Chem Biol Drug Des 2013; **81:** 41–49
Special Issue-Review
Virtual Screening and Estimation of Binding Energies

Accounting for Receptor Flexibility and Enhanced Sampling Methods in Computer-Aided Drug Design

William Sinko[1,*], Steffen Lindert[2,3] and J. Andrew McCammon[1,2,3,4]

[1]*Biomedical Sciences Program, University of California San Diego, La Jolla, CA 92093-0365, USA*
[2]*Department of Chemistry & Biochemistry, NSF Center for Theoretical Biological Physics, University of California San Diego, La Jolla, CA 92093-0365, USA*
[3]*Department of Pharmacology, NSF Center for Theoretical Biological Physics, University of California San Diego, La Jolla, CA 92093-0365, USA*
[4]*Howard Hughes Medical Institute, University of California San Diego, La Jolla, CA 92093-0365, USA*
**Corresponding author: William Sinko, wsinko@ucsd.edu*

Protein flexibility plays a major role in biomolecular recognition. In many cases, it is not obvious how molecular structure will change upon association with other molecules. In proteins, these changes can be major, with large deviations in overall backbone structure, or they can be more subtle as in a side-chain rotation. Either way the algorithms that predict the favorability of biomolecular association require relatively accurate predictions of the bound structure to give an accurate assessment of the energy involved in association. Here, we review a number of techniques that have been proposed to accommodate receptor flexibility in the simulation of small molecules binding to protein receptors. We investigate modifications to standard rigid receptor docking algorithms and also explore enhanced sampling techniques, and the combination of free energy calculations and enhanced sampling techniques. The understanding and allowance for receptor flexibility are helping to make computer simulations of ligand protein binding more accurate. These developments may help improve the efficiency of drug discovery and development. Efficiency will be essential as we begin to see personalized medicine tailored to individual patients, which means specific drugs are needed for each patient's genetic makeup.

Key words: accelerated molecular dynamics, computer-aided drug design, ensemble docking, free energy calculation, molecular dynamics, receptor flexibility, relaxed complex scheme, structure-based drug design

Abbreviations: CYP, cytochrome P450s; RMSD, root mean square deviation; MD, molecular dynamics; MC, Monte Carlo; IFD, induced fit docking; RCS, relaxed complex scheme; GPU, graphics processing unit; HPC, high-performance computing; CV, collective variable; aMD, accelerated molecular dynamics; GET, guided entry of tail-anchored proteins; FEP, free energy perturbation; TI, thermodynamic integration; MBAR, multistate Bennett acceptance ratio; MM/GBSA, molecular mechanics/generalized born surface area.

The advent of pharmacogenomics has opened up many new and exciting avenues to optimize drug treatment for a variety of diseases (1). While often used to correlate drug efficiency or toxicity with specific mutations, it has great potential in combination with structure-based drug discovery as well. Particularly in the realm of computer-aided drug design, it is very conceivable that studies on individual mutant receptors can yield 'personalized drugs' tailored to target specific mutant protein forms only. These studies ideally will include full receptor and ligand flexibility to find the most optimally suited inhibitors. This review will focus on techniques that are used to account for receptor flexibility in computer-aided drug discovery studies.

Introduction to Receptor Flexibility

Often it is of great interest to know the manner in which different molecules interact. Gibbs free energy can be related to the favorability of one state over another such as state A, a drug bound to its protein target, or state B, unbound. The free energy can be related to the concentration bound and unbound, and affinity can be quantitatively expressed in this way. A large number of methods have been proposed to calculate the free energy of binding in biomolecular recognition, because this property is of great interest to those that develop drugs for biological targets. If one only needs to know the structure of a compound to calculate its binding affinity to a target of interest, it would save a great effort in screening compounds and synthesizing them. In addition, the vast majority of the chemical space possible for drug-like molecules has not been probed by chemists, but could in theory be more efficiently explored via computation if rapid and accurate methods are developed. In fact, in terms of drug-sized molecules, it is estimated that only a very small percentage of the chemical space available has been synthesized (2). Of great relevance and application have been docking and scoring functions, which provide rapid

 doi: 10.1111/cbdd.12051

ranking of compounds, sometimes estimating binding energies, by generating and scoring poses, which are three-dimensional orientations of ligands interacting with receptors. There are numerous examples of docking programs; the first widely used method was DOCK (3) and more recent versions are still in use today (4). Other commonly used programs are AUTODOCK (5), which is freely available, and GLIDE (6) distributed with the Schrödinger suite of tools[a]. These methods have proved of great use (7) because of their speed and simplicity in setup, but suffer from numerous errors because they do not account well for entropy, water effects, or protein structural changes upon ligand binding (8).

Often proteins adopt different conformations upon ligand binding than they do in their unbound form. Significant changes can be seen in the backbone structure of some proteins, while minor changes like side-chain rotations or subtle rearrangements are often important in terms of binding affinity as well. Some cytochrome P450s (CYP) including CYP2B4 undergo alpha-helical rearrangements to allow different substrates to bind (9). The infamous CYP3A4 isoform adopts major changes in conformation during binding, which is believed to be necessary for metabolism of many substrates, which vary greatly in size (1000 daltons variation), geometry, and composition. These changes have been verified by ligand co-crystallography (10–12). With all the degrees of freedom and flexibility inherent in protein structure, the changes in receptor conformation are very important in calculating the binding free energy. If the basic geometric shape of the binding site is not described correctly, then it will be quite difficult to predict compound binding correctly using methods like docking and scoring. Two theories have been proposed to describe receptor configurational change upon ligand interaction. The first is conformational selection (13,14) in which all conformations are present in the unbound receptor, but the populations of each configuration change in the bound form, meaning that the average structure is different from bound to unbound. The second theory is induced fit (15), in which one configuration is forced into another configuration by the ligand-binding event. See Figure 1 for a schematic of these two theories of biomolecular recognition and many methods to account for these conformational changes in protein–ligand interactions.

