

Rational Prediction with Molecular Dynamics for Hit Identification

Sara E. Nichols^{*,+1,2,3}, Robert V. Swift⁺² and Rommie E. Amaro²

¹University of California, San Diego Department of Pharmacology; ²University of California, San Diego Department of Chemistry and Biochemistry; ³Howard Hughes Medical Institute

Abstract: Although the motions of proteins are fundamental for their function, for pragmatic reasons, the consideration of protein elasticity has traditionally been neglected in drug discovery and design. This review details protein motion, its relevance to biomolecular interactions and how it can be sampled using molecular dynamics simulations. Within this context, two major areas of research in structure-based prediction that can benefit from considering protein flexibility, binding site detection and molecular docking, are discussed. Basic classification metrics and statistical analysis techniques, which can facilitate performance analysis, are also reviewed. With hardware and software advances, molecular dynamics in combination with traditional structure-based prediction methods can potentially reduce the time and costs involved in the hit identification pipeline.

Keywords: Computational methods, docking, drug design, flexibility, molecular dynamics, structure-based prediction, validation, statistical performance analysis.

INTRODUCTION

Proteins are flexible macromolecular structures that can undergo a wide range of motions considered fundamental for their function. Movement can be subtle, including bond vibrations and side chain reorientation. Larger motion is also typical and includes hinge-like rigid domain translations and rotations, backbone rearrangements that may result in alternative secondary structure and loop region flexibility. Flexibility can also entail an equilibrium distribution among isoenergetic conformations that are intrinsically disordered. This hierarchy of motion is relevant in structure based discovery and design. For example, many known proteins have been co-crystallized in alternative conformations, distinct from the unbound crystallized state. This implies motion is fundamental to the co-crystallized complex, and suggests alternative conformational states may exist that are relevant, and imperative, to the discovery of other potential ligands. While historically, protein flexibility has been largely ignored in high-throughput methods used in drug discovery and design, recent software and hardware advances have allowed for the consideration of implicit or explicit protein motion [1-6].

Ideally, the use of computational methods should reduce the cost and time involved in structure-based drug discovery. We highlight here the general concepts behind protein motion and its relation to molecular interactions. Two major types of predictions are discussed including binding site classification, and virtual screening. During hit discovery,

prediction, carried out in a high-throughput fashion, assists practically in the process of research and discovery. Methods should be carefully evaluated on a system-dependent basis, and quantifiable validation should be established at the onset to enable the user to critically decide between several methods and convey the results to potential collaborators with confidence. This review discusses these concepts and how molecular dynamics conformational diversity can enhance the structure-based drug design process.

MOLECULAR FLEXIBILITY AND ITS ROLE IN DRUG DISCOVERY AND DESIGN

Protein motion is key to functional mechanisms that require biomolecular association, such as catalysis, chaperone-assisted folding, protein-protein and protein-nucleic acid interactions, and allosteric regulation. Ligand binding is designed to enhance or interrupt these functions, and therefore may be tightly coupled to protein motion. Ligand association can be as simple as substrate binding prior to enzyme turnover, or as complex as the signal transduction cascades that orchestrate life on the cellular level. It has been shown that most ligands are buried upon association with a protein, suggesting fundamental rearrangement of the receptor is required for complexation [7]. These motions can involve a single side chain movement, such as the repositioning of a catalytic residue in an enzyme, or a coordinated isomerization that occurs concertedly among a group of residues. The timescales for relevant biophysical phenomena cover a wide range, from femtosecond bond vibrations to collective motions requiring seconds to occur. Methyl rotation, loop motion and side-chain rotamer relaxation all happen on the picosecond to microsecond time scale [8]. Frequently, there is a compensation of enthalpy changes, or changes of the average internal energy of a system, and entropy changes, or changes in the measure of favorable disorder of a system. For example, one might observe a favorable decrease in en-

*Address correspondence to this author at the University of California, San Diego, Howard Hughes Medical Institute, 9500 Gilman Drive, Urey Hall 4202, M/C 0365, La Jolla, CA 92093-0365, USA; Tel: (858) 822-1469; Fax: (858) 534-4974; E-mail: senichols@ucsd.edu

+ These authors contributed equally to this work.

thalpy upon binding, as strong hydrogen bonds form between the ligand and receptor. This offsets the entropic penalty that results from rotational and translational ligand confinement. Often enthalpy is the primary consideration in potency optimization, but entropy can also be factored into the strategy [9]. Furthermore, the free energy of the water at the binding site may also play an important, even active role in molecular association [10]. Flexibility-function studies can lead to novel design strategies, and broaden the idea of drug action, through correlated networks of residues [11].

Flexible degrees of freedom, in addition to titratable side chains and non-bonded coulombic and dispersive interactions, produce a rugged, multidimensional energy landscape. The relative populations of the energy wells along this landscape define the thermodynamic properties of the system, while the transitions between them yield the system's kinetics. Both thermodynamics and kinetics are fundamentally important considerations to understand proteins in motion. Developing efficient sampling methods for understanding minima and the transitions between them are current active areas of research [12]. This multidimensional complexity is a challenge for computational modeling with respect to drug discovery and development. Often, an approximate, rapid, high-throughput method is more appropriate for fast-paced pharmaceutical settings than a potentially more accurate, but time-consuming, technique. Among current theories of molecular association, two main concepts prevail, namely, induced fit and conformational selection. Induced fit theory assumes that the ligand influences the shape the protein receptor takes upon binding. Conformational selection hypothesizes that the bound-state conformation exists in the apo protein ensemble, and that upon ligand binding, the population shifts such that the bound-state conformation becomes the most populated conformer [13, 14]. For different targets, these two theories are not necessarily mutually exclusive and can be considered simultaneously, sequentially, or independently [15, 16].

Protein motion can be quantified with a number of experimental techniques. These include hydrogen deuterium exchange, Förster and bioluminescence energy transfer (FRET/BRET), nuclear magnetic resonance, solution X-ray scattering, circular dichroism, electron microscopy, and X-ray crystallography. One classic scenario from early crystallographic studies, that implies the requirement of protein flexibility in ligand association, is the experimentally observed conformations of the heme-based oxygen binding proteins, myoglobin and hemoglobin. Space-filled models of these proteins revealed no clear oxygen binding pathway to the heme [17]. It was clear that thermal fluctuations were necessary to facilitate the ligand binding process. Other experimental methods demonstrate the relationship between conformational reorganization and biomolecular function. A second, more contemporary, example is the use of NMR to classify dynamics relevant for adenylate kinase catalysis. Kern and co-workers proposed a linkage between picosecond to nanosecond motions and larger amplitude domain rearrangements that are important in facilitating catalytic activity [8]. Observations such as these inextricably link motion to function, and should be exploited for discovery and design purposes.

PREDICTION OF MOLECULAR RECOGNITION SITES

Molecular recognition site detection can be used early on in the pipeline of structure-based drug design (SBDD) to identify where a drug could possibly bind. Here we refer to three of the most commonly targeted types of binding sites, active sites, allosteric sites and protein-protein interaction sites. Most drugs are discovered or designed to bind the known active site, where an enzymatic reaction happens. Allosteric sites, distinct from the active site, are where a molecule binds and influences receptor activity and can also be targeted. These types of interactions have the advantage that they are often not conserved, and therefore can be more unique, or selective for the desired target, circumventing off-target interactions. Finally small molecules can be designed to target protein-protein interfaces, to prevent protein association into functional complexes. Due to crystallographic conditions, a binding site on a protein may at some point exist in an alternative shape, extend to interact with crystallographically undetected residues, or be completely distinct from a known binding site where a small molecule has already been observed.

Binding pockets are characterized by shape and complementarily to prospective binders, but traditionally would include detection of active and allosteric sites. The nature of a binding site is constantly in fluctuation, including the contour, the side-chain dihedral angles, and potential polar and charged contacts. High affinity and specificity is a careful balance between opposing forces and interactions. There are a variety of different binding site detection methods. Broadly, these can be characterized into three main classification algorithms: geometric methods based on the surface of the site, energetic methods based on physical and chemical properties of the binding site, and knowledge-based methods that incorporate data gathered from other sources besides the protein structure [18-20]; methods that fall into these groups have been recently reviewed [19].

