# 1

# Predictive Power of Biomolecular Simulations

*Vojtěch Spiwok*

*University of Chemistry and Technology, Prague, Department of Biochemistry and Microbiology, Technická 3, 166 28 Prague 6, Czech Republic*



Biomolecular simulations are becoming routine in structure-based drug design and related fields. This chapter briefly presents the history of molecular simulations, basic principles and approximations, and the most common designs of computational experiments. I also discuss statistical analysis of simulation results together with possible limits of accuracy.

The history of computational modeling of molecular structure and dynamics goes back to 1953, to the work of Rosenbluth and coworkers [1]. It introduced the Markov chain Monte Carlo as a method to study a simplified model of the fluid system. Atoms of the studied system were perfectly inelastic and the system was two-dimensional (2D) instead of three-dimensional (3D), so the analogy with real molecular systems was not perfect. The first molecular dynamics simulation (i.e. modeling of motions) on the same system was done by Alder and Wainwright in 1957 [2] using perfectly elastic collision between 2D particles. The first molecular simulation with specific atom types was done by Rahman in 1964 [3]. Rahman used a CDC 3600 computer to simulate dynamics of 864 argon atoms modeled using Lennard-Jones potential. The first simulation of liquid water was published by Rahman and Stillinger in 1971 [4].

Another big milestone was the first biomolecular simulation. McCammon, Gelin, and 2013 Nobel Prize winner Karplus simulated 9.2 ps of the life of the bovine pancreatic trypsin inhibitor (BPTI, also known as aprotinin) in vacuum [5]. The simulation was performed during the CECAM (Centre Européen de Calcul Atomic et Moléculaire) workshop "Models of Protein Dynamics" in Orsay, France on CECAM computer facilities [6]. It was one of the first works showing proteins as a dynamic species with fluid-like internal motions, even though in the native state.

Biomolecular simulations have undergone a huge progress in terms of accuracy, size of simulated systems, and simulated times since their pioneer times. However, the question arises whether this progress is enough for their practical application in drug discovery, protein engineering, and related applied fields. To address this issue, let me present here the concept of the hype cycle [7] developed

*Biomolecular Simulations in Structure-Based Drug Discovery,*
First Edition. Edited by Francesco L. Gervasio and Vojtech Spiwok.

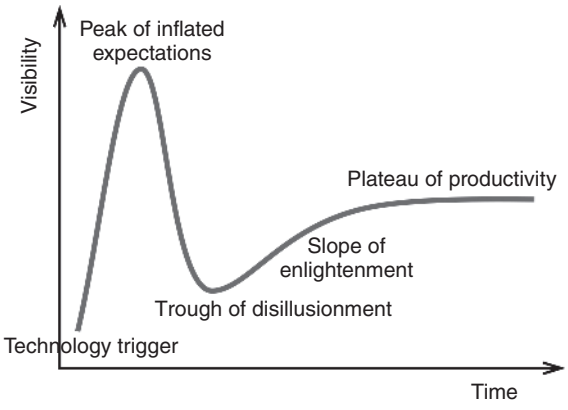


**Figure 1.1** Gartner hype cycle of inventions.

by Gartner Inc. and depicted in Figure 1.1. According to this concept, every new invention starts by a *Technology Trigger*. Visibility of the invention grows until it reaches the *Peak of Inflated Expectations*. At this point, failures of the invention start to dominate over its benefits and the invention falls into the phase of *Trough of Disillusionment*. From this phase a new and slower progress starts in the phase of *Slope of Enlightenment* toward the *Plateau of Productivity*. Biomolecular simulation passed the *Technology Trigger* and *Peak of Inflated Expectations* as many expected that biomolecular simulation would become routine and an inexpensive alternative to experimental testing of compounds for biological activity. Now, in my opinion, biomolecular simulations are located on the *Slope of Enlightenment* with a slow but steady progress toward the *Plateau of Productivity*.

## 1.1 Design of Biomolecular Simulations

Biomolecular simulations can follow different designs. I use the term design to describe the setup of the simulation procedure chosen in order to answer the research hypothesis. There are three major designs of molecular simulation. The first design starts from a predicted structure of the molecular system, which we want to evaluate, for example, a protein–ligand complex predicted by a simple protein–ligand docking. I refer to this as the *evaluative design* (Figure 1.2). The research hypothesis is: Does the predicted structure represent real structure? The basic assumption behind this design is that an accurately predicted structure of the system, for example, an accurately modeled structure of the complex, is lower in free energy than an inaccurately predicted one. The system therefore tends to be stable in a simulation starting from an accurately modeled structure and tends to be unstable in a simulation starting from an inaccurate structure. The evaluative design can be represented by the study of Cavalli et al. [8]. This study was published in 2004, and simulated times are therefore significantly shorter (typically 2.5 ns) than those available today. Nevertheless, the same length of simulations can be used today with much higher throughput in terms of the number of tested compounds or their binding poses; therefore, the study is still highly actual. Docking of propidium into human acetylcholine esterase (Alzheimer disease target) by

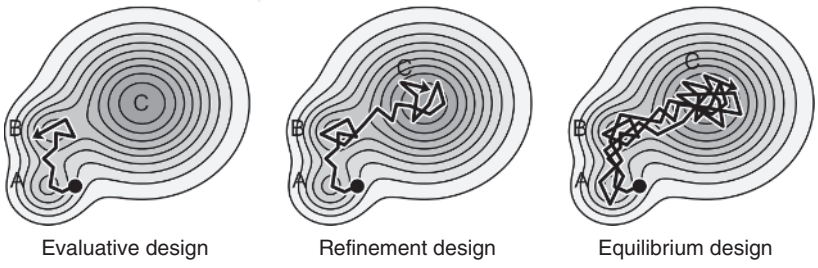


**Figure 1.2** Schematic illustration of designs of biomolecular simulations. Horizontal dimensions correspond to coordinates of the system, and contours correspond to the free energy.

the program Dock resulted in the prediction of 36 possible binding poses (clusters of docked binding poses). Six of them were then subjected to 2.5-ns simulation. Evolution of these systems was analyzed in terms of root-mean-square deviation (RMSD). Binding poses with high stability in simulations were similar to experimentally determined binding poses for a homologous enzyme.

The second design is referred to as *refinement design* (Figure 1.2). It uses an assumption similar to the evaluative design, i.e. that molecular simulations tend to evolve from high-free energy states to low-free energy states. In the refinement design, it is hoped that the dynamics can drive the system from the predicted structure, even though incorrectly predicted, to global free energy minimum, the correct structure, or at least close to it. Naturally, shorter simulation times are necessary to demonstrate correctness or incorrectness of a model by the evaluative design. Longer simulation times are necessary to drive the system from the incorrect to the correct state by the refinement design. In the previous paragraph, I used the study of Cavalli et al. from 2004 [8] as an example of evaluative design. I can present the refinement design on the work published by the same author 11 years later [9]. They used unbiased simulation to predict the binding pose of picomolar inhibitor 4′-deaza-1′-aza-2′-deoxy-1′-(9-methylene)-immucillin-H in human purine nucleoside phosphorylase. They carried out 14 simulations (500 ns each) of the system containing the trimeric enzyme, 9 ligand molecules (to increase its concentration) placed outside the protein molecule, solvent, and ions. From these simulations, 11 evolved toward binding with a good agreement with the experimentally determined structure of the complex. RMSD from the experimentally determined structure of the complex dropped during these simulations from approximately 6 to 0.2–0.3 nm.

The last design introduced here is referred to as *equilibrium design* (Figure 1.2). In this design, we hope that the simulation is sufficiently long (or sampling is sufficiently enhanced) to explore all relevant free energy minima and to sample them according to their distribution in the real system. Naturally, the equilibrium design requires longest simulation times or highest sampling enhancement from all three simulation designs. As an example I can present the study by D.E. Shaw Research [10]. The authors simulated systems containing the protein FK506 binding protein (FKBP) with one of six fragment ligands, water, and ions. They

carried out 10-μs simulations for each ligand. The dissociation constant of a complex can be calculated from its association kinetics as $K_D = k_{off}/k_{on}$. Weak binding (high $K_D$) together with reasonably fast binding kinetics therefore implies that unbinding is also sufficiently fast. For this reason, microsecond timescales were enough to observe multiple binding and unbinding events for millimolar ligands. The fragments identified by these simulations as relatively strong binders can be selected and combined into larger compounds with higher affinity in the manner of fragment-based drug design [11]. Fragment-based drug design and molecular dynamics simulation seem to be a good combination. Fragment-based design requires testing of a low number of weak ligands. This is good, since biomolecular simulations are computationally expensive. Reciprocally, weak binding enables to use molecular dynamics simulations in available timescales. Moreover, unlike some experimental methods of fragment-based drug design, molecular simulations provide binding pose prediction that can be used to combine fragments.

