ACCOUNTS of chemical research

Article

pubs.acs.org/accounts

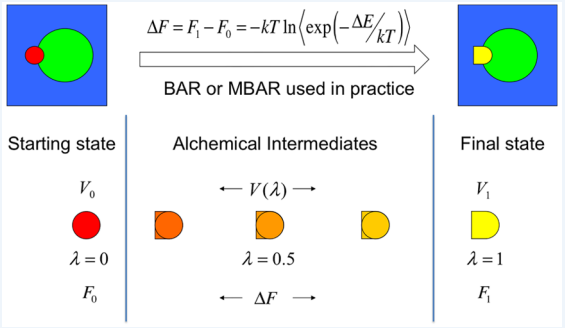
Advancing Drug Discovery through Enhanced Free Energy Calculations

Robert Abel,[†,§] Lingle Wang,[†,§] Edward D. Harder,[†,§] B. J. Berne,[‡] and Richard A. Friesner*[,‡]

[†]Schrödinger, Inc., 120 West 45th Street, New York, New York 10036, United States

[‡]Department of Chemistry, Columbia University, 3000 Broadway, New York, New York 10027, United States

CONSPECTUS: A principal goal of drug discovery project is to design molecules that can tightly and selectively bind to the target protein receptor. Accurate prediction of protein−ligand binding free energies is therefore of central importance in computational chemistry and computer aided drug design. Multiple recent improvements in computing power, classical force field accuracy, enhanced sampling methods, and simulation setup have enabled accurate and reliable calculations of protein−ligands binding free energies, and position free energy calculations to play a guiding role in small molecule drug discovery. In this Account, we outline the relevant methodological advances, including the REST2 (Replica Exchange with Solute Temperting) enhanced sampling, the incorporation of REST2 sampling with convential FEP (Free Energy Perturbation) through FEP/REST, the OPLS3 force field, and the advanced simulation setup that constitute our FEP+ approach, followed by the presentation of extensive comparisons with experiment, demonstrating sufficient accuracy in potency prediction (better than 1 kcal/mol) to substantially impact lead optimization campaigns. The limitations of the current FEP+ implementation and best practices in drug discovery applications are also discussed followed by the future methodology development plans to address those limitations. We then report results from a recent drug discovery project, in which several thousand FEP+ calculations were successfully deployed to simultaneously optimize potency, selectivity, and solubility, illustrating the power of the approach to solve challenging drug design problems. The capabilities of free energy calculations to accurately predict potency and selectivity have led to the advance of ongoing drug discovery projects, in challenging situations where alternative approaches would have great difficulties. The ability to effectively carry out projects evaluating tens of thousands, or hundreds of thousands, of proposed drug candidates, is potentially transformative in enabling hard to drug targets to be attacked, and in facilitating the development of superior compounds, in various dimensions, for a wide range of targets. More effective integration of FEP+ calculations into the drug discovery process will ensure that the results are deployed in an optimal fashion for yielding the best possible compounds entering the clinic; this is where the greatest payoff is in the exploitation of computer driven design capabilities.



A key conclusion from the work described is the surprisingly robust and accurate results that are attainable within the conventional classical simulation, fixed charge paradigm. No doubt there are individual cases that would benefit from a more sophisticated energy model or dynamical treatment, and properties other than protein−ligand binding energies may be more sensitive to these approximations. We conclude that an inflection point in the ability of MD simulations to impact drug discovery has now been attained, due to the confluence of hardware and software development along with the formulation of "good enough" theoretical methods and models.

INTRODUCTION

All atom, explicit solvent molecular dynamics (MD) simulations have become a powerful tool for modeling biomolecular systems. Interesting results have been obtained in studying a wide range of biological processes, including protein folding, ion channel transport, conformational change in G-protein coupled receptors, and ligand binding kinetics, with simulations times reported in the millisecond range.[1] The advent of inexpensive GPU hardware has made extensive MD simulations routinely available in academic and industrial laboratories.[2−4]

In the present Account, we focus on the application of free energy perturbation (FEP) methods utilizing MD for the calculation of protein−ligand binding affinities in structure based drug discovery projects. This problem differs from many of those enumerated above in that a very high degree of accuracy (on the order of 1 kcal/mol) and reliability, for a wide range of ligand chemistries, is required in the calculation of relative ligand binding affinities to substantively impact hit-to-lead and lead optimization efforts.[5,6] FEP calculations in principle provide a rigorous evaluation of the free energy difference $\Delta\Delta G_{AB}$ between the binding affinity of two ligands A and B. However, the accuracy is critically dependent upon both

Received: February 10, 2017
Published: July 5, 2017

DOI: 10.1021/acs.accounts.7b00083
Acc. Chem. Res. 2017, 50, 1625−1632

a series of heuristic approximations inherent in the classical simulation methodology, and the details of the model parametrization and sampling algorithms.

