JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

pubs.acs.org/jcim

Perspective

Rigorous Free Energy Simulations in Virtual Screening

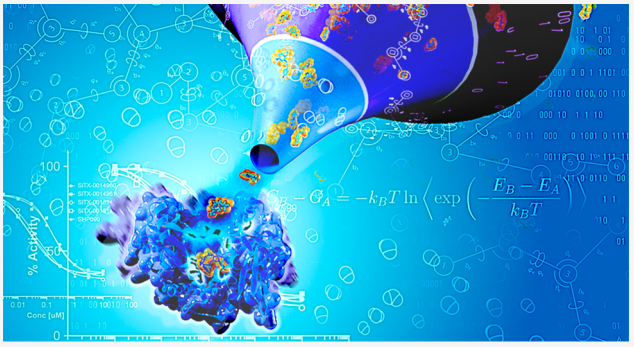
Zoe Cournia, Bryce K. Allen, Thijs Beuming, David A. Pearlman, Brian K. Radak, and Woody Sherman*

Cite This: J. Chem. Inf. Model. 2020, 60, 4153−4169

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: Virtual high throughput screening (vHTS) in drug discovery is a powerful approach to identify hits: when applied successfully, it can be much faster and cheaper than experimental high-throughput screening approaches. However, mainstream vHTS tools have significant limitations: ligand-based methods depend on knowledge of existing chemical matter, while structure-based tools such as docking involve significant approximations that limit their accuracy. Recent advances in scientific methods coupled with dramatic speedups in computational processing with GPUs make this an opportune time to consider the role of more rigorous methods that could improve the predictive power of vHTS workflows. In this Perspective, we assert that alchemical binding free energy methods using all-atom molecular dynamics simulations have matured to the point where they can be applied in virtual screening campaigns as a final scoring stage to prioritize the top molecules for experimental testing. Specifically, we propose that alchemical absolute binding free energy (ABFE) calculations offer the most direct and computationally efficient approach within a rigorous statistical thermodynamic framework for computing binding energies of diverse molecules, as is required for virtual screening. ABFE calculations are particularly attractive for drug discovery at this point in time, where the confluence of large-scale genomics data and insights from chemical biology have unveiled a large number of promising disease targets for which no small molecule binders are known, precluding ligand-based approaches, and where traditional docking approaches have foundered to find progressible chemical matter.



INTRODUCTION

Identifying hits for therapeutic targets of interest is an ongoing effort in drug discovery. Due to the continual discovery of disease targets through mining of an ever-increasing amount of genomic data, coupled with our improved understanding of chemical biology, the need is greater than ever to discover molecules that bind to these targets. High-throughput screening (HTS) has been a traditional approach used in large pharmaceutical companies to find hits for a target of interest, where libraries of millions of compounds that have been accumulated over decades (from internal research, company acquisitions, and publicly available chemical sources) are screened. Still, HTS remains very expensive and time-consuming.[1] Even when HTS is performed, there can be a dearth of hits for challenging targets (e.g., KRAS,[2] MYC,[3] STING,[4] PPIs,[5] and others) because the existing chemical libraries have been built around more traditional targets (e.g., kinases, proteases, and GPCRs[6−9] and might not contain any molecules that bind to the target of interest.

In recent years, it has become increasingly clear that virtual high-throughput screening (vHTS), which involves computationally screening libraries of molecules to discover hits, can yield hit rates that are significantly higher than random screens.[10−12] The computational methods deployed in vHTS vary widely, owing to an array of approximations, implementation choices, and availability of data. Some methods rely on similarity to known active molecules (ligand-based), while others use information about the protein target (structure-based). These methods can differ greatly in speed, from seconds on a desktop machine (e.g., Fast ROCS[13] or fingerprints[14]) to days on thousands of processors (e.g., docking[15]) for a typical chemical library of millions of molecules. Recently, ultralarge virtual libraries containing billions of molecules have been screened in less than a day using up to 45 000 CPUs.[16,17] When both ligand and structural information exists, it is possible to combine methods to yield improved results.[18,19]

The ability of a given method to find active compounds depends greatly on the complexity of the protein target of interest. For example, for a target that is well-characterized with many known active compounds (e.g., a kinase or

Special Issue: New Trends in Virtual Screening

Received: January 31, 2020
Published: June 16, 2020

https://dx.doi.org/10.1021/acs.jcim.0c00116
J. Chem. Inf. Model. 2020, 60, 4153−4169

protease), it might be possible to find additional active molecules from screening millions of compounds in a few seconds with ligand-based methods. In other cases, when the target structure is known but there are no known actives, even the most advanced docking methods being deployed today can yield few hits, especially when the target binding site is unique compared with previously drugged targets on which the docking scoring functions have been trained. These more challenging cases require improved computational methods that can "find a needle in the haystack".

Methods have been proposed to improve upon docking scoring methods, such as MM/GBSA,[20] where some of the energy terms missing in docking methods have been added (e.g., treating water with an implicit solvent model or including limited protein dynamics), but the value of such post-docking rescoring has been inconclusive, improving results considerably in some cases[21] but not in others.[22] This discrepancy is likely the result of the thermodynamic terms that are either missing or poorly treated, such as explicit water molecules, protein dynamics, and entropy, plus unreliable approximations in some of the other terms (i.e., implicit solvent and ligand strain). While it remains computationally and technically challenging to compute these terms accurately, the free energy simulation field has clearly defined rigorous statistical thermodynamic frameworks to achieve this objective and considerable work has been carried out to develop binding free energy methods that can better evaluate the aforementioned energy terms.

Here, we propose that alchemical free energy methods are the most computationally efficient of the rigorous all-atom methods for predicting binding free energies and therefore the most suitable for deployment in virtual screening. Rigorous binding free energy methods, as described in the work here, account for the primary energetic contributions to binding, including water (represented explicitly), protein dynamics, and entropy of the system, in addition to the more commonly treated energetic terms such as protein−ligand interactions and ligand strain. Additionally, rigorous physics-based free energy methods can find more diverse hits than other methods, owing to the unbiased nature of the simulations, accounting for protein flexibility, and the avoidance of reference ligands needed for ligand-based methods. While it is not practical at the time of this article to perform rigorous alchemical simulations on millions of molecules, it is feasible to run on thousands, which can provide significant benefits in the final triaging of molecules to be purchased for experimental testing. Even in cases of docking ultralarge libraries of billions of drug-like molecules to sample broad chemical space, benefits can be gained from incorporating binding free energy simulations to assess top scoring hits and augment the traditional virtual screening workflow with higher quality scoring. A hierarchical virtual screening funnel is shown in Figure 1, with absolute binding free energy (ABFE) simulations inserted as the final computational stage before human selection of compounds for purchase and experimental testing.

In this Perspective, we first summarize the thermodynamic framework and challenges associated with predicting protein−ligand binding energies, which is the primary objective of most virtual screening campaigns. We then briefly cover some of the traditional virtual screening methods being deployed today, such as ligand-based approaches and docking, discussing their strengths and limitations. Finally, we review rigorous binding free energy methods, which include alchemical (e.g., free energy perturbation, thermodynamic integration, MBAR, etc.), path-based (metadynamics, potential of mean force, string methods, etc.), and brute force approaches (unbiased simulations, where the thermodynamics of binding are computed directly based on the statistical sampling of the system). While many methods are rigorously derived from statistical thermodynamics, there are significant differences in efficiency that render some methods more practical than others for virtual screening. Of the rigorous methods, we propose that alchemical absolute binding free energy (ABFE) methods offer the most computationally efficient means to predict the binding free energy of diverse molecules in virtual screening. We conclude with a discussion of the future of virtual screening, including the promise of new hardware platforms such as quantum computers and limitations of physics-based methods in virtual screening.

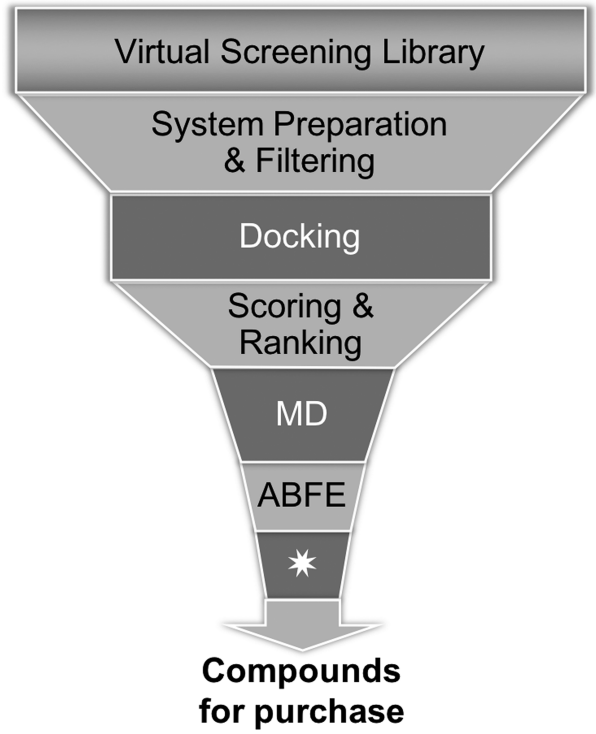


Figure 1. Example of a virtual screening funnel, where a large virtual screening library (typically millions to billions of compounds) is filtered in a hierarchical fashion to a manageable number of compounds for purchase (typically hundreds to thousands). The star (*) denotes any additional filtering, such as human selection, which is typical in drug discovery where experienced chemists and modelers make the final determination of the compounds to purchase based on a shortlist from the best scoring molecules. In this Perspective, we propose that ABFE be inserted as the final computational step because it is the most computationally efficient of the rigorous binding free energy method but still significantly more computationally expensive than the previous stages in the virtual screening funnel.

THERMODYNAMICS OF BINDING

Before discussing the plethora of virtual screening methods, we briefly describe the thermodynamics underlying biomolecular recognition (e.g., protein−ligand binding), which is essential to understanding the strengths/weaknesses of various vHTS methods. A solid understanding of the thermodynamic processes that govern the phenomena of biomolecular recognition is central both to improving virtual screening

https://dx.doi.org/10.1021/acs.jcim.0c00116
J. Chem. Inf. Model. 2020, 60, 4153−4169

methods and to more rigorously treating the underlying physics associated with protein−ligand binding. The ultimate goal of rigorous binding free energy methods is to completely recapitulate the underlying physics associated with biomolecular recognition using computation, thereby replacing the need for large-scale experimental screening and focusing experimental assays on a limited number of top-scoring virtual hits.

If computational resources were not an issue, then an ideal representation of protein−ligand binding would involve simulating the entire protein−ligand system solvated by water (typically tens of thousands of atoms) and all relevant cofactors, using a quantum mechanical energy model for time scales sufficiently long to capture the biological time scales of important protein motions. While attractive in theory, this approach would require inordinate amounts of time to complete even a single simulation with all accessible computations resources, so we must first approximate the quantum physics with simplified models (typically molecular mechanics force fields).[23] While we will not discuss force fields in detail here, it suffices to note that while an accurate force field is paramount to achieving robust and precise binding free energy predictions, a quality force field alone is insufficient: It is also necessary to sample the system adequately to obtain converged statistical thermodynamic quantities.

In Figure 2, we show a particular thermodynamic decomposition of protein−ligand binding, which has been separated in a way that is both rigorous in nature and practical in application (i.e., it separates energy terms in a way that should be familiar to computational and medicinal chemists in drug discovery). The initial state (top images of Figure 2) is one in which the protein and ligand are separated and in aqueous solution, which is analogous to the real-world situation of an unbound ligand. The ligand exists in an ensemble of conformations (Boltzmann weighted by the free energy of each conformation) and the protein similarly exhibits fluctuations due to side chain and other thermally accessible motions. The protein first adopts the induced-fit conformation appropriate to the particular ligand in question (A) while the ligand adopts the bioactive conformation while still in solution (B). Next, the protein and ligand are desolvated in a way that is consistent with the binding process (C and D). Finally, the bioactive conformation of the desolvated ligand comes together with the induced-fit conformation of the desolvated protein to form the protein−ligand complex (E). In this process, we can identify the important components to the binding free energy, including the change in configurational entropy, energetic strain of the protein and ligand, the desolvation energy associated with transfer of water from the ligand and protein surfaces into bulk solvent, and the nonbonded (enthalpic) interactions between the protein and ligand.

In the following section, we discuss the most common methods used in virtual screening, referencing their relationships to the thermodynamic decomposition shown above, so that the strengths/weaknesses of the methods can be more clearly understood. We feel this assessment is important for the field, not to point out deficiencies, but to help researchers understand the approximations involved with different methods, so that they can make the best decisions when pursuing a virtual screening campaign. We assert that rigorous free energy methods can now play an important role in virtual screening efforts, and can improve the chances of success for such efforts. While computational resources, force field accuracy, and technical challenges may still limit the applicability of rigorous free energy methods today, we anticipate that they will play an ever-growing role of importance as we embark on the discovery of medicines for challenging disease targets, especially in cases where insufficient known chemical matter limits our ability to deploy ligand-based models.