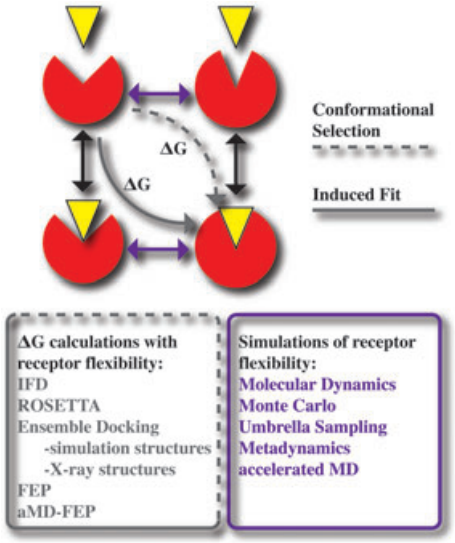

Figure 1: Schematic representation of the possible pathways taken in ligand binding. The induced fit pathways and conformational selection pathways are shown via dashed and solid lines. Various methods for binding energy calculations, which account for receptor flexibility, are listed in the gray box. Simulation methods for sampling receptor flexibility are listed in the purple box.

Docking and Scoring Functions that Account for Receptor Flexibility

Docking combined with a scoring function is a fast method for ranking compound binding or complementarity to a target of interest. First, the poses are generated by performing a three-dimensional conformational search of the binding pocket with a molecule of interest (docking), and then poses are evaluated based on a scoring function (scoring). Docking functions may perform three-dimensional searches, randomly, systematically, or by performing MD simulation of ligand conformations and then mapping to the three-dimensional structure of the receptor. Many scoring functions have been developed, and they may score using a force field-based, empirically based, knowledge-based, or consensus-based algorithm. A fine review of docking and scoring functions is given by Kitchen *et al.* (8). Often only a single rigid receptor conformation is used to dock and score each pose for computational efficiency; however, there are ways in which docking programs try to account for receptor flexibility. Soft docking, or the softening of van der Waals potentials, can allow for small overlaps between the ligand and receptor without large steric penalties (16,17). However, this may increase the rate of false positives because more diverse structures are allowed to bind. It also does not allow for larger conformational changes like side-chain rotations or protein backbone motions. Allowing for certain side chains near the active site to rotate can also account partially for receptor flexibility (18).

Induced Fit Docking

Additionally, one can use induced fit docking where the receptor is modeled with flexibility to accommodate the induced fit associated with ligand binding. One example of induced fit docking is the procedure outlined by Sherman *et al.* and implemented in the popular Schrödinger suite of tools as an option in the GLIDE docking program (19,20). Here, residue side chains are changed to alanine residues to prevent steric clashes in the initial docking. Then, side-chain predictions are used to generate possible

conformations, and the binding site and ligand are energy minimized. This protocol allows for local movements around the protein upon ligand binding, and significant reductions of the root mean square deviation (RMSD) of some binding poses as compared with crystal structures in flexible receptors (19). Nonetheless, it remains a challenge to predict larger-scale motions that may lead to different binding conformations.

RosettaLigand

An alternate method that accounts for receptor flexibility is RosettaLigand. RosettaLigand is based on the popular modeling software Rosetta (21). In contrast to the aforementioned methods, RosettaLigand does not split the receptor and ligand ensemble generation into separate steps that are combined during docking. Rather it allows for both receptor and ligand structural changes during the docking stage. Also, in contrast to the previously mentioned methods, Rosetta uses knowledge-based scoring functions to assess protein conformations while the protein–ligand interactions are modeled employing first principles. In its original implementation, RosettaLigand allowed for ligand flexibility as well as receptor side-chain flexibility (22). In a subsequent release, the developers went beyond just accounting for side-chain flexibility in the active site to incorporating full protein backbone and side-chain flexibility (23). An extensive blind docking benchmark revealed that the performance of RosettaLigand is comparable with that of other commercial software (24). In its latest release, RosettaLigand can dock multiple ligands simultaneously, allowing for the redesign of the binding interface during docking, and it has a more user-friendly xml script interface (25).