Defining what constitutes a binding site is problematic, particularly for the detection of druggable pockets in protein-protein interaction sites. Protein-protein interaction sites, tend to be shallow, hydrophobic, and extended surfaces [21]. Clackson and Wells defined hot spots as residues at a protein-protein interaction (PPI) site that are essential for association, and dominate the binding free energy [22]. These hot spots can be utilized as starting points for molecular docking and design optimization protocols. [23-25]. Computational prediction methods for these crucial areas have also been recently reviewed [26]. While targeting PPIs with small molecule inhibitors has been challenging, the concept of identifying epitopes important for binding is relevantly extended to other types of molecular recognition where clear binding cavities are not obvious, such as nucleic acid-protein or lipid-protein binding sites.

SMALL MOLECULE DOCKING AND VIRTUAL SCREENING

Docking and virtual screening can be used to identify potential small molecules binders, and most docking programs reduce the search space and computational time by requiring the user to specify the putative binding site. Small

molecule docking is the process of computationally fitting a ligand into a biomolecule, typically held rigid in a single conformation. To expedite the required calculations, early docking programs also fixed the ligand conformations and often relied on steric complementarities to rank the plausibility of the predicted compound orientations [27]. As Moore's law [28] evolved and processor speed increased exponentially, ligand flexibility was introduced [29], and force fields were used to provide a classical estimate of ligand-receptor interactions. Flexibility allowed the ligand to relax in response to the binding pocket requirements, which were still held rigid, while the force field guided physically reasonable interactions, such as hydrogen bonds. Today, these force field descriptions are generally known as "scoring functions" [30, 31] and may be augmented by a variety of energetic terms, such as solvation and entropy; these typically provide a rough estimate of binding affinity. To further speed the calculations and facilitate the rapid docking of large compound databases, scoring function interactions can be pre-calculated and stored on grids for speedy look up [31-33]. Many docking programs are grid-based, and current state-of-the-art methods are very fast, which allows routine screening of large compound databases [34]. Moreover, the growth in docking technology has resulted in a large number of high-quality, freely available docking programs. Many of these programs can be found at click2drug.com, an excellent compendium resource of valuable *in silico* drug design tools.

Docking is an important tool in contemporary lead discovery. Its value lies in the potential to improve the efficiency and reduce the costs of modern high throughput screens (HTS) and ultra high throughput screens (uHTS) during hit discovery. Since its initial development in the early to mid 1990's, modern HTS has grown to dominate lead discovery efforts and has undergone an increasing trend toward miniaturization [35]. Miniaturization resulted in smaller well volumes and a larger number of wells on each assay plate. Still, the cost of reagents and other consumables can be significant, particularly when large corporate or commercial databases, with compounds numbering in the millions, are considered [35]. To reduce costs, rather than randomly screening an entire database, docking can be used to rationally prioritize a subset for testing. This prioritization process is known as virtual high throughput screening (vHTS). In a typical vHTS, database compounds are docked into a single crystal structure and ranked according to their predicted binding affinities. Ideally, all of the binders would be ranked ahead of all of the non-binders, tremendously reducing experimental HTS expenditures. In practice, these results are never achieved, and even carefully validated docking protocols may fail to identify novel hits.[36, 37] Nevertheless, vHTS has demonstrated success across a range of targets, from dengue virus protease [38], to HIV-1 reverse transcriptase[39] to the RNA editing ligase from *Trypanosoma brucei*, the causative agent of African sleeping sickness [40, 41], among others [34]. Despite successes, efforts to improve vHTS performance, including ensemble-based methods [42-44], are a vibrant area of research today.

PREDICTION AND PERFORMANCE EVALUATION

As virtual screening method performance can vary from system to system, a sound, rigorous method of evaluation is

important when developing a protocol. While we focus on virtual screening examples here, the concepts introduced are general and apply to other classification problems that occur in SBDD. Regardless of the classification method, we want to employ the very best protocol at our disposal to ensure project success. In the case of a virtual screen, this means that the list of compounds we recommend for testing contains as many true binders as possible. In the next section, we introduce the Receiver Operating Characteristic (ROC) curve [45] and the area under the curve, or AUC, as classification metrics and gauges of virtual screening performance. We also discuss the origins and statistical characterization of performance variability.

ROCing Performance

Numerous virtual screening performance metrics have been published in the literature and have been discussed at length [46, 47]. Some popular examples include the enrichment factor EF, robust initial enhancement (RIE) [48], and the Boltzmann-Enhanced Discrimination of ROC (BED-ROC) [46]. However, the AUC of a ROC curve is perhaps the most frequently applied and has been championed as *the* virtual screening performance metric [49, 50]. Below, we characterize the properties of ROC curves, and their corresponding AUC values, through several numerical illustrations.

For the purposes of assessment, virtual screening is a binary classification problem [45], and the ROC curve is an easily interpreted graphical representation of docking protocol performance. During a virtual screen the estimated binding affinities, or binding free energies, can be used to rank compounds. Compounds with more favorable binding affinities will receive a better rank and are more likely to be recommended for subsequent experimental validation. To understand how a ROC curve is constructed, and why it reflects docking performance, it is useful to plot the distribution of predicted binding affinities for both the binders and non-binders. For example, Fig. (1A) represents a hypothetical virtual screening experiment in which the predicted distributions of the binders and non-binders are represented by black and gray curves, respectively. These normal distributions were constructed to have means of -5 and -3 kcal/mol, respectively, and identical standard deviations of 2 kcal/mol. A binding affinity threshold can be chosen such that all of the compounds with higher binding affinities will be tested. For example, the vertical black dashed line represents a -5 kcal/mol threshold. All of the compounds to the left of the dashed black line will be tested and all of the compounds to the right will not. Since the compounds represented by the black curve are binders, integrating from negative infinity up to the threshold gives the fraction of the binders we'd expect to discover if we chose a -5 kcal/mol threshold. These are the "true positives," *i.e.* the compounds predicted to bind that actually bind. The value of the integral is often called the "true positive rate," or TPR. On the other hand, if the gray curve is integrated from negative infinity up to the threshold, the fraction of non-binders we'd expect to recommend for experimental validation is given. These compounds are "false positives." That is, the compounds that were predicted to bind that actually do not. Similarly, the value of this integral is often called the "false positive rate," or FPR. By re-

peating the integration process while continuously varying the threshold, the true and false positive rates can be determined as functions of the threshold value Fig. (1B). In the language of probability theory and statistics, these curves are called cumulative distribution functions. Plotting the TPR as a function of the FPR yields a ROC curve Fig. (1C). This information is useful, as docking protocols that yield higher TPRs at a fixed FPR are expected to produce a greater number of actives at a given threshold. Indeed, the TPR at FPR of 0.5%, 1%, 2%, and 5% have been suggested as standard measure of virtual screening enrichment [49].