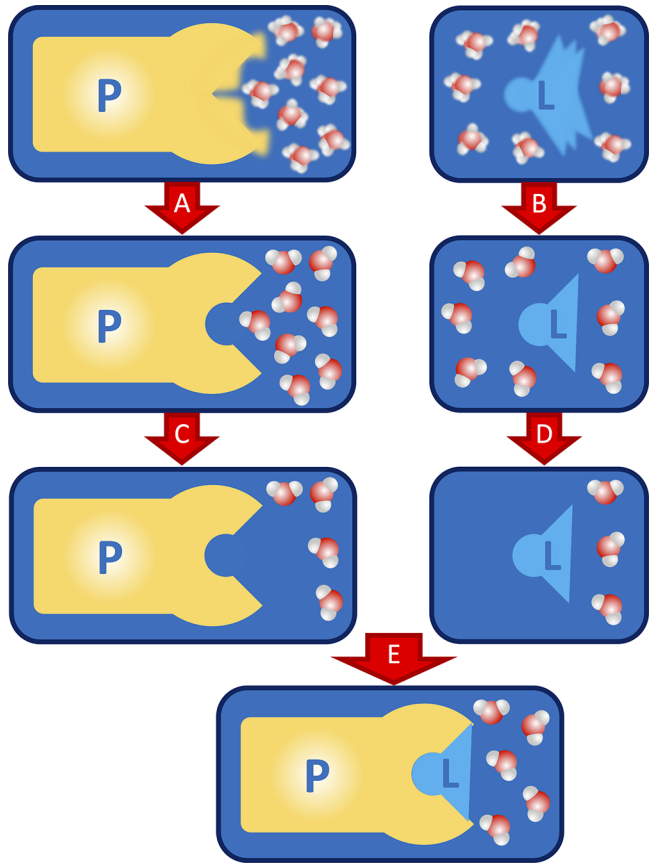


Figure 2. An example thermodynamic decomposition of the protein−ligand binding process into constituent components. The protein and ligand initially exist as a Boltzmann-weighted ensemble of conformational states in solution (top). Both protein and ligand then adopt bioactive conformations (arrow A for protein reorganization and arrow B for ligand conformational focusing). The protein and ligand are then desolvated (arrows C and D, respectively) as water molecules are expelled from their surface. Finally, the desolvated protein and ligand come together for binding (arrow E). The thermodynamic processes associated with arrows A and B typically have unfavorable entropy and enthalpy. The thermodynamic processes associated with arrows C and D typically have favorable entropy due to the release of confined water molecules and either favorable or unfavorable enthalpy, depending on the difference in interactions with the protein or ligand relative to water in bulk solvent. Finally, the thermodynamic process associated with arrow E is enthalpically favorable due to protein−ligand interactions but entropically unfavorable, as two entities come together. Note that there are other ways that protein−ligand binding can be decomposed into constituent parts: the example shown here is chosen because it maps nicely onto common drug discovery terminology such as strain, desolvation, and interactions.

TRADITIONAL VIRTUAL SCREENING METHODS

Overview. Virtual screening has been performed for decades,[24] with a large number of successful applications. Traditional virtual screening approaches can be broadly

https://dx.doi.org/10.1021/acs.jcim.0c00116
J. Chem. Inf. Model. 2020, 60, 4153−4169

classified into "ligand-based" and "structure-based", where the former includes similarity methods based on a representation of one or more known binders through chemical fingerprints, shape, shape plus electrostatics, pharmacophore features, or other representations that can be used with machine learning algorithms. The latter approach (structure-based) is dominated by docking methods that first identify possible ligand conformations, and then assess the scores of these ligand poses when bound to the protein (we call this approach "sample-and-score"). Below we will highlight the most common ligand-based and structure-based approaches, discussing the strengths and weaknesses of each.

Ligand-Based Methods. Ligand-based virtual screening methods use features of molecules derived from chemical structures of known binders to score new molecules from a chemical library. The scoring is typically based on a similarity metric using a representation amenable to computer algorithms, such as chemical fingerprints, shape, or pharmacophore features. Ligand-based methods can perform exceptionally well when there are known molecules that bind to the target of interest and library molecules are sufficiently similar to the known binders. Many studies have recapitulated the notion that molecules similar in structure share similar experimental attributes[25−28] reinforcing the validity and practical nature of such methods. Yet in a significant number of instances, this assumption is invalid and is dependent on the featurization method in use. For example, while two molecules with chiral methyl groups can be chemically very similar, the three-dimensional conformational preferences of the molecules can differ substantially, thereby contributing to arrow B (and possibly others) in Figure 2. Similarly, ortho- and meta-fluoro substitutions on a biaryl molecule will generally be deemed chemically similar, but the torsional preference of the biaryl system will shift significantly and may disrupt the binding mode of the molecule (possibly influencing arrows A and E, in addition to B of Figure 2). Thus, it is important to understand the limitations and applicability domain of each ligand-based virtual screening method. Furthermore, for many of the most interesting and therapeutically relevant targets, there are no known ligands on which to perform a ligand-based similarity search, and therefore such methods cannot be applied. In this section, we will briefly discuss some of the most commonly used ligand-based vHTS methods before moving to structure-based methods in the following sections.

Chemical fingerprints are widely used topological descriptors, with Morgan Extended-connectivity fingerprints perhaps being the most common.[14,29] Descriptors such as these define molecules based on their atom-type based connectivity to build a "fingerprint" of the molecule, which can then be compared using many different similarity metrics.[30−32] The representation is typically hashed to an integer value and the topological distance is defined as the number of bonds in the shortest path between each pair of atoms. Alternatively, topological torsion fingerprints encode fragments consisting of a path of four atoms that are differentiated by atom type.[33] Rather than leveraging topological descriptors, MACCS keys encode specific structural patterns via SMARTS (166 such patterns in the original implementation), which encode a count vector for each molecule of the occurrence of each substructure.[34] Continuing along these lines, physicochemical properties can be computed from molecular structures using first principle calculations or equations to derive numerical attributes. As opposed to fingerprints, which are often binary or count-based, physicochemical properties are often presented as a list (i.e., vector) of floating-point values. Many such physiochemical descriptors can be quickly generated using open source tools, such as Mordred, which is capable of calculating 1825 descriptors.[35] Chemical descriptors are continuously evolving, and new ones are being generated with the hope that there will be some latent relationship to binding properties.[36]

In addition to similarity searching, chemical descriptors can be used to train supervised machine learning models that construct nonlinear mappings of input features to experimental end points such as chemical bioactivity data. Unlike similarity searching, machine learning algorithms require training data that must be sufficient in quantity to generalize to unseen molecules within an applicability domain. Previous studies have demonstrated that generalization capacity increases with the amount and diversity of training data available. In addition, new representation learning techniques such as graph convolutional neural networks have shown significant improvements in predictive performance when the amount of data available for training exceeds 10k molecules. Machine learning methods have shown promise in numerous applications related to virtual screening and protein−ligand binding.[37−41]

If three-dimensional structural information is available for active molecules, other approaches such as shape or pharmacophore similarity can be utilized to discover new compounds. For shape-based screening, one searches for molecules that have similar shape to a known molecule of interest, with the possibility of adding additional information such as electrostatic properties.[42−44] Shape-based methods have the advantage of being easy to use (a reference ligand is all one needs to run a calculation) and very fast, especially using newer GPU implementations; although incomplete representation of the molecular conformational space may be a bottleneck. Therefore, to ensure reliable results using shape-based methods, it is important to balance speed with accuracy of the conformational sampling process, taking into consideration the size of the database to be screened.[45]

In pharmacophore modeling, the fundamental premise is that related chemical groups, such as hydrogen bond donors/acceptors and aromatic/hydrophobic centers, if oriented in spatially and geometrically similar arrangements, can facilitate comparable intramolecular interactions with a target receptor, thereby conferring similar biological activity.[46] If many active ligands are available, a consensus or common pharmacophore model can be generated by aligning the active ligands and identifying a consensus of pharmacophore elements. In general, pharmacophore modeling is most useful when information is known about active ligands such that important chemical features can be differentiated from inconsequential features. This type of information facilitates its use in phenotypic-based drug discovery projects or against targets that present difficulties in obtaining structures via X-ray crystallography or other approaches.

Docking. Of the structure-based virtual screening methods, docking is the most commonly used approach due to its speed and simplicity. In general, docking involves two stages: (1) sampling the pose of the ligand in the protein binding pocket and (2) scoring the binding energy of the ligand to the receptor. For this reason, we will refer to this approach as "sample-and-score", where sampling and scoring will be described separately below together with the different methods that can be applied for each.

https://dx.doi.org/10.1021/acs.jcim.0c00116
J. Chem. Inf. Model. 2020, 60, 4153−4169

Docking: Sampling. To properly score a compound for virtual screening, it is first necessary to predict the pose of each ligand binding to the target. For some systems, this sampling problem is relatively well-solved: predicting the correct binding pose in the receptor site can be determined accurately and rapidly enough that it is possible to reliably screen millions of potential binders in a relatively short amount of time. Still, significant challenges still exist when protein flexibility is involved or when ligands are complex (e.g., large macrocyclic compounds or peptides).

The basic methodology of docking was first described more than 40 years ago by the Kuntz group, who released their widely used DOCK program shortly thereafter.[47] While a large variety of docking methods have since been developed (see refs.[48,49] for comprehensive reviews), most approaches still rely on the rather simplistic treatment established in the early methods (flexible ligand, simplified energy model, rigid receptor, no explicit solvent, and no accounting for entropy). In this section, we review the basic machinery of docking, and critically analyze the underlying approximations that are established for this approach to be useful for virtual screening. While we propose in this work that docking is insufficient for highly accurate scoring in the final stages of a vHTS workflow, it is still important to quickly filter large virtual screening libraries and to predict poses that can be used as input structures for rigorous free energy methods, which will be presented later in this work.

All sample-and-score docking programs face the same fundamental problem—how to sample and score an almost infinite set of possible ligand conformations and orientations in a binding site of a protein on a reasonable time scale (i.e., a few seconds or minutes). Due to the large and growing chemical databases that are typically used in virtual screening calculations (traditionally in the millions, and now growing to billions with the availability of virtual compounds),[16] docking algorithms require some drastic approximations to evaluate the molecular recognition event while maintaining the throughput needed to impact drug discovery projects. These approximations pertain to both sampling and scoring. Most docking programs achieve speed through a funnel-like approach, where at the outset of the calculations, when millions of possible candidate poses need to be evaluated, the structural complexity of the ligand binding site is significantly reduced and/or the treatment of the molecules is greatly simplified. When the number of reasonable poses is sufficiently reduced, increased structural detail can be included to more accurately determine the energetics of interactions. Note, however, that increasing the accuracy of the protein−ligand interactions is necessary but insufficient to accurately predict the binding free energy, because energetic terms such as entropy, solvation, and dynamics are still largely omitted (see Figure 2). Additionally, it should be emphasized that sampling of the ligand pose will be futile if the protein is not in the correct conformation,[50] and even small protein conformational changes can make it impossible to predict the correct protein−ligand pose (e.g., if a protein side chain is blocking the binding site).

Docking programs like Glide[51] and Fred[52] use an anchor and growth strategy to gradually build up the ligand in the binding site whereas docking programs such as AutoDock[53] and Gold[54] use a Genetic Algorithm (GA) approach. The GA algorithm addresses the high computational cost associated with stochastic methods by applying concepts derived from the theory of evolution and natural selection, iteratively optimizing low energy conformations found in random searches across the energy landscape.[54]

A simple approach that offers a crude approximation of protein rearrangement upon ligand binding is to make scoring functions relatively "soft", i.e. allowing for minor overlap of the van der Waals radii of atoms between the protein and the ligand. The advantage of this approach relative to less forgiving scoring functions is that poses with minor clashes can still get a reasonable score, but the drawbacks are reduced accuracy and an increase in false positives (inactive compounds that also score well). An alternative approach to indirectly deal with receptor flexibility is to dock into multiple structures (obtained via crystallography or from a molecular dynamics simulation) via ensemble docking.[55]

While more robust methods for treating protein rearrangement in docking exist (Rosetta, IFD),[56−58] they have found limited application in virtual screening due to the thus-far prohibitive computational cost required to generate poses, and perhaps more importantly, due to their inability to adequately quantify the energetic effects of protein strain on the binding energy. We are not aware of any studies where such approaches have been used for virtual screening. In the absence of a rigorous method to estimate protein reorganization energies, there will always remain an upper limit on the accuracy that can be obtained with these sample-and-score type docking approaches.

Another major shortcoming of classical sample-and-score docking experiments is their inability to treat the structural and energetic effects of water. A water molecule in a protein binding site can be displaced, replaced, bridged, or altogether avoided, depending on the energetics of the water molecule in question and the nature of the ligand in the binding site. Ideally, a docking algorithm would sample the positions/orientations of water molecules on the fly and calculate the effect of the interaction with the ligand (see arrows C and D in Figure 2). This type of calculation, however, is beyond the capability of most methods. The effects of water on binding energetics are complex and require a full characterization of the thermodynamics for the method to be quantitatively predictive. Attempts have been made to include a minimal set of water molecules (either explicit or implicit) in docking calculations,[59] but these treatments are incomplete, and the benefits have not been consistently demonstrated in the context of virtual screening.

Docking: Scoring. Scoring functions, in general, have two separate roles in virtual screening. First, for each ligand, a scoring function is required to predict the most likely pose within the set of all generated poses. Second, scoring functions are then used to rank-order the top predicted poses for each compound to compare energies across compounds. These two scoring functions are not necessarily identical—due to cancellation of errors with regard to many terms in the binding free energy (i.e., ligand desolvation; arrow D in Figure 2) scoring functions for pose determination can be significantly simpler than those for rank-ordering. Here, we consider the former scoring function (to predict the best pose for a given molecule) as part of "sampling", whereas the latter (comparing the energies across different molecules) as part of "scoring".

Even if a docking algorithm can provide a pose of sufficient accuracy to manifest all the critical interactions of a complex (hydrogen bonds, salt bridges, hydrophobic interactions, induced-fit effects), the fact remains that estimating the energy

https://dx.doi.org/10.1021/acs.jcim.0c00116
J. Chem. Inf. Model. 2020, 60, 4153−4169

from a static representation of the ligand-protein complex is inaccurate, and the degree of inaccuracy will increase with the amount of protein flexibility and importance of other terms shown in Figure 2 that are not accounted for in the scoring function. For example, docking approaches generally neglect arrows A and C in Figure 2, which can be a reasonable approximation when the target protein is rigid and has a buried binding site that is equally desolvated by all ligands. Nevertheless, situations like this are an exception, not the rule.