Over the past 5 years, advances in both computer hardware and FEP methodology have enabled large-scale testing of the accuracy and robustness of FEP methods in both retrospective and prospective studies.[5,7−13] We discuss below the progress that has been made in enhanced sampling, force field development, and automation of system setup, and report results comparing to experimental data for a wide range of ligand−receptor complexes. An illustrative application of FEP in an industrial drug discovery project is then presented. Finally, the implications of these developments for drug discovery efforts going forward, as the calculations continue to become more efficient and reliable, are considered.

FEP METHODOLOGY

Free energy perturbation (FEP) refers to an ensemble of rigorous statistical mechanical methods enabling the calculation the free energy change of an alchemical process by slowly morphing the potential energies, such as the transformation of ligand A to ligand B, thus giving the relative binding free energy of the ligands to the same receptor. The thermodynamic cycle depicted in Figure 1 illustrates how the binding free energy

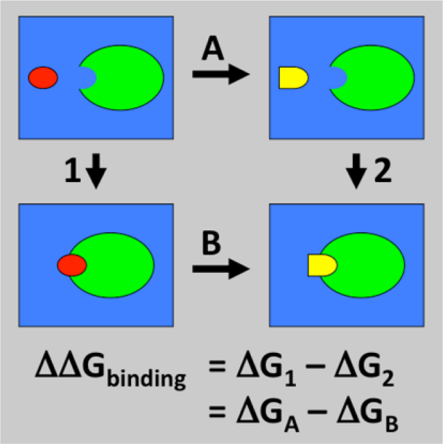


Figure 1. Thermodynamic pathway used for relative binding free energy calculations. The protein is depicted in green, the aqueous solvent in blue, the initial ligand "1" in red, and the final ligand "2" in yellow. The relative binding free energy is calculated via two distinct alchemical transformations where first alchemical transformation "A" is used to determine the free energy of transforming ligand 1 to ligand 2 in the solvent; and second alchemical transformation "B" is used to determine the free energy of transforming ligand 1 to ligand 2 in the receptor. The difference between the free energies obtained from alchemical transformations A and B can be rigorously related to the binding free energy difference of the two ligands 1 and 2.

difference, $\Delta\Delta G_{AB}$, is typically computed in practice. The Zwanzig exponential average[16] (also called FEP in some literature) is a representative way among the various formulations to relate the free energy difference between the two physical states A and B to the changes in their energy distributions:[14−18]

$$e^{-\beta\Delta F_{A\to B}} = \langle e^{-\beta\Delta U_{A\to B}(x)}\rangle_A \quad (1)$$

where $\Delta U_{A\to B}(x)$ is the potential energy difference between the two states at configuration x, and the average is taken over the ensemble of configurations sampled for state A. In practice, a number of intermediate states (also called lambda windows) are introduced such that the neighboring windows have sufficient overlapped regions in phase space to enable converged free energy calculations.

Since the first FEP calculations of protein−ligand binding were carried out in 1980s,[19] a standardized approach, incorporating a series of heuristic approximations, has been developed which accounts for the great majority of FEP simulations performed to date.[20,21] First, the configurations are sampled through classical molecular dynamics simulations, as opposed to a quantum mechanical treatment of nuclear motion.[22,23] Second, a molecular mechanics force field based on atom-centered fixed charges is employed.[7,24−33] Typical functional form of the force field is given by

$$U = \sum_{\text{bonds}} \frac{1}{2}k_b(b - b_0)^2 + \sum_{\text{angles}} \frac{1}{2}k_\theta(\theta - \theta_0)^2 + \sum_{\text{torsions}} K_\varphi[1 - \cos(n\varphi - \delta)] + \sum_{\text{nonbonded}} \left\{\frac{332q_iq_i}{r_{ij}} + \varepsilon_{ij}\left[\left(\frac{r_{ij,0}}{r_{ij}}\right)^{12} - \left(\frac{r_{ij,0}}{r_{ij}}\right)^{6}\right]\right\} \quad (2)$$

The use of fixed charges instead of an explicit representation of polarization effects, and other limitations of the details of the functional form, potentially limit the accuracy and robustness of the model. Third, typical FEP simulation times are of limited duration; as the potential energy surface of the protein−ligand complex exhibits a huge number of local minima, the system can become trapped and fail to execute ergodic sampling across configuration space.[34−40]