Ensemble-Based Screening Methods

Ensemble-based screening methods rely on using varied receptor conformations in a docking protocol. The conformations can be determined from crystallographic or NMR structures, Monte Carlo simulation, molecular dynamics simulation, or enhanced sampling methods. See Figure 2 for a schematic of ensemble-based docking methods. The relaxed complex scheme (RCS) (26,27) is a type of ensemble docking, which relies on previously determined conformations from molecular dynamics simulation to perform docking studies against. This ensemble of structures is then used in conjunction with a docking and scoring function rather than a single structure, with the idea that an ensemble of low energy structures will bind a larger variety of compounds, and thus more hits will be obtained from a compound library (26–28). While one could just extract structures at equidistant time-points, it seems more logical to perform a clustering-type analysis and use a single representative of related structures (28) to cover a large conformational space without redundant structures. In the final

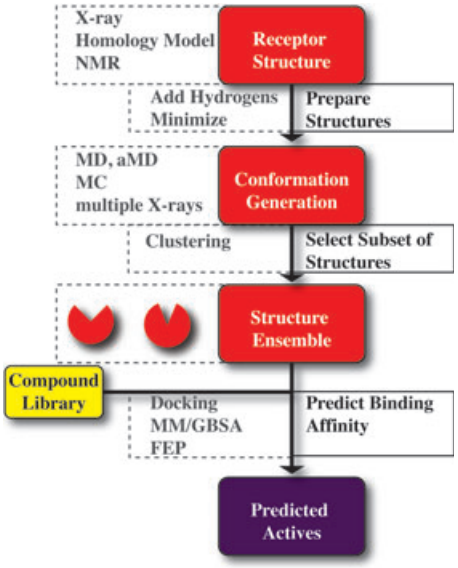

Figure 2: Schematic representation of ensemble-based docking. Examples of methods or data used at each step are given inside the dashed gray lined squares.

scoring of compounds, it may be useful to weight the score from each structure based on how often that structure was present in the original simulation, or one can simply take the highest score given. The problem remains, though, that perhaps very rare structures will be the most important in compound binding. The protein structure can be simulated in the apo state with the hope that without any compounds bound more states will be sampled as the structure will not be biased by ligand–protein interactions. However, in some structures the binding pocket may adopt a structure that does not bind ligands or occludes itself completely without a ligand bound. In this case, it is best to simulate with a ligand bound; however, the structures will be biased toward the ligand-bound state and may fail to explore a larger variety of potential ligand-binding structures.

This method has been successfully used in a number of studies to find compounds for various targets. McCammon *et al.* described a novel binding trench in HIV integrase using RCS methods and docking, which helped inspire the discovery of the FDA-approved drug raltegravir (29). Other diketo acid HIV integrase inhibitors have been modeled and designed using RCS (30). In addition, inhibitors of essential enzymes for *Trypanosoma brucei* have been found using the RCS method (31). The selection of appropriate configurations based on compound class and active site volume has been described for the proposed antibacterial target undecaprenyl pyrophosphate synthase (32). New compounds have been described, which inhibit neuraminidase in avian influenza (33). Although molecular dynamics (MD) structures can enrich active compound predictions, some structures may perform worse than X-ray structures, and a broadly

applicable protocol for *a priori* determining the most predictive structures from the simulations has not been determined (34). Indeed, ensemble-based screening can generate more hits with more chemical diversity than a single structure, but if many crystal structures are available with varied binding site geometries, the enrichment may be better for ensembles of crystal structures than for ensembles of simulation structures (35). This leads us to the conclusion that ensemble-based screening is superior to a single structure screening, but simulation-based ensemble screening may be more applicable when extensive crystallographic data is not available. Generally, the configurational changes observed with these RCS methods and induced fit methods are smaller changes. The short timescale upon which most molecular simulations are run does not fully map out biomolecular phase space. Greater sampling of biomolecular phase space as well as a broadly applicable method for determining the most predictive structures *a priori* would both be big steps forward in ensemble-based docking.

Enhanced Sampling: Hardware Improvements

Over the course of MD simulations if there is a high energy barrier between two low-energy states, it is unlikely that the simulation will cross this barrier and describe the second low-energy state, unless the simulation is very long. In addition, the free energy surface of proteins is vast and rugged and exploration is slow. The timescales often simulated with MD using today's computers are nanoseconds to microseconds and sometimes even milliseconds (36). However, many interesting events occur on the timescales of milliseconds to seconds. In addition, if one wishes to recover a Boltzmann ensemble of structures, then the crossing of energy barriers must be performed many times to generate converged statistics. It is of interest then to speed the crossing of high energy barriers, and numerous methods have been proposed to perform this.

Conceptually, the simplest improvements come from increased speed and efficiency of computer hardware and software. Two recent hardware advances have been graphics processing unit (GPU) accelerated computing and the construction of a special purpose machine Anton for running long MD simulations. GPU computing has shown impressive benchmarks for MD[b]. The common test case DHFR runs with AMBER11 was benchmarked at 30.79 ns/day on 48 Intel X5670 2.93 GHz processors, while a single NVIDIA GTX580 GPU ran 40.74 ns/day, and the GPU performance could be improved by running in parallel[c]. Also, GPUs may provide a more reasonably priced solution compared with traditional CPU-based high-performance computing (HPC). Popular simulation packages like NAMD (37), AMBER[c], GROMACS, and ACEMD[b] have released code, which runs on GPUs. A fine review of GPU computing in the context of molecular modeling is given by Stone *et al.* (38). In addition, special purpose machines like Anton have been built for running fast MD simulations and extending brute force MD simulation into the millisecond timescale (36). These extremely long simulations have been used to show drug binding events to their protein target (39). They have also been used to show in full atomic resolution how small proteins fold and that in many cases a very accurate representation of the folded state can be obtained from MD simulation (40). The accuracy of force fields on these millisecond timescales has never truly been tested before, and these new hardware developments have been an exciting development in molecular modeling. Protein flexibility on longer timescales is now being accessed much more routinely. In addition to the hardware changes, simulation techniques that help improve sampling have been developed and continue to show promise and efficiency over brute force descriptions of conformational flexibility.