The three designs described are not without pitfalls. Most of these pitfalls are caused by limitations of simulated timescales. It is often difficult or impossible to simulate timescales long enough to destabilize the structure in the evaluation design, reach the global free energy minimum in the refinement design, or obtain the equilibrium distribution in the equilibrium design. This problem can be addressed by enhanced sampling techniques discussed later in this chapter.

The main problem of the evaluative design is that many correct structures of proteins or protein–ligand complexes are relatively flexible. It is therefore difficult to decide whether high flexibility (in terms of RMSD or ligand displacement) indicates a wrong model or not.

This is not the only problem of biomolecular simulation designs. Figure 1.2 shows three minima A, B, and C. Even an incorrect model A may be separated by a large energy barrier from the structure B and from the correct structure C. This can make A stable in the timescales of an evaluative simulation. Similarly, when a refinement simulation evolves from structure A to structure B and stays there, it is not guaranteed that B is the correct structure. Finally, even if a perfect equilibrium sampling is reached between A and B, the unexplored structure C can still exist.

## 1.2 Collective Variables and Trajectory Clustering

When the system is fully sampled and equilibrium distribution of states is achieved in the equilibrium design, it is possible to calculate a free energy profile of the studied system. For this it is necessary to classify states along the trajectory. In other than equilibrium design, it is necessary to monitor the progress of the simulation. These analyses often employ the concept of collective variables (CVs). A CV is a parameter that can be calculated from the atomic coordinates of the studied system. It can be calculated in every simulation snapshot, so it can be viewed as a function of time (i.e. $\mathbf{s}(t)$). It has to be chosen so that its value changes with the progress of the simulated process. Finally, CVs should be relevant to the experiment. There are simple CVs such as distances between

atoms or geometrical centers or 3-point (valence) and 4-point (torsion) angles. RMSD from the reference structure often used to monitor stability during simulation is also an example of CV. Other more sophisticated CVs include those specifically developed for studying intermolecular interactions [12] and protein folding [13], principal component analysis (PCA), and related methods [14, 15], machine-learning-based CVs [16–18], and others.

Once values of some CV (or CVs) are calculated for all snapshots along the trajectory, it is possible to calculate one-dimensional (1D), 2D, or multidimensional histograms. These histograms can be expressed in energy units as estimated free energy surface:

$$F(\mathbf{s}) = -kT \log(P(\mathbf{s})) \tag{1.1}$$

where $F$ is a (relative) free energy surface, $\mathbf{s}$ is a multidimensional vector of CVs, $P$ is its probability distribution (histogram), $k$ is the Boltzmann constant, and $T$ is temperature. Calculation of an accurate free energy surface requires complete sampling of all relevant states of the simulated system. Its accuracy is addressed later.

A discontinuous alternative to CVs is trajectory clustering. Cluster analysis of simulation coordinates (usually preprocessed by fitting to a reference structure to remove translational and rotational motions) makes it possible to place each simulation snapshot to a certain cluster. Similar to CVs, it is possible to estimate free energy surface as

$$F_i = -kT \log(P_i) \tag{1.2}$$

where $F_i$ and $P_i$ are free energy and probability, respectively, of the $i$th cluster. Several clustering algorithms, general as well as tailored for molecular simulations, have been tested in the analysis of molecular simulations. Several packages and tools have been developed for trajectory clustering, namely, the gmx cluster from Gromacs package [19], Gromos tools [20], CPPTRAJ from Amber package [21], and stand-alone packages Bio3D (for R) [22], MDAnalysis (for Python) [23] and MDTraj (for Python) [24]. Many of these tools make it possible to analyze trajectories in terms of both clusters and CVs. Popular algorithms for trajectory clustering are nonhierarchical $K$-means [25], $K$-medoids [26], and Gromos algorithm by Daura and coworkers [27]. Hierarchical methods can be used for a tree-based representation of free energy surfaces [28], but they are often used together with nonhierarchical methods to reduce the number of clusters.

A key question in application of nonhierarchical clustering methods, such as the $K$-means or $K$-medoids algorithm, is the choice of the value of $K$ – the number of clusters. This question is general, not related only to the analysis of molecular dynamics trajectories. Interestingly, the solution of this problem by "Clustering by fast search and find of density peaks," was developed by molecular scientists, namely, by Laio and Rodriguez, and became widely used in nonmolecular sciences [29]. This method automatically chooses a suitable number of clusters on the basis of density of points.

The result of a CV-based analysis of a molecular trajectory is a one-, two-, or multidimensional probability distribution or a free energy surface. The result of cluster analysis is a list of clusters with representative structures or

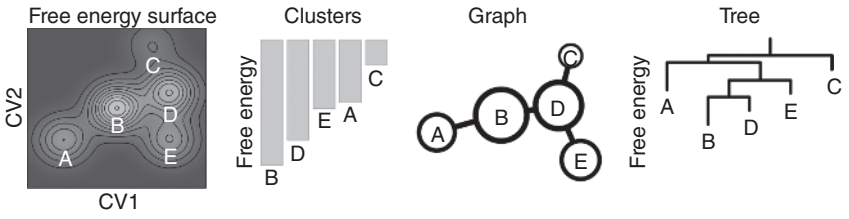


**Figure 1.3** Alternative representations of free energy relationships (schematic views).

centroids and with corresponding probabilities or free energies. Alternatively, it is possible to represent clusters in graph-based or tree-based representations. The graph-based representation [30] shows free energy minima as graph nodes. Connection of two nodes by edges usually indicates that a transition between these nodes is kinetically favorable. The tree-based representation [28] shows free energy minima as nodes and transitions as branches. Finally, the Markov chain model is another elegant way to represent free energy surface. This approach is presented in Chapter 4. Different representations of free energy relationships in molecular systems are depicted in Figure 1.3.

## 1.3 Accuracy of Biomolecular Simulations

The predictive power of molecular simulations depends on their accuracy. The accuracy is influenced by accuracy of simulation methods, molecular mechanics (MM) potentials (also referred to as force fields, mathematical models used to calculate potential energy, and forces based on atomic coordinates) and on completeness of sampling of all relevant states of the studies system. Accuracy of simulation methods has been assured by the development of sophisticated thermostats, barostats, and electrostatics models in the past decades. Application of these models and methods nowadays avoids most simulation artifacts. Nowadays one of the few important method-related artifacts in biomolecular simulations is self-interaction in the periodic boundary condition because many researchers tend to minimize the simulated system to increase the simulation speed.

The second ingredient in biomolecular simulations is the MM force field. Exciting quantum mechanical (QM) or mixed QM/MM simulations are not discussed here. Force fields have been the subject of intensive development focused on their accuracy. Evaluation of the accuracy of molecular simulations is not trivial. For example, force field accuracy can be simply tested by comparing energies calculated by the force field and by an accurate reference method, for example, by some quantum chemistry method. However, this evaluation approach is tricky. Individual bonded and nonbonded force field terms differ significantly in their magnitudes. For example, a small change in a bond angle can be associated with high change of energy. In contrast, formation of non-covalent interactions is usually associated with much lower energy changes. Both these terms can contribute differently to overall accuracy of predictions made by molecular simulations. As a result, a force field that seems to be inaccurate by comparison of energies may be, in fact, pretty accurate in practical application and vice versa.

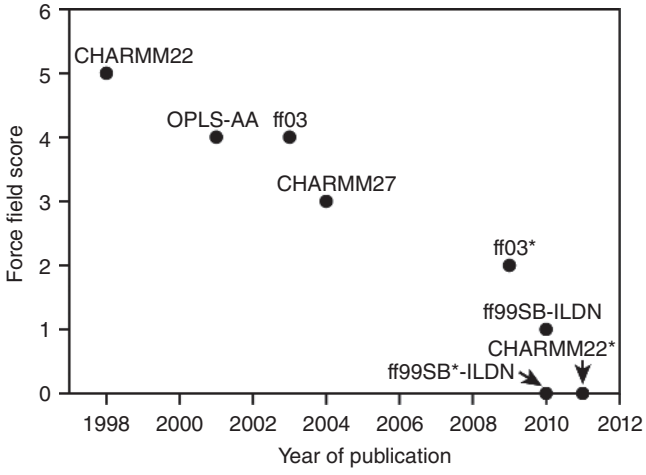


**Figure 1.4** Improvement of force fields over time. Each force field was evaluated in three simulation tasks and awarded 0–2 points per task depending on the agreement with experimental data. Low scores indicate good agreement with experiments. Source: Taken from Lindorff-Larsen et al. [31], Creative Commons Attribution License.

The progress in accuracy of MM potential can be illustrated by Figure 1.4 from the work of Lindorff-Larsen et al. [31]. These authors systematically tested MM potentials for proteins developed from 1998 to 2011. These potentials were tested by very long simulations of a folded protein and protein folding process. Each potential was given a score from 0 to 6 depending on agreement of simulations with experimental data (0 for the best agreement). Figure 1.4 shows a steady progress in accuracy, with no major accuracy issues in two force fields published in 2010 and 2011. This progress fits well into the picture of the hype cycle with a slow but steady and systematic improvement in the field in the *Slope of Enlightenment*.