The use of the classical equations of motion and neglect of explicit polarization effects constitute major approximation to the exact physics, adopted because of the large increase in complexity and computational cost associated with more realistic treatments. Over the past 5 years, we have endeavored to answer a relatively straightforward question: what sort of accuracy can be achieved with the standard classical simulation, fixed charge FEP methodology, if a large engineering effort is made to improve the parametrization of the force field, apply enhanced sampling methods that are better able to overcome barriers, and ensure that the initial system setup is as precise as possible? Below we briefly outline the improvements in the force field, sampling algorithms, and implementation that constitute our current approach, which we call FEP+.[5] Comparisons with extensive and diverse experimental data sets are presented to address the key issues of accuracy and reliability of the FEP calculations.

THE OPLS3 FORCE FIELD

OPLS3 is based on the OPLS force field developed over the past 30 years by Jorgensen and co-workers.[7,24−26] The functional form is that of eq 2, although some off center charges are employed for ring nitrogens and halogens, based on investigations showing that asymmetries in the atomic charge distribution play a particularly important role in these cases (a similar modification may be required for sulfur; this is currently under investigation).[7,26] van der Waals parameters, and some atomic charges, are obtained from fitting to liquid state thermodynamic data; valence force field parameters such as

DOI: 10.1021/acs.accounts.7b00083
Acc. Chem. Res. 2017, 50, 1625−1632

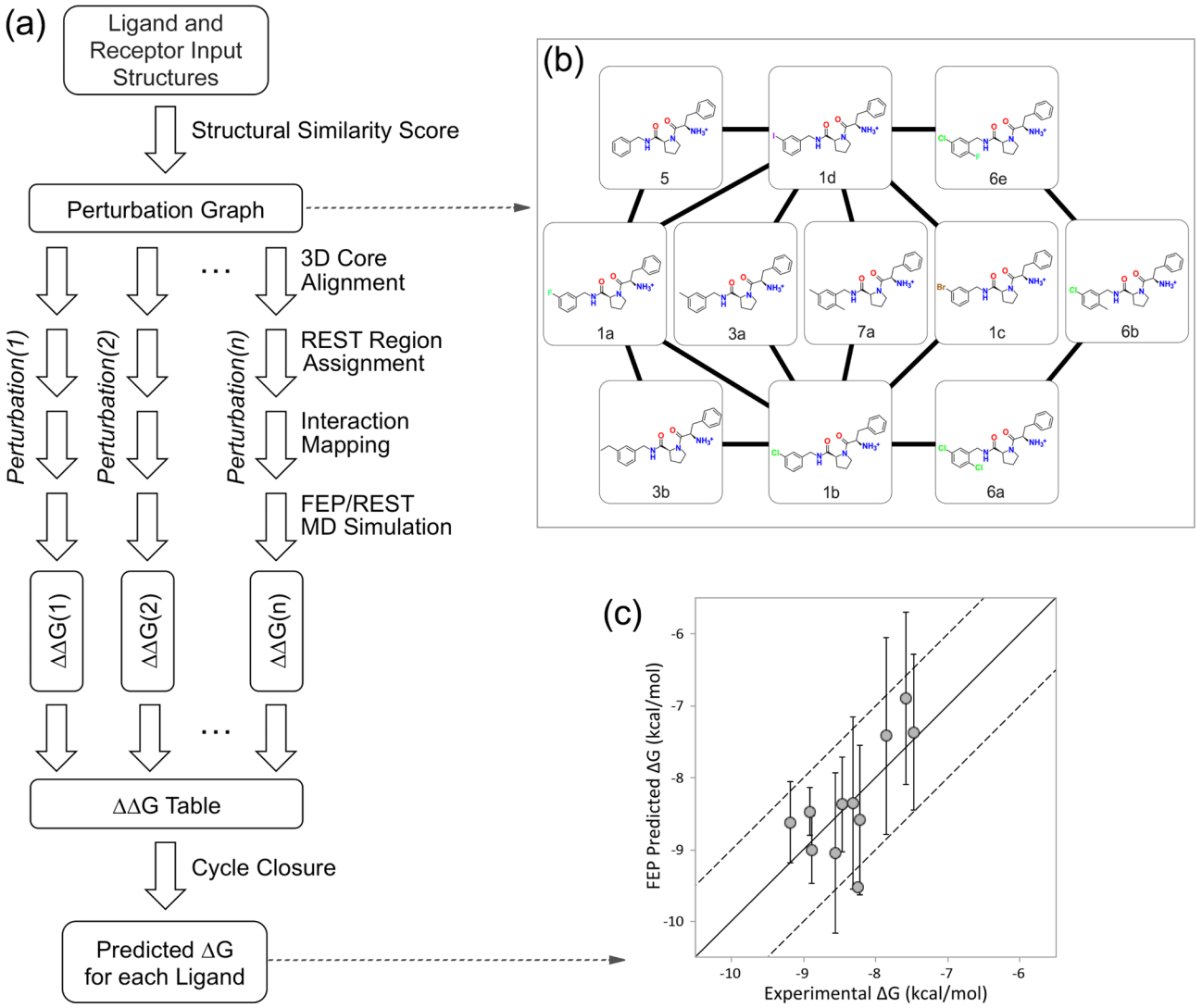


Figure 2. (a) FEP+ workflow for protein−ligand binding affinities calculations. (b) Example of a mapping of a perturbation space onto a set of pathways for thrombin ligands generated from the workflow. Each line represents two FEP+ calculations, one conducted in the receptor and one in solution, each perturbing between the two connected ligands. (c) Correlation plot of FEP+ predicted and experimental binding affinities for thrombin ligands. Reproduced from ref 5. Copyright 2015 American Chemical Society.

torsions are primarily determined by fitting to high level quantum chemical data, although parameters for proteins are modified based on angular distributions for both the backbone and side chain as found in the Protein Data Bank (PDB). The CM1a-BCC model is used to determine molecular charges; bond charge corrections are obtained from fitting to aqueous solvation free energy data, and from quantum chemical calculations. This overall approach to parametrization is similar to that used for other widely used force fields such as CHARMm and AMBER.[27−31]