Enhanced Sampling: Methodologies

One can speed sampling in a number of ways by introducing artificial biases into the model upon which the simulation is based. The simplest of artificial biases is raising the temperature, which causes more rapid fluctuations in structure by increasing the average velocity of all atoms. Temperature accelerated replica exchange uses many replicas at varying temperatures to increase sampling, and these replicas exchange with each other based on a Metropolis criterion, finally recovering the canonical ensemble (41). In addition, one may use a Hamiltonian modifying technique like accelerated molecular dynamics (aMD) in a replica exchange framework to modify only the potential energy surface at a given temperature (42). Other methods like umbrella sampling (43–45) and metadynamics (46,47) have also been used to enhance sampling. However, many of these methods suffer from the need to define a reaction coordinate *a priori* to simulating the system. This is not advantageous if one is looking for new configurations to bind a drug-like compound to a protein active site without knowledge of what that configuration is, but can be quite useful in determining the free energy change between known conformations.

Umbrella Sampling

The calculation of free energy differences from molecular dynamics simulations requires enhanced sampling techniques to probe the energy landscape effectively. Free energy difference calculations are facilitated by the use of a reaction coordinate, a parameter that measures the degree to which the system is near each of the two or more thermodynamic states of interest. One major challenge is the finite simulation time that results in regions close to energy minima being sampled well, whereas regions of higher energy are rarely or never sampled (48).

Sufficient sampling of the entire space along the reaction coordinate is necessary to derive the free energy difference. Umbrella sampling is one such technique that enforces adequate sampling of high energy regions. In umbrella sampling (or biased MD) (43–45), the underlying energy potential is modified to allow for an easier transition over the energy barrier. For this, an additional energy term, often referred to as bias or bias potential, is applied to the system to ensure efficient sampling along the entire reaction coordinate (48). This is carried out in separate simulations or windows, which overlap. Bias potentials should be chosen to ensure that sampling along the reaction coordinate is as uniform as possible. Umbrella sampling frequently employs harmonic biasing potentials in a series of windows along the reaction coordinate. An alternative to this is the use of a single window in combination with an adaptive biasing potential aimed at matching the free energy profile along the reaction coordinate as well as possible. This method is referred to as adaptive bias umbrella sampling. The sampling in each individual window can be performed using conventional molecular dynamics (cMD) or employing enhanced sampling techniques such as Hamiltonian replica exchange (41,49). Careful choice of the reaction coordinate is crucial for correct umbrella sampling results (50). The free energy curves are combined using techniques such as the weighted histogram analysis method [WHAM, (51,52)].

Metadynamics

Metadynamics is another technique designed to accelerate rare events and reconstruct the free energy profile (47). Metadynamics (46) relies on the introduction of a set of collective variables (CVs), which best describe the process of interest. Within the space of these CVs, a history-dependent potential is built up by dropping Gaussians along the sampled trajectory effectively discouraging the system from revisiting configurations that have already been sampled (53). The overall sum of these Gaussians is then used to compute a free energy profile along the CVs. For an excellent example of a metadynamics simulation, see Figure 1 in (47). The main difference to umbrella sampling is the non-systematic sampling along the collective variable(s). Important user-defined parameters are the height, width, and dropping frequency of Gaussians. Choosing the right set of CVs is crucial to the success of the simulation, and criteria for picking good CVs have been outlined in the literature (47,53). An instructive example of what happens if the correct CVs are not chosen is given in (54). Unfortunately owing to the complexity of large biological systems, determining a set of appropriate collective variables is challenging. However, this has not deterred a large number of applications using the method. Most notably to mention for this review is work that used metadynamics to characterize small molecule–protein interactions accounting for full receptor flexibility (55–59).

Accelerated Molecular Dynamics

Accelerated molecular dynamics (60), an extension of hyperdynamics (61), is a simulation method that does not require the selection of a reaction coordinate or CVs and is ideally suited for efficiently exploring configurational space. Accelerated MD modifies the potential energy surface based on the difference between a user-defined reference energy E and the normal potential energy of the underlying MD force field at each point. Accelerated MD speeds sampling by decreasing the size of energy barriers and smoothing the energetic ruggedness of phase space exploration according to a parameter α. See Figure 3 for the equations used to modify the potential energy surface and a hypothetical two-dimensional representation of the effect of aMD on a potential energy landscape. Accelerated MD has been used to simulate small benchmark systems like alanine dipeptide and reproduce the Ramachandran plot from MD (60). Originally, aMD was only applied to torsional angles (60) but was subsequently extended to all force field terms including explicit solvent (62). These two forms were combined in the dual-boost approach (63). Accelerated MD has been used for complex systems and validated with experimental results such as the improved prediction of experimental NMR observables in large protein systems (64). Grant *et al.* (65) suggested that conformational selection is the dominant mechanism for nucleotide binding-dependent conformational changes in G proteins, while induced fit plays a smaller role based on data from aMD simulations.