One problematic feature of most MM force fields is the absence of polarizability. Conventional force fields model atoms as charged points. In reality, charge distribution changes dynamically as a response to the environment. Polarizable versions of CHARMM [32] and special AMOEBA force fields [33] were developed.

Main developers of protein force field also develop compatible general force fields for ligands, either under the same title (such as OPLS3 [34]) or under an alternative name (General Amber Force Field, or GAFF [35], for the Amber force field series or CHARMM General Force Field, or CGenFF [36] for the CHARMM force field series). Some force field developers also provide online tools for generation of force field parameters for an uploaded compound in *mol2* or *pdb* format, such as CGenFF web [36] and SwissParm [37] for CHARMM or LigParGen [38] for OPLS-AA. A web-based graphical user interface for CHARMM, known as CHARMM-GUI [39], also provides this functionality, besides other features such as membrane setup for membrane protein simulations.

When comparing protein and general molecule force fields, the situation is not so bright for general molecules. General druglike molecules are much

more diverse than 20 amino acid residues. Therefore, at least early force fields for general small molecules contained utterly erroneous terms, for example, wrong hybridization types. Evolution of general force fields corrected most of these errors; nevertheless, development of force fields applicable for all druglike molecules is challenging and these force fields are still inaccurate for many classes of compounds.

Systematic evaluation of force fields by comparison of energies calculated by force fields and by quantum chemistry methods for optimized structures [40] revealed that most problematic molecules are flexible multitorsion molecules or molecules with unusual conjugation of double bonds; however, the relationship between the structure and force field inaccuracy is not clear.

Also, modeling of interactions between a protein and a ligand can be affected by ligand force field inaccuracies or incompleteness. Widely discussed in this context is a halogen bond C—X···A, where X is a halogen (usually other than fluorine) and A is a conventional hydrogen bond acceptor, typically oxygen [41]. It has been shown that this type of bond is common in recognition of druglike molecules [42]. Classical D—H···A hydrogen bond is modeled by most force fields as a combination of electrostatic attraction and van der Waals repulsion between H and A. Since halogens in organic molecules as well as hydrogen bond acceptors are partially negatively charged, interactions between these two groups are rather repulsive. The origin of the halogen bond is in unusual distribution of electrons, referred to as sigma hole, in halogens bound in organic molecules. This phenomenon is usually not modeled by conventional force fields. A new atom type of halogen bond donor atoms has been introduced into the ligand version of optimized potentials for liquid simulations (OPLS) force field and this force field was successfully applied in computational prediction of binding free energies of HIV reverse transcriptase inhibitors [42].

It is possible to improve the accuracy of an individual modeled molecule instead of trying to improve the force field as a whole. Several approaches and tools have been developed for this purpose. For example, it is possible to improve CHARMM force fields using the Force Field Toolkit (ffTK) [43], which is a plugin for a popular visual molecular dynamics (VMD) viewer [44]. Another effort to improve accuracy of simulation of protein–ligand complexes is a repository of ligand parameters. At the website www.ligandbook.org it is possible to find parameters of approximately 3000 molecules in different force fields and for different program packages [45].

## 1.4 Sampling

The necessity to use femtosecond integration steps together with the fact that each atom in a condensed biomolecular system interacts with another approximately 5000 atoms (considering 2 nm as an interaction cutoff) causes biomolecular simulations that are extremely computationally expensive. The history of biomolecular simulations is tightly connected with availability of computer power. The 1980s were characterized by the introduction of personal computers and a boom in academic supercomputers. The 1990s were

characterized by parallelization, i.e. joining of inexpensive computers to larger clusters. Other ideas, such as distributed computing projects using computer power of volunteers' PCs [46], use of GPUs [47], and special purpose computers [48], were introduced later. As a result of the progress in computer power, the first biomolecular simulations studied picosecond timescales, nanosecond simulations became available in the early 1990s, the first microsecond simulations were carried out in the late 1990s, and the milliseconds milestone was reached in around 2010. However, it must be kept in mind that these timescales were typically reached for small molecular systems on cutting-edge hardware and at the time of their publication were far from routine.

Sampling of a biomolecular system can be compared to the situation when a department store manager wants to evaluate the "affinity" of customers to different parts of the department store he manages. It is possible to choose a certain customer and follow his or her route through the department store. It is then possible to calculate probability for individual departments as a ratio of time spent in the department divided by the total time. It is also possible to use Eq. (1.1) to express this probability as free energy (temperature is discussed later). However, this approach, equivalent to the classical molecular dynamics simulation, is inefficient because the customer may stay for a long time in some department and it can take a very long time to sample all departments.

An alternative in the molecular world to running very long simulations is application of enhanced sampling techniques. These techniques were designed to provide equivalent information as several orders of magnitude longer conventional (unenhanced) simulations. There is a group of enhanced sampling techniques that use a bias force or bias potential to accelerate the studied process. Other methods use elevated temperature or other principles. Several hybrid sampling enhancement methods combining multiple principles have been also developed.

Simulations using a bias potential or a bias force, further referred to as biased simulations, include the umbrella sampling method [49], metadynamics [50], steered molecular dynamics [51], local elevation [52], local elevation umbrella sampling [53], adaptively biased molecular dynamics [54], variationally enhanced sampling [55], flying Gaussian method [56], and others. These methods can be divided into two groups depending on whether the bias potential or force is static or dynamic.

The method known as umbrella sampling uses a static bias potential. In the analogy to the department store presented, it is possible to represent it by organizing sales in some unattractive departments and hiking prices in attractive ones. This will make sampling much more efficient. Provided that it is possible to quantify the effect of sales and price elevations, it is possible to calculate the equilibrium probabilities (probabilities under condition of regular prices) from sampling and from price modifications.

Umbrella sampling introduced by Torrie and Valleau in 1977 [49], originally in connection with the Monte Carlo method, represents methods with a static bias potential (some scientists use the term umbrella sampling as a synonym for any simulation with a static bias potential). In the most common design, it is used to enhance sampling along certain CVs (e.g. protein–ligand distance) to predict the corresponding free energy surface. Umbrella sampling is done by running

a series of simulations, each with a bias potential $k(s - S_i)^2/2$, where $k$ determines strength of the bias potential, $s$ is the CV, and $S_i$ (for $i$th simulation) ranges from the initial $S_0$ and the final state $S_N$ of the simulated process (e.g. bound and unbound state) and is usually uniformly distributed. This potential forces the ligand to sample all states along the binding pathway. Free energy surface can be calculated by, for example, weighted histogram analysis method (WHAM) [57] or by the reweighting formula [58–60]. These methods are explained later; so, briefly, it is possible to calculate unbiased sampling from the knowledge of the biased sampling and the bias potential. An example of umbrella sampling in drug discovery is the study of Bennion et al. [61]. They simulated permeation of drug molecules through the membrane. They used a coordinate perpendicular to the membrane as the CV. This CV was ranging from 0 to 10 nm in 0.1-nm windows (i.e. 100 simulations). They correctly ranked tested compounds as impermeable; low, medium or highly permeable; and in a good quantitative agreement with parallel artificial membrane permeability assays (PAMPA).

Biased simulation with a time-dependent bias potential can be represented by the metadynamics method [50]. In the department store example, it is possible to carry out metadynamics using a device that, at regular intervals, releases some stinky compound. Such a device must be installed onto a customer's shopping basket. If the customer stays for a long time in some department, the device causes the stinky compound to accumulate there. This forces the customer to escape the department and to visit other departments. This makes sampling much more efficient. The free energy surface can be estimated from the amount of the stinky compound, i.e. deep minima require a high amount of the stinky compound.

In the molecular world, that application of metadynamics starts with choice of CVs, typically two. The system is then simulated by conventional simulation for 1 or 2 ps. Then, values of CVs are calculated and recorded as $\mathbf{S}_1$. From this point, a bias potential in the form of a Gaussian hill centered in $\mathbf{S}_1$ is added to the simulated system. The system evolves for another 1 or 2 ps, then another hill is added to $\mathbf{S}_2$, and so forth. The bias potential accumulates in certain free energy minima until this minimum is flooded and the simulation can escape it. This allows for complete sampling of the free energy surface. The free energy surface can be estimated as the negative value of the bias potential [50, 62, 63], because the deeper the free energy minimum, the more hills it needs to flood.

The accuracy of metadynamics (and other biased simulations) is critically dependent on the choice of CVs. Ideally, the CVs must account for all slow degrees of freedom in the simulated system. Existence of some slow degree of freedom not addressed by CVs may cause a significant drop of accuracy. Imagine a simulation of protein–ligand interaction. Naturally, one of the CVs for protein–ligand interaction modeling can be the protein–ligand distance to accelerate binding and unbinding. The second CV should address other slow motions. Imagine the situation that the entrance to the binding site may be occasionally blocked by some amino acid side chain. If the site is blocked, the ligand cannot move inside or outside the binding site. This leads to a huge overestimation or underestimation of the predicted binding free energy.