Where OPLS3 differs from these alternative force fields is in the degree of parametrization that has been applied, particularly to the valence force field and the charge model. First, OPLS3 contains more than 15 000 torsional parameters, as well as thousands of stretching and bending parameters.[7] Second, the BCC component of the charge model has been explicitly optimized to improve agreement with aqueous solvation free energies for a database of small organic molecules with known experimental data. Detailed comparisons with alternative force fields, demonstrating substantial improvement in both areas, are provided in refs 7 and 41.

Despite the vastly increased coverage of torsional parameter space in OPLS3, the constant search for new compounds and chemistries in drug discovery projects inevitably yields compounds with torsions not accurately represented by parameters in the force field database. In such cases, development of a customized torsional parameter set is needed to accurately describe the interaction energies of the unparameterized chemical groups. This issue is addressed in OPLS3 by an automated algorithm, here denoted FFBuilder, that detects the lack of a good match and initiates quantum chemical calculations, followed by torsional fitting, to obtain the missing parameters.[7]

■ THE FEP/REST SAMPLING METHOD

Numerous sampling algorithms have been proposed to enable biomolecular MD simulations to escape from being trapped in local minima. The REST2 method (Replica Exchange with Solvent Tempering 2) employs multiple parallel simulations in

DOI: 10.1021/acs.accounts.7b00083
Acc. Chem. Res. 2017, 50, 1625−1632

Table 1. Relative Binding Free Energy Results for the OPLS 2005, OPLS2.1, and OPLS3 Force Fields[a]

		OPLS_2005		OPLS2.1		OPLS3	
system	no. cmpds	R2	RMSE	R2	RMSE	R2	RMSE
BACE	36	0.01	1.35	0.56	1.03	0.64	0.89
CDK2	16	0.48	0.98	0.07	1.27	0.51	0.86
JNK1	21	0.75	0.87	0.74	0.87	0.76	0.62
MCL1	42	0.46	1.77	0.62	1.44	0.37	1.4
P38	34	0.32	0.95	0.54	0.97	0.57	1.05
PTP1B	23	0.55	1.55	0.5	1.05	0.79	0.57
thrombin	11	0.21	1.35	0.4	0.97	0.38	0.83
Tyk2	16	0.86	0.75	0.8	0.98	0.84	0.98
weighted avg			1.28 ± 0.06		1.11 ± 0.05		0.95 ± 0.04

[a]This table is reproduced from ref 7. Copyright 2016 American Chemical Society.

which the potential energies of a selected subsystem (selected because conformational changes in this region might not be sampled efficiently otherwise) have been scaled in a way that mimics the application of a locally higher temperature as one ascends the replica ladder, but leaves the rest of the system at the desired temperature.[36] This has the effect of reducing the number of replicas needed and greatly increasing the acceptance probability of the Monte Carlo replica exchange, thereby accelerating the sampling while maintaining detailed balance.[36] The lowest replica has no scaling and thus represents the thermodynamics of interest.