It is, however, important to note that recovery of Boltzmann statistics in large simulations remains a challenge because of sampling limitations in simulating large biomolecules. A non-Boltzmann simulation like aMD requires that each configuration be reweighted in a way that recovers the canonical ensemble (60). Often when longer timescales are accessed with aMD, the reweighting procedure is subject to statistical error in the estimate of the weighting factor for each point. An excellent review of the statistical issues with reweighting is given by Shen and Hamelberg

$$V^*(r) = \begin{cases} V(r), & V(r) \geq E \\ V(r) + \Delta V(r), & V(r) < E \end{cases}$$

$$\Delta V(r) = \frac{(E - V(r))^2}{\alpha + (E - V(r))}$$

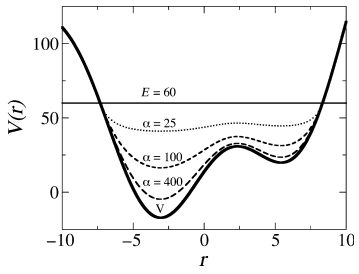

Figure 3: Accelerated Molecular Dynamics. Equations used to calculate the boost energy and modified potential energy surface in aMD (61). A two-dimensional representation of the modified potential, V*(r) (dashed lines), and the unmodified potential energy surface, V (thick black line). α was varied as indicated E was always fixed at 60 and is indicated by a thin solid black line.

(66). However, as the underlying shape of the energy surface is preserved, although flattened, one can consider the most common configurations from aMD to be likely structures although the calculation of the exact ratios of their populations is still challenging. In simulations with the goal of conformational exploration where the exact free energy of the conformation is not desired, simulations are often described without reweighting, with the risk that a high-energy state may have been over-represented. Recently, Wereszczynski *et al.* explored the conformational space of GET3 (a protein involved in the guided entry of tail-anchored proteins into the membrane) with aMD and then performed rigorous potential of mean force calculations to determine the energetics of conformational change in the presence of various bound nucleotides. GET3 simulation results agreed with experiments and were able to capture the conformational biases between open, semi-open, and closed states associated with nucleotide binding. Additionally, simulations with aMD showed much greater conformational exploration than classical MD simulations and helped suggest that the apo GET3 may explore a semi-open state when free in solution (67). Accelerated MD was also recently used to study the dynamics of the important cardiomyocyte calcium-binding protein troponin C (68). Indeed, enhanced sampling methods can play a critical role in the generation of structural ensembles with larger conformational changes, and these methods are also seeing increased use in combination with rigorous free energy calculations.

Accelerated MD in Free Energy Calculations

Free energy calculations based on computer simulations have been pursued for a few decades now, although they are not as ubiquitous in pharmaceutical settings as quicker methods of structure-based compound binding prediction like docking combined with scoring functions. Part of the problem is that they are not well automated like many docking and scoring algorithms. Another major issue is that the computational power required for these calculations is great, and a large investment in computing resources is needed before one can predict the affinity of a single ligand let alone a large compound library. We now have the computational power to make predictions of free energy, for pharmaceutically relevant ligand–receptor interactions, using methods like free energy perturbations (FEP) and thermodynamic integration (TI). These methods rely on defining a thermodynamic cycle where alchemical transitions can be used to change between states and then one can calculate a free energy difference between two states. Often the states will be the bound and unbound form of the ligand, but variations allow calculation of other properties such as the free energy of solvation, or the difference in free energy of binding between two ligands (69). One of the reasons behind the slow speed of these calculations is that the alchemical change must be made so that there is appropriate phase space overlap between the successive states along the reaction coordinate between the different end states. In thermodynamic integration calculations, similar convergence challenges arise. This requires considerable sampling of any conformational changes around the modified portion of the system, as well as accurate convergence of the energy of those conformations, which can lead to large amounts of simulation. As free energy methods rely on sampling, each simulation will be different and one can generate a simple estimate of error by running simulations in replicate, further increasing the need for computation.

Because rapid generation of low energy configurations is necessary for accurate free energy estimations, it is logical that enhanced sampling methods may be beneficially applied to these calculations. Accelerated MD has been applied in a number of ways to free energy calculations. Fajer *et al.* (42) used a replica exchange framework where the difference between replicas was the level of acceleration applied via varied boost parameters. As the ground-state replica was run on an unaltered potential energy surface, there were no issues with reweighting it as long as the replicas were spaced closely enough that they would exchange rapidly. There are, however, many ways to incorporate data from replicas at different levels of acceleration, and multistate Bennett acceptance ratio (MBAR) was determined to be roughly four times more efficient at recovering data than the ground state alone (70).

Oliveira created upside down aMD to overcome some of the issues of reweighting and calculate free energies with a single replica. The simple test system of a butane to butane symmetric transformation was used to show that one can increase the accuracy and speed of convergence by enhancing sampling with aMD in free energy calculations (71). This method allows the simulation to populate statistics by running at low to no acceleration in low energy regions, and then jumping energy barriers to rapidly move between different configurations. While quite promising in simple systems, this method proved challenging to parameterize so that high energy barriers were not flattened more than low energy barriers in simulations with many degrees of freedom. To ameliorate this issue, we proposed a boost limiting factor and created windowed aMD. This method efficiently reproduced the free energy surface of alanine dipeptide. In addition, the new boost equation was used to calculate the free energy difference between the antibiotic vancomycin and two of its glycopeptide-binding partners using TI. The overall results demonstrated more rapid convergence of free energy calculations and easily reweighted statistics; however, more parameters were required (72).