An ideal solution to this problem would be a second CV that fully addresses side chain motions. It is difficult to design such CVs due to the complexity of

the molecular system because there could be multiple problematic side chains or other degrees of freedom. Instead, most researchers rely on sampling. Simulations in timescales of hundreds of nanoseconds or microseconds are usually not long enough to simulate binding and unbinding events, but it is often sufficient to sample such problematic degrees of freedom once binding and unbinding is enhanced.

However, in classical metadynamics, this may cause the problem of hysteresis in the predicted free energy surface due to altering overestimation of the bound and unbound state. This problem can be addressed by well-tempered metadynamics [64]. Well-tempered metadynamics is metadynamics with variable hill heights. The height set by user is scaled by $\exp(-V_{\text{bias}}(s)/k\Delta T)$, where $\Delta T$ is the difference between sampling temperature and the temperature of the simulation. Classical metadynamics corresponds to $\Delta T =$ infinity and unbiased simulation to $\Delta T = 0$. Flooding of the free energy surface in well-tempered metadynamics slows down until its convergence. The free energy can be calculated as a negative value of the bias potential scaled by $(T + \Delta T)/\Delta T$. The fact that the biasing slows down reduces the hysteresis and increases the accuracy. For this reason, well-tempered metadynamics replaced classical metadynamics in the past decade. Well-tempered metadynamics, together with a funnel method (described later), was used to simulate binding and unbinding and to accurately predict binding free energies for ligands of GPCR, including cannabinoid $CB_1$ [65], $\beta_2$ adrenergic [66], chemokine CXCR3 [67], and vasopressin [68] receptors.

In the previous paragraph I assumed that a single CV cannot address all slow degrees of freedom. However, it is possible to address many slow degrees of freedom by multiple CVs. It has been shown that metadynamics with more than two or three CVs is not efficient [69]. A special variant called bias exchange metadynamics [70] was developed to run metadynamics with multiple CVs. The system is simulated in multiple ($N$) replicas (usually one per processor CPU), where $N$ is the number of CVs. Metadynamics biases a single CV in each replica (or there could be some unbiased replicas). Occasionally (every few picoseconds) coordinates are exchanged on the basis of an exchange criteria calculated from potential energies and bias potentials in each system. This makes it possible to predict a one-dimensional free energy surface for each CV. Calculation of a multidimensional free energy surface requires a special reweighting procedure [71]. The bias exchange metadynamics has been applied in predicting the binding mode of the compound SSR128129E to fibroblast growth factor receptor [72].

Sampling can be also enhanced by elevated temperature. In the department store example, it is possible to find an analogy between temperature and the music played in the store. It has been shown experimentally that a tempo of music in a supermarket influences the pace of shoppers [73]. It is therefore possible to enhance sampling by playing a fast-paced music. However, by this we would obtain a different free energy surface from the normal music played in the department store. For example, fast moving customers would prefer easy-to-find departments and shelves and would ignore difficult-to-find ones.

Similarly, in a high-temperature molecular simulation, we would obtain a free energy surface different from the normal temperature. Such a free energy surface is usually not interesting. For example, the "native" structure of a protein at a temperature higher than its melting temperature is the unfolded structure.

There is a method that makes it possible to use elevated temperature to enhance sampling and at the same time to obtain normal-temperature free energy surfaces. This method is known as parallel tempering and belongs to the family of replica exchange methods. In the department store analogy, it would be necessary to distribute radios with headphones to multiple customers. Customers would listen to music differing in the tempo. In periodic intervals, their music would be exchanged based on the special criteria. Normal-tempo free energy surface would be obtained by the analysis of trajectories of only those customers who listen to the normal-tempo music.

In a molecular system, it is possible to run parallel tempering by simulation of multiple replicas of the system at different temperatures. These temperatures are chosen so that the lowest is slightly lower than the normal temperature and the highest is high enough to significantly enhance sampling. Replica exchange attempts are evaluated usually every 1 or 2 ps. The potential energy of the $i$th replica is compared with the potential energy of the $i+1$th replica. If the potential energy of colder replica is lower, the coordinates of replicas are swapped. If not, the Metropolis criterion is calculated as $\exp((E_i - E_{i+1})(1/kT_i - 1/kT_{i+1}))$. If a random number (with a uniform distribution from 0 to 1) is lower than the Metropolis criterion, the coordinates in the replicas are also swapped. If the simulated system adopts an unfavorable (high-energy) structure, it tends to be exchanged for higher temperature replicas and to climb on the temperature ladder. There it can adopt some nice structure with low energy. Once this happens, it would tend to descend on the temperature ladder. Structures sampled at the temperature of interest can be analyzed by Eq. (1.1) to obtain the corresponding free energy surface.

Parallel tempering is a very powerful method for folding of mini-proteins. It is particularly suitable for simulation of small systems because large systems require a huge number of replicas to reach reasonable exchange rates (with a low number of exchanges, the method would behave as a series of independent unbiased simulations). I see the highest potential of parallel tempering in drug design in combination with other sampling enhancement methods. Parallel tempering in combination with metadynamics [74] has been applied to compare wild-type and oncogenic mutants of the epidermal growth factor receptor [75].

An interesting multiple replica method that enhances sampling by cloning and merging replicas is WExplore [76]. This method simulates the system in a constant number of replicas. When two or more replicas sample similar states, they are merged. If a single replica samples some distant state, it is cloned. The free energy method can be obtained from sampling and from cloning and merging history. This method was successfully applied in modeling of the interaction between 1-(1-propanoylpiperidin-4-yl)-3-[4-(trifluoromethoxy)phenyl]urea (TPPU) and its enzyme target epoxide hydrolase [77].

## 1.5 Binding Free Energy

So far I have presented methods that can be used for general prediction of free energy relationships. Here I present special issues of modeling of protein–ligand

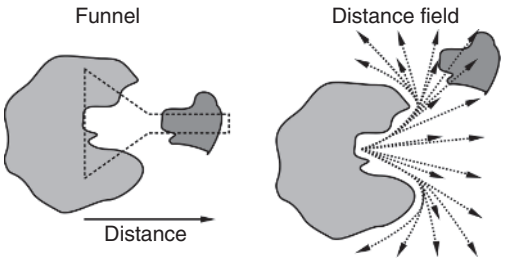


**Figure 1.5** Schematic representation of funnel techniques and distance field CV.

interactions and molecular simulation methods especially suited for modeling molecular recognition. An important issue of such simulations is the fact that the entrance into the binding site usually represents only a small part of the overall protein surface. It may seem like a good idea to enhance simulation of protein–ligand binding by biasing the distance between the binding site and the ligand. However, it may happen that the ligand chooses a wrong entrance into the binding site. This significantly slows down the simulation and makes the predicted free energy surface inaccurate. There are two major approaches developed to address this problem. First, is the abovementioned application of a funnel (Figure 1.5), first introduced as funnel metadynamics [78]. The ligand is restrained into a funnel-shaped space outside the binding site by means of an artificial potential. This prevents the ligand from exploring other entrances into the binding site. The result of such a simulation is the free energy difference between the bound state and the state when the ligand resides at the tip of the funnel. A simple correction can be applied on this value to obtain the absolute binding free energy, considering ligand concentration, volume of the system, and the volume of the funnel [78]. The method has been successfully applied in G protein–coupled receptor (GPCR) research [65–68].

An alternative to a funnel is a distance field (Figure 1.5) [79]. Instead of the Euclidean distance between the binding site and the ligand, it is possible to measure the shortest path from the binding site and the ligand without their collisions. At the beginning, a three-dimensional grid is constructed in the simulation box. For each point on the grid (except those inside the protein) a collision-free distance between the binding site and the ligand is calculated. Next, in the simulation it is possible to estimate this distance from grid points close to the ligand position. This approach has been applied together with Hamiltonian replica exchange simulation to study binding of 14-3-3ζ domains with phosphopeptides [80].

The so-called Alchemistic methods can be used to predict binding free energy without simulating the binding process. The term "Alchemistic" indicates that some elements change into other elements, similarly to medieval alchemists attempting to produce gold from inexpensive metals. These methods typically do not provide absolute binding free energies. Instead, they make it possible to predict an outcome of a modification of the ligand, for example, change of hydrogen to halogen, addition of a small group, or other minor modifications. More complex modifications can be studied by combination of multiple Alchemistic simulations.