Scoring functions can be grouped into three major types: force field, empirical, and knowledge-based. Force field-based functions consist of a sum of energy terms derived from simple molecular mechanical approximations, accounting for both bonded and nonbonded terms. Noncovalent interactions are treated with simple Coulombic (for electrostatics) or Lennard−Jones (for van der Waals) interaction potentials, plus some representation of the internal energetics of the ligands. The advantage of this approach is the transferability across targets and the reduced need for target-specific training/tuning. The drawback is that the approximations in these scoring functions (namely protein reorganization, explicit waters, and entropy) can lead to inaccurate results. Even with the most sophisticated force fields, such as those that incorporate polarizability, the lack of protein sampling remains a limiting factor in the applicability of these scoring functions. Even a quantum mechanical energy model would suffer the same limitations with regard to protein flexibility and all of the other important factors in binding that are neglected in sampling-and-score docking approaches. Empirical scoring functions such as ChemScore[60] and GlideScore[61] are obtained by fitting scoring function terms against known experimental binding affinity data. These methods combine terms based on force field models with additional empirically derived terms that have been empirically fit to experimental data and generally provide better virtual screening results than the force field scoring functions.

All of these scoring functions suffer in many ways from inaccurate descriptions of terms shown in Figure 2:

1. Desolvation (ligand and protein)
2. Entropic penalties (ligand, protein, and water)
3. Conformational strain (ligand and protein)

Here, we briefly elaborate on the above energetic terms and describe some approaches that have been employed to address them at various levels of accuracy.

Desolvation. Some efforts have been made to incorporate protein desolvation effects into docking algorithms. For example, Wscore[62] is an extension of Glide that uses a description of the water thermodynamics in a binding site (obtained through an MD calculation)[63,64] to score poses. Other work has been reported that includes empirical terms for ligand desolvation, such as the HYDE (HYdrogen bonds and DEsolvation) scoring function.[65,66] Limited work has been carried out to account for protein desolvation in docking programs, although post-docking scoring methods such as MM/PBSA and MM/GBSA account for protein desolvation through an implicit solvent model.

Entropy. Conformational focusing (the entropic cost associated with loss of degrees of freedom upon complex formation) is incorporated implicitly for the ligand in most scoring functions by penalizing each rotatable bond in a ligand. This is a highly simplistic approach that works only when comparing molecules of similar sizes and/or numbers of rotatable bonds and neglects entropic contributions from the protein and water.

Ligand Strain. The inability to incorporate induced fit effects into rapid docking algorithms often causes ligands to be docked with suboptimal and nonrealistic internal geometries. For this reason, it is difficult to accurately determine the amount of strain energy that a given ligand incurs. A common practice is to characterize the conformation landscape of a ligand separately using molecular mechanics and quantum chemical approaches and exclude from consideration poses that correspond to a conformation with high internal energy. Needless to say, this becomes intractable in cases where large numbers of ligands are to be considered, and still neglects the protein reorganization energy (i.e., strain) that would be required to allow for a relaxed ligand conformation.

Protein Strain. Adjustment of the protein to adopt the induced-fit (holo) conformation appropriate to a given ligand will always lead to a higher (less favorable) free energy as compared to the apo structure of the protein. Quantifying this protein strain energy requires extensive characterization of the conformational landscape via molecular mechanics simulations. Hence, this term is typically ignored in scoring functions.

There have been some attempts to address the aforementioned limitations in various ways, primarily using approximate free energy methods that are based on continuum solvation methods, including the well-known MM/GBSA and MM/PBSA methods.[67] These methods are based on the concept that a linear combination of molecular mechanics (MM) energies, polar and nonpolar solvation terms, and an entropy term can approximate the free energy of binding of a ligand to a receptor. The free energy for each species (ligand, receptor, and complex) is decomposed into a gas-phase MM energy, polar, and nonpolar solvation terms, and an entropy term. In some cases, MM/GBSA postprocessing has been shown to modestly improve enrichment in virtual screening calculations.[68,69] In other cases, the value of this approach has not been so clear.[70]

Lastly, knowledge-based scoring functions circumvent the limitation imposed from the fixed functional form of force-field and empirical scoring functions by learning directly from training data using machine learning. Knowledge-based scoring functions could thus implicitly capture intermolecular binding interactions that are hard to model explicitly.[71] Many different implementations of knowledge-based scoring functions have been previously reported, including ML approaches such as random forests (RF),[72] support vector machines (SVM),[73,74] deep neural networks (DNN),[75] and convolutional neural networks (CNN).[39]

■ RIGOROUS BINDING FREE ENERGY METHODS

The fundamental premise of this Perspective is that we are at a point where science (decades of research from laboratories around the world) and technology (recent advances in computing technologies, namely GPUs) have culminated such that it is now possible to compute binding free energies with high accuracy and with sufficient throughput to be used in virtual screening campaigns as a high-resolution scoring method after large virtual libraries have been filtered with faster approximate methods and before the final selection of compounds for purchase. This paradigm shift in virtual screening methodology could not come at a better time from the standpoint of treating unmet medical needs, as there has been an explosion of genetically and biologically validated

https://dx.doi.org/10.1021/acs.jcim.0c00116
J. Chem. Inf. Model. 2020, 60, 4153−4169

targets that do not have any small molecule therapeutics modulators. While screening millions of compounds is not yet tractable using rigorous free energy methods, the framework for deploying these methods in the context of virtual screening is now in place, and with the ever-growing availability of computational resources and ongoing reduction in computing prices, the number of compounds that can be processed with rigorous free energy methods will continue to grow. We suggest that even with currently available resources these methods can be applied today in virtual screening toward the end of the funnel for the final compound selection, as shown in Figure 1.

Below, we discuss a variety of rigorous binding free energy methods. We start with alchemical methods, which we propose are the most efficient for the purpose of virtual screening, because the computational efforts are focused on the two states of interest for vHTS (ligand bound to the target versus unbound in solution). Specifically, we assert that ABFE is the best solution for finding diverse hits from a virtual screening library because it offers the most computationally efficient approach to converge binding free energies. We then discuss path-based methods, which are also rigorous in formulation but require sampling of the compete transition from bound to unbound (or vice versa) and therefore call for significantly more computational resources to assess the rare thermodynamic states associated with the binding transition state(s). We conclude with a brief discussion of unbiased sampling methods (i.e., the brute force approach), where simulations are run in an analogous fashion to how binding takes place in nature (e.g., simulations are run for long enough time to witness the binding and unbinding event enough times to compute the partition function needed to calculate a binding free energy). We focus on molecular dynamics (MD) methods due to their broad use and advantages in sampling concerted motions, but successes have been published using Monte Carlo approaches as well.[76−79] For additional details, we refer the reader to several excellent reviews on alchemical free energy methods for drug discovery, which cover the thermodynamic framework and practical considerations of MD-based and MC-based alchemical binding free energy methods.[80−85]

General Alchemical Free Energy Overview. The statistical mechanical framework for free energy calculations was developed many decades ago in the seminal contributions of Kirkwood[86] and Zwanzig,[87] who are credited with the thermodynamic integration (TI) and free energy perturbation (FEP) methods, respectively. Zwanzig's ideas were later expanded to a formulation in terms of a finite molecular "perturbation" in McQuarrie.[88] Both approaches can be described in terms of a transformation between two states, A and B, and the quantity of interest is the relative free energy, ΔG between those states. The definition of a "state" can be fairly loose and only requires the definition of energy functions, U, used to model the end points (e.g. bound and unbound ligand states). In the FEP formalism this can be written as

$$\Delta G = G_B - G_A = \Delta H - T\Delta S = -RT\ln\langle e^{(-(U_B-U_A)/RT)}\rangle_A \quad (1)$$

where the subscripts indicate the relevant state and the angle brackets denote an average value over the sampled ensemble. R is the ideal gas constant, and T is temperature. Zwanzig obtained an analytic series solution to eq 1 by assuming that the (known) free energy of an ideal gas is only nominally different from the nonideal case. Modern simulation methods can now be used to evaluate eq 1 directly for any arbitrary system but the efficiency/effectiveness of the approach is often highly contingent on the change between the two states being small. In an unfortunate mixing of parlances, the difference between states is now referred to as a perturbation (in the sense that the chemical structure is "perturbed") and is generally small. The solutions resulting from these "free energy perturbation" (FEP) simulations are theoretically *exact* and thus have only a conceptual relation to the actual approach pioneered by Zwanzig.

Below, we begin by describing Relative Binding Free Energy (RBFE) calculations because it is the most commonly used alchemical free energy approach in drug discovery applications and provide a useful framework for further discussions on free energy methods for vHTS even though RBFE calculations are not well-suited for virtual screening because they depend on commonalities between molecules. In virtual screening libraries such commonalities generally do not exist because diverse molecules are explored. As such, for virtual screening ABFE calculations are more directly applicable and in our opinion ABFE is the best approach in terms of speed and accuracy for vHTS.

RBFE Simulations. Calculating relative free energy differences between two molecules involves evaluation of the bracketed term in eq 1 by generating Boltzmann distributed samples using the energy function of state A. Note that the value for which the ensemble is being evaluated depends on the potential energies of both states. As such, one needs to sample a very long time to pick up not only configurations where U_A is favorable but also states where U_B is favorable (there must be at least some degree of *overlap*). If the energy functions for the states are very different (e.g., A and B represent very different bound ligands), then the chances of ever sampling low energy states for B while evolving under the potential U_A are quite small, and so it will be difficult to reliably evaluate eq 2. This issue is also somewhat exacerbated by explicit solvent models, which include the physical time scales of not only molecules A and B but also their differing interactions with solvent (e.g., different hydrogen bonding patterns). Explicit solvent also makes the sampling of protein domain motions difficult, due to water diffusion, which lengthens the time scale of such motions. Implicit solvent models have been proposed as one way to circumvent these latter problems, but implicit solvent models still suffer from the overlap issue and recent studies have shown that the neglect of explicit waters can lead to considerable artifacts.[89]

At the time Zwanzig published his equation, there was no way to evaluate the ensemble average for a molecular system of interest. But by the mid-1980s, molecular dynamics and Monte Carlo methods and the evolution of computers made it possible to apply the equation to real systems, as first demonstrated in the laboratories of McCammon[90] and Jorgensen.[91] These laboratories demonstrated how Zwanzig's equation could be applied to a closed thermodynamic cycle for protein−ligand binding, as shown in Figure 3.

The free energy differences that can be evaluated in the laboratory are represented by the vertical lines ΔG_1 and ΔG_2. These can either be assessed directly through calorimetric analysis (e.g., ITC) or estimated from the relationship shown in eq 2.

$$\Delta G = -RT\ln K_i \quad (2)$$

https://dx.doi.org/10.1021/acs.jcim.0c00116
J. Chem. Inf. Model. 2020, 60, 4153−4169

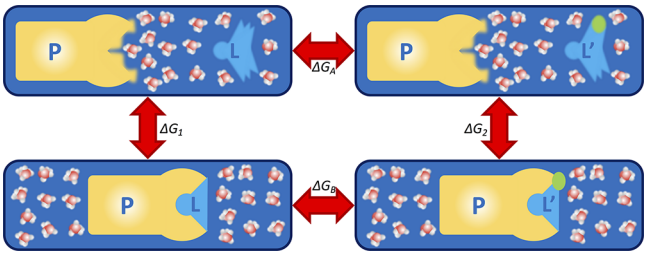


Figure 3. Closed thermodynamic cycle for RBFE calculations, where ligand L and modified ligand L′ bind to a protein target P. The experimental measurable quantities of interest are the binding energy of reference ligand L and modified ligand L′ to protein P, with associated binding free energies of ΔG_1 and ΔG_2, respectively. While these quantities can be computed, it is much more computationally efficient to compute the modification of L to L′ in solution (ΔG_A) and in the binding site (ΔG_B). Because this is a closed thermodynamic cycle, the relative binding free energy difference between the two ligands binding ($\Delta\Delta G_{bind}$) computed by $\Delta G_2 - \Delta G_1$ is identical to $\Delta G_B - \Delta G_A$.

where K_i is the experimentally measured binding constant (in units of molarity, as per standard state requirements).

The experimental binding processes (ΔG_1 and ΔG_2) are not easy to simulate directly. Computationally, it is much easier to simulate the horizontal processes in this cycle, ΔG_A and ΔG_B. Fortunately, because free energy is a state function (not dependent on the path, only on the endpoints), we can relate the difference in the experimental free energy changes to those calculated as shown in eq 3.

$$\Delta\Delta G_{bind} = \Delta G_2 - \Delta G_1 = \Delta G_B - \Delta G_A \Delta\Delta G_{bind} = \Delta G_2 - \Delta G_1 = \Delta G_B - \Delta G_A \quad (3)$$

As noted above, a fundamental limitation of free energy calculations is that one must determine the ensemble average of a quantity that depends on the difference of the energy functions of both endpoints from a molecular dynamics trajectory created by describing only one of them. Perturbations that are too large result in significant clashes with other molecules (e.g., if ligand B has a long solvent-exposed side chain, where ligand A has none, then the chances that there will be sufficient space in the surrounding solvent for B will be quite small). To circumvent this issue, a family of intermediate states is introduced by defining a coupling parameter λ such that the "generalized" energy function $U(\lambda)$ interpolates between the physical end points (i.e., $U(0) = U_A$ and $U(1) = U_B$). These intermediates constitute new "alchemical" endpoints for smaller perturbations. The fact that free energy is a state function means that, for any choice of this "λ pathway", the overall free energy can be simply computed as a sum over intermediate states:

$$\Delta G_{A\to B} = \Delta G_{A\to\lambda(1)} + \Delta G_{\lambda(1)\to\lambda(2)} + \ldots + \Delta G_{\lambda(N-1)\to B} \quad (4)$$

The TI formalism can be written using the same generalized energy function definition above and supposing an arbitrary set of states between A and B. An (exact) alternative to the free energy expression from FEP can then be written in terms of an integral

$$\Delta G = \int_{\lambda=0}^{1} <\partial H/\partial\lambda>_\lambda d\lambda \quad (5)$$

In this equation, λ has the same meaning as above, representing a continuous variable that takes the system from U_A to U_B as λ progresses from 0 to 1. In practice, the integral is evaluated via some form of quadrature by calculating the integrand at a finite number of values of λ. The difficulty emphasized with this free energy implementation arises from approximating the continuous integral from a finite number of points. This becomes a particular issue near the endpoints, where the integrand can be very noisy or even discontinuous. This "endpoint catastrophe" often arises from an excessively large perturbation such as the appearance of a large chemical moiety. It is especially significant in explicit models of condensed phase systems but is also an issue for implicit models based on a continuum approximation (e.g., generalized Born models).