As the reweighting statistical error is directly related to the amount of acceleration and the complexity of the system, it seems reasonable to limit the portion of the simulation to which boost is applied. Acceleration of just the dihedral angles was the original way aMD was applied, and in

biomolecular systems, these dihedral terms provide much of the restraint in conformational exploration (60). This also provided a convenient limitation of the energetic terms to which acceleration was applied because water molecules, which make up the majority of the atoms in explicit solvent simulations, have no dihedral angles. Selectively applied aMD took this idea one step further and limited acceleration only to predefined dihedral angles in alanine dipeptide, and in a free energy simulation of decoupling oseltamivir's binding to neuraminidase. This work demonstrated that reweighting a small subset of dihedral angles helped overcome reweighting issues and that the free energy results converged to the same answer as cMD simulations, but in less time (73). Although this provides for better statistical recovery, it requires the prediction of which dihedral angles are important in ligand recognition or protein flexibility, which may be non-trivial.

Concluding Remarks

There has been tremendous progress in the field of personalized sequencing of genetic code within the past 20 years. The '$100 genome' is within reach in our generation. The possibility of personalized genetic knowledge of specific diseases has unprecedented potential for targeted drug treatment. This exciting development goes hand in hand with advances in computer-aided drug design that can be used to discover novel leads targeting specific mutant receptors. Here, we presented recent developments in computational algorithms and hardware used in drug discovery with a focus on using enhanced protein dynamics sampling techniques to aid in the incorporation of full receptor flexibility in structure-based drug discovery. In the near future, it should be possible to include the effects of genetic variation in models of drug targets and speed the choice of therapies appropriate for individual patients. The effects of genetic variation not only lead to sequence changes, but structural and dynamical changes too. Thus, we anticipate that computer-aided drug discovery, which accommodates receptor flexibility, will be an important component of pharmacogenomics in the near future opening many new and exciting opportunities for combining the two techniques.

Acknowledgments

This work was supported by the Molecular Biophysics Training Grant GM08326 (WS), the National Science Foundation Grant MCB-1020765, NBCR, CTBP, Howard Hughes Medical Institute, and National Institutes of Health Grant GM31749 (JAM).

Conflict of Interest

The authors declare that there is no conflicts of interests.

References

1. Wang L. (2010) Pharmacogenomics: a systems approach. Wiley Interdiscip Rev Syst Biol Med;2:3–22.
2. Editorial (2004) New horizons in chemical space. Nat Rev Drug Discov;3:375.
3. Kuntz I.D., Blaney J.M., Oatley S.J., Langridge R., Ferrin T.E. (1982) A geometric approach to macromolecule-ligand interactions. J Mol Biol;161:269–288.
4. Lang P.T., Brozell S.R., Mukherjee S., Pettersen E.F., Meng E.C., Thomas V., Rizzo R.C., Case D.A., James T.L., Kuntz I.D. (2009) DOCK 6: combining techniques to model RNA-small molecule complexes. RNA;15:1219–1230.
5. Morris G.M., Huey R., Lindstrom W., Sanner M.F., Belew R.K., Goodsell D.S., Olson A.J. (2009) AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. J Comput Chem;30:2785–2791.
6. Friesner R.A., Banks J.L., Murphy R.B., Halgren T.A., Klicic J.J., Mainz D.T., Repasky M.P., Knoll E.H., Shelley M., Perry J.K., Shaw D.E., Francis P., Shenkin P.S. (2004) Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. J Med Chem;47:1739–1749.
7. McInnes C. (2007) Virtual screening strategies in drug discovery. Curr Opin Chem Biol;11:494–502.
8. Kitchen D.B., Decornez H., Furr J.R., Bajorath J. (2004) Docking and scoring in virtual screening for drug discovery: methods and applications. Nat Rev Drug Discov;3:935–949.
9. Gay S.C., Sun L., Maekawa K., Halpert J.R., Stout C.D. (2009) Crystal structures of cytochrome P450 2B4 in complex with the inhibitor 1-biphenyl-4-methyl-1H-imidazole: ligand-induced structural response through alpha-helical repositioning. Biochemistry;48:4762–4771.
10. Yano J.K., Wester M.R., Schoch G.A., Griffin K.J., Stout C.D., Johnson E.F. (2004) The structure of human microsomal cytochrome P450 3A4 determined by X-ray crystallography to 2.05-A resolution. J Biol Chem;279:38091–38094.
11. Ekroos M., Sjögren T. (2006) Structural basis for ligand promiscuity in cytochrome P450 3A4. Proc Natl Acad Sci U S A;103:13682–13687.
12. Sevrioukova I.F., Poulos T.L. (2010) Structure and mechanism of the complex between cytochrome P4503A4 and ritonavir. Proc Natl Acad Sci U S A;107:18422–18427.
13. Monod J., Wyman J., Changeux J.P. (1965) On the nature of allosteric transitions: a plausible model. J Mol Biol;12:88–118.
14. Ma B., Kumar S., Tsai C.J., Nussinov R. (1999) Folding funnels and binding mechanisms. Protein Eng;12:713–720.