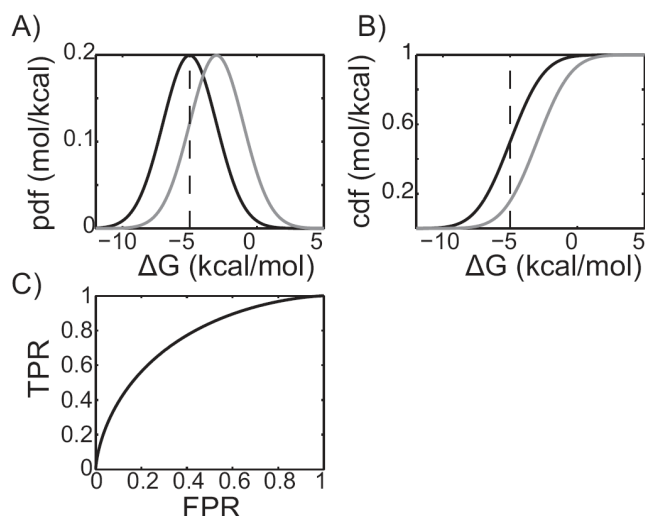


Fig. (1). Receiver Operating Characteristic (ROC) plots. A) Predicted binding affinity probability distribution functions (pdfs) of the actives, shown in black, and inactives, shown in grey, for a hypothetical virtual screening experiment. A -5 kcal/mol threshold is shown as a grey, vertical dashed line. Compounds whose scores lie to the left of the line are experimentally assayed. Compounds whose scores lie to the right are not. B) The integrals of the pdfs, or cumulative distribution functions (cdfs). The black cdf is equivalent to the true positive rate (TPR), while the grey cdf is equivalent to the false positive rate (FPR). C) A ROC plot, generated by plotting the grey cdf along the X-axis versus the black cdf along the Y-axis.

The area under the ROC curve is a valuable metric of virtual screening performance. To understand why, it is again useful to consider the distribution of predicted binding affinities, their corresponding cumulative distribution functions, and ROC curves. If a docking protocol has no discriminatory ability, then the distribution of predicted binding affinities, as well as their cumulative distribution functions, will be identical for the binders and the non-binders. Since the TPR and FPR are identical, this will result in a ROC curve that bisects the unit square Fig. (2A). In this case, the area under the ROC curve is simply one-half the unit square, or 0.5. Thus, a protocol that has no discriminatory power (equivalent to random selection), yields an AUC of 0.5. As discriminatory power improves, the binding affinity distributions and cumulative distribution functions are more widely separated. This shifts the ROC curve toward the upper left hand corner, and the AUC increases to a value between 0.5 and 1 Fig. (2B & C). When discrimination is perfect, the

binders and non-binders are completely separated and the AUC value is 1. Significantly, it can be shown that the AUC value is identical to the probability that a randomly selected binder will be ranked ahead of a randomly selected non-binder, which is equivalent to the Wilcoxon statistic [51].

Performance Variability

Conceptually, a null-hypothesis test can determine whether a docking protocol performs better than random in a statistically significant way. As a result of performance variability, a virtual screening protocol that performs randomly, on average, will also yield results that are better or worse than random. For example, Fig. (3A) shows the distribution of AUC values that result from a hypothetical randomly performing docking protocol. While, on average, the distribution is centered on an AUC of 0.5, performance variability results in much higher and much lower AUC values. For virtual screening purposes, it is desirable to create a method that on average performs better than random and to simultaneously assess the confidence of the results. Confidence can be quantified through probability, *i.e.* assuming the method performs randomly, what is the probability of observing an AUC value as large, or larger, than the value we estimated? The smaller the probability, the less likely it is that the value we observed was due to a randomly performing protocol.

In practice, a null hypothesis test can be carried out using the recipe outlined here. First, the null hypothesis: “the virtual screening protocol performs randomly” is assumed. The second step is to determine the corresponding null distribution, which can be determined using the central limit theorem and an analytic estimate of the standard error as in [45, 48, 49] (see, for example, Fig. (3B)); note that the gray curve is the null distribution, centered on an AUC value of 0.5). In the third step, the null distribution is used to determine the so-called “p-value”, or the probability of observing an AUC value as large, or larger, than the value that was estimated (see the shaded region in Fig. 3B). The fourth step is to consider the size of the p-value to decide whether the null hypothesis should be rejected in favor of an alternative hypothesis. If the p-value is smaller than some threshold, or significance level, we reject the null-hypothesis, replace it with an alternative hypothesis, and deem the result statistically significant. Popular significance levels for rejecting the null-hypothesis are 5% and 1% [52], though we note that these levels are arbitrary [53]. If the null hypothesis was rejected during the fourth step, the fifth step is to make the alternative hypothesis: “the protocol of interest performs better than random.” Subsequently, the alternative distribution must be determined, again using the central limit theorem and an analytic estimate of the standard error [45, 48, 49]. The average of the alternative distribution is typically taken as the AUC value of the virtual screening experiment (see, for example, the black curve in Fig. 3B). The alternative distribution is important because it allows us to estimate confidence intervals. In the event that the null hypothesis was not rejected during the fourth step, one might consider alternative virtual screening protocols and repeat steps one through four until the null hypothesis can be rejected at the desired significance level.

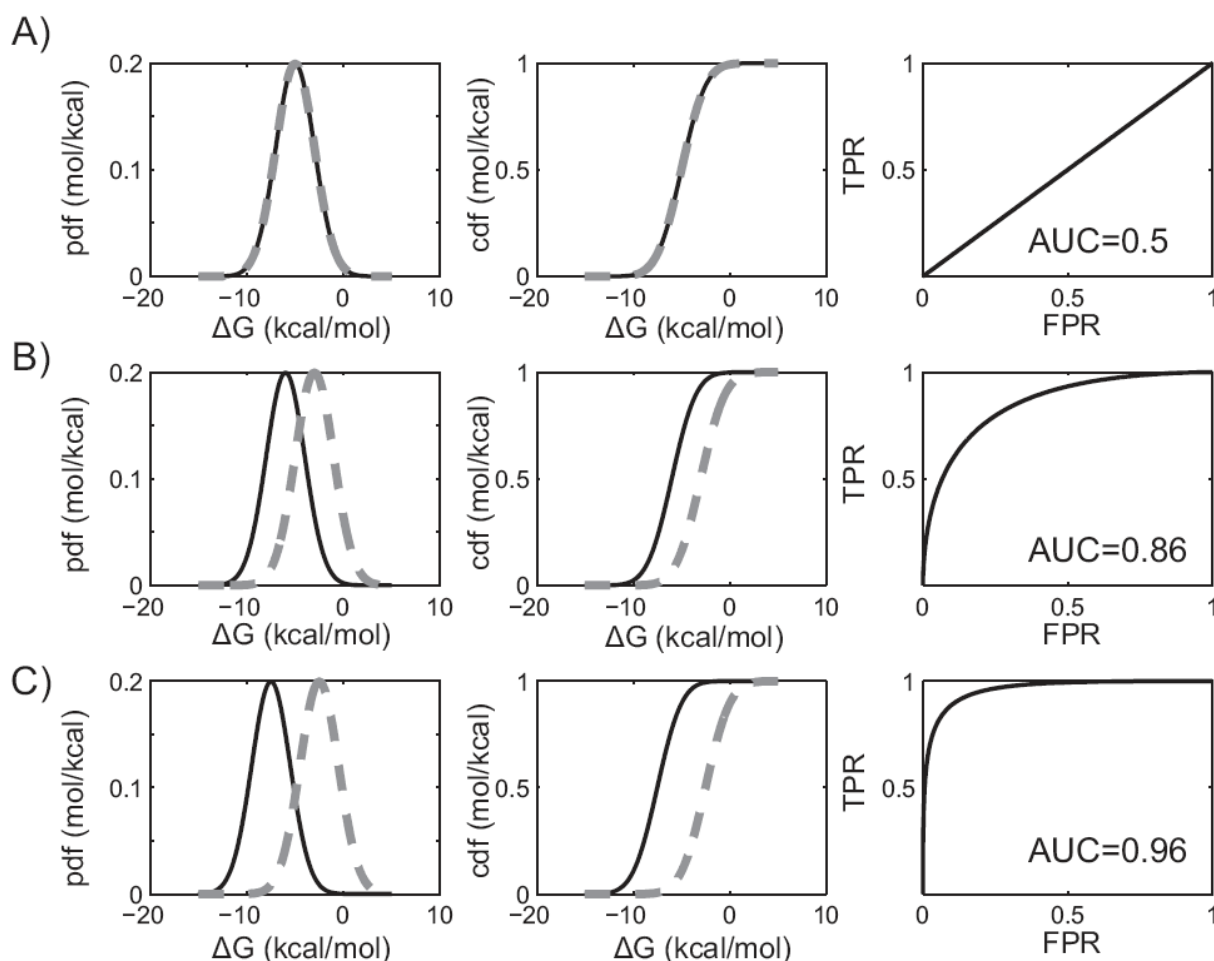


Fig. (2). The area under the ROC curve and virtual screening performance. In figures A) through C), plots in the leftmost column describe the predicted binding affinity probability distributions. Plots in the middle column give the corresponding cumulative distribution functions. The binders are shown in solid black, while the non-binders are shown in dashed grey. Plots in the right column show the corresponding ROC curves. AUC values are shown for three different hypothetical virtual screening protocols with A) no discriminatory power, B) improved discriminatory power, or C) near perfect discrimination.

