. Åqvist J, Medina C, Samuelsson J-E. 1994. A new method for predicting binding affinity in computer-aided drug design. *Protein Eng.* 7:385–91
52. Carlson HA, Jorgensen WL. 1995. An extended linear response method for determining free energies of hydration. *J. Phys. Chem.* 99:10667–673
53. Paulsen MD, Ornstein RL. 1996. Binding free energy calculations for P450cam-substrate complexes. *Protein Eng.* 9:567–71
54. McDonald NA, Carlson HA, Jorgensen WL. 1997. Free energies of solvation in chloroform and water from a linear response approach. *J. Phys. Org. Chem.* 10: 563–76
55. Smith RH Jr, Jorgensen WL, Tirado-Rives J, Lamb ML, Janssen PA, et al. 1998. Prediction of binding affinities for TIBO inhibitors of HIV-1 reverse transcriptase using Monte Carlo simulations in a linear response method. *J. Med. Chem.* 41:5272–86
56. Hansson T, Marelius J, Aqvist J. 1998. Ligand binding affinity prediction by linear interaction energy methods. *J. Comput.-Aided Molec. Des.* 12:27–35
57. Marelius J, GraffnerNordberg M, Hansson T, Hallberg A, Aqvist J. 1998. Computation of affinity and selectivity: binding of 2,4-diaminopteridine and 2,4-diaminoquinazoline inhibitors to dihydrofolate reductases. *J. Comput.-Aided Molec. Des.* 12:119–31
58. Carlson HA, Masukawa KM, McCammon JA. 1999. Method for including the dynamic fluctuations of a protein in computer-aided drug design. *J. Phys. Chem. A* 103:10213–19
59. Carlson HA, McCammon JA. 2000. Accommodating protein flexibility in computational drug design. *Mol. Pharmacol.* 57:213–18
60. Carlson HA, Masukawa KM, Rubins K, Bushman FD, Jorgensen WL, et al. 2000.

Developing a dynamic pharmacophore model for HIV-1 integrase. *J. Med. Chem.* 43:2100–14

61. Jorgensen WL. 1997. *BOSS*. New Haven: Yale Univ. Press
62. Lin JH, Perryman A, Schames J, McCammon JA. 2002. Computational drug design accommodating receptor flexibility: the relaxed complex scheme. *J. Am. Chem. Soc.* 124:5632–33

62a. Lin JH, Perryman A, Schames J, McCammon JA. 2002. The Relaxed Complex Method: accommodating receptor flexibility for drug design with an improved scoring scheme. *Biopolymers.* In press

63. Shuker SB, Hajduk PJ, Meadows RP, Fesik SW. 1996. Discovering high-affinity ligands for proteins—SAR by NMR. *Science* 274:1531–34
64. Erlanson DA, Braisted AC, Raphael DR, Randal M, Stroud RM, et al. 2000. Site-directed ligand discovery. *Proc. Natl. Acad. Sci. USA* 97:9367–72
65. Morris GM, Goodsell DS, Huey R, Olson AJ. 1996. Distributed automated docking of flexible ligands to proteins—parallel applications of Autodock 2.4. *J. Comput.-Aided Molec. Des.* 10:293–304
66. Goodsell DS, Morris GM, Olson AJ. 1996. Automated docking of flexible ligands: applications of AutoDock. *J. Mol. Recognit.* 9:1–5
67. Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, et al. 1998. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* 19:1639–62
68. Wang W, Donini O, Reyes CM, Kollman PA. 2001. Biomolecular simulations: recent developments in force fields, simulations of enzyme catalysis, protein-ligand, protein-protein, and protein-nucleic acid noncovalent interactions. *Annu. Rev. Biophys. Biomolec. Struct.* 30:211–43
69. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. 2001. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. USA* 98:10037–41