15. Koshland D.E. (1958) Application of a theory of enzyme specificity to protein synthesis. Proc Natl Acad Sci U S A;44:98–104.
16. Jiang F., Kim S.H. (1991) ''Soft docking'': matching of molecular surface cubes. J Mol Biol;219:79–102.
17. Ferrari A.M., Wei B.Q., Costantino L., Shoichet B.K. (2004) Soft docking and multiple receptor conformations in virtual screening. J Med Chem;47:5076–5084.
18. Leach A.R. (1994) Ligand docking to proteins with discrete side-chain flexibility. J Mol Biol;235:345–356.
19. Sherman W., Day T., Jacobson M.P., Friesner R.A., Farid R. (2006) Novel procedure for modeling ligand/receptor induced fit effects. J Med Chem;49:534–553.
20. Sherman W., Beard H.S., Farid R. (2006) Use of an induced fit receptor structure in virtual screening. Chem Biol Drug Des;67:83–84.
21. Simons K.T., Kooperberg C., Huang E., Baker D. (1997) Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. J Mol Biol;268:209–225.
22. Meiler J., Baker D. (2006) ROSETTALIGAND: protein-small molecule docking with full side-chain flexibility. Proteins;65:538–548.
23. Davis I.W., Baker D. (2009) RosettaLigand docking with full ligand and receptor flexibility. J Mol Biol;385:381–392.
24. Davis I.W., Raha K., Head M.S., Baker D. (2009) Blind docking of pharmaceutically relevant compounds using RosettaLigand. Protein Sci;18:1998–2002.
25. Lemmon G., Meiler J. (2012) Rosetta ligand docking with flexible XML protocols. Methods Mol Biol;819:143–155.
26. Lin J.-H., Perryman A.L., Schames J.R., McCammon J.A. (2002) Computational drug design accommodating receptor flexibility: the relaxed complex scheme. J Am Chem Soc;124:5632–5633.
27. Lin J.-H., Perryman A.L., Schames J.R., McCammon J.A. (2003) The relaxed complex method: accommodating receptor flexibility for drug design with an improved scoring scheme. Biopolymers;68:47–62.
28. Amaro R.E., Baron R., 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.
29. Schames J.R., Henchman R.H., Siegel J.S., Sotriffer C.A., Ni H., McCammon J.A. (2004) Discovery of a novel binding trench in HIV integrase. J Med Chem;47:1879–1881.
30. Di Santo R., Costi R., Roux A., Artico M., Lavecchia A., Marinelli L., Novellino E., Palmisano L., Andreotti M., Amici R., Galluzzo C.M., Nencioni L., Palamara A.T., Pommier Y., Marchand C. (2006) Novel bifunctional quinolonyl diketo acid derivatives as HIV-1 integrase inhibitors: design, synthesis, biological activities, and mechanism of action. J Med Chem;49:1939–1945.
31. Durrant J.D., Urbaniak M.D., Ferguson M.A.J., McCammon J.A. (2010) Computer-aided identification of *Trypanosoma brucei* uridine diphosphate galactose 4′-epimerase inhibitors: toward the development of novel therapies for African sleeping sickness. J Med Chem;53:5025–5032.
32. Sinko W., de Oliveira C., Williams S., Van Wynsberghe A., Durrant J.D., Cao R., Oldfield E., McCammon J.A. (2011) Applying molecular dynamics simulations to identify rarely sampled ligand-bound conformational states of undecaprenyl pyrophosphate synthase, an antibacterial target. Chem Biol Drug Des;77:412–420.
33. Cheng L.S., Amaro R.E., Xu D., Li W.W., Arzberger P.W., McCammon J.A. (2008) Ensemble-based virtual screening reveals potential novel antiviral compounds for avian influenza neuraminidase. J Med Chem;51:3878–3894.
34. Nichols S.E., Baron R., Ivetac A., McCammon J.A. (2011) Predictive power of molecular dynamics receptor structures in virtual screening. J Chem Inf Model;51:1439–1446.
35. Osguthorpe D.J., Sherman W., Hagler A.T. (2012) Exploring protein flexibility: incorporating structural ensembles from crystal structures and simulation into virtual screening protocols. J Phys Chem B;116:6952–6959.
36. Shaw D.E., Deneroff M.M., Dror R.O., Kuskin J.S., Larson R.H., Salmon J.K., Young C. *et al.* (2008) Anton, a special-purpose machine for molecular dynamics simulation. Commun ACM;51:91–97.
37. Stone J.E., Phillips J.C., Freddolino P.L., Hardy D.J., Trabuco L.G., Schulten K. (2007) Accelerating molecular modeling applications with graphics processors. J Comput Chem;28:2618–2640.
38. Stone J.E., Hardy D.J., Ufimtsev I.S., Schulten K. (2010) GPU-accelerated molecular modeling coming of age. J Mol Graph Model;29:116–125.
39. Shan Y., Kim E.T., Eastwood M.P., Dror R.O., Seeliger M.A., Shaw D.E. (2011) How does a drug molecule find its target binding site? J Am Chem Soc;133:9181–9183.
40. Lindorff-Larsen K., Piana S., Dror R.O., Shaw D.E. (2011) How fast-folding proteins fold. Science; 334:517–520.
41. Sugita Y., Okamoto Y. (1999) Replica-exchange molecular dynamics method for protein folding. Chem Phys Lett;314:141–151.
42. Fajer M., Hamelberg D., McCammon J.A. (2008) Replica-Exchange Accelerated Molecular Dynamics (REXAMD) applied to thermodynamic integration. J Chem Theory Comput;4:1565–1569.
43. Torrie G.M., Valleau J.P. (1974) Monte-Carlo free-energy estimates using non-Boltzmann sampling – application to subcritical Lennard-Jones fluid. Chem Phys Lett;28:578–581.
44. Torrie G.M., Valleau J.P. (1977) Non-physical sampling distributions in Monte-Carlo free-energy estimation – umbrella sampling. J Comput Phys;23:187–199.
45. Patey G.N., Valleau J.P. (1975) Monte-Carlo method for obtaining interionic potential of mean force in ionic solution. J Chem Phys;63:2334–2339.