Alchemistic methods such as free energy perturbation, thermodynamic integration, or Bennett acceptance ratio method use a series of nonphysical processes to study a physical process. For example, it is possible to predict the outcome of a replacement of a hydrogen atom in a ligand by chlorine, i.e. the difference between binding free energy of the ligand L–Cl and the ligand L–H. First, the complex protein – L–H is simulated and its force field parameters are gradually changed (linearly or nonlinearly) into parameters of L–Cl; that is, the increase in mass from 1 to 35.45, the increase in bond length from ∼1 to ∼1.8 Å, etc. The response of energy of the system is monitored. This response makes it possible to predict the free energy difference of a nonphysical (experimentally unfeasible) process of changing of H to Cl on a protein-bound ligand. In addition, it is possible to do the same calculation for an unbound ligand and to construct a thermodynamic cycle comprising (i) binding of L–H to protein, (ii) change of bound L–H to L–Cl, (iii) unbinding of L–Cl, and (iv) change of unbound L–Cl to L–H. Despite the fact that two of these processes are nonphysical, the overall free energy change of the thermodynamic cycle is zero. It is therefore possible to predict $\Delta\Delta G$ (the difference of binding $\Delta G$ of L–Cl versus L–H). This can give an answer to whether the change of H to Cl strengthens or weakens the binding to the protein. A good example of application of Alchemistic simulation is the campaign leading from a weakly binding docking hit to a picomolar inhibitor of HIV integrase by Jorgensen's group [81–84].

Finally, several methods have been developed to predict binding free energies from molecular simulations without simulating the binding process. These methods assume that the affinity is determined by the strength of non-covalent intermolecular interactions. The ligand is simulated as a complex in the target and, in parallel, in a solvent. Non-covalent interactions are monitored in both simulations and they are used to predict binding free energy and the effect of ligand desolvation. Examples are linear interaction energy [85] and methods combining MM with implicit solvent models (molecular mechanics/Poisson–Boltzmann surface area (MM/PBSA) and molecular mechanics/generalized born surface area (MM/GBSA)) [86]. Wright et al. used the MM/PBSA and MM/GBSA method to predict binding free energies of nine HIV-protease inhibitors approved for HIV treatment [87]. This study is an example of replications in simulations. The authors used short simulations (4 ns) done in 50 independent replicates for each molecule to obtain a robust model with a good agreement with experiment.

## 1.6 Convergence of Free Energy Estimates

Experimental researchers use replication to assess and improve accuracy of their predictions. In the spirit of the central limit theorem, measurements done in multiple replicates can be averaged to estimate the mean value. Standard deviation or standard error of the mean can be used to assess the accuracy. Measurements done in replicates are also used to statistically test research hypotheses.

In principle, replications can also be used in biomolecular simulations; however, most researchers prefer prolonging their simulations rather than replicating

them. It is possible to use experiment, such as nuclear magnetic resonance (NMR) measurement, to determine a dissociation constant of a protein–ligand complex. By a properly designed NMR experiment it is possible to determine concentrations of the free ligand, free protein, and the protein–ligand complex (or at least ratios of their concentrations). In the other words, it is possible to determine the number of molecules in different states (free protein, free ligand, and free complex) in the studied system at a certain moment.

Instead, biomolecular simulations study a single biomolecule as a sample of the whole biomolecular system. They calculate how long the single studied system spends in different forms. A dissociation constant of the protein–ligand complex can be calculated as the ratio of times spent in the ligand-bound and the unbound form. Both concepts, concentration and time ratios, can be generalized in the way that both quantities are proportional to probabilities of states, i.e. dissociation constant can be determined as the ratio of equilibrium probabilities of different forms of the studied systems.

The main reason why replication is rarely used in biomolecular simulation is the fact that it is difficult to generate independent starting conditions. Basic molecular dynamics simulation is a deterministic method. Running two simulations from the same starting coordinates with the same starting velocity vectors should give identical trajectories. Random initialization by different starting velocities usually does not provide the satisfactory level of independence. The second reason is that many biologically interesting quantities, such as dissociation constants, require sampling of multiple transitions between the relevant states of the system.

Nevertheless, errors of some quantities of the molecular system can be calculated by a "standard" way used by experimental scientists who average results of independent experiments. These quantities include temperature, pressure, membrane surface tension, number of non-covalent interactions, experiment-related properties (e.g. fluorescence resonance energy transfer (FRET), pair and radial distribution functions or NMR quantities), molecular surface, forces acting on selected molecular degree of freedom, and others. Calculation of these properties requires that the system exist only in one form whose property we want to calculate or the transitions between forms are rapid enough.

Most interesting from the point of view of drug design is prediction of thermodynamic and kinetic quantities, especially association/dissociation constants and binding/unbinding rates of protein–ligand complexes. Calculation of these quantities requires sampling of multiple transitions between the forms of the molecular system. The equilibrium constant of the transition from form A to B can be predicted as the time spent in form B divided by the time spent in form A. A 1-μs simulation with a single transition from A to B at ∼0.5 μs would give free energy difference estimate around 0 kJ mol$^{-1}$ (i.e. $-kT \log(0.5/0.5)$). However, it is possible that the system would have stayed in state B for another 100 μs, so the real free energy difference is approximately −13 kJ mol$^{-1}$ (i.e. $-kT \log(100.5/0.5)$). On the other hand, a simulation with many A to B and B to A transitions provides good confidence that the calculated binding free energy is accurate, at least in terms of sampling.

This phenomenon can be addressed by a block analysis [88–91]. Simulation trajectories are separated into $M$ equivalently sized blocks with $n = 1$ to $N$, where

$M = N/n$, $N$ is the number of samples in the trajectory, and $n$ is the number of samples in a block. The calculated value, for example, the population of the state B $P_B$, is averaged in each $i$th block yielding $\langle P_B \rangle_i$. Next, standard deviation and standard error (block standard error, BSE) is estimated for each value of $n$ from $\langle P_B \rangle_i$ as

$$\sigma(P, n) = \sqrt{\frac{\sum (\langle P_B \rangle_i - \langle\langle P_B \rangle\rangle)^2}{M - 1}} \tag{1.3}$$

$$\mathrm{BSE}(P, n) = \frac{\sigma(P, n)}{\sqrt{M}} \tag{1.4}$$

where $\langle\langle\ \rangle\rangle$ is average across the block size $n$. This procedure can be demonstrated on sampling of a model one-dimensional energy profile with two minima at CV equal to approximately 3 (minimum A) and 7 (minimum B). These minima have the same depth, so the free energy difference is 0 and equilibrium constant is 1. It was sampled by the Monte Carlo method with CV profiles depicted in Figure 1.6. The top profile shows sampling at low temperature with few A to B and B to A transitions. A block analysis with $n = 1$–$100$ gives a divergent estimate of BSE($P$, $n$). The value for $n = 1$ corresponds to classical standard error of the mean calculated for independent samples in many fields of experimental sciences. This value is strongly underestimated due to autocorrelation of values of the CV in the trajectory. If the system is in state A, it is highly probable that it will be in state A in the next step or 10 steps later. The value of $P_B$ was calculated as 0.503 (equilibrium constant 1.01). The block analysis shows that the value of BSE($P$, $n$) rises for $n = 1$–$100$ and is not convergent (extending $n$ does not help; data not shown). It would be therefore necessary to prolong the simulation in order to obtain a convergent estimate of standard error. The situation is different in the simulation at a higher temperature depicted in the bottom profile. The $P_B$ was calculated as 0.453 and the number of A to B and B to A transitions was higher. The result of the higher number of transitions is a convergent profile of BSE. Highest BSE value (0.08) can be used as a standard error estimate, i.e. $P_B$ is equal to $0.45 \pm 0.08$ (mean $\pm$ BSE).

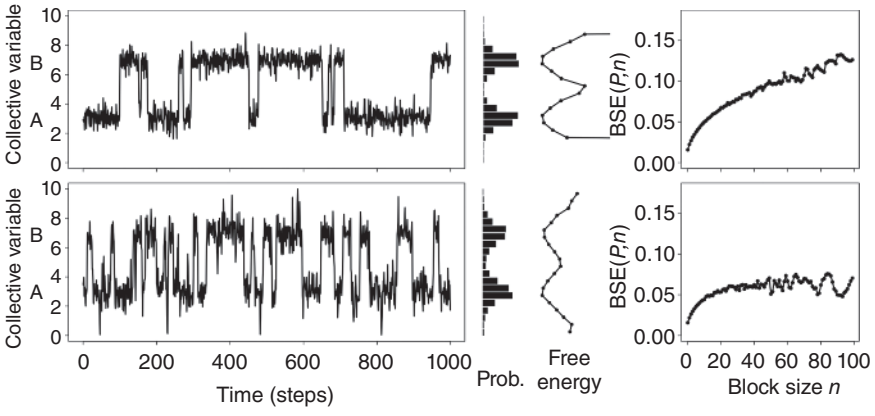


**Figure 1.6** Demonstration of block error analysis on sampling of a model energy profile.

A similar analysis can be applied on biased simulations. The easiest free energy estimation can be done from metadynamics simulations. In classical metadynamics [50, 69], it is possible to use a negative value of the bias potential as an estimate of the free energy surface [62, 63]. In well-tempered metadynamics, the free energy can be predicted as a negative value of the bias potential scaled by a constant factor $((T + \Delta T)/\Delta T)$. However, this approach does not provide an estimate of its accuracy. Simple averaging of free energy differences $\Delta G_{A\to B}$ along the simulation suffers the problem of autocorrelation in simulation trajectories. It is possible to plot the profile $\Delta G_{A\to B}$ along a metadynamics simulation with a nice convergence, but the converged $\Delta G_{A\to B}$ can be completely wrong due to a low number of A to B and B to A transitions.