In the application of REST2 to FEP calculations for protein−ligand complexes, which we call FEP/REST, the elevated effective local temperatures are focused in the region of the ligand where an alchemical change is being performed, and protein residues close to the binding pocket may also be included in the enhanced sampling region if needed.[5,8,10,11,34,35,37,39,42] In FEP/REST, the effective temperature of the enhanced sampling region, as a function of lambda, gradually increases from room temperature with lambda = 0 for the initial physical state to a much higher temperature with an intermediate lambda value equal to 0.5 (the highest effective temperature is about 1000 K for a typical perturbation with about 20 heavy atoms in the hot region); then, the temperature is gradually lowered to room temperature while lambda is increased to 1 corresponding to the final physical state. In this way, the potential energies for the two end points reach the correct physical states, and enhanced sampling can be achieved through the increased effective temperatures of the intermediate lambda windows. This effective local heating via the scaling of the Hamiltonian significantly improves the efficiency of exchanging the configurations across the temperature ladder over other alternatives, such as temperature replica exchange method.[35,36]

THE FEP+ IMPLEMENTATION

The above-mentioned FEP/REST sampling method using the OPLS3 force field has been implemented in the Desmond GPU molecular dynamics simulation package which is called FEP+.[5] Within FEP+, FFbuilder can be used to obtain customized torsional parameters to extend the torsional coverage,[7] and calculation setup and cycle closure convergence analysis[37] has also been fully automated through a graphical user interface. An entire suite of calculations for a series of compounds can be launched with a graph that automatically enumerates the transformations needed for prediction of the molecules in the specified ensemble of ligands.[5,43] Multiple pathways can be run for each calculation, enabling a superior convergence error estimate (via "cycle closure" formulas) to be produced.[37,43,44] The workflow of the FEP+ calculations for a series of thrombin ligands is shown in Figure 2 as an example.

REQUIREMENTS FOR ACHIEVING A SUCCESSFUL FEP+ SIMULATION; LIMITATIONS AND PITFALLS OF THE CURRENT IMPLEMENTATION

FEP+ is a physics-based method, and a sufficiently accurate initial structure for the complex is required to obtain reasonable results. The method is fairly tolerant of relatively low-resolution crystal structures; resolution below 2.5 Å is generally sufficient, and good results have been obtained in the 2.5−3.0 Å resolution range. Some successes have also been achieved using homology models rather than crystal structures, although more extensive testing is needed before a definitive guide to this type of application can be produced.[45] On the other hand, if there are a substantial number of missing residues, or unresolved loops, in contact with the ligand, it is going to be difficult to achieve high accuracy unless the missing or unresolved structures can be accurately constructed using computational methods.[46] In addition, while the binding enthalpy and entropy can also be obtained by multiple free energy simulations at different temperatures through Van 't Hoff equation, they are much difficult to converge and more extensive testing is needed to assess the accuracy with which these quantities can be routinely computed.

An important issue that affects the utility of any FEP methodology in practical applications is the magnitude of the perturbation that can be handled without unacceptable loss of accuracy. The FEP+ protocol has been found to yield robust results for perturbations up to 10 heavy atoms,[5] a significant advance as compared to much of the prior work in the literature, even though perturbations inducing significant protein motion are still challenging for the current technology.[34,46] Further, perturbations larger than 10 heavy atoms are routinely pursued with project needs, and a similar level of accuracy of 1 kcal/mol RMSE (root-mean-square error) is most typically obtained. When pursuing such very large perturbations, it is crucial to closely monitor the cycle closure hysteresis values to ensure that the predicted free energies remain largely independent of the particular sampled alchemical path.

Finally, there are a number of specific problems that can adversely affect the accuracy of FEP results. For example, classical force fields for transition metals are generally not as accurate as for organic molecules; if a perturbation involves a change in the metal−ligand interaction, the FEP results may have larger errors than are normally observed. Water molecules

DOI: 10.1021/acs.accounts.7b00083
Acc. Chem. Res. 2017, 50, 1625−1632

can become kinetically trapped in the interior of the protein when a perturbation should push them out into bulk solution. Likewise, if the ligand or the protein changes protonation state upon binding, the accuracy of the prediction may suffer. We expect these issues will diminish with the next round of methodological improvements, including enhanced sampling of water equilibration, constant pH simulation, broader coverage on chemical space by the force field, as well as more efficient and convergent enhanced sampling of the protein motion related degrees of freedom. Lastly, the current implementation of FEP+ protocol does not support the change of the total charge on the ligand, i.e., transforming a neutral ligand into an ionic species, which we hope to resolve this limitation in the future.