The elegance of both the FEP and TI formalisms is quite appealing, but as we have suggested, practical applications are not entirely straightforward due to scientific, technical, and computational challenges. Indeed, in the wake of the wave of excitement that followed the series of free energy publications in the mid-1980s, it was subsequently determined that for many systems of interest (such as ligand/protein) the amount of conformational sampling required to reliably calculate ΔG via FEP or TI was beyond what was available at that time: CPU speeds have increased by roughly 100 000× between the time when the first FEP calculations were published in 1984 and today. On top of that, effective improvements made possible by massive parallelism (GPU and/or CPU-based) provide an additional factor of 50× or more. As such, a free energy calculation today can access 5 million times more compute power than was available at the start of the alchemical free energy prediction era.

The initial reports of FEP calculations for ligand/protein systems were carried out for a total of roughly 30−40 ps on the supercomputers of the time.[90,92] Today, where total simulation times of >1 μs are not unusual for RBFE calculations (5 orders of magnitude longer), those total simulation times would not be considered sufficient for even pre-equilibration of a single λ-window. As a comparison focusing only on the past decade, consider the performance of the modernized MD engines ca. 2010, which, for a solvated protein system, barely obtained 1 ns/day on a single (relatively) modern CPU and contrast that to the latest AMBER MD engine which can approach ∼1000 ns/day on the most recent line of NVIDIA GPUs (atom counts are similar, ∼25k).

While the vast majority of the published applications of alchemical free energy calculations for protein−ligand binding have been to determine the RBFE between related molecules, there remains interest in approaches to calculate ABFEs, especially in the context of virtual screening. The main advantage of ABFE simulations in this context is the ability to calculate free energies that compare directly to binding constants, even across many dissimilar compounds. Because the results are on an absolute scale, we are able to determine whether the binding of a compound is "good" or "bad". This contrasts to approximate scoring functions, where we are mostly limited, in the best case, to determining whether binding or relatively "better" or "worse" than similar compounds. Ignoring computational cost/speed issues, for rigorous free energy calculations to be incorporated in vHTS, one would want a method allowing the calculation of absolute free energies, which can be readily applied to diverse structural matter, as discussed below.

https://dx.doi.org/10.1021/acs.jcim.0c00116
J. Chem. Inf. Model. 2020, 60, 4153−4169

Absolute Binding Free Energy Calculations. The first alchemical free energy calculation and associated thermodynamic formalism to predict ABFEs was published in the late-1980s, referred to as double-annihilation simulations. In this early work, Jorgensen and coworkers studied the binding of two methane particles to each other in water.[93] Soon thereafter, the ABFE formalism was applied to protein−ligand complexes, such as the binding of carbon dioxide to carbonic anhydrase by Merz[94] and of biotin to streptavidin by Miyamoto and Kollman.[95] Although in some cases a simplified description of the protein−ligand complex was used and the simulations were relatively short, reflecting the limited computational resources of the time, the experimental binding constants were convincingly reproduced.

In Figure 4, we can see that the calculation of absolute binding affinities is a specific case of relative free energy

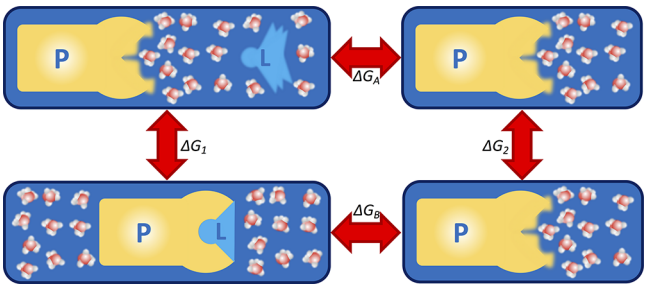


Figure 4. Closed thermodynamic cycle for ABFE calculations, where ligand L binds to a protein target P. As described in Figure 3, the experimental measurable quantity of interest is the binding energy of ligand L to protein P with associated binding free energy of ΔG_1. While this quantity can be computed, it is more computationally efficient to compute the annihilation of L in solution (ΔG_A) and in the binding site (ΔG_B). ΔG_2 is exactly zero by definition. Because this is a closed thermodynamic cycle, the absolute binding free energy of ligand L binding to protein P ($\Delta G_{bind} = \Delta G_1$) is identical to $\Delta G_A - \Delta G_B$.

calculations, where the second ligand is nothing. These calculations depend on the free energy cycle, similar to that described above in Figure 3, but with the mutated ligand (L′) replaced by "nothing" (the ligand no longer interacting with the system).

In this case, the experimental value ΔG_1 is the ABFE of ligand L binding to protein P. ΔG_2 is 0 by definition, and therefore

$$\Delta G_1 = \Delta G_A - \Delta G_B \quad (6)$$

ΔG_A is calculated by "disappearing" the ligand in solvent, while ΔG_B is calculated by "disappearing" the ligand in the receptor binding site.

This ABFE approach incorporates a considerably different perturbation from the RBFE method and carries its own set of difficulties and special considerations. For example, the "end point catastrophe" can be especially pronounced due to the larger size of the perturbation and may involve solvent molecules entering the binding cavity. This is especially difficult to model successfully if the binding pocket is buried and/or the kinetics of the unbinding process are slow (e.g., involves a large-scale opening between two protein domains). Water diffusion can be a slower process than protein side-chain reorganization (nanoseconds vs hundreds of picoseconds) and thus ABFE generally demands longer simulation times than RBFE to attain good accuracy. Fortunately, the level of accuracy required to separate binders from nonbinders for virtual screening is lower than for typical hit-to-lead and lead optimization efforts, where small chemical modifications to a molecule typically result in energetic changes of less than a kcal/mol, which offset some of the additional computational requirements of ABFE relative to RBFE. From a drug-design perspective, RBFE and ABFE represent two considerably different domains of applicability. RBFE is more appropriate for hit-to-lead and lead optimization, when near chemical accuracy (e.g., 1 kcal/mol) is critical to making correct predictions (does molecule A have a superior binding affinity than molecule B, where B represents a small modification to A), while for virtual screening ABFE accuracy need only be sufficient to separate binders from nonbinders (a much larger energy separation).

The idea of using ABFE in virtual screening has been proposed previously,[96,97] although there is a paucity of actual applications, likely due to a combination of the computational resources and technical expertise required to successfully execute an ABFE screen. There has only been a limited number of recent examples, where ABFE calculations have been performed to rank the binding of diverse molecules to a target, which lays the foundation for ABFE calculations in vHTS. For example, Aldeghi et al. performed ABFE calculations on a set of 11 diverse bromodomain-containing protein 4 (BRD4) inhibitors that yielded an absolute error of 0.6 kcal/mol for a set of validation molecules and 1.0 kcal/mol for blinded predictions.[98] More recent work by Araujo et al. presented a structure-based virtual screening protocol that culminates with ABFE calculations.[99] The authors found promising hits for ALK-5 inhibitors that were initially identified from docking, but further characterized with ABFE calculations to add confidence in the predictions, although these molecules have yet to be confirmed as ALK-5 inhibitors. In another study, virtual screening was performed on FGFR1 to find a number of novel hits, which were characterized with ABFE calculations.[100] While it does not appear that ABFE was actually used in the selection of the molecules, ABFE provided a richer understanding of the molecular recognition process of these molecules. In a more recent study, and perhaps the largest study of ABFE in vHTS, 77 α-hydroxytropolone derivatives were screened against HIV reverse transcriptase.[101] The work involved more than 300 individual ABFE calculations and over 1 million CPU hours, yet this work was performed using an implicit solvent model rather than explicit water molecules and therefore is a faster and less accurate representation than the other examples provided here. That being said, for virtual screening implicit solvent models may be sufficient for some cases.

While the above examples are relatively small in scale and number, they demonstrate the potential for ABFE calculations to be used in vHTS. With reasonable settings (1 μs cumulative sampling per ligand), it is possible to screen a single compound in two GPU-days for a 25 000-atom system. Therefore, 1000 ligands could be processed in a day on 2000 GPUs. GPU resources of this magnitude are readily available, either on the cloud or with on-premise resources. While the computational costs for ABFE calculations of this scale are not insignificant, it is still substantially cheaper than purchasing library compounds and performing experimental assays. At the time of writing this paper, the hardware cost for running an ABFE calculation with typical settings is less than $10/compound ($0.15/GPU-hour

https://dx.doi.org/10.1021/acs.jcim.0c00116
J. Chem. Inf. Model. 2020, 60, 4153−4169

for preemptible instances of Nvidia K80 GPUs on Google Cloud), although additional costs associated with software licensing are not accounted for here due to the significant variability (ranging from free open source software to relatively expensive commercial solutions). Significant savings in hardware costs can be realized with in-house resources (approximately 4× cheaper based on our internal GPU cluster costs for ∼500 GPUs, which includes square footage, power, cooling, bandwidth, and security from a high-performance computing (HPC) colocation facility, plus maintenance), and computational costs will continue to drop. While higher-value applications of ABFE likely exist when expensive chemical syntheses are involved (e.g., selectivity optimization, core hopping, or any nonlibrary chemical synthesis), significant value can be realized from deploying ABFE in vHTS workflows.

Another advantage of ABFE calculations is the ability to compare predicted binding energies across different targets, because the binding energies computed are on an absolute scale.[98] This allows for *in silico* selectivity screening, where one could optimize narrow selectivity (binding to a single target and not undesirable off-targets) or broad promiscuity (multiple targets for polypharmacology or multiple protein variants to evade drug resistant mutations). While there are limited examples where this approach has been applied in the context of virtual screening, we foresee this as a high-value application area for ABFE in future vHTS campaigns.

The significant computational demand coupled with the high technical expertise required for successfully running ABFE calculations has resulted in attempts to calculate free energies more quickly and easily, albeit at a lower level of accuracy. Such approaches include implicit solvent endpoint methods (MM/GBSA and MM/PBSA),[20] solvent/fragment mapping (estimating free energies from fragment occupancies),[102] linear interaction energy (LIE) methods,[103] and methods that attempt to predict free energies using precalculated free energy grids.[104] In our experience, the approximations employed by these methods render them too inaccurate or system-dependent for broad use in vHTS and therefore we will not discuss approximate free energy methods in this Perspective, although they were covered briefly here prior to the rigorous free energy methods and have been discussed elsewhere.[105,106]

Path-Based Simulations. Path-based binding free energy methods compute the complete binding free energy profile between the ligand bound and unbound states. As such, these methods can address a set of problems that cannot be tackled with standard alchemical methods, such as computing the binding transition state and thereby binding kinetics. As a consequence of computing the full binding path, these methods require significantly more computational resources to attain the same level of accuracy as alchemical binding free energy methods, where the objective is the binding free energy of a molecule to a target protein (which maps directly to the vHTS problem). In this work, we only briefly cover path-based methods, but when large scale conformational changes are required for binding, then path-based methods may be more appropriate. Furthermore, when the biological process of interest goes beyond binding (e.g., allosteric modulation), then path-based methods could be employed to assess the conformational changes associated with binding.

One of the most widely used path-based techniques is umbrella sampling, developed by Torrie and Valleau in 1975,[107] where a non-Boltzmann weighting function is applied to the potential energy. The bias is then removed at the end of the simulation to yield unbiased potential of mean force (i.e., an energy surface). Early Monte Carlo applications of conformational equilibria using umbrella sampling e.g. studying Lennard-Jones liquids and subsequently water were published by Berne[108] and Jorgensen.[109] Path-based methods were more widely used in the late 1990s to 2000s to describe ligand binding and unbinding pathways using random expulsion molecular dynamics (REMD),[110] using a memory-dependent biasing potential such as local elevation[111] and conformational flooding.[112] In the early 2000s, several new path-based methods were developed to aid in sampling complex free energy states and landscapes, such as accelerated molecular dynamics (aMD) by the McCammon group,[113,114] Metadynamics by the Parrinello group,[115] Gaussian boost potentials,[116,117] improvements to random expulsion molecular dynamics,[110] and adaptive biasing force (ABF) method.[118,119] The field of path-based methods is continuing to evolve with emerging methods as, for example, meta-extended ABF (meta-eABF)[120] and the use of the string method to compute binding free energies.[121] More recent work demonstrated a generally applicable metadynamics scheme for predicting the free energy profiles (and therefore binding energies), combining well-tempered multiple-walker metadynamics with a funnel-like boundary, which was used to predict binding free energies with a root-mean-square error (RMSE) of 12 diverse ligands in 5 receptors to less than 1 kcal/mol.[122] An extensive review highlights recent advances and discusses challenges in predicting protein−ligand binding using metadynamics-based approaches, with an emphasis on structure-based design.[123]

Unbiased Simulations. The above alchemical and path-based free energy approaches offer a clear route for improving virtual screening hit rates using existing computational resources. With significantly more computing power it will be possible to run unbiased, brute force molecular dynamics simulations to assess protein−ligand binding affinities. The setup and execution for such a calculation is relatively straightforward because there is no need for alchemical variables, transition paths, etc. The user would simply run a molecular dynamics simulation for a sufficient duration and assess the amount of time that the ligand is bound and unbound to compute a binding affinity, as per eq 7:

$$\Delta G_{\text{bind}} = -RT\ln\left(\frac{P_{\text{bound}}}{P_{\text{unbound}}}\right) \tag{7}$$

where P_{bound} and P_{unbound} are the probabilities that the ligand is bound and unbound to the target, respectively. This approach has been attempted with the most powerful MD supercomputer (Anton),[124] where the authors demonstrated that an unbiased simulation can be used to detect where a ligand binds on a protein as well as the binding path. For example, using this unbiased approach the authors were able to observe drug molecules spontaneously binding to a target (a GPCR in this case) and predict a pose that closely matched the experimental crystal pose.[125] Follow-up work with a similar approach predicted how the cancer drug dasatinib binds to Src kinase, forming a complex virtually identical to the experimental crystal structure.[126] A particularly interesting finding from these studies was the importance of water molecules during the association/dissociation process. These promising results suggest that increases in computational power will facilitate

https://dx.doi.org/10.1021/acs.jcim.0c00116
J. Chem. Inf. Model. 2020, 60, 4153−4169

the application of unbiased simulations in future virtual screening efforts.