The problem can be addressed by block analysis also in biased simulations [92]. As an alternative to calculating free energy surface from the bias potential, it is possible to calculate it from the combination of the bias potential and sampling. Equilibrium (unbiased) probabilities can be predicted from biased sampling by reweighting formula [58–60]:

$$P(\mathbf{S}) = \frac{\sum \delta(\mathbf{s}(t) - \mathbf{S}) \exp(+V_{\text{bias}}(t)/kT)}{\sum \exp(+V_{\text{bias}}(t)/kT)} \tag{1.5}$$

where $\mathbf{S}$ is a multidimensional vector of CVs and $\mathbf{s}(t)$ is the vector of CVs sampled at time $t$. In other words, equilibrium probabilities from biased simulations are calculated in the same way as from unbiased except that they are scaled by the factor $\exp(+V_{\text{bias}}(t)/kT)$. This is a generalization of Eq. (1.1), where $\exp(+V_{\text{bias}}(t)/kT) = 1$ in the absence of the biased potential. Similarly, the bias potential in a non-well-tempered metadynamics is constructed to make sampling of all values of $\mathbf{S}$ with the same probability, i.e. $\delta(\mathbf{s}(t) - \mathbf{S})$ is constant. This is true only if $P(\mathbf{S}) = \exp(-F(\mathbf{S})/kT)$, i.e. $F = -V_{\text{bias}}$. This idea can be extended for well-tempered metadynamics. Prediction of $P(\mathbf{S})$ using reweighting formula makes it possible to analyze the data by block analysis to predict BSE.

The problem of reweighting formula is that it should be used together with a static (time-independent) bias potential. The bias potential of metadynamics is time-dependent. With caution it is possible to use reweighting formula and considering the metadynamics bias potential as quasi-static. Alternatively, it is possible to apply corrections developed by Tiwary and Parrinello [93].

Another possibility to predict the free energy surface is application of WHAM. It should be noted that some researchers use the term umbrella sampling for any biased simulation with a static bias potential. The same researchers would call the reweighting formula in Eq. (1.5) as WHAM. However, most scientists use the term umbrella sampling for biased simulations carried out in multiple windows, where different bias potentials are used in each window and all windows cover the whole range of the CV [57]. The pair of WHAM equations

$$P(\mathbf{S}) = \frac{\sum \delta(\mathbf{s}(t) - \mathbf{S})}{\sum \exp([F_i - V_{\text{bias}}(t)]/kT)} \tag{1.6}$$

$$F_i = -kT \log \sum [P(\mathbf{S}) \exp(-V_{\text{bias}}(\mathbf{S})/kT)] \tag{1.7}$$

is solved iteratively to self-consistency to obtain $P(\mathbf{S})$ and $F_i$. It can be intuitively explained that the free energy is calculated in small fragments using reweighting

formula for each window. Simultaneously, free energy shifts $F_i$ of these fragments are calculated. The whole free energy surface is reconstructed by merging fragments of the free energy surface shifted by $F_i$. Block analysis has also been applied in WHAM [94].

Prediction of kinetics of drug binding and unbinding has become attractive for drug design [95]. Partially, this is because it is easier to sample a single drug binding or single unbinding event compared to sampling of numerous binding and unbinding events, so researchers make a virtue of necessity. Beside this, numerous experimental results show that binding or unbinding kinetics can be equivalently or even more useful in drug design compared to thermodynamics. Prediction of kinetics from unbiased and biased simulations and assessment of the accuracy of these predictions is not as developed as for thermodynamics, but there are several examples of extraction of kinetic information from unbiased simulations [96] or metadynamics [97, 98]. The Markov chain model made from biomolecular simulations is presented in Chapter 4.

## 1.7 Future Outlook

The question is: What are the limits in the predictive power of biomolecular simulations? In 1998, Xie and coworkers measured the rate of enzymatic reaction at the single-molecule level using a special fluorescence microscopy technique [99]. They found significant heterogeneity in kinetic parameters in individual enzyme molecules (standard deviation of 70% for $k_{cat}$). These heterogeneities were explained by conformational heterogeneities. A similar heterogeneity was observed in signaling by the $\beta_2$ adrenergic receptor [100]. In other words, individual enzyme or receptor molecules are highly heterogeneous in their ligand-binding or catalytic properties not only *in vivo* but also *in vitro*. The experimentally measured kinetic or thermodynamic parameters represent an averaged value across all molecules in the system. Biomolecular simulations study a single molecule. It is therefore natural that predicted parameters of biomolecular simulation may differ from experimental results due to the heterogeneity in target molecules. This problem can be, in principle, solved by replication of simulations or by enhancement of sampling of degrees of freedom associated with such heterogeneity, but none of these approaches is simple.

At the beginning of this chapter I introduced three designs of molecular simulations: evaluative, refinement, and equilibrium. The examples of studies presented later in this chapter follow almost always the equilibrium design. This can be explained by the fact that biomolecular simulations in drug design are mostly the domain of physical chemists. A typical physical chemist approaches the problem from the bottom-up perspective. This starts with a precise development and tuning of simulation methods, force fields, and sampling enhancement tools, walking stepwise from simple systems to complicated ones. Other approaches in computational drug design such as protein–ligand docking or pharmacophore modeling are the domain of chemoinformaticians. Chemoinformaticians are more open to heuristic approaches. They typically train a model on a training set and validate

it on a validation set of data. If the model helps distinguish between good and bad ligands with statistical significance, it is acceptable for drug design, no matter how solid is its physical basis. I believe that more and more researchers will use biomolecular simulations in such chemoinformatics spirit. Instead of tuning methods on low number of systems, they will test practical impacts on large numbers of systems. This is the area where evaluative and refinement design of molecular simulations can be used.

Predictive power of biomolecular simulations is determined by availability of computer power, partially because longer computational times provide better sampling, partially because long computational times make it possible to identify and correct other limiting factors of biomolecular simulations, such as force field inaccuracies. In the area of DNA sequencing, there has been an enormous jump in performance due to introduction of parallel sequencing machines. The question is whether we can expect a similar jump in biomolecular simulations or whether we can expect evolution rather than revolution. Two emerging technologies have a certain potential to cause such a jump in performance of biomolecular simulations; these technologies are machine learning and quantum computing.

## References


1 Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N. et al. (1953). Equation of state calculations by fast computing machines. *J. Chem. Phys.* 21: 1087–1092.

2 Alder, B.J. and Wainwright, T.E. (1957). Phase transition for a hard sphere system. *J. Chem. Phys.* 27: 1208–1209.

3 Rahman, A. (1964). Correlations in the motion of atoms in liquid argon. *Phys. Rev.* 136: A405–A410.

4 Rahman, A. and Stillinger, F.H. (1971). Molecular dynamics study of liquid water. *J. Chem. Phys.* 55: 3336–3359.

5 McCammon, J.A., Gelin, B.R., and Karplus, M. (1977). Dynamics of folded proteins. *Nature* 267: 585–590.

6 Karplus, M. (2013). Development of multiscale models for complex chemical systems from $H+H_2$ to biomolecules. Nobel Prize lecture.

7 www.gartner.com.

8 Cavalli, A., Bottegoni, G., Raco, C. et al. (2014). A computational study of the binding of propidium to the peripheral anionic site of human acetylcholinesterase. *J. Med. Chem.* 47 (16): 3991–3999.

9 Decherchi, S., Berteotti, A., Bottegoni, G. et al. (2015). The ligand binding mechanism to purine nucleoside phosphorylase elucidated via molecular dynamics and machine learning. *Nat. Commun.* 6: 6155.

10 Pan, A.C., Xu, H., Palpant, T., and Shaw, D.E. (2017). Quantitative characterization of the binding and unbinding of millimolar drug fragments with molecular dynamics simulations. *J. Chem. Theory Comput.* 13 (7): 3372–3377.

11 Erlanson, D.A. and Jahnke, W. (eds.) (2016). *Fragment-Based Drug Discovery: Lessons and Outlook*. Wiley VCH. ISBN: 978-3-527-33775-0.

12 Iannuzzi, M., Laio, A., and Parrinello, M. (2003). Efficient exploration of reactive potential energy surfaces using Car-Parrinello molecular dynamics. *Phys. Rev. Lett.* 90: 238302.

13 Pietrucci, F. and Laio, A. (2009). A collective variable for the efficient exploration of protein beta-sheet structures: application to SH3 and GB1. *J. Chem. Theory Comput.* 5 (9): 2197–2201.