BENCHMARK RESULTS FOR FEP+ BINDING AFFINITY PREDICTION

To evaluate the performance of the FEP+ methodology in a statistically meaningful fashion, we have assembled a large, diverse data set of test cases, based on the results of medicinal chemistry studies reported in the literature. Each data set contains a series of related ligands, and their binding free energies, to a specified protein target of pharmaceutical interest. Details of the data set have been presented previously in ref 5.

Table 1 reports the FEP+ predictions of the binding affinity, RMS errors, and correlation coefficients for this data set versus the experimental binding affinities, using the most recent FEP+ implementation and the OPLS3 force field. The first three columns in Table 1 enumerate results using three different versions of the OPLS force field (OPLS2005, OPLS2.1, OPLS3), while employing the same sampling methodology. Performance is robust across the various data sets, with the error following an approximately Gaussian distribution, as shown in Table 2. Furthermore, the RMS error decreases systematically with improvement in the force field. Importantly, this improvement comes from functional form modification and fitting to additional quantum chemical data; no direct fitting to protein−ligand binding affinities was utilized.

Table 2. Histogram of the Error Distribution of the OPLS3 Relative Binding Free Energy Results versus the Experimental Data Adapted from Ref 7[a]

error (kcal/mol)	% obsd	% expected
% < 0.5	42%	38%
% < 1.0	73%	68%
% < 1.5	87%	86%
% < 2.0	94%	95%
% < 3.0	99.4%	99.70%

[a]The error distribution closely follows the expected distribution for a prediction method with Gaussian error distribution of 1 kcal/mol.

The RMS error for current best practices FEP+ reported in Table 1, 0.95 kcal/mol, includes errors in both computational and experimental results. Taking the estimate in ref 6 for the experimental target data RMSE to be approximately 0.5 kcal/mol, and using the relation:

$$\sigma_{tot} = \sqrt{\sigma_i^2 + \sigma_j^2} \quad (3)$$

where σ_{tot} is the total apparent error and σ_i and σ_j are individual contributions from experimental and computational results, respectively. We would further observe for this particular case

$$0.95\ \text{kcal/mol} = \sqrt{0.5\ \text{kcal/mol}^2 + \sigma_{\text{FEPintrinsic}}{}^2} \quad (4)$$

where $\sigma_{\text{FEPintrinsic}}$ is the expected intrinsic error of the FEP+ prediction if perfectly accurate experimental data were available, which in this case implies an intrinsic RMSE of 0.8 kcal/mol. This is relatively close to the experimental RMS error itself, suggesting further improvement in the force field and sampling algorithms may yield another ∼0.1−0.2 kcal/mol improvement; but unless experimental measurements start to be routinely made with much higher precision, further progress in the near future is going to be very difficult, as it becomes increasingly challenging to separate true computational outliers from experimental noise. Furthermore, the motivation for increasing the precision of experimental assays is limited, as what ultimately matters in a drug discovery project is the in vivo efficacy; there is often a good correlation between in vitro assays and in vivo affinity, but not in most cases beyond the level of 0.5 kcal/mol.

To further substantiate this finding, we have recently undertaken a large-scale review of FEP+ scoring accuracy across projects both within our groups, and within industrial and academic collaborators. At the time of this writing, we are aware of 92 distinct applications of FEP+ to score small molecule ligand series, both prospective and retrospective, where the number of ligands was sufficiently large to make judgments regarding the accuracy of the scoring for the series. This data set includes the scoring of more than 3000 ligands where the resulting computed affinity could be directly compared to the experimental data either prior to or subsequent to the FEP+ calculations. Key statistics of this large-scale analysis are presented on Table 3. Quite encouragingly, the average and median RMS errors across this very large test set (including extensive prospective studies) agrees well with the earlier reported estimates, suggesting such accuracy should be reliably observed in future projects.

Table 3. Key Statistics of a Recent Large-Scale Review of FEP+ Scoring Accuracy

total no. of projects	92
total no. of ligands	3021
no. of academic collaborations	4
no. of internal projects	26
no. of discovery collaboration	24
no. of industrial projects	38
no. of prospective projects	27
avg RMSE of all projects	1.1 kcal/mol
median RMSE of all projects	1.0 kcal/mol
avg R^2 value of all projects	0.57
median R^2 of all projects	0.62
avg RMSE of prospective projects	1.1 kcal/mol
median RMSE of prospective projects	1.0 kcal/mol
avg R^2 value of prospective projects	0.66
median R^2 value of prospective projects	0.68