If two virtual screening protocols perform identically, on average, then the distribution of AUC difference values (ΔAUC) will have an average of zero, as the gray curve in Fig. (3C) illustrates. To insure that the calculated ΔAUC is not due to the variability of two methods that perform identically (on average), a null hypothesis test must be carried out. This can be accomplished by following the recipe outlined in the previous paragraph, but assuming a different null hypothesis and distribution. The correct null hypothesis is: “the two different protocols perform identically.” The corresponding ΔAUC null distribution is centered on zero and can be estimated using the central limit theorem. Importantly, the standard error estimate must account for compound ranking correlation between methods [45, 48, 49]. Neglecting correlation may artificially inflate the standard error causing the null hypothesis to be inappropriately retained. The p-values are then determined as the shaded regions under the gray, null distribution, curve in Fig. (3C). The alternative distribution determined is shown in black in Fig. (3C).

Assessing how much performance variability to expect is necessary when considering a virtual screening protocol. For example, given an initial AUC value of 0.90, when a second

database is screened, should an AUC value of 0.65 be surprising? Confidence intervals, a range of values that includes the true value with a given probability (confidence level), help address this question. The confidence level is typically taken at 95% [52], which implies that the true average AUC will be contained within identically constructed intervals in 95 of 100 identically performed virtual screens (with different databases). If we assume that the central limit theorem holds, then we can use the analytic estimates of the standard error to estimate the confidence intervals [53]. For example, Fig. (3D) illustrates a 95% confidence interval [54]. Typically, confidence intervals are estimated on the alternative distributions, discussed in the paragraphs above.

GENERATING RECEPTOR CONFORMATION VARIABILITY WITH MOLECULAR DYNAMICS

Molecular docking, virtual screening and the other computational structure-based drug design methods discussed in previous sections have traditionally been performed on a single protein conformation. Despite this precedence, it is widely agreed that macromolecular structures have heterogeneity that cannot be represented by a single conformation,

such as the crystallographic models deposited into the Protein Data Bank [55, 56]. In solution, proteins move and interact with their environment, and a crystallographic coordinate model can be represented as a network of physically interacting atoms by an equation and parameters called a force field. A force field allows us to calculate the potential energy of a system. Force field models are based on a variety of assumptions; most notably, the Born-Oppenheimer approximation, which allows us to describe the system solely as a function of nuclear coordinates. The simplified interactions are described by mathematical functions with few parameters, such as a harmonic potential and its corresponding spring constant. Making the physics simple allows parameters for the model, such as spring constants for bond vibrations, to be transferable to many different types of systems.

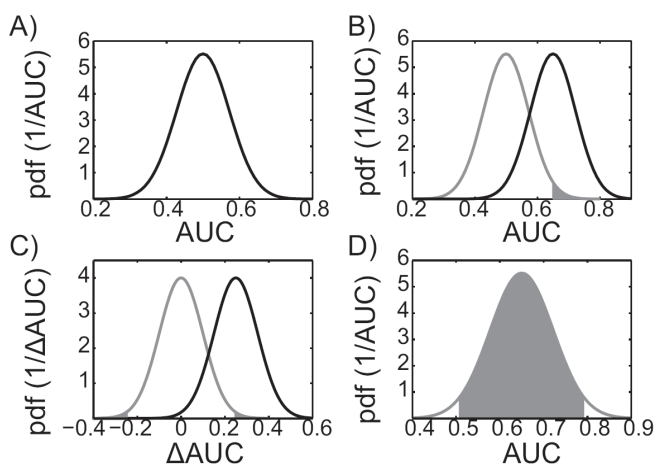


Fig. (3). Performance evaluation statistics. **A)** Distribution of a virtual screening protocol that performs randomly, on average. **B)** An example p-value for a hypothetical virtual screening experiment with an AUC of 0.65 is illustrated as the shaded area under the null distribution, shown in grey. The alternative distribution, corresponding to the alternative hypothesis, is shown in black. **C)** The null distribution corresponding to the null hypothesis, “the two docking protocols perform identically,” is shown in grey. The p-value corresponding to two virtual screening experiments whose Δ AUC value is 0.25 is illustrated as the shaded areas under the grey null-distribution curve. The corresponding alternative distribution is shown in black. **D)** The 95% confidence interval of a hypothetical virtual screening protocol with an observed AUC value of 0.65. The 95% confidence interval is bounded by the grey shaded region, which contains 95% of the distribution.

While much insight is gleaned from crystallographic models, many complex interactions and reactions require more dynamic detail. Molecular dynamics (MD) is a simulation method that propagates the system described by a potential function and numerical integration of Newton’s second law. Structural coordinates are propagated through time using fixed integration steps (typically 1 or 2 fs), according to the forces acting on the system. After each integration step, the forces on the atoms are recalculated and combined with the current velocity and positions resulting in new velocities and positions. The underlying physics of the fastest motions in the system, bond stretching, dictate a time step limitation of 1-2 fs. The positions calculated during a MD simulation

are saved at periodic intervals generating a trajectory of snapshots in time, which approximate a statistical sample from an ensemble of configurations of the protein system at thermal equilibrium. Using this sample, time-averaged properties can be determined and connected to ensemble-averaged properties through the concept of ergodicity and the laws of statistical mechanics. While the accuracy of these averages depends heavily on proper equilibration and sampling of the system, qualitatively useful estimates can be determined in relatively short simulation times for small systems.

Initially, one must decide on the type of force field to use when modeling a pharmaceutically relevant system. The energy models consistently used when examining atomic interactions are classical fixed-charge force fields such as AMBER, CHARMM, or OPLS [57-60]. These are typically used with one of several water models, such as TIP3P or SPC, and a generalized organic force field for ligand parameters, such as CGenFF or GAFF [61-63]. It is best to review the literature on the compatibility of water models with protein and generalized force fields being selected for use, as some energy functions were parameterized to be compatible only in specific circumstances [64, 65].

With most computational methods, a dangerous outcome is that results can be produced regardless of the expertise of the user. The fundamental learning curve is in setting up the system correctly. This entails taking into account relevant structures, force field parameters, protonation and ionization states, solvent and system sizes. A basic understanding of the underlying theory is critical for interpretation of the results produced. MD is essentially numerical statistical mechanics, therefore, users who have not been formally trained, should at least read some introductory text to familiarize themselves with such concepts before proceeding too far. Additionally, the highly technical nature of running and analyzing molecular dynamics simulations usually requires an apprenticeship in a laboratory specializing in these computational techniques. The choice of ionic strength and temperature of the simulation cell should be guided by relevant experimental biological conditions where possible, although current MD simulation standard practices do not come close to standard conditions in a cell [66]. With this construct established, the system is equilibrated until several metrics of stability are achieved. Structural stability is commonly measured by monitoring kinetic and potential energy time series, as well as two coordinate metrics, the root mean square deviation (RMSD) and root mean square fluctuation (RMSF). The RMSD is the coordinate position deviations averaged over particles at a fixed time point, and the RMSF is the variation in atomic position averaged over time. Typically, the RMSD is plotted as a function of time, and is a geometric measure of how different a protein conformation is from a reference conformation, while the RMSF is plotted on a per residue basis to describe local structural fluctuations. In some cases, particularly with proteins that undergo large structural rearrangements, many transitions between conformational macrostates will not occur, and the system will not reach equilibrium. However, meaningful information and structures near the starting conformation can still be extracted. Additionally, to improve conformational sampling local to a starting structure, multiple independent trajectories with different velocity

distributions are often used [67], and the additional information can be integrated in the design process.