The HPC Landscape and How It Will Affect Free Energy Calculations for Virtual Screening. While the above studies were able to accurately predict binding poses, to predict binding energies in an unbiased way would require simulations long enough to see multiple binding/unbinding events, which can range from milliseconds for weak binders to minutes or even hours for tight binders.[127,128] Given that Anton II (the most recent published version) can run on the order of 100 μs of simulations per day on a typical 50 000 atom protein/water system, it is clear that running an unbiased calculation long enough to get sufficient statistics on binding/unbinding for an accurate binding energy prediction would be intractable for even a single molecule, let alone on thousands of molecules, as would be needed to have a significant impact on virtual screening. Yet there will be a time when the above approach is possible: below is a short listing of some promising computational hardware technologies that will enable future free energy simulations.

First, it is possible that Moore's Law (i.e., that the number of transistors on a microchip doubles every two years, thus computing power doubles every two years) will continue indefinitely to a point where rigorous free energy simulations can be used routinely in virtual screening campaigns. With this growth, computers will be 1000× faster in 20 years, which would enable large-scale free energy simulations in virtual screening.

Alternatively, custom supercomputers designed specifically for MD, such as Anton, may become more accessible, thereby making screening using unbiased simulations tractable. Along these lines, commodity hardware such as Field-Programmable Gate Arrays (FPGA), which already have high bandwidth and low latency connections, could prove to be a viable path forward to build custom MD supercomputers. Another approach involves building a very large single silicon wafer. For example, Cerebras Systems[112] has unveiled the world's largest microprocessor, a chip roughly the size of a laptop computer, custom-built for machine learning. The silicon wafer contains 1.2 trillion transistors and incorporates 400 000 cores on a single chip, and a network fabric with over 100 Pb/sec of aggregate bandwidth. Integrated together, those components provide what is essentially a giant supercomputer cluster on a chip, with far faster interconnectivity than any chip-to-chip interconnects expected to be available in the foreseeable future.

Finally, it is pertinent to briefly discuss quantum computers. Google recently announced "quantum supremacy" for a specific algorithm, showing that its 53-qubit Sycamore processor was able to perform a calculation in 200 s that would have taken a state-of-the-art classical supercomputer 10 000 years.[129] Quantum computers seem to be well-suited for running quantum chemistry calculations, as has been discussed for almost a decade[130] and more recently demonstrated.[131−133] Nonetheless, the chemical systems that have been successfully run on quantum computers contain only a few atoms; scaling to thousands of atoms, as required for running protein−ligand binding simulations in explicit water, are likely to need thousands of qubits while the field is currently stuck at less than one hundred qubits. Still, we can expect that eventually quantum computers will surpass traditional silicon-based chips for chemistry-based applications, which may facilitate virtual screening with quantum mechanical energy models to conduct binding free energy calculations.

Limitations of Rigorous Binding Free Energy Methods. The most significant sources of error for rigorous binding free energy methods are poor quality force field and insufficient sampling (if no errors have been made in setting up the simulations, which should not be assumed without careful inspection). High quality force field parameters are required for all components of the system to ensure reliable and accurate results, including the protein, solvent, ligand molecules, and any cofactors. Force fields that can accurately reproduce quantum mechanical conformational energies and electrostatic potentials are now available for standard protein amino acids and water molecules, yet each ligand is unique and therefore requires its own parameters. The most common force fields currently used in drug discovery are fixed-charge, atom-centered models—these have proven to be useful, although they neglect important terms such as nonspherical charge distributions, polarization, and charge flux. Recent improvements to fixed-charge models have added off-centered charges to account for phenomena such as sigma holes on halogens and lone pairs.[134,135] Additionally, many force fields rely on simple lookup tables to get parameters for new molecules, based on existing parameters for related molecules, which oftentimes produces significant errors. For example, ortho-substitutions on a biphenyl system will alter the torsional profile of the bond connecting the two rings and change the electron distribution among the rings and therefore the quantum mechanical torsional profile, while the force field parameters assigned from a torsional lookup would remain unchanged, emphasizing why a simple lookup table is often insufficient. As such, it is recommended to generate high-quality force field parameters based on quantum mechanics for all molecules to be run in free energy simulations. Force field parametrization takes significantly less time than free energy simulations on GPU hardware (<1 h/molecule for high-quality force field parametrization versus 1−2 days/compound for free energy simulations on a modern GPU) and therefore should be a routine practice for binding free energy workflows.

Other force field work has focused on polarizable models, which have great promise and are starting to show value in practical applications.[136,137] Such models should be more accurate than traditional fixed-charge force fields, especially in cases of highly charged or polarizable environments, such as interactions with metals or soft polarizable atoms (like sulfur). To our knowledge, a direct comparison of polarizable force fields has yet to show improvements over state-of-the-art fixed charge models for binding free energy predictions, possibly because of the higher computational costs of polarizable models and the additional effort required to parametrize ligands, but we believe eventually polarizable force fields will prove superior to fixed charged models due to a more accurate representation of the underlying quantum physics.

Regarding sampling, one never knows when enough is enough because there could always be unexplored minima around the corner. Some systems, due to the nature of the free energy surface and associated energetic barriers for conformational transitions, may require more sampling to converge than is possible even with the high-powered GPUs of today. For example, it was recently reported that simulations using the same parameters with different random seeds can produce significantly different results,[138] although there appears to be improvement in this area.[139] We know that many protein motions of biological interest can take well over a microsecond to occur even once, and typical alchemical binding free energy

https://dx.doi.org/10.1021/acs.jcim.0c00116
J. Chem. Inf. Model. 2020, 60, 4153−4169

simulations are run for significantly less than this. Furthermore, for alchemical calculations, convergence at the endpoint λ intervals (where groups are appearing or disappearing) can be particularly problematic. In some cases it is possible to apply enhanced sampling of specific degrees of freedom to improve convergence, but such approaches typically require knowledge of the degrees of freedom to be sampled, which is challenging in the case of virtual screening with diverse ligands that differentially impact the free energy landscape. Applying more general enhanced sampling techniques may speed convergence but likely lower accuracy by introducing unrealistic biases.

Other limitations come from the setup of the simulations. For example, in general there is no sampling of the topology of the system during the free energy simulations, so incorrect states (e.g., tautomers and protonation states) will persist throughout the simulations. Even if the "correct" state is chosen, properly accounting for the energetics of that state can be challenging (e.g., the most favorable state when the ligand is bound to the protein may differ from the most favorable state in solution). When estimates are available for the energetics of the different states, it is possible to use methods like constant pH MD to sample the states during the free energy simulations.[140,141] Other system setup problems arise from not including all of the biologically relevant parts of the system. For example, if cyclin-dependent kinase (CDK2) is dependent on cyclin, then it is likely to be important in the simulations, especially if the experimental assay results are different with and without cyclin. Finally, even if everything is performed perfectly in the simulations, experimental assays reflect their own set of errors, and these can lead to false positives/negatives during the vHTS analysis (even if the computational results are correct). A thorough discussion of protein and ligand preparation in the context of virtual screening enrichments has been presented elsewhere.[142]

Finally, and perhaps most significantly from the full perspective of vHTS, high quality free energy calculations depend on suitable starting models for both the receptor and the ligand pose within the binding site (along with any metals, cofactors, and other biologically relevant molecules). The receptor structure typically comes from X-ray crystallography, but could come from Cryo-EM, full structure assignment with NMR, or even high-quality homology models. The bound ligand structure in the case of virtual screening must come from molecular docking or related approaches, which carry all of the limitations described in the docking sections above. Moreover, in the context of rigorous free energy calculations, the error for computed energies quasi-linearly increases with the increasing number of interactions present in a protein−ligand complex.[143] Indeed, in a recent comparison of seven free energy methodologies on a set of three host−guest systems parametrized by the same force field, the largest system showed larger uncertainty and slower convergence.[139] Methodologies to overcome this problem are emerging such as BLUES, which uses nonequilibrium candidate Monte Carlo (NCMC) method to improve sampling of ligand binding modes.[144] Other challenges in ABFE calculations include the treatment of trapped water molecules, performing bidirectional calculations to increase efficiency, and in general the correct setup for ABFE calculations.

It is certainly true that many challenges exist to obtain highly reliable ABFE results, yet ABFE has notable advantages relative to other approaches, as described in this Perspective, for post-docking scoring of a final set of hits. In fact, a significant advantage of rigorous free energy simulations is that the system can relax through molecular dynamics, making free energy simulations less sensitive to the exact coordinates of the starting structure, but structures more distant from the correct structure will require additional sampling and generally yield larger errors.

While accurate ABFE predictions are only possible with high-quality docking poses, some aspects of vHTS mitigate this issue. To begin, the ABFE vHTS paradigm implies a funnel in which docking precedes ABFE. The first function of docking is thus to eliminate a large number of inactive compounds while retaining some fraction of active compounds. The second function is to predict a binding pose or poses. Only one of the poses need to be near correct for ABFE to succeed. Multiple poses can be assessed as an ensemble of ABFE calculations or an additional pose assessment could be injected (e.g., short MD to assess ligand stability in the pocket).[145] Furthermore, MD refinement can improve pose prediction, as demonstrated in multiple D3R blind pose prediction challenges,[146,147] as well as other recent works.[58] As such, ABFE (which includes MD) should be less sensitive to the input pose than other rescoring approaches.

Even if the pose prediction is inaccurate for some ligands, the main effect would be to incorrectly reject active compounds due to a poor pose (a *false negative*), while the main goal of vHTS is to maximize the probability of true positives (i.e., find actual active compounds). As such, incorrect poses are less of a concern than using an inaccurate scoring function, which would create *false positives*. Put another way, of all of the methods discussed here, ABFE has the best chance of obtaining the right answer for the right reason and no worse of a chance of obtaining the wrong answer for the wrong reason. While ABFE cannot correct for all cases of poor docking, it can recover from some of them−unlike other scoring functions used with docking−and can therefore be expected to maximize the number of true positives obtained at the end of the screen. The degree of sensitivity of ABFE to the input pose will likely depend on many factors, including the nature of the protein binding site and conformational flexibility of the ligand. This is an important area of research that requires further investigation.

DISCUSSION

Finding molecules that bind to a target of interest is of the utmost importance when starting a drug discovery project. Experimental approaches to screen large chemical libraries are still very expensive, so accurate vHTS approaches are of considerable value. In our discussion, we have placed a forward-looking emphasis on the application of rigorous binding free energy methods in virtual screening because we feel these have the strongest statistical mechanical foundation and will yield the greatest success in finding novel chemical matter for yet-to-be-drugged targets. While the rigorous methods are significantly more computationally expensive than simpler methods such as ligand-based similarity searching or protein−ligand docking, the reward will be worth the costs if an advantage can be gained in finding hits for therapeutically relevant targets with unmet medical needs. Additionally, rigorous free energy methods are not biased by existing chemical matter and therefore should be able to retrieve more diverse hits, even in the cases where ligand-based methods can be applied.

https://dx.doi.org/10.1021/acs.jcim.0c00116
J. Chem. Inf. Model. 2020, 60, 4153−4169

While there are many rigorous binding free energy approaches that could be used for virtual screening, we propose that alchemical ABFE methods are the most appropriate for vHTS because they encapsulate all of the relevant physics for computing binding energies while avoiding the costly and challenging calculations associated with the binding/unbinding pathway. Assuming that an ABFE calculation on a single ligand can be completed in a day on a modern GPU for a typically sized protein system, it is possible to virtually screen thousands of compounds per day with GPUs available on the cloud (and some companies can do this with internal resources).

To be sure, there are many challenges involved in applying ABFE in vHTS workflows that extend beyond access to compute infrastructure. These include correctly preparing the system, optimizing simulation protocols, employing enhanced sampling methods, automating the generation of high-quality ligand force field parameters, and automating the process of running ABFE simulations on many ligands. However, these are issues that can be addressed with the appropriate effort and expertise. Interestingly, just before submitting the final version of this Perspective an excellent research paper by Heinzelmann and Gilson entitled, "Automated docking refinement and virtual compound screening with absolute binding free energy calculations", was published.[148] The authors acknowledge that absolute binding free energy calculations can be complex to implement and perform, yet they implemented a fully automated workflow using Python to run AMBER ABFE calculations on a series of ligands. They present encouraging results, both for rescoring docked poses and estimating the absolute binding free energies.

We believe that the value of rigorous vHTS methods like ABFE is greatest for genetically validated, yet-to-be-drugged targets where ligand-based methods (including machine learning) have limited value. As computational resources continue to increase in capacity and decline in price, rigorous binding free energy methods will become more attractive for virtual screening and eventually will be the methods of choice for the field. Until then, we anticipate an increase in the application of these methods by researchers in industry and academia who see the value and have the resources to perform such screens. We look forward to seeing how the field evolves in the coming years as more rigorous methods are adopted in virtual screening and our success in drugging challenging disease targets increases.