It is important to note that, in addition to using an accurate force field, accurate predictions are also critically dependent on the efficiency of the sampling. If ordinary FEP is used without REST2 enhanced sampling, a variety of perturbations can not be converged on even longer time scales resulting in much worse free energy predictions.[35,37,40,47] We have found in the context of our discovery collaborations this is especially important for prospective work where the binding modes of

DOI: 10.1021/acs.accounts.7b00083
Acc. Chem. Res. 2017, 50, 1625−1632

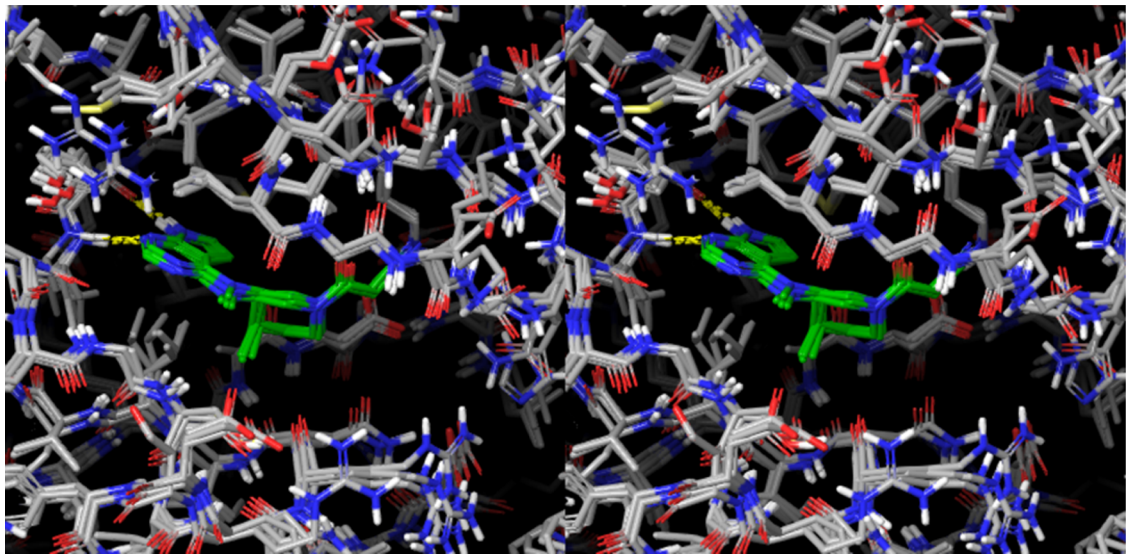


Figure 3. Superimposed crystal structures of Tyk2, JAK1, JAK2, and JAK3 cocrystallized with tofacitinib in stereo representation (PDB IDs 3LXK, 3LXN, 3FUP, and 3EYG, respectively).

the R-groups may not yet be fully understood from the project crystallography and structure−activity relationship.

MASSIVE FEP+ SCREENING IN DRUG DISCOVERY PROJECTS

The greatest impact of FEP calculations in a drug discovery project is manifested when a very small number of compounds in a large set of plausible candidates will optimally advance the project. In such a situation, conventional approaches will generally be unable to locate the optimal molecules to make without incurring extraordinarily large expenses in synthetic chemistry. Situations of this type occur routinely in drug discovery efforts. For example, if a high degree of selectivity is required against multiple, closely related family members, it is likely to be very difficult to simultaneously achieve these objectives, along with other properties like potency, solubility, metabolic stability, and membrane permeability. A second factor that can contribute significantly to the degree of difficulty is the nature of the target; some targets possess very challenging binding sites, for which designing a drug-like yet potent molecule constitutes a major hurdle.

We briefly outline a recent project in which very large numbers of molecules were computationally assayed via FEP+ calculations. In this drug discovery program, described in detail in ref 48, the focus of the project was the development of a selective inhibitor of Tyk2, a kinase involved in control of immune response. Inhibition of Tyk2 has been shown to modulate autoimmune disease, for example, in animal models of psoriasis. However, Tyk2 is a member of the JAK family of kinases, which includes JAK1, JAK2, and JAK3. Overly strong inhibition of other members of the family can result in side effects including anemia and enhanced susceptibility to infection. Therefore, the project goal was to design a molecule with 100× selectivity of Tyk2 against the JAK kinases, which is quite challenging due to the high degree of active site similarity between these proteins, as depicted in Figure 3.[48] FEP+ calculations for the project incorporated all three selectivity criteria, as well as potency and solubility. Other properties, such as membrane permeability and metabolic stability, were modeled using more approximate computational methods.