Free software is available to run molecular dynamics simulations, such as the widely used Desmond, Gromacs, or NAMD packages [68-70]. Other commonly used dynamics packages, such as AMBER and CHARMM, are available at minimal cost to non-profit organizations [71, 72]. Many of these packages can be operated with limited experience, and have tutorials, user manuals and forums where users can have questions answered by developers and experts fairly rapidly. While there are numerous force field comparisons noted in the literature, there are limited direct comparisons of MD packages that are implemented for parallelization and scalability. While good packages read several force fields and the results should not depend on specific software implementation, often the processor number and type available to the user may dictate which package to use.

Prior to using MD snapshots in SBDD pipelines, evaluation of the trajectories should be performed to assess the ensemble sampled. In particular, one question must be addressed when considering the use of MD snapshots in a drug design context: is the sampled phase space relevant to inhibiting the targeted mechanism? Phase space is a high dimensional concept representing all of the degrees of freedom of the system (i.e. for a system of N atoms, each atom can be described with 3 coordinates and 3 components of momentum, resulting in $6N$ phase space). One way to understand how a simulation is changing is by quantifying the phase space that is being sampled. For example, principal component analysis is a popular technique that decomposes the variance of atomic positions into principal modes [73]. Visualizing these modes is useful in identifying the collective motion responsible for large conformational events, such as binding pocket expansion. Alternatively, clustering and population analysis of a trajectory identify highly populated conformational regions of phase space and are other techniques that may be employed [24, 74]. By carefully analyzing the distribution of conformational states in the context of the project goals, useful insights can be identified. To help with such analyses, many tools freely available to visualize and analyze molecular dynamics trajectories exist, including, but certainly not limited to, the widely used VMD, AmberTools and Gromacs programs [69, 72, 75]. Open source tools for commonly used interpreted programming languages and packages such as Python and R are also available [76-79].

Currently routine time-dependent molecular simulations can access microsecond timescales, although it is more common to find studies publishing hundreds of nanoseconds of simulation [80]. Recent advances in software and hardware have thrust MD timescales forward. Some software has been implemented to run on graphics processing units (GPUs), which have more arithmetic logic units than conventional CPUs, but are not as complex. The clock speed is slower on the GPU but the access to memory is faster, and this, combined with GPU programming language development, has led to potential for routinely achieving longer MD timescales. The promise of GPUs is further enhanced by massively parallel GPU hardware [81, 82]. Impressively, GPU hardware was recently used to sample a binding asso-

ciation event [83]. Previously, these time-scales were only accessible on a specialized hardware machine [83, 84].

DEALING WITH BIG DATA: HOW TO GET THE MOST FROM DOCKING WITH MD

With ever improving hardware, molecular dynamics simulations will continue to be a ubiquitous biophysical technique, and rightly so: in the absence of multiple crystal structures, it is a good way to quickly and realistically generate conformational diversity. While other methods, such as normal mode analysis [85], can be used to generate conformational diversity, MD has the benefit of approximating the conformational ensemble expected at thermal equilibrium, which may facilitate virtual screening success. Contemporary MD simulations can easily result in tens of thousands to hundreds of thousands of receptor conformations, and a virtual screen can be performed against anyone of these [86], all of them [87, 88], or any subset [40, 41, 89]. Moreover, there are multiple ways to combine the docking scores from each receptor. As each choice of protein receptor(s) and score combination method can perform differently, a large hurdle in MD-based ensemble virtual screens is deciding which set of protein conformations to pair with which score combination method. In this section, we outline a set of best practices to systematically evaluate ensemble performance.

Typically, the first step in performing a virtual screen using an MD trajectory is ensemble reduction. Ensemble docking scales linearly with the number of receptor conformations, and since only finite resources are available, culling a tractable conformational subset is essential. Several methods have been suggested in the literature, including RMSD-based clustering [44], QR factorization [40, 41] and binding-site volumetric clustering [90, 91]. While the RMSD-based clustering and QR factorization methods can be performed with freely available software, the binding-site volumetric clustering is currently implemented in commercial software. However, there are several freely available pocket-volume calculation programs [92, 93] whose output could be used to cluster in a similar fashion. With the variety of ensemble reduction techniques available, one may naturally wonder which performs best. Unfortunately, to the best of the authors' knowledge, a set of conformational descriptors that correlate strongly to virtual screening performance has not been definitively determined [86]. Moreover, since the best ensemble reduction technique may be system dependent, methodically evaluating the performance of several techniques before performing a large-scale screen is a sensible approach, assuming available time and resources. With a set of structures in hand, an ensemble-based virtual screen can be carried out. Faced with designing an ensemble virtual screening protocol, which receptor conformations should be used? The simplest answer is to use them all. For example, given 20 receptor conformations and a database of 100 compounds, each compound would be docked into each of the 20 conformations, for a total of 2000 ligand scores. Then each ligand score can be a combined in some fashion yielding an ensemble score. While this simple approach has been used successfully in some cases [40, 41], it overlooks a combinatorial subtlety, and performance may suffer as a result. Alternatively, given 20 conformations, a smaller subset of m conformations can be chosen to constitute the ensemble [94].

The number of unique ensembles that can be constructed this way is given by the binomial coefficient. This can quickly result in a large number of possibilities to consider, each with a different performance. The number of unique ensembles increases as a function of size. For example, when ensembles are made from any of 20 unique receptor conformations, a maximum of 184,800 combinations are possible using 10-member ensembles Fig. (4A). The total number of possible ensembles that can be considered is determined by summing the number of unique ensembles that can be constructed at each ensemble size. For example, when ensembles are constructed from 20 unique receptor conformations, there are a total of 1,048,575 ensembles to consider. In practice, performance can be plotted as a function of size by extracting the top-performing member for each size [89, 94]. For example, Fig. (4B) plots the AUC of the top-performing ensemble as a function of ensemble size for a hypothetical vHTS experiment. While the data is contrived, the trend follows our own experiences and is similar to those that can be found in the literature [94]. Clearly, optimal performance occurs for a five-member ensemble, and this ensemble would then be considered against other top-performers, as described below.

Once the top-performing ensemble has been identified, the process can be repeated using different score combination techniques. For example, one might consider arithmetically averaging ligand scores. Alternatively, ligand scores could be combined in a weighted average, where the weights may be extracted from clustering and population analysis [44], or may be taken as the Boltzmann weights [95]. Another popular method selects only the best ligand score for ranking purposes [94]. Additionally, as docking performance varies from one target to another [96], different docking programs should also be considered. Each method can then be compared using the simple statistics presented in the "Prediction and Performance Evaluation" section of this review to arrive at the optimal docking protocol for the system of

interest. In principal, identifying the optimally performing ensemble method on a smaller validation database should increase the success of virtual screens performed on much larger databases. While this optimization scheme represents a significant resource investment, much of the analysis can be automated to reduce the required "hands on" time. Additionally, while this method is limited to targets with known actives, inexpensive semi-high through put techniques, such as differential scanning fluorimetry [97], coupled with free databases, like the National Cancer Institutes Diversity set, may facilitate validation database development [98].