AUTHOR INFORMATION

Corresponding Author

Woody Sherman − *Silicon Therapeutics, Boston, Massachusetts 02210, United States;* orcid.org/0000-0001-9079-1376; Email: woody@silicontx.com

Authors

Zoe Cournia − *Biomedical Research Foundation, Academy of Athens, 11527 Athens, Greece;* orcid.org/0000-0001-9287-364X

Bryce K. Allen − *Silicon Therapeutics, Boston, Massachusetts 02210, United States;* orcid.org/0000-0002-0804-8127

Thijs Beuming − *Latham BioPharm Group, Cambridge, Massachusetts 02142, United States*

David A. Pearlman − *QSimulate Incorporated, Cambridge, Massachusetts 02139, United States;* orcid.org/0000-0002-2574-6066

Brian K. Radak − *Silicon Therapeutics, Boston, Massachusetts 02210, United States;* orcid.org/0000-0001-8628-5972

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jcim.0c00116

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We would like to thank T. Dwight McGee for helpful discussions related to the application of ABFE in virtual screening and for demonstrating at Silicon Therapeutics that it is possible to run ABFE in real-world drug discovery projects, where we have discovered experimentally validated hits for challenging targets with no previously published active molecules. We would also like to thank Vipin Sachdeva for determining the costs associated with performing ABFE calculations on the cloud and on in-house GPU resources and for building the HPC infrastructure that has allowed us to run such calculations at scale in drug discovery projects. Finally, we would like to acknowledge Darrin M. York and Tai-Sung Lee for a productive and ongoing collaboration to improve alchemical free energy simulations in the AMBER molecular dynamics package. Many of our improvements will be available in the upcoming AMBER20 release. Z.C. has been cofinanced by the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH − CREATE − INNOVATE (project code :T1EDK-03532).

REFERENCES

(1) Macarron, R.; Banks, M. N.; Bojanic, D.; Burns, D. J.; Cirovic, D. A.; Garyantes, T.; Green, D. V.; Hertzberg, R. P.; Janzen, W. P.; Paslay, J. W.; Schopfer, U.; Sittampalam, G. S. Impact of High-Throughput Screening in Biomedical Research. *Nat. Rev. Drug Discovery* **2011**, *10*, 188−195.

(2) Kessler, D.; Gmachl, M.; Mantoulidis, A.; Martin, L. J.; Zoephel, A.; Mayer, M.; Gollner, A.; Covini, D.; Fischer, S.; Gerstberger, T.; Gmaschitz, T.; Goodwin, C.; Greb, P.; Haring, D.; Hela, W.; Hoffmann, J.; Karolyi-Oezguer, J.; Knesl, P.; Kornigg, S.; Koegl, M.; Kousek, R.; Lamarre, L.; Moser, F.; Munico-Martinez, S.; Peinsipp, C.; Phan, J.; Rinnenthal, J.; Sai, J.; Salamon, C.; Scherbantin, Y.; Schipany, K.; Schnitzer, R.; Schrenk, A.; Sharps, B.; Siszler, G.; Sun, Q.; Waterson, A.; Wolkerstorfer, B.; Zeeb, M.; Pearson, M.; Fesik, S. W.; McConnell, D. B. Drugging an Undruggable Pocket on KRAS. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116*, 15823−15829.

(3) Allen-Petersen, B. L.; Sears, R. C. Mission Possible: Advances in Myc Therapeutic Targeting in Cancer. *BioDrugs* **2019**, *33*, 539−553.

(4) Sheridan, C. Drug Developers Switch Gears to Inhibit Sting. *Nat. Biotechnol.* **2019**, *37*, 199−201.

(5) Verdine, G. L.; Walensky, L. D. The Challenge of Drugging Undruggable Targets in Cancer: Lessons Learned from Targeting Bcl-2 Family Members. *Clin. Cancer Res.* **2007**, *13*, 7264−7270.

(6) Shelat, A. A.; Guy, R. K. Scaffold Composition and Biological Relevance of Screening Libraries. *Nat. Chem. Biol.* **2007**, *3*, 442−446.

(7) Gregori-Puigjane, E.; Mestres, J. Coverage and Bias in Chemical Library Design. *Curr. Opin. Chem. Biol.* **2008**, *12*, 359−365.

(8) Saha, A.; Varghese, T.; Liu, A.; Allen, S. J.; Mirzadegan, T.; Hack, M. D. An Analysis of Different Components of a High-Throughput Screening Library. *J. Chem. Inf. Model.* **2018**, *58*, 2057−2068.

(9) Follmann, M.; Briem, H.; Steinmeyer, A.; Hillisch, A.; Schmitt, M. H.; Haning, H.; Meier, H. An Approach Towards Enhancement of a Screening Library: The Next Generation Library Initiative (NGLI) at Bayer - against All Odds? *Drug Discovery Today* **2019**, *24*, 668−672.

(10) Dingjan, T.; Gillon, E.; Imberty, A.; Perez, S.; Titz, A.; Ramsland, P. A.; Yuriev, E. Virtual Screening against Carbohydrate-Binding Proteins: Evaluation and Application to Bacterial Burkholderia Ambifaria Lectin. *J. Chem. Inf. Model.* **2018**, *58*, 1976−1989.
(11) Perez-Castillo, Y.; Helguera, A. M.; Cordeiro, M.; Tejera, E.; Paz, Y. M. C.; Sanchez-Rodriguez, A.; Borges, F.; Cruz-Monteagudo, M. Fusing Docking Scoring Functions Improves the Virtual Screening Performance for Discovering Parkinson's Disease Dual Target Ligands. *Curr. Neuropharmacol* **2017**, *15*, 1107−1116.
(12) Damm-Ganamet, K. L.; Bembenek, S. D.; Venable, J. W.; Castro, G. G.; Mangelschots, L.; Peeters, D. C.; McAllister, H. M.; Edwards, J. P.; Disepio, D.; Mirzadegan, T. A Prospective Virtual Screening Study: Enriching Hit Rates and Designing Focus Libraries to Find Inhibitors of PI3Kδ and PI3Kγ. *J. Med. Chem.* **2016**, *59*, 4302−4313.
(13) FastROCS. OpenEye Scientific Software: 2019, http://www.eyesopen.com/fastrocs.
(14) Rogers, D.; Hahn, M. Extended-Connectivity Fingerprints. *J. Chem. Inf. Model.* **2010**, *50*, 742−754.
(15) Shoichet, B. K.; McGovern, S. L.; Wei, B.; Irwin, J. Lead Discovery Using Molecular Docking. *Curr. Opin. Chem. Biol.* **2002**, *6*, 439−446.
(16) Lyu, J.; Wang, S.; Balius, T. E.; Singh, I.; Levit, A.; Moroz, Y. S.; O'Meara, M. J.; Che, T.; Algaa, E.; Tolmachova, K.; Tolmachev, A. A.; Shoichet, B. K.; Roth, B. L.; Irwin, J. J. Ultra-Large Library Docking for Discovering New Chemotypes. *Nature* **2019**, *566*, 224−229.
(17) OpenEye Giga Docking - Structure Based Virtual Screening of over 1 Billion Molecules. https://www.eyesopen.com/news/giga-docking-structure-based-virtual-screening-webinar, accessed May 1, 2020.
(18) Svensson, F.; Karlen, A.; Skold, C. Virtual Screening Data Fusion Using Both Structure- and Ligand-Based Methods. *J. Chem. Inf. Model.* **2012**, *52*, 225−232.
(19) Sastry, G. M.; Inakollu, V. S.; Sherman, W. Boosting Virtual Screening Enrichments with Data Fusion: Coalescing Hits from Two-Dimensional Fingerprints, Shape, and Docking. *J. Chem. Inf. Model.* **2013**, *53*, 1531−1542.
(20) Genheden, S.; Ryde, U. The MM/PBSA and MM/GBSA Methods to Estimate Ligand-Binding Affinities. *Expert Opin. Drug Discovery* **2015**, *10*, 449−461.
(21) Lyne, P. D.; Lamb, M. L.; Saeh, J. C. Accurate Prediction of the Relative Potencies of Members of a Series of Kinase Inhibitors Using Molecular Docking and MM-GBSA Scoring. *J. Med. Chem.* **2006**, *49*, 4805−4808.
(22) Pearlman, D. A. Evaluating the Molecular Mechanics Poisson-Boltzmann Surface Area Free Energy Method Using a Congeneric Series of Ligands to p38 MAP Kinase. *J. Med. Chem.* **2005**, *48*, 7796−7807.
(23) Hagler, A. T. Force Field Development Phase II: Relaxation of Physics-Based Criteria Or Inclusion of More Rigorous Physics into the Representation of Molecular Energetics. *J. Comput.-Aided Mol. Des.* **2019**, *33*, 205−264.
(24) Walters, W. P.; Stahl, M. T.; Murcko, M. A. Virtual Screening − an Overview. *Drug Discov. Today* **1998**, *3*, 160−178.
(25) Muchmore, S. W.; Debe, D. A.; Metz, J. T.; Brown, S. P.; Martin, Y. C.; Hajduk, P. J. Application of Belief Theory to Similarity Data Fusion for Use in Analog Searching and Lead Hopping. *J. Chem. Inf. Model.* **2008**, *48*, 941−948.
(26) Yan, X.; Liao, C.; Liu, Z.; Hagler, A. T.; Gu, Q.; Xu, J. Chemical Structure Similarity Search for Ligand-Based Virtual Screening: Methods and Computational Resources. *Curr. Drug Targets* **2016**, *17*, 1580−1585.
(27) Muegge, I.; Mukherjee, P. An Overview of Molecular Fingerprint Similarity Search in Virtual Screening. *Expert Opin. Drug Discovery* **2016**, *11*, 137−148.
(28) Eckert, H.; Bajorath, J. Molecular Similarity Analysis in Virtual Screening: Foundations, Limitations and Novel Approaches. *Drug Discov. Today* **2007**, *12*, 225−233.
(29) Morgan, H. L. The Generation of a Unique Machine Description for Chemical Structures−A Technique Developed at Chemical Abstracts Service. *J. Chem. Doc.* **1965**, *5*, 107−113.
(30) Duan, J.; Dixon, S. L.; Lowrie, J. F.; Sherman, W. Analysis and Comparison of 2D Fingerprints: Insights into Database Screening Performance Using Eight Fingerprint Methods. *J. Mol. Graphics Modell.* **2010**, *29*, 157−170.
(31) Sastry, M.; Lowrie, J. F.; Dixon, S. L.; Sherman, W. Large-Scale Systematic Analysis of 2D Fingerprint Methods and Parameters to Improve Virtual Screening Enrichments. *J. Chem. Inf. Model.* **2010**, *50*, 771−784.
(32) Carhart, R. E.; Smith, D. H.; Venkataraghavan, R. Atom Pairs as Molecular Features in Structure-Activity Studies: Definition and Applications. *J. Chem. Inf. Model.* **1985**, *25*, 64−73.
(33) Nilakantan, R.; Bauman, N.; Dixon, J. S.; Venkataraghavan, R. Topological Torsion: A New Molecular Descriptor for SAR Applications. Comparison with Other Descriptors. *J. Chem. Inf. Model.* **1987**, *27*, 82−85.
(34) Durant, J. L.; Leland, B. A.; Henry, D. R.; Nourse, J. G. Reoptimization of MDL Keys for Use in Drug Discovery. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1273−1280.
(35) Moriwaki, H.; Tian, Y.-S.; Kawashita, N.; Takagi, T. Mordred: A Molecular Descriptor Calculator. *J. Cheminf.* **2018**, *10*, 4.
(36) Todeschini, R.; Consonni, V. *Handbook of Molecular Descriptors*; WILEY-VCH Verlag GmbH: 2000.
(37) Shen, C.; Ding, J.; Wang, Z.; Cao, D.; Ding, X.; Hou, T. From Machine Learning to Deep Learning: Advances in Scoring Functions for Protein−Ligand Docking. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2020**, *10*, 10.
(38) Hassan, M.; Mogollon, D. C.; Fuentes, O.; Sirimulla, S. DLSCORE: A Deep Learning Model for Predicting Protein-Ligand Binding Affinities. *ChemRxiv* **2018**, DOI: 10.26434/chemrxiv.6159143.v1.
(39) Jimenez, J.; Skalic, M.; Martinez-Rosell, G.; De Fabritiis, G. KDEEP: Protein-Ligand Absolute Binding Affinity Prediction Via 3D-Convolutional Neural Networks. *J. Chem. Inf. Model.* **2018**, *58*, 287−296.
(40) Stepniewska-Dziubinska, M. M.; Zielenkiewicz, P.; Siedlecki, P. Development and Evaluation of a Deep Learning Model for Protein-Ligand Binding Affinity Prediction. *Bioinformatics* **2018**, *34*, 3666−3674.
(41) Bitencourt-Ferreira, G.; de Azevedo, W. F. Development of a Machine-Learning Model to Predict Gibbs Free Energy of Binding for Protein-Ligand Complexes. *Biophys. Chem.* **2018**, *240*, 63−69.
(42) Grant, J. A.; Gallardo, M. A.; Pickup, B. T. A Fast Method of Molecular Shape Comparison: A Simple Application of a Gaussian Description of Molecular Shape. *J. Comput. Chem.* **1996**, *17*, 1653−1666.
(43) Hawkins, P. C. D.; Skillman, A. G.; Nicholls, A. Comparison of Shape-Matching and Docking as Virtual Screening Tools. *J. Med. Chem.* **2007**, *50*, 74−82.
(44) Sastry, G. M.; Dixon, S. L.; Sherman, W. Rapid Shape-Based Ligand Alignment and Virtual Screening Method Based on Atom/Feature-Pair Similarities and Volume Overlap Scoring. *J. Chem. Inf. Model.* **2011**, *51*, 2455−2466.
(45) Poli, G.; Seidel, T.; Langer, T. Conformational Sampling of Small Molecules with iCon: Performance Assessment in Comparison with Omega. *Front. Chem.* **2018**, *6*, 229−229.
(46) Yang, S.-Y. Pharmacophore Modeling and Applications in Drug Discovery: Challenges and Recent Advances. *Drug Discov. Today* **2010**, *15*, 444−450.
(47) Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. A Geometric Approach to Macromolecule-Ligand Interactions. *J. Mol. Biol.* **1982**, *161*, 269−288.
(48) Ferreira, L. G.; Dos Santos, R. N.; Oliva, G.; Andricopulo, A. D. Molecular Docking and Structure-Based Drug Design Strategies. *Molecules* **2015**, *20*, 13384−13421.