46. Laio A., Parrinello M. (2002) Escaping free-energy minima. Proc Natl Acad Sci U S A;99:12562–12566.
47. Laio A., Gervasio F.L. (2008) Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. Rep Prog Phys;71:126601.
48. Kastner J. (2011) Umbrella sampling. WIRES Comput Mol Sci;1:932–942.
49. Sugita Y., Kitao A., Okamoto Y. (2000) Multidimensional replica-exchange method for free-energy calculations. J Chem Phys;113:6042–6051.
50. Rosta E., Woodcock H.L., Brooks B.R., Hummer G. (2009) Artificial reaction coordinate "tunneling" in free-energy calculations: the catalytic reaction of RNase H. J Comput Chem;30:1634–1641.
51. Kumar S., Bouzida D., Swendsen R.H., Kollman P.A., Rosenberg J.M. (1992) The weighted histogram analysis method for free-energy calculations on biomolecules. 1. The method. J Comput Chem;13:1011–1021.
52. Souaille M., Roux B. (2001) Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations. Comput Phys Commun;135:40–57.
53. Barducci A., Bonomi M., Parrinello M. (2011) Metadynamics. WIRES Comput Mol Sci;1:826–843.
54. Branduardi D., Gervasio F.L., Cavalli A., Recanatini M., Parrinello M. (2005) The role of the peripheral anionic site and cation-pi interactions in the ligand penetration of the human AChE gorge. J Am Chem Soc;127:9147–9155.
55. Gervasio F.L., Laio A., Parrinello M. (2005) Flexible docking in solution using metadynamics. J Am Chem Soc;127:2600–2607.
56. Limongelli V., Bonomi M., Marinelli L., Gervasio F.L., Cavalli A., Novellino E., Parrinello M. (2010) Molecular basis of cyclooxygenase enzymes (COXs) selective inhibition. Proc Natl Acad Sci U S A;107:5411–5416.
57. Provasi D., Bortolato A., Filizola M. (2009) Exploring molecular mechanisms of ligand recognition by opioid receptors with metadynamics. Biochemistry;48:10020–10029.
58. Masetti M., Cavalli A., Recanatini M., Gervasio F.L. (2009) Exploring complex protein-ligand recognition mechanisms with coarse metadynamics. J Phys Chem B;113:4807–4816.
59. Pietrucci F., Marinelli F., Carloni P., Laio A. (2009) Substrate binding mechanism of HIV-1 protease from explicit-solvent atomistic simulations. J Am Chem Soc;131:11811–11818.
60. Hamelberg D., Mongan J., McCammon J.A. (2004) Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. J Chem Phys, American Institute of Physics;120:11919–11929.
61. Voter A.F. (1997) Hyperdynamics: accelerated molecular dynamics of infrequent events. Phys Rev Lett;78:3908–3911.
62. de Oliveira C.A., Hamelberg D., McCammon J.A. (2006) On the application of accelerated molecular dynamics to liquid water simulations. J Phys Chem B;110:22695–22701.
63. Hamelberg D., de Oliveira C.A.F., McCammon J.A. (2007) Sampling of slow diffusive conformational transitions with accelerated molecular dynamics. J Chem Phys;127:155102.
64. Markwick P.R., Cervantes C.F., Abel B.L., Komives E.A., Blackledge M., McCammon J.A. (2010) Enhanced conformational space sampling improves the prediction of chemical shifts in proteins. J Am Chem Soc;132:1220–1221.
65. Grant B.J., McCammon J.A., Gorfe A.A. (2010) Conformational selection in G-proteins: lessons from Ras and Rho. Biophys J;99:L87–L89.
66. Shen T., Hamelberg D. (2008) A statistical analysis of the precision of reweighting-based simulations. J Chem Phys;129:034103.
67. Wereszczynski J., McCammon J.A. (2012) Nucleotide-dependent mechanism of Get3 as elucidated from free energy calculations. Proc Natl Acad Sci;109(20):7759–7764.
68. Lindert S., Kekenes-Huskey P.M., Huber G., Pierce L., McCammon J.A. (2012) Dynamics and calcium association to the N-terminal regulatory domain of human cardiac troponin C: a multiscale computational study. J Phys Chem B;116(29):8449–8459.
69. Simonson T., Archontis G., Karplus M. (2002) Free energy simulations come of age: protein-ligand recognition. Acc Chem Res;35:430–437.
70. Fajer M., Swift R.V., McCammon J.A. (2009) Using multistate free energy techniques to improve the efficiency of replica exchange accelerated molecular dynamics. J Comput Chem;30:1719–1725.
71. de Oliveira C.A., Hamelberg D., McCammon J.A. (2008) Coupling accelerated molecular dynamics methods with thermodynamic integration simulations. J Chem Theory Comput;4:1516–1525.
72. Sinko W., de Oliveira C.A.F., Pierce L.C.T., McCammon J.A. (2012) Protecting high energy barriers: a new equation to regulate boost energy in accelerated molecular dynamics simulations. J Chem Theory Comput;8:17–23.
73. Wereszczynski J., McCammon J.A. (2012) Using selectively applied accelerated molecular dynamics to enhance free energy calculations. J Chem Theory Comput;6:3285–3292.

Notes

[a]Glide, version 5.8, Schrödinger, LLC New York, NY 2012.
[b]http://www.nvidia.com/object/molecular_dynamics.html.
[c]http://ambermd.org/gpus/benchmarks.htm.