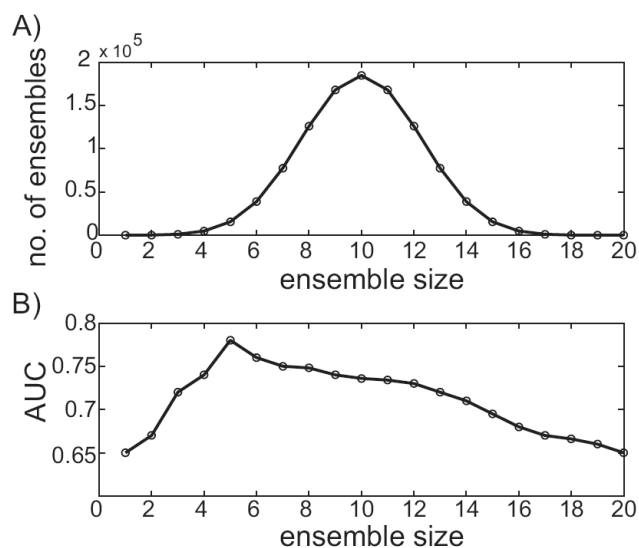


Fig. (4). Ensemble size and virtual screening performance variability. A) For 20 receptor conformations, the number of ensembles that can be constructed is plotted as a function of the ensemble size. B) For a hypothetical ensemble-based virtual screening experiment, the AUC of the top-performing ensemble is plotted as a function of ensemble size.

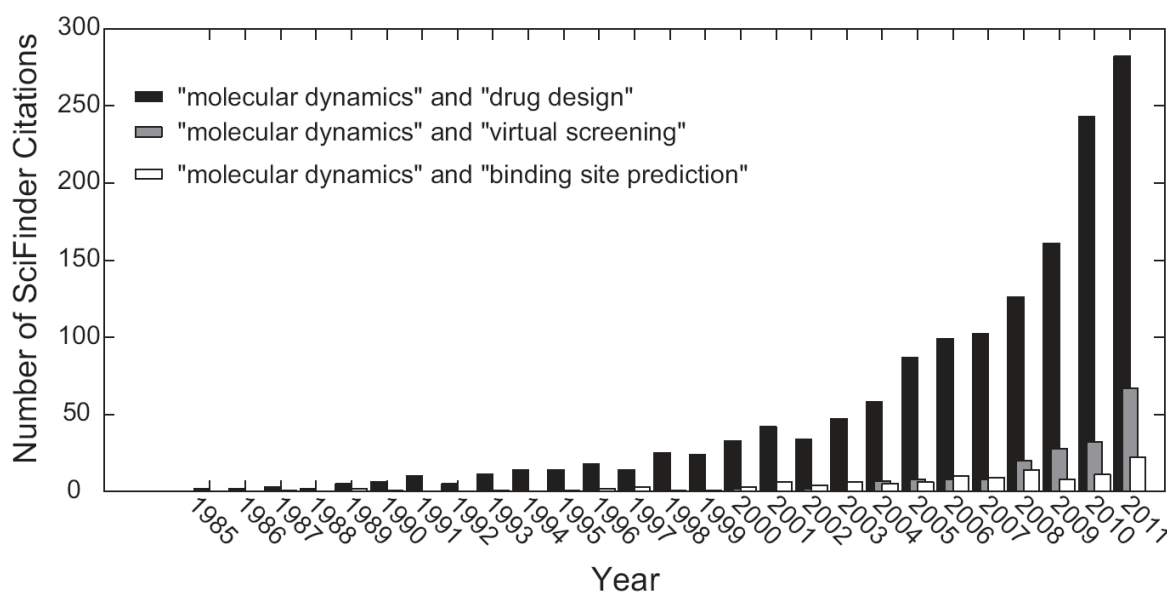


Fig. (5). Exponential increase in molecular dynamics citations. Searches using the quoted keywords were performed using SciFinder Scholar. The number of citations returned is plotted as a function of year, indicating the increasing use of molecular dynamics in drug discovery.

CONCLUSIONS

Molecular dynamics is increasingly being included in academic high-throughput efforts for pharmaceutically relevant prediction research, and the future is bright. For example, using SciFinder Scholar to track citations involving molecular dynamics research, one can see in Fig. (5) that molecular dynamics simulations carried out with respect to drug design has been exponentially increasing since the 1980s, and has been increasing in the context of virtual screening and binding site prediction since the early 2000s [99]. While a gold standard practice for selection of MD snapshots for virtual screening and binding site prediction has yet to be established, the growing number of academic research laboratories carefully considering this as part of their discovery toolkit improves future prospects of structure-based drug design. Similar optimism can be found in related reviews discussing methods of incorporating receptor flexibility into the drug design process [2, 81, 100, 101]. Molecular dynamics is becoming an increasingly practical tool to enhance drug discovery and design efforts. The performance of hardware and software is continuing to improve, and a growing number of researchers are turning their minds toward tackling the most pressing challenges facing the field. The authors' are optimistic that these factors will synergize to make the successful application of molecular dynamics simulations to structure based drug discovery and design problems routine.

CONFLICT OF INTEREST

The author(s) confirm that this article content has no conflicts of interest.

ACKNOWLEDGEMENTS

We thank Professor J. Andrew McCammon for support and critical discussions in the initial planning of this review. This work has been supported in part by grants from NIH, NSF, the Center for Theoretical Biological Physics, the National Biomedical Computing Resource, and Howard Hughes Medical Institute. RVS and REA are supported in part by the National Institutes of Health through the NIH Director's New Innovator Award Program DP2-OD007237.