https://dx.doi.org/10.1021/acs.jcim.0c00116
J. Chem. Inf. Model. 2020, 60, 4153−4169

(49) Torres, P. H. M.; Sodero, A. C. R.; Jofily, P.; Silva-Jr, F. P. Key Topics in Molecular Docking for Drug Design. *Int. J. Mol. Sci.* **2019**, *20*, 20.
(50) Athanasiou, C.; Vasilakaki, S.; Dellis, D.; Cournia, Z. Using Physics-Based Pose Predictions and Free Energy Perturbation Calculations to Predict Binding Poses and Relative Binding Affinities for FXR Ligands in the D3R Grand Challenge 2. *J. Comput.-Aided Mol. Des.* **2018**, *32*, 21−44.
(51) Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *J. Med. Chem.* **2004**, *47*, 1750−1759.
(52) McGann, M. FRED and HYBRID Docking Performance on Standardized Datasets. *J. Comput.-Aided Mol. Des.* **2012**, *26*, 897−906.
(53) Morris, G. M.; Goodsell, D. S.; Huey, R.; Olson, A. J. Distributed Automated Docking of Flexible Ligands to Proteins: Parallel Applications of AutoDock 2.4. *J. Comput.-Aided Mol. Des.* **1996**, *10*, 293−304.
(54) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and Validation of a Genetic Algorithm for Flexible Docking. *J. Mol. Biol.* **1997**, *267*, 727−748.
(55) Craig, I. R.; Essex, J. W.; Spiegel, K. Ensemble Docking into Multiple Crystallographically Derived Protein Structures: An Evaluation Based on the Statistical Analysis of Enrichments. *J. Chem. Inf. Model.* **2010**, *50*, 511−524.
(56) Sherman, W.; Day, T.; Jacobson, M. P.; Friesner, R. A.; Farid, R. Novel Procedure for Modeling Ligand/Receptor Induced Fit Effects. *J. Med. Chem.* **2006**, *49*, 534−553.
(57) DeLuca, S.; Khar, K.; Meiler, J. Fully Flexible Docking of Medium Sized Ligand Libraries with Rosettaligand. *PLoS One* **2015**, *10*, No. e0132508.
(58) Miller, E.; Murphy, R.; Sindhikara, D.; Borrelli, K.; Grisewood, M.; Ranalli, R.; Dixon, S.; Jerome, S.; Boyles, N.; Day, T.; Ghanakota, P.; Mondal, S.; Rafi, S. B.; Troast, D. M.; Abel, R.; Friesner, R. A Reliable and Accurate Solution to the Induced Fit Docking Problem for Protein-Ligand Binding. *ChemRxiv*. **2020**, DOI: 10.26434/chemrxiv.11983845.v1.
(59) Lenselink, E. B.; Beuming, T.; van Veen, C.; Massink, A.; Sherman, W.; van Vlijmen, H. W.; AP, I. J. In Search of Novel Ligands Using a Structure-Based Approach: A Case Study on the Adenosine A2A Receptor. *J. Comput.-Aided Mol. Des.* **2016**, *30*, 863−874.
(60) Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P. Empirical Scoring Functions: I. The Development of a Fast Empirical Scoring Function to Estimate the Binding Affinity of Ligands in Receptor Complexes. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 425−445.
(61) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* **2004**, *47*, 1739−1749.
(62) Murphy, R. B.; Repasky, M. P.; Greenwood, J. R.; Tubert-Brohman, I.; Jerome, S.; Annabhimoju, R.; Boyles, N. A.; Schmitz, C. D.; Abel, R.; Farid, R.; Friesner, R. A. WScore: A Flexible and Accurate Treatment of Explicit Water Molecules in Ligand-Receptor Docking. *J. Med. Chem.* **2016**, *59*, 4364−4384.
(63) Abel, R.; Young, T.; Farid, R.; Berne, B. J.; Friesner, R. A. Role of the Active-Site Solvent in the Thermodynamics of Factor Xa Ligand Binding. *J. Am. Chem. Soc.* **2008**, *130*, 2817−2831.
(64) Young, T.; Abel, R.; Kim, B.; Berne, B. J.; Friesner, R. A. Motifs for Molecular Recognition Exploiting Hydrophobic Enclosure in Protein-Ligand Binding. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 808−813.
(65) Reulecke, I.; Lange, G.; Albrecht, J.; Klein, R.; Rarey, M. Towards an Integrated Description of Hydrogen Bonding and Dehydration: Decreasing False Positives in Virtual Screening with the HYDE Scoring Function. *ChemMedChem* **2008**, *3*, 885−897.
(66) Schneider, N.; Lange, G.; Hindle, S.; Klein, R.; Rarey, M. A Consistent Description of Hydrogen Bond and Dehydration Energies in Protein−Ligand Complexes: Methods Behind the HYDE Scoring Function. *J. Comput.-Aided Mol. Des.* **2013**, *1*, 15−29.
(67) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E. Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. *Acc. Chem. Res.* **2000**, *33*, 889−897.
(68) Greenidge, P. A.; Kramer, C.; Mozziconacci, J. C.; Sherman, W. Improving Docking Results Via Reranking of Ensembles of Ligand Poses in Multiple X-ray Protein Conformations with MM-GBSA. *J. Chem. Inf. Model.* **2014**, *54*, 2697−2717.
(69) Greenidge, P. A.; Kramer, C.; Mozziconacci, J. C.; Wolf, R. M. MM/GBSA Binding Energy Prediction on the PDBbind Data Set: Successes, Failures, and Directions for Further Improvement. *J. Chem. Inf. Model.* **2013**, *53*, 201−209.
(70) Virtanen, S. I.; Niinivehmas, S. P.; Pentikainen, O. T. Case-Specific Performance of MM-PBSA, MM-GBSA, and SIE in Virtual Screening. *J. Mol. Graphics Modell.* **2015**, *62*, 303−318.
(71) Ballester, P. J.; Mitchell, J. B. O. A Machine Learning Approach to Predicting Protein−Ligand Binding Affinity with Applications to Molecular Docking. *Bioinformatics* **2010**, *26*, 1169−1175.
(72) Afifi, K.; Al-Sadek, A. F. Improving Classical Scoring Functions Using Random Forest: The Non-Additivity of Free Energy Terms' Contributions in Binding. *Chem. Biol. Drug Des.* **2018**, *92*, 1429−1434.
(73) Li, L.; Khanna, M.; Jo, I.; Wang, F.; Ashpole, N. M.; Hudmon, A.; Meroueh, S. O. Target-Specific Support Vector Machine Scoring In Structure-Based Virtual Screening: Computational Validation, in Vitro Testing in Kinases, and Effects on Lung Cancer Cell Proliferation. *J. Chem. Inf. Model.* **2011**, *51*, 755−759.
(74) Nogueira, M. S.; Koch, O. The Development of Target-Specific Machine Learning Models as Scoring Functions for Docking-Based Target Prediction. *J. Chem. Inf. Model.* **2019**, *59*, 1238−1252.
(75) Pereira, J. C.; Caffarena, E. R.; Dos Santos, C. N. Boosting Docking-Based Virtual Screening with Deep Learning. *J. Chem. Inf. Model.* **2016**, *56*, 2495−2506.
(76) Gao, J. Absolute Free Energy of Solvation from Monte Carlo Simulations Using Combined Quantum and Molecular Mechanical Potentials. *J. Phys. Chem.* **1992**, *96*, 537−540.
(77) Jorgensen, W. L.; Tirado-Rives, J. Molecular Modeling of Organic and Biomolecular Systems Using BOSS and MCPRO. *J. Comput. Chem.* **2005**, *26*, 1689−1700.
(78) Woods, C. J.; Malaisree, M.; Hannongbua, S.; Mulholland, A. J. A Water-Swap Reaction Coordinate for the Calculation of Absolute Protein-Ligand Binding Free Energies. *J. Chem. Phys.* **2011**, *134*, 054114.
(79) Takahashi, R.; Gil, V. A.; Guallar, V. Monte Carlo Free Ligand Diffusion with Markov State Model Analysis and Absolute Binding Free Energy Calculations. *J. Chem. Theory Comput.* **2014**, *10*, 282−288.
(80) Christ, C. D.; Mark, A. E.; van Gunsteren, W. F. Basic Ingredients of Free Energy Calculations: A Review. *J. Comput. Chem.* **2009**, *31*, 1569−1582.
(81) Chodera, J. D.; Mobley, D. L.; Shirts, M. R.; Dixon, R. W.; Branson, K.; Pande, V. S. Alchemical Free Energy Methods for Drug Discovery: Progress and Challenges. *Curr. Opin. Struct. Biol.* **2011**, *21*, 150−160.
(82) Mobley, D. L.; Klimovich, P. V. Perspective: Alchemical Free Energy Calculations for Drug Discovery. *J. Chem. Phys.* **2012**, *137*, 230901.
(83) Hansen, N.; van Gunsteren, W. F. Practical Aspects of Free-Energy Calculations: A Review. *J. Chem. Theory Comput.* **2014**, *10*, 2632−2647.
(84) Mobley, D. L.; Gilson, M. K. Predicting Binding Free Energies: Frontiers and Benchmarks. *Annu. Rev. Biophys.* **2017**, *46*, 531−558.

https://dx.doi.org/10.1021/acs.jcim.0c00116
J. Chem. Inf. Model. 2020, 60, 4153−4169