14 Amadei, A., Linssen, A.B., and Berendsen, H.J. (1993). Essential dynamics of proteins. *Proteins Struct. Funct. Bioinf.* 17 (4): 412–425.

15 Spiwok, V., Lipovová, P., and Králová, B. (2007). Metadynamics in essential coordinates: free energy simulation of conformational changes. *J. Phys. Chem. B* 111 (12): 3073–3076.

16 Das, P., Moll, M., Stamati, H. et al. (2006). Low-dimensional, free-energy landscapes of protein-folding reactions by nonlinear dimensionality reduction. *Proc. Natl. Acad. Sci. U.S.A.* 103 (26): 9885–9890.

17 Spiwok, V. and Králová, B. (2011). Metadynamics in the conformational space nonlinearly dimensionally reduced by isomap. *J. Chem. Phys.* 135 (22): 224504.

18 Ceriotti, M., Tribello, G.A., and Parrinello, M. (2011). Simplifying the representation of complex free-energy landscapes using sketch-map. *Proc. Natl. Acad. Sci. U.S.A.* 108 (32): 13023–13028.

19 Abraham, M.J., Murtola, T., Schulz, R. et al. (2015). GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1–2: 19–25.

20 Christen, M., Hünenberger, P.H., Bakowies, D. et al. (2005). The GROMOS software for biomolecular simulation: GROMOS05. *J. Comput. Chem.* 26: 1719–1751.

21 Salomon-Ferrer, R., Case, D.A., and Walker, R.C. (2013). An overview of the Amber biomolecular simulation package. *WIREs Comput. Mol. Sci.* 3: 198–210.

22 Skjærven, L., Yao, X.Q., Scarabelli, G., and Grant, B.J. (2014). Integrating protein structural dynamics and evolutionary analysis with Bio3D. *BMC Bioinf.* 15: 399.

23 Michaud-Agrawal, N., Denning, E.J., Woolf, T.B., and Beckstein, O. (2011). MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem.* 32: 2319–2327.

24 McGibbon, R.T., Beauchamp, K.A., Harrigan, M.P. et al. A modern open library for the analysis of molecular dynamics trajectories. *Biophys. J.* 109 (8): 1528–1532.

25 Lloyd, S.P. (1982). Least square quantization in PCM. *IEEE Trans. Inf. Theory* 28 (2): 129–137.

26 Kaufman, L. and Rousseeuw, P.J. (1987). Clustering by means of medoids. In: *Statistical Data Analysis Based on the L1-Norm and Related Methods* (ed. Y. Dodge), 405–416. North-Holland: Springer.

27 Daura, X., Gademann, K., Jaun, B. et al. (1999). Peptide folding: when simulation meets experiment. *Angew. Chem. Int. Ed.* 38: 236–240.

**28** Wales, D. (2004). *Energy Landscapes: Applications to Clusters, Biomolecules and Glasses*. Cambridge University Press.

**29** Rodriguez, A. and Laio, A. (2014). Clustering by fast search and find of density peaks. *Science* 344 (6191): 1492–1496.

**30** Gfeller, D., De Los Rios, P., Caflisch, A., and Rao, F. (2007). Complex network analysis of free-energy landscapes. *Proc. Natl. Acad. Sci. U.S.A.* 104 (6): 1817–1822.

**31** Lindorff-Larsen, K., Maragakis, P., Piana, S. et al. (2012). Systematic validation of protein force fields against experimental data. *PLoS One* 7 (2): e32131.

**32** Vanommeslaeghe, K. and MacKerell, A.D. Jr., (2015). CHARMM additive and polarizable force fields for biophysics and computer-aided drug design. *Biochim. Biophys. Acta* 1850 (5): 861–871.

**33** Shi, Y., Xia, Z., Zhang, J.H. et al. (2013). Polarizable atomic multipole-based AMOEBA force field for proteins. *J. Chem. Theory Comput.* 9 (9): 4046–4063.

**34** Harder, E., Damm, W., Maple, J. et al. (2015). OPLS3: a force field providing broad coverage of drug-like small molecules and proteins. *J. Chem. Theory Comput.* 12 (1): 281–296.

**35** Wang, J., Wolf, R.M., Caldwell, J.W. et al. (2004). Development and testing of a general AMBER force field. *J. Comput. Chem.* 25: 1157–1174.

**36** Vanommeslaeghe, K., Hatcher, E., Acharya, C. et al. (2010). CHARMM general force field: a force field for drug-like molecules compatible with the CHARMM all-atom additive biological force field. *J. Comput. Chem.* 31: 671–690.

**37** Zoete, V., Cuendet, M.A., Grosdidier, A., and Michielin, O. (2011). SwissParam, a fast force field generation tool for small organic molecules. *J. Comput. Chem.* 32 (11): 2359–2368.

**38** Dodda, L.S., Cabeza de Vaca, I., Tirado-Rives, J., and Jorgensen, W.L. (2017). LigParGen web server: an automatic OPLS-AA parameter generator for organic ligands. *Nucleic Acids Res.* 45 (W1): W331–W336.

**39** Jo, S., Kim, T., Iyer, V.G., and Im, W. (2008). CHARMM-GUI: a web-based graphical user interface for CHARMM. *J. Comput. Chem.* 29 (11): 1859–1865.

**40** Kanal, I.Y., Keith, J.A., and Hutchison, G.R. (2018). A sobering assessment of small-molecule force field methods for low energy conformer predictions. *Int. J. Quantum Chem.* 118: e25512.

**41** Hobza, P. and Muller-Dethlefs, K. (2009). *Non-covalent Interactions: Theory and Experiment*. RSC Publishing. ISBN: 978-1-84755-853-4.

**42** Jorgensen, W.L. and Schyman, P. (2012). Treatment of halogen bonding in the OPLS-AA force field: application to potent anti-HIV agents. *J. Chem. Theory Comput.* 8 (10): 3895–3901.

**43** Mayne, C.G., Saam, J., Schulten, K. et al. (2013). Rapid parameterization of small molecules using the force field toolkit. *J. Comput. Chem.* 34: 2757–2770.

**44** Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD – visual molecular dynamics. *J. Mol. Graphics* 14: 33–38.

**45** Domanski, J., Beckstein, O., and Iorga, B.I. (2017). Ligandbook: an online repository for small and drug-like molecule force field parameters. *Bioinformatics* 33 (11): 1747–1749.

**46** Shirts, M. and Pande, V.S. (2000). Screen savers of the world unite. *Science* 290 (5498): 1903–1904.

**47** Kutzner, C., Páll, S., Fechner, M. et al. (2015). Best bang for your buck: GPU nodes for GROMACS biomolecular simulations. *J. Comput. Chem.* 36 (26): 1990–2008.

**48** Shaw, D.E., Deneroff, M.M., Dror, R.O. et al. (2008). Anton, a special-purpose machine for molecular dynamics simulation. *Commun. ACM* 51 (7): 91–97.

**49** Torrie, G.M. and Valleau, J.P. (1977). Nonphysical sampling distributions in Monte Carlo free-energy estimation – umbrella sampling. *J. Comput. Phys.* 23: 187–199.

**50** Laio, A. and Parrinello, M. (2002). Escaping free-energy minima. *Proc. Natl. Acad. Sci. U.S.A.* 99 (20): 12562–12566.

**51** Colizzi, F., Perozzo, R., Scapozza, L. et al. (2010). Single-molecule pulling simulations can discern active from inactive enzyme inhibitors. *J. Am. Chem. Soc.* 132 (21): 7361–7371.

**52** Huber, T., Torda, A.E., and van Gunsteren, W.F. (1994). Local elevation: a method for improving the searching properties of molecular dynamics simulation. *J. Comput.-Aided Mol. Des.* 8: 695–708.

**53** Hansen, H.S. and Hünenberger, P.H. (2010). Using the local elevation method to construct optimized umbrella sampling potentials: calculation of the relative free energies and interconversion barriers of glucopyranose ring conformers in water. *J. Comput. Chem.* 31: 1–23.

**54** Babin, V., Roland, C., and Sagui, C. (2008). Adaptively biased molecular dynamics for free energy calculations. *J. Chem. Phys.* 128: 134101.

**55** Valsson, O. and Parrinello, M. (2014). Variational approach to enhanced sampling and free energy calculations. *Phys. Rev. Lett.* 113 (9): 090601.

**56** Šućur, Z. and Spiwok, V. (2016). Sampling enhancement and free energy prediction by flying Gaussian method. *J. Chem. Theory Comput.* 12 (9): 4644–4650.

**57** Kumar, S., Bouzida, D., Swendsen, R.H. et al. (1992). The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* 13: 1011–1021.

**58** Dickson, B.M. (2011). Approaching a parameter-free metadynamics. *Phys. Rev. E: Stat. Nonlinear Soft Matter Phys.* 84: 037701.