REFERENCES

- [1] Amaro, R.E.; and W.W. Li, Emerging methods for ensemble-based virtual screening. *Curr. Top. Med. Chem.*, **2010**, *10*(1): p. 3-13.
- [2] Lexa, K.W.; and H.A. Carlson, Protein flexibility in docking and surface mapping. *Q. Rev. Biophys.*, **2012**, *1*(1): p. 1-42.
- [3] B-Rao, C.; J. Subramanian.; and S.D. Sharma., Managing protein flexibility in docking and its applications. *Drug Discovery Today*, **2009**, *14*(7-8): p. 394-400.
- [4] Henzler, A.M.; and M. Rarey, In pursuit of fully flexible protein-ligand docking: Modeling the bilateral mechanism of binding. *Mol. Inform.*, **2010**, *29*(3): p. 164-173.
- [5] Lill, M.A., Efficient incorporation of protein flexibility and dynamics into molecular docking simulations. *Biochemistry*, **2011**, *50*(28): p. 6157-6169.
- [6] Kokh, D.B.; R.C. Wade.; and W. Wenzel., Receptor flexibility in small-molecule docking calculations. *Wiley Interdiscip Rev: Comput Mol Sci*, **2011**, *1*(2): p. 298-314.
- [7] Smith, R.D., et al., Exploring protein-ligand recognition with binding moad. *J Mol Graphics Modell* **2006**, *24*(6): p. 414-425.
- [8] Henzler-Wildman, K.; and D. Kern., *Dynamic personalities of proteins. Nature*, **2007**, *450*(7172): p. 964-972.
- [9] Fernandez, A.; C. Fraser.; and L.R. Scott., *Purposely engineered drug-target mismatches for entropy-based drug optimization. Trends Biotechnol*, **2012**, *30*(1): p. 1-7.
- [10] Baron, R.; P. Setny.; and J.A. McCammon., *Water in cavity-ligand recognition. J Am Chem Soc*, **2010**, *132*(34): p. 12091-12097.
- [11] Peng, J.W., Communication breakdown: Protein dynamics and drug design. *Structure*, **2009**, *17*(3): p. 319-320.
- [12] Schlick, T., et al., Biomolecular modeling and simulation: A field coming of age. *Q Rev Biophys*, **2011**, *44*(2): p. 191-228.
- [13] Silva, D.A., et al., A role for both conformational selection and induced fit in ligand binding by the lao protein. *PLoS Comp Biol*, **2011**, *7*(5): p. e1002054.
- [14] Zhou, H.X., From induced fit to conformational selection: A continuum of binding mechanism controlled by the timescale of conformational transitions. *Biophys J*, **2010**, *98*(6): p. L15-L17.
- [15] Hammes, G.G.; Y.C. Chang.; and T.G. Oas., Conformational selection or induced fit: A flux description of reaction mechanism. *Proc Natl Acad Sci USA*, **2009**, *106*(33): p. 13737-13741.
- [16] Vertessy, B.G.; and F. Orosz., From "fluctuation fit" to "conformational selection": Evolution, rediscovery, and integration of a concept. *Bioessays*, **2011**, *33*(1): p. 30-34.
- [17] McCammon, J.A., *Gated diffusion-controlled reactions. BMC Biophys*, **2011**, *4*: p. 4.
- [18] Perot, S., et al., Druggable pockets and binding site centric chemical space: A paradigm shift in drug discovery. *Drug Discovery Today*, **2010**, *15*(15-16): p. 656-667.
- [19] Nisius, B.; F. Sha.; and H. Gohlke., Structure-based computational analysis of protein binding sites for function and druggability prediction. *J Biotechnol*, **2012**, *159*(3): p. 123-134.
- [20] Lahti, J.L., et al., Bioinformatics and variability in drug response: A protein structural perspective. *J R Soc, Interface*, **2012**, *9*(72): p. 1409-1437.
- [21] Bienstock, R.J., Computational drug design targeting protein-protein interactions. *Curr Pharm Des* **2012**, *18*(9): p. 1240-1254.
- [22] Clackson, T.; and J.A. Wells., A hot-spot of binding-energy in a hormone-receptor interface. *Science*, **1995**, *267*(5196): p. 383-386.
- [23] DeLano, W.L., Unraveling hot spots in binding interfaces: Progress and challenges. *Curr Opin Struct Biol*, **2002**, *12*(1): p. 14-20.
- [24] Zhao, H.Y.; Y.D. Yang.; and Y.Q. Zhou., Structure-based prediction of rna-binding domains and rna-binding sites and application to structural genomics targets. *Nucleic Acids Res*, **2011**, *39*(8): p. 3017-3025.
- [25] Lee, A.G., *Lipid-protein interactions. Biochem Soc Trans* **2011**, *39*: p. 761-766.
- [26] Morrow, J.K.; and S. Zhang., Computational prediction of protein hot spot residues. *Curr Pharm Des*, **2012**, *18*(9): p. 1255-1265.
- [27] Kuntz, I.D., et al., A geometric approach to macromolecule-ligand interactions. *J Mol Biol*, **1982**, *161*(2): p. 269-288.
- [28] Moore, G.E., *Cramming more components onto integrated circuits. Electronics Magazine*, **1965**, *38*: p. 4.
- [29] Leach, A.R.; and I.D. Kuntz., Conformational analysis of flexible ligands in macromolecular receptor sites. *J Comput Chem*, **1992**, *13*(6): p. 730-748.
- [30] Huang, S.Y.; S.Z. Grinter.; and X. Zou., Scoring functions and their evaluation methods for protein-ligand docking: Recent advances and future directions. *Phys Chem Chem Phys* **2010**, *12*(40): p. 12899-12908.
- [31] Huey, R., et al., A semiempirical free energy force field with charge-based desolvation. *J Comput Chem*, **2007**, *28*(6): p. 1145-1152.
- [32] Meng, E.C.; B.K. Shoichet.; and I.D. Kuntz., Automated docking with grid-based energy evaluation. *J Comput Chem*, **1992**, *13*(4): p. 505-524.
- [33] Friesner, R.A., et al., Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem*, **2004**, *47*(7): p. 1739-1749.
- [34] Stumpfe, D.; P. Ripphausen.; and J. Bajorath., Virtual compound screening in drug discovery. *Future Med Chem* **2012**, *4*(5): p. 593-602.
- [35] Mayr, L.M.; and D. Bojanic., Novel trends in high-throughput screening. *Curr Opin Pharmacol* **2009**, *9*(5): p. 580-588.
- [36] Barreiro, G., et al., Search for non-nucleoside inhibitors of hiv-1 reverse transcriptase using chemical similarity, molecular docking,