(85) Cournia, Z.; Allen, B.; Sherman, W. Relative Binding Free Energy Calculations in Drug Discovery: Recent Advances and Practical Considerations. *J. Chem. Inf. Model.* **2017**, *57*, 2911−2937.
(86) Kirkwood, J. G. Statistical Mechanics of Fluid Mixtures. *J. Chem. Phys.* **1935**, *3*, 300−313.
(87) Zwanzig, R. W. High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *J. Chem. Phys.* **1954**, *22*, 1420−1426.
(88) McQuarrie, D. A. *Statistical Mechanics*; Harper & Row: New York, 1976.
(89) Pal, R. K.; Gadhiya, S.; Ramsey, S.; Cordone, P.; Wickstrom, L.; Harding, W. W.; Kurtzman, T.; Gallicchio, E. Inclusion of Enclosed Hydration Effects in the Binding Free Energy Estimation of Dopamine D3 Receptor Complexes. *PLoS One* **2019**, *14*, No. e0222902.
(90) Wong, C. F.; McCammon, J. A. Dynamics and Design of Enzymes and Inhibitors. *J. Am. Chem. Soc.* **1986**, *108*, 3830−3832.
(91) Jorgensen, W. L.; Ravimohan, C. Monte Carlo Simulation of Differences in Free Energies of Hydration. *J. Chem. Phys.* **1985**, *83*, 3050−3054.
(92) Bash, P. A.; Singh, U. C.; Brown, F. K.; Langridge, R.; Kollman, P. A. Calculation of the Relative Change in Binding Free Energy of a Protein-Inhibitor Complex. *Science* **1987**, *235*, 574.
(93) Jorgensen, W. L.; Buckner, J. K.; Boudon, S.; Tirado-Rives, J. Efficient Computation of Absolute Free Energies of Binding by Computer Simulations. Application to the Methane Dimer in Water. *J. Chem. Phys.* **1988**, *89*, 3742−3746.
(94) Merz, K. M. Carbon Dioxide Binding to Human Carbonic Anhydrase II. *J. Am. Chem. Soc.* **1991**, *113*, 406−411.
(95) Miyamoto, S.; Kollman, P. A. What Determines the Strength of Noncovalent Association of Ligands to Proteins in Aqueous Solution? *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90*, 8402−8406.
(96) Waszkowycz, B. Towards Improving Compound Selection in Structure-Based Virtual Screening. *Drug Discov. Today* **2008**, *13*, 219−226.
(97) Amaro, R. E.; Li, W. W. Emerging Methods for Ensemble-Based Virtual Screening. *Curr. Top. Med. Chem.* **2010**, *10*, 3−13.
(98) Aldeghi, M.; Heifetz, A.; Bodkin, M. J.; Knapp, S.; Biggin, P. C. Accurate Calculation of the Absolute Free Energy of Binding for Drug Molecules. *Chem. Sci.* **2016**, *7*, 207−218.
(99) Araujo, S. C.; Maltarollo, V. G.; Almeida, M. O.; Ferreira, L. L. G.; Andricopulo, A. D.; Honorio, K. M. Structure-Based Virtual Screening, Molecular Dynamics and Binding Free Energy Calculations of Hit Candidates as ALK-5 Inhibitors. *Molecules* **2020**, *25*, 25.
(100) Ravindranathan, K. P.; Mandiyan, V.; Ekkati, A. R.; Bae, J. H.; Schlessinger, J.; Jorgensen, W. L. Discovery of Novel Fibroblast Growth Factor Receptor 1 Kinase Inhibitors by Structure-Based Virtual Screening. *J. Med. Chem.* **2010**, *53*, 1662−1672.
(101) Zhang, B.; D'Erasmo, M. P.; Murelli, R. P.; Gallicchio, E. Free Energy-Based Virtual Screening and Optimization of RNASE H Inhibitors of HIV-1 Reverse Transcriptase. *ACS Omega* **2016**, *1*, 435−447.
(102) Raman, E. P.; Lakkaraju, S. K.; Denny, R. A.; MacKerell, A. D., Jr. Estimation of Relative Free Energies of Binding Using Pre-Computed Ensembles Based on the Single-Step Free Energy Perturbation and the Site-Identification by Ligand Competitive Saturation Approaches. *J. Comput. Chem.* **2017**, *38*, 1238−1251.
(103) Aqvist, J.; Marelius, J. The Linear Interaction Energy Method for Predicting Ligand Binding Free Energies. *Comb. Chem. High Throughput Screening* **2001**, *4*, 613−626.
(104) Pearlman, D. A.; Charifson, P. S. Improved Scoring of Ligand-Protein Interactions Using OWFEG Free Energy Grids. *J. Med. Chem.* **2001**, *44*, 502−511.
(105) Boresch, S.; Tettinger, F.; Leitgeb, M.; Karplus, M. Absolute Binding Free Energies: A Quantitative Approach for Their Calculation. *J. Phys. Chem. B* **2003**, *107*, 9535−9551.
(106) Mobley, D. L.; Graves, A. P.; Chodera, J. D.; McReynolds, A. C.; Shoichet, B. K.; Dill, K. A. Predicting Absolute Ligand Binding Free Energies to a Simple Model Site. *J. Mol. Biol.* **2007**, *371*, 1118−1134.
(107) Torrie, G. M.; Valleau, J. P. Nonphysical Sampling Distributions in Monte Carlo Free-Energy Estimation: Umbrella Sampling. *J. Comput. Phys.* **1977**, *23*, 187−199.
(108) Rao, M.; Berne, B. J. On the Force Bias Monte Carlo Simulation of Simple Liquids. *J. Chem. Phys.* **1979**, *71*, 129−132.
(109) Bigot, B.; Jorgensen, W. L. Sampling Methods for Monte Carlo Simulations of N-Butane in Dilute Solution. *J. Chem. Phys.* **1981**, *75*, 1944−1952.
(110) Ludemann, S. K.; Lounnas, V.; Wade, R. C. How Do Substrates Enter and Products Exit the Buried Active Site of Cytochrome P450cam? 1. Random Expulsion Molecular Dynamics Investigation of Ligand Access Channels and Mechanisms. *J. Mol. Biol.* **2000**, *303*, 797−811.
(111) Huber, T.; Torda, A. E.; van Gunsteren, W. F. Local Elevation: A Method for Improving the Searching Properties of Molecular Dynamics Simulation. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 695−708.
(112) Grubmuller, H. Predicting Slow Structural Transitions in Macromolecular Systems: Conformational Flooding. *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.* **1995**, *52*, 2893−2906.
(113) Hamelberg, D.; Mongan, J.; McCammon, J. A. Accelerated Molecular Dynamics: A Promising and Efficient Simulation Method for Biomolecules. *J. Chem. Phys.* **2004**, *120*, 11919−11929.
(114) Voter, A. F. Hyperdynamics: Accelerated Molecular Dynamics of Infrequent Events. *Phys. Rev. Lett.* **1997**, *78*, 3908−3911.
(115) Laio, A.; Parrinello, M. Escaping Free-Energy Minima. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12562−12566.
(116) Miao, Y.; Feher, V. A.; McCammon, J. A. Gaussian Accelerated Molecular Dynamics: Unconstrained Enhanced Sampling and Free Energy Calculation. *J. Chem. Theory Comput.* **2015**, *11*, 3584−3595.
(117) Huang, Y. M.; McCammon, J. A.; Miao, Y. Replica Exchange Gaussian Accelerated Molecular Dynamics: Improved Enhanced Sampling and Free Energy Calculation. *J. Chem. Theory Comput.* **2018**, *14*, 1853−1864.
(118) Henin, J.; Chipot, C. Overcoming Free Energy Barriers Using Unconstrained Molecular Dynamics Simulations. *J. Chem. Phys.* **2004**, *121*, 2904−2914.
(119) Darve, E.; Rodriguez-Gomez, D.; Pohorille, A. Adaptive Biasing Force Method for Scalar and Vector Free Energy Calculations. *J. Chem. Phys.* **2008**, *128*, 144120.
(120) Fu, H.; Zhang, H.; Chen, H.; Shao, X.; Chipot, C.; Cai, W. Zooming across the Free-Energy Landscape: Shaving Barriers, and Flooding Valleys. *J. Phys. Chem. Lett.* **2018**, *9*, 4738−4745.
(121) Suh, D.; Jo, S.; Jiang, W.; Chipot, C.; Roux, B. String Method for Protein-Protein Binding Free-Energy Calculations. *J. Chem. Theory Comput.* **2019**, *15*, 5829−5844.
(122) Saleh, N.; Ibrahim, P.; Saladino, G.; Gervasio, F. L.; Clark, T. An Efficient Metadynamics-Based Protocol to Model the Binding Affinity and the Transition State Ensemble of G-Protein-Coupled Receptor Ligands. *J. Chem. Inf. Model.* **2017**, *57*, 1210−1217.
(123) Cavalli, A.; Spitaleri, A.; Saladino, G.; Gervasio, F. L. Investigating Drug-Target Association and Dissociation Mechanisms Using Metadynamics-Based Algorithms. *Acc. Chem. Res.* **2015**, *48*, 277−285.
(124) Shaw, D. E.; Deneroff, M. M.; Dror, R. O.; Kuskin, J. S.; Larson, R. H.; Salmon, J. K.; Young, C.; Batson, B.; Bowers, K. J.; Chao, J. C.; Eastwood, M. P.; Gagliardo, J.; Grossman, J. P.; Ho, C. R.; Ierardi, D. J.; Kolossváry, I.; Klepeis, J. L.; Layman, T.; McLeavey, C.; Moraes, M. A.; Mueller, R.; Priest, E. C.; Shan, Y.; Spengler, J.; Theobald, M.; Towles, B.; Wang, S. C. Anton, a Special-Purpose Machine for Molecular Dynamics Simulation. *Commun. ACM* **2008**, *51*, 91−97.
(125) Dror, R. O.; Pan, A. C.; Arlow, D. H.; Borhani, D. W.; Maragakis, P.; Shan, Y.; Xu, H.; Shaw, D. E. Pathway and Mechanism

https://dx.doi.org/10.1021/acs.jcim.0c00116
J. Chem. Inf. Model. 2020, 60, 4153−4169

of Drug Binding to G-Protein-Coupled Receptors. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 13118−13123.
(126) Shan, Y.; Kim, E. T.; Eastwood, M. P.; Dror, R. O.; Seeliger, M. A.; Shaw, D. E. How Does a Drug Molecule Find Its Target Binding Site? *J. Am. Chem. Soc.* **2011**, *133*, 9181−9183.
(127) Tummino, P. J.; Copeland, R. A. Residence Time of Receptor-Ligand Complexes and Its Effect on Biological Function. *Biochemistry* **2008**, *47*, 5481−5492.
(128) Copeland, R. A. The Drug-Target Residence Time Model: A 10-Year Retrospective. *Nat. Rev. Drug Discovery* **2016**, *15*, 87−95.
(129) Arute, F.; Arya, K.; Babbush, R.; Bacon, D.; Bardin, J. C.; Barends, R.; Biswas, R.; Boixo, S.; Brandao, F. G. S. L.; Buell, D. A.; Burkett, B.; Chen, Y.; Chen, Z.; Chiaro, B.; Collins, R.; Courtney, W.; Dunsworth, A.; Farhi, E.; Foxen, B.; Fowler, A.; Gidney, C.; Giustina, M.; Graff, R.; Guerin, K.; Habegger, S.; Harrigan, M. P.; Hartmann, M. J.; Ho, A.; Hoffmann, M.; Huang, T.; Humble, T. S.; Isakov, S. V.; Jeffrey, E.; Jiang, Z.; Kafri, D.; Kechedzhi, K.; Kelly, J.; Klimov, P. V.; Knysh, S.; Korotkov, A.; Kostritsa, F.; Landhuis, D.; Lindmark, M.; Lucero, E.; Lyakh, D.; Mandrà, S.; McClean, J. R.; McEwen, M.; Megrant, A.; Mi, X.; Michielsen, K.; Mohseni, M.; Mutus, J.; Naaman, O.; Neeley, M.; Neill, C.; Niu, M. Y.; Ostby, E.; Petukhov, A.; Platt, J. C.; Quintana, C.; Rieffel, E. G.; Roushan, P.; Rubin, N. C.; Sank, D.; Satzinger, K. J.; Smelyanskiy, V.; Sung, K. J.; Trevithick, M. D.; Vainsencher, A.; Villalonga, B.; White, T.; Yao, Z. J.; Yeh, P.; Zalcman, A.; Neven, H.; Martinis, J. M. Quantum Supremacy Using a Programmable Superconducting Processor. *Nature* **2019**, *574*, 505−510.
(130) Kassal, I.; Whitfield, J. D.; Perdomo-Ortiz, A.; Yung, M.-H.; Aspuru-Guzik, A. Simulating Chemistry Using Quantum Computers. *Annu. Rev. Phys. Chem.* **2011**, *62*, 185−207.
(131) Ryabinkin, I. G.; Yen, T. C.; Genin, S. N.; Izmaylov, A. F. Qubit Coupled Cluster Method: A Systematic Approach to Quantum Chemistry on a Quantum Computer. *J. Chem. Theory Comput.* **2018**, *14*, 6317−6326.
(132) Arguello-Luengo, J.; Gonzalez-Tudela, A.; Shi, T.; Zoller, P.; Cirac, J. I. Analogue Quantum Chemistry Simulation. *Nature* **2019**, *574*, 215−218.
(133) Genin, S. N., Quantum Chemistry on Quantum Annealers. *arXiv* **2019** 1901.04715.
(134) Yan, X. C.; Schyman, P.; Jorgensen, W. L. Cooperative Effects and Optimal Halogen Bonding Motifs for Self-Assembling Systems. *J. Phys. Chem. A* **2014**, *118*, 2820−2826.
(135) Yan, X. C.; Robertson, M. J.; Tirado-Rives, J.; Jorgensen, W. L. Improved Description of Sulfur Charge Anisotropy in OPLS Force Fields: Model Development and Parameterization. *J. Phys. Chem. B* **2017**, *121*, 6626−6636.
(136) Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A. J.; Head-Gordon, M.; Clark, G. N. I.; Johnson, M. E.; Head-Gordon, T. E. Current Status of the AMOEBA Polarizable Force Field. *J. Phys. Chem. B* **2010**, *114*, 2549−2564.
(137) Halgren, T. A.; Damm, W. Polarizable Force Fields. *Curr. Opin. Struct. Biol.* **2001**, *11*, 236−242.
(138) Loeffler, H. H.; Bosisio, S.; Duarte Ramos Matos, G.; Suh, D.; Roux, B.; Mobley, D. L.; Michel, J. Reproducibility of Free Energy Calculations across Different Molecular Simulation Software Packages. *J. Chem. Theory Comput.* **2018**, *14*, 5567−5582.
(139) Rizzi, A.; Jensen, T.; Slochower, D. R.; Aldeghi, M.; Gapsys, V.; Ntekoumes, D.; Bosisio, S.; Papadourakis, M.; Henriksen, N. M.; de Groot, B. L.; Cournia, Z.; Dickson, A.; Michel, J.; Gilson, M. K.; Shirts, M. R.; Mobley, D. L.; Chodera, J. D. The Sampl6 Sampling Challenge: Assessing the Reliability and Efficiency of Binding Free Energy Calculations. *J. Comput.-Aided Mol. Des.* **2020**, *34*, 601.
(140) Mongan, J.; Case, D. A.; McCammon, J. A. Constant pH Molecular Dynamics in Generalized Born Implicit Solvent. *J. Comput. Chem.* **2004**, *25*, 2038−2048.
(141) Kim, M. O.; Blachly, P. G.; Kaus, J. W.; McCammon, J. A. Protocols Utilizing Constant pH Molecular Dynamics to Compute pH-Dependent Binding Free Energies. *J. Phys. Chem. B* **2015**, *119*, 861−872.
(142) Sastry, M. G.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W. Protein and Ligand Preparation: Parameters, Protocols, and Influence on Virtual Screening Enrichments. *J. Comput.-Aided Mol. Des.* **2013**, *27*, 221−234.
(143) Merz, K. M. Limits of Free Energy Computation for Protein-Ligand Interactions. *J. Chem. Theory Comput.* **2010**, *6*, 1769−1776.
(144) Gill, S. C.; Lim, N. M.; Grinaway, P. B.; Rustenburg, A. S.; Fass, J.; Ross, G. A.; Chodera, J. D.; Mobley, D. L. Binding Modes of Ligands Using Enhanced Sampling (BLUES): Rapid Decorrelation of Ligand Binding Modes Via Nonequilibrium Candidate Monte Carlo. *J. Phys. Chem. B* **2018**, *122*, 5579−5598.
(145) Clark, A. J.; Tiwary, P.; Borrelli, K.; Feng, S.; Miller, E. B.; Abel, R.; Friesner, R. A.; Berne, B. J. Prediction of Protein-Ligand Binding Poses Via a Combination of Induced Fit Docking and Metadynamics Simulations. *J. Chem. Theory Comput.* **2016**, *12*, 2990−2998.
(146) Fradera, X.; Verras, A.; Hu, Y.; Wang, D.; Wang, H.; Fells, J. I.; Armacost, K. A.; Crespo, A.; Sherborne, B.; Wang, H.; Peng, Z.; Gao, Y. D. Performance of Multiple Docking and Refinement Methods in the Pose Prediction D3R Prospective Grand Challenge 2016. *J. Comput.-Aided Mol. Des.* **2018**, *32*, 113−127.
(147) Ignatov, M.; Liu, C.; Alekseenko, A.; Sun, Z.; Padhorny, D.; Kotelnikov, S.; Kazennov, A.; Grebenkin, I.; Kholodov, Y.; Kolosvari, I.; Perez, A.; Dill, K.; Kozakov, D. Monte Carlo on the Manifold and MD Refinement for Binding Pose Prediction of Protein-Ligand Complexes: 2017 D3R Grand Challenge. *J. Comput.-Aided Mol. Des.* **2019**, *33*, 119−127.
(148) Heinzelmann, G.; Gilson, M. K., Automated Docking Refinement and Virtual Compound Screening with Absolute Binding Free Energy Calculations. *bioRxiv* **2020** 2020.04.15.043240.

https://dx.doi.org/10.1021/acs.jcim.0c00116
J. Chem. Inf. Model. 2020, 60, 4153−4169