**59** Tribello, G.A., Ceriotti, M., and Parrinello, M. (2012). Using sketch-map coordinates to analyze and bias molecular dynamics simulations. *Proc. Natl. Acad. Sci. U.S.A.* 109: 5196–5201.

**60** Bonomi, M., Barducci, A., and Parrinello, M. (2009). Reconstructing the equilibrium Boltzmann distribution from well-tempered metadynamics. *J. Comput. Chem.* 30: 1615–1621.

**61** Bennion, B.J., Be, N.A., McNerney, M.W. et al. (2017). Predicting a drug's membrane permeability: a computational model validated with in vitro permeability assay data. *J. Phys. Chem. B* 121 (20): 5228–5237.

**62** Bussi, G., Laio, A., and Parrinello, M. (2006). Equilibrium free energies from nonequilibrium metadynamics. *Phys. Rev. Lett.* 96: 090601.

**63** Hošek, P. and Spiwok, V. (2016). Metadyn view: fast web-based viewer of free energy surfaces calculated by metadynamics. *Comput. Phys. Commun.* 198: 222–229.

**64** Barducci, A., Bussi, G., and Parrinello, M. (2008). Well-tempered metadynamics: a smoothly converging and tunable free-energy method. *Phys. Rev. Lett.* 100: 020603.

**65** Saleh, N., Hucke, O., Kramer, G. et al. (2018). Multiple binding sites contribute to the mechanism of mixed agonistic and positive allosteric modulators of the cannabinoid $CB_1$ receptor. *Angew. Chem. Int. Ed.* doi: 10.1002/anie.201708764.

**66** Saleh, N., Ibrahim, P., and Clark, T. (2017). Differences between G-protein-stabilized agonist-GPCR complexes and their nanobody-stabilized equivalents. *Angew. Chem. Int. Ed.* 56 (31): 9008–9012.

**67** Milanos, L., Saleh, N., Kling, R.C. et al. (2016). Identification of two distinct sites for antagonist and biased agonist binding to the human chemokine receptor CXCR3. *Angew. Chem. Int. Ed. Engl.* 55 (49): 15277–15281.

**68** Saleh, N., Saladino, G., Gervasio, F.L. et al. (2016). A three-site mechanism for agonist/antagonist selective binding to vasopressin receptors. *Angew. Chem. Int. Ed. Engl.* 55 (28): 8008–8012.

**69** Laio, A., Rodriguez-Fortea, A., Gervasio, F.L. et al. (2005). Assessing the accuracy of metadynamics. *J. Phys. Chem. B* 109 (14): 6714–6721.

**70** Piana, S. and Laio, A. (2007). A bias-exchange approach to protein folding. *J. Phys. Chem. B* 111: 4553–4559.

**71** Marinelli, F., Pietrucci, F., Laio, A., and Piana, S. (2009). A kinetic model of Trp-cage folding from multiple biased molecular dynamics simulations. *PLoS Comput. Biol.* 5 (8): e1000452.

**72** Herbert, C., Schieborr, U., Saxena, K. et al. (2013). Molecular mechanism of SSR128129E, an extracellularly acting, small-molecule, allosteric inhibitor of FGF receptor signaling. *Cancer Cell* 23 (4): 489–501.

**73** Milliman, R.E. (1982). Using background music to affect the behavior of supermarket shoppers. *J. Mark.* 46 (3): 86–91.

**74** Bussi, G., Gervasio, F.L., Laio, A., and Parrinello, M. (2006). Free-energy landscape for β hairpin folding from combined parallel tempering and metadynamics. *J. Am. Chem. Soc.* 128 (41): 13435–13441.

**75** Sutto, L. and Gervasio, F.L. (2013). Effects of oncogenic mutations on the conformational free-energy landscape of EGFR kinase. *Proc. Natl. Acad. Sci. U.S.A.* 110 (26): 10616–10621.

**76** Dickson, A. and Brooks, C.L. (2014). WExplore: hierarchical exploration of high-dimensional spaces using the weighted ensemble algorithm. *J. Phys. Chem. B* 118 (13): 3532–3542.

77 Lotz, S.D. and Dickson, A. (2018). Unbiased molecular dynamics of 11 min timescale drug unbinding reveals transition state stabilizing interactions. *J. Am. Chem. Soc.* 140 (2): 618–628.

78 Limongelli, V., Bonomi, M., and Parrinello, M. (2013). Funnel metadynamics as accurate binding free-energy method. *Proc. Natl. Acad. Sci. U.S.A.* 110 (16): 6358–6363.

79 de Ruiter, A. and Oostenbrink, C. (2013). Protein-ligand binding from distancefield distances and Hamiltonian replica exchange simulations. *J. Chem. Theory Comput.* 9 (2): 883–892.

80 Nagy, G., Oostenbrink, C., and Hritz, J. (2017). Exploring the binding pathways of the 14-3-3ζ protein: structural and free-energy profiles revealed by Hamiltonian replica exchange molecular dynamics with distance field distance restraints. *PLoS One* 12 (7): e0180633.

81 Bollini, M., Domaoal, R.A., Thakur, V.V. et al. (2011). Computationally-guided optimization of a docking hit to yield catechol diethers as potent anti HIV agents. *J. Med. Chem.* 54 (24): 8582–8591.

82 Lee, W.G., Gallardo-Macias, R., Frey, K.M. et al. (2013). Picomolar inhibitors of HIV reverse transcriptase featuring bicyclic replacement of a cyanovinylphenyl group. *J. Am. Chem. Soc.* 135 (44): 16705–16713.

83 Frey, K.M., Puleo, D.E., Spasov, K.A. et al. (2015). Structure-based evaluation of non-nucleoside inhibitors with improved potency and solubility that target HIV reverse transcriptase variants. *J. Med. Chem.* 58 (6): 2737–2745.

84 Kudalkar, S.N., Beloor, J., Quijano, E. et al. (2018). From in silico hit to long-acting late-stage preclinical candidate to combat HIV-1 infection. *Proc. Natl. Acad. Sci. U.S.A.* 115 (4): E802–E811.

85 Gutiérrez-de-Terán, H. and Åqvist, J. Linear interaction energy: method and applications in drug design. *Methods Mol. Biol.* 819: 305–323.

86 Genheden, S. and Ryde, U. (2015). The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin. Drug Discovery* 10 (5): 449–461.

87 Wright, D.W., Hall, B.A., Kenway, O.A. et al. (2014). Computing clinically relevant binding free energies of HIV-1 protease inhibitors. *J. Chem. Theory Comput.* 10 (3): 1228–1241.

88 Flyvbjerg, H. and Petersen, H.G. (1998). Error estimates on averages of correlated data. *J. Chem. Phys.* 91: 461–466.

89 Romo, T.D. and Grossfield, A. (2011). Block covariance overlap method and convergence in molecular dynamics simulation. *J. Chem. Theory Comput.* 7 (8): 2464–2472.

90 Grossfield, A. and Zuckerman, D.M. (2009). Quantifying uncertainty and sampling quality in biomolecular simulations. *Annu. Rep. Comput. Chem.* 5: 23–48.

91 Klimovich, P.V., Shirts, M.R., and Mobley, D.L. (2015). Guidelines for the analysis of free energy calculations. *J. Comput.-Aided Mol. Des.* 29 (5): 397–411.

92 https://plumed.github.io/doc-v2.4/user-doc/html/trieste-2.html.

93 Tiwary, P. and Parrinello, M. (2015). A time-independent free energy estimator for metadynamics. *J. Phys. Chem. B* 119 (3): 736–742.

94 Zhu, F. and Hummer, G. (2012). Convergence and error estimation in free energy calculations using the weighted histogram analysis method. *J. Comput. Chem.* 33 (4): 453–465.

95 Erlanson, D.A. and Jahnke, W. (2016). *Fragment-Based Drug Discovery: Lessons and Outlook*. Wiley-VCH 9783527337750.

96 Zhang, Y. and McCammon, J.A. (2003). Studying the affinity and kinetics of molecular association with molecular-dynamics simulation. *J. Chem. Phys.* 118: 1821.

97 Tiwary, P. and Parrinello, M. (2013). From metadynamics to dynamics. *Phys. Rev. Lett.* 111: 230602.

98 Tiwary, P., Limongelli, V., Salvalaglio, M., and Parrinello, M. (2015). Kinetics of protein-ligand unbinding: predicting pathways, rates, and rate-limiting steps. *Proc. Natl. Acad. Sci. U.S.A.* 112 (5): E386–E391.

99 Lu, H.P., Xun, L., and Xie, X.S. (1998). Single-molecule enzymatic dynamics. *Science* 282 (5395): 1877–1882.

100 Peleg, G., Ghanouni, P., Kobilka, B.K., and Zare, R.N. (2001). Single-molecule spectroscopy of the $\beta_2$ adrenergic receptor: observation of conformational substates in a membrane protein. *Proc. Natl. Acad. Sci. U.S.A.* 98 (15): 8469–8474.