- and mm-gb/ssa scoring. *J Chem Inf Model*, **2007**, 47(6): p. 2416-2428.
- [37] Barreiro, G., *et al.*, From docking false-positive to active anti-hiv agent. *J Med Chem*, **2007**, 50(22): p. 5324-9.
- [38] Tomlinson, S.M., *et al.*, Structure-based discovery of dengue virus protease inhibitors. *Antiviral Res* **2009**, 82(3): p. 110-114.
- [39] Nichols, S.E., *et al.*, Discovery of wild-type and y181c mutant non-nucleoside hiv-1 reverse transcriptase inhibitors using virtual screening with multiple protein structures. *J Chem Inf Model*, **2009**, 49(5): p. 1272-1279.
- [40] Amaro, R.E., *et al.*, Discovery of drug-like inhibitors of an essential rna-editing ligase in trypanosoma brucei. *Proc Natl Acad Sci USA*, **2008**, 105(45): p. 17278-17283.
- [41] Durrant, J.D., *et al.*, Novel naphthalene-based inhibitors of trypanosoma brucei rna editing ligase 1. *PLoS Neglected Trop Dis*, **2010**, 4(8): p. e803.
- [42] Bottegoni, G., *et al.*, Systematic exploitation of multiple receptor conformations for virtual ligand screening. *Plos One*, **2011**, 6(5): p. e18845.
- [43] Rueda, M.; G. Bottegoni.; and R. Abagyan., Recipes for the selection of experimental protein conformations for virtual screening. *J Chem Inf Model*, **2010**, 50(1): p. 186-193.
- [44] Amaro, R.E.; R. Baron.; and J.A. McCammon., An improved relaxed complex scheme for receptor flexibility in computer-aided drug design. *J Comput-Aided Mol Des*, **2008**, 22(9): p. 693-705.
- [45] Fawcett, T., An introduction to roc analysis. *Pattern Recognition Letters*, **2006**, 27(8): p. 861-874.
- [46] Truchon, J.F.; and C.I. Bayly., Evaluating virtual screening methods: Good and bad metrics for the "early recognition" problem. *J Chem Inf Model*, **2007**, 47(2): p. 488-508.
- [47] Zhao, W., *et al.*, A statistical framework to evaluate virtual screening. *BMC Bioinf*, **2009**, 10: p. 225.
- [48] Sheridan, R.P., *et al.*, Protocols for bridging the peptide to nonpeptide gap in topological similarity searches. *J Chem Inf Comput Sci*, **2001**, 41(5): p. 1395-1406.
- [49] Jain, A.N.; and A. Nicholls., Recommendations for evaluation of computational methods. *J Comput-Aided Mol Des*, **2008**, 22(3-4): p. 133-139.
- [50] Nicholls, A., What do we know and when do we know it? *J Comput-Aided Mol Des*, **2008**, 22(3-4): p. 239-255.
- [51] Hanley, J.A.; and B.J. McNeil., The meaning and use of the area under a receiver operating characteristic (roc) curve. *Radiology*, **1982**, 143(1): p. 29-36.
- [52] du Prel, J.B., *et al.*, Confidence interval or p-value?: Part 4 of a series on evaluation of scientific publications. *Dtsch Arztebl Int*, **2009**, 106(19): p. 335-339.
- [53] Nicholls, A., What do we know?: *Simple statistical techniques that help. Methods Mol Biol*, **2011**, 672: p. 531-581.
- [54] Snedecor, G.W.; and C. W.G., *Statistical methods*. 7 ed1980, Ames: The Iowa State University Press.
- [55] Furnham, N., *et al.*, Is one solution good enough? *Nat Struct Mol Bio*, **2006**, 13(3): p. 184-185.
- [56] Berman, H.M., *et al.*, The protein data bank. *Nucleic Acids Res*, **2000**, 28(1): p. 235-242.
- [57] Ponder, J.W.; and D.A. Case., Force fields for protein simulations. *Adv Protein Chem*, **2003**, 66: p. 27-+.
- [58] Mackerell, A.D.; M. Feig, and C.L. Brooks., Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J Comput Chem*, **2004**, 25(11): p. 1400-1415.
- [59] MacKerell, A.D., *et al.*, All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B*, **1998**, 102(18): p. 3586-3616.
- [60] Jorgensen, W.L.; D.S. Maxwell.; and J. TiradoRives., Development and testing of the opls all-atom force field on conformational energetics and properties of organic liquids. *J Am Chem Soc*, **1996**, 118(45): p. 11225-11236.
- [61] Vanommeslaeghe, K., *et al.*, Charmm general force field: A force field for drug-like molecules compatible with the charmm all-atom additive biological force fields. *J Comput Chem*, **2010**, 31(4): p. 671-690.
- [62] Wang, J.M., *et al.*, Development and testing of a general amber force field. *J Comput Chem*, **2004**, 25(9): p. 1157-1174.
- [63] Jorgensen, W.L.; and J. Tirado-Rives., Potential energy functions for atomic-level simulations of water and organic and biomolecular systems. *Proc Natl Acad Sci USA*, **2005**, 102(19): p. 6665-6670.
- [64] Beauchamp, K.A., *et al.*, Are protein force fields getting better? A systematic benchmark on 524 diverse nmr measurements. *J Chem Theory Comput*, **2012**, 8(4): p. 1409-1414.
- [65] Mark, P., and L. Nilsson., Structure and dynamics of the tip3p, spc, and spc/e water models at 298 k. *J Phys Chem A*, **2001**, 105(43): p. 9954-9960.
- [66] Ellis, R.J., Macromolecular crowding: An important but neglected aspect of the intracellular environment. *Curr Opin Struct Biol*, **2001**, 11(1): p. 114-119.
- [67] Caves, L.S.D.; J.D. Evanseck.; and M. Karplus., Locally accessible conformations of proteins: Multiple molecular dynamics simulations of crambin. *Protein Sci*, **1998**, 7(3): p. 649-666.
- [68] Bowers, K.J., *et al.* Scalable algorithms for molecular dynamics simulations on commodity clusters. in Proceedings of the ACM/IEEE Conference on Supercomputing (SC06). **2006**, Tampa, Florida.
- [69] Hess, B., *et al.*, Gromacs 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J Chem Theory Comput*, **2008**, 4(3): p. 435-447.
- [70] Phillips, J.C., *et al.*, Scalable molecular dynamics with namd. *J Comput Chem*, **2005**, 26(16): p. 1781-802.
- [71] Brooks, B.R., *et al.*, Charmm: The biomolecular simulation program. *J Comput Chem*, **2009**, 30(10): p. 1545-1614.
- [72] Case, D.A., *et al.*, 2012, University of California, San Francisco.
- [73] Balsara, M.A., *et al.*, Principal component analysis and long time protein dynamics. *J Phys Chem*, **1996**, 100(7): p. 2567-2572.
- [74] Hensen, U., *et al.*, Exploring protein dynamics space: The dynamo as the missing link between protein structure and function. *Plos One*, **2012**, 7(5).
- [75] Humphrey, W.; A. Dalke.; and K. Schulten., Vmd: Visual molecular dynamics. *J Mol Graphics Modell*, **1996**, 14(1): p. 33-38.
- [76] Grant, B.J., *et al.*, Bio3d: An r package for the comparative analysis of protein structures. *Bioinformatics*, **2006**, 22(21): p. 2695-2696.
- [77] Michaud-Agrawal, N., *et al.*, Mdanalysis: A toolkit for the analysis of molecular dynamics simulations. *J Comput Chem*, **2011**, 32(10): p. 2319-2327.
- [78] Wriggers, W., *et al.*, Automated event detection and activity monitoring in long molecular dynamics simulations. *J Chem Theory Comput*, **2009**, 5(10): p. 2595-2605.
- [79] Grunberg, R.; M. Nilges.; and J. Leckner., Biskit--a software platform for structural bioinformatics. *Bioinformatics*, **2007**, 23(6): p. 769-70.
- [80] Vendruscolo, M.; and C.M. Dobson., Protein dynamics: Moore's law in molecular biology. *Curr Biol*, **2011**, 21(2): p. R68-70.
- [81] Harvey, M.J.; and G. De Fabritiis., High-throughput molecular dynamics: The powerful new tool for drug discovery. *Drug Discovery Today*, **2012**.
- [82] Baker, J.A.; and J.D. Hirst., Molecular dynamics simulations using graphics processing units. *Mol Inf*, **2011**, 30(6-7): p. 498-504.
- [83] Buch, I.; T. Giorgino.; and G. De Fabritiis., Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations. *Proc Natl Acad Sci USA*, **2011**, 108(25): p. 10184-10189.
- [84] Shan, Y.B., *et al.*, How does a drug molecule find its target binding site? *J Am Chem Soc*, **2011**, 133(24): p. 9181-9183.
- [85] Cavasotto, C.N.; J.A. Kovacs.; and R.A. Abagyan., Representing receptor flexibility in ligand docking through relevant normal modes. *J Am Chem Soc*, **2005**, 127(26): p. 9632-40.
- [86] Nichols, S.E., *et al.*, Predictive power of molecular dynamics receptor structures in virtual screening. *J Chem Inf Model*, **2011**, 51(6): p. 1439-46.
- [87] Lin, J.H., *et al.*, Computational drug design accommodating receptor flexibility: The relaxed complex scheme. *J Am Chem Soc*, **2002**, 124(20): p. 5632-3.
- [88] Lin, J.H., *et al.*, The relaxed complex method: Accommodating receptor flexibility for drug design with an improved scoring scheme. *Biopolymers*, **2003**, 68(1): p. 47-62.

- [89] Xu, M.; and M.A. Lill., Utilizing experimental data for reducing ensemble size in flexible-protein docking. *J Chem Inf Model*, **2012**, 52(1): p. 187-98.
- [90] Osguthorpe, D.J.; W. Sherman.; and A.T. Hagler., Exploring protein flexibility: Incorporating structural ensembles from crystal structures and simulation into virtual screening protocols. *J Phys Chem B*, **2012**, 116(23): p. 6952-9.
- [91] Osguthorpe, D.J.; W. Sherman.; and A.T. Hagler., Generation of receptor structural ensembles for virtual screening using binding site shape analysis and clustering. *Chem Biol Drug Des*, **2012**,
- [92] Durrant, J.D.; C.A. de Oliveira.; and J.A. McCammon., Povme: An algorithm for measuring binding-pocket volumes. *J Mol Graphics Modell*, **2011**, 29(5): p. 773-6.
- [93] Le Guilloux, V.; P. Schmidtke.; and P. Tuffery., Fpocket: An open source platform for ligand pocket detection. *BMC Bioinf*, **2009**, 10: p. 168.
- [94] Korb, O., *et al.*, Potential and limitations of ensemble docking. *J Chem Inf Model*, **2012**, 52(5): p. 1262-74.
- [95] Paulsen, J.L.; and A.C. Anderson., Scoring ensembles of docked protein:Ligand interactions for virtual lead optimization. *J Chem Inf Model*, **2009**, 49(12): p. 2813-9.
- [96] Warren, G.L., *et al.*, A critical assessment of docking programs and scoring functions. *J Med Chem*, **2006**, 49(20): p. 5912-31.
- [97] Niesen, F.H.; H. Berglund.; and M. Vedadi., The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat Protoc*, **2007**, 2(9): p. 2212-21.
- [98] Chang, M.W., *et al.*, Virtual screening for hiv protease inhibitors: A comparison of autodock 4 and vina. *Plos One*, **2010**, 5(8): p. e11955,
- [99] *Scifinder*, Chemical Abstracts Service: Columbus, OH.
- [100] Tuffery, P.;and P. Derreumaux., Flexibility and binding affinity in protein-ligand, protein-protein and multi-component protein interactions: Limitations of current computational approaches. *J. R. Soc. Interface*, **2012**, 9(66): p. 20-33.
- [101] Morra, G., *et al.*, Molecular recognition and drug-lead identification: What can molecular simulations tell us? *Curr. Med. Chem.*, **2010**, 17(1): p. 25-41.