A total of 4000 design ideas, starting from a number of different lead compounds, were computationally screened via FEP+. Of these, 46 compounds were prioritized for synthesis on the basis of the calculations, and 9 were found to meet the target potency, selectivity, and solubility criteria after experimental testing. A number of these compounds have been shown to potently inhibit targeted immune cell cytokine signaling, and demonstrate outstanding efficacy in ameliorating disease in mouse models of psoriasis.[48]

DISCUSSIONS AND CONCLUSIONS

We have established, via extensive retrospective and prospective testing, that FEP+ is capable of potency, and selectivity predictions that are beginning to approach the limit of experimental accuracy. FEP+ calculations have a cost and speed advantage that is 100−1000× as compared to a brute force experimental evaluation of all of the proposed candidate molecules. The ability to effectively carry out projects evaluating tens of thousands, or hundreds of thousands, of proposed drug candidates is potentially transformative in enabling hard to drug targets to be attacked.

Continued developments of both experimental and computational technology will enhance both the efficacy and domain of applicability of FEP-enabled drug discovery. Increasing numbers of high-resolution protein structures, accelerated by the emergence of cryo-electron microscopy methods for structure determination, and augmented by increasingly capable homology modeling approaches, will increase the fraction of targets amenable to structure-based drug design. Improvements in sampling, GPU hardware, and molecular mechanics force fields will enhance the reliability of FEP predictions while systematically reducing the computational cost per calculation, following the Moore's law curve in that regard.

Possibly the most exciting, although more speculative, opportunity is in the potential expansion of chemical space, beyond the ∼100 000−300 000 compounds that could comfortably be evaluated via FEP+ in a project to the billions or trillions of compounds accessible in a de novo design approach. Improvements in more approximate methods, such as docking, empirical scoring functions, and continuum solvent

DOI: 10.1021/acs.accounts.7b00083
Acc. Chem. Res. 2017, 50, 1625−1632

based energy models, are needed to make such an approach practical; but free energy calculations have a fundamental role to play in this type of workflow, serving to provide benchmark evaluation of molecules emerging from earlier stages, generating new low energy receptor conformations, and recycling these conformations, along with their reorganization energies, back into the earlier stages. The ability to access an ultralarge chemical space could enable highly challenging drug design problems to be solved with precision molecules in a fashion that is currently not possible, enabling a renaissance in the potential of small molecule drug discovery.

We are optimistic about the future of MD-based atomistic simulation, not only in the area of biomolecular modeling, but also in describing a wide variety of materials and chemical processes. Further advances will be needed in quantum chemistry, force field development, and simulation technology to reach the point of reliable quantitative prediction in these related fields, but these can be expected in the next several decades. For the present, FEP-enabled drug discovery applications are poised at an exciting historical moment, with the opportunity for extensive validation in the clinic over the next 5−10 years.

AUTHOR INFORMATION

Corresponding Author

*E-mail: raf8@columbia.edu.

ORCID

Lingle Wang: 0000-0002-8170-6798
Richard A. Friesner: 0000-0002-1708-9342

Author Contributions

§R.A., L.W., and E.D.H. contributed equally to this work.

Notes

The authors declare the following competing financial interest(s): R.A.F. has a significant financial stake in, is a consultant for, and is on the Scientific Advisory Board of Schrodinger, Inc. B.J.B. is consultant to Schrodinger, Inc. and is on its Scientific Advisory Board.

Biographies

Robert Abel is the Vice President of Scientific Development at Schrödinger Inc. His research interests include statistical mechanics, biomolecular modeling, and drug discovery.

Lingle Wang is a senior principal scientist at Schrödinger Inc. His research interests include computational biophysics and computer aided drug design.

Edward D. Harder is a senior principal scientist at Schrödinger Inc. His research interests are in molecular modeling and simulation.

Bruce J. Berne is Higgins Professor of Chemistry and Professor of Chemical Engineering at Columbia University. His research interests are in statistical mechanics and biomolecular simulations.

Richard A. Friesner is William P. Schweitzer Professor of Chemistry at Columbia University. His research interests include quantum chemistry, biomolecular modeling, and structure based drug discovery.

ACKNOWLEDGMENTS

The authors thank William Jorgensen, Vijay Pande, David Mobley, and John Chodera for many helpful discussions.

REFERENCES

(1) De Vivo, M.; Masetti, M.; Bottegoni, G.; Cavalli, A. Role of Molecular Dynamics and Related Methods in Drug Discovery. *J. Med. Chem.* **2016**, *59*, 4035−4061.

(2) Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* **2013**, *9*, 3878−3888.

(3) Bergdorf, M.; Baxter, S.; Rendleman, C. A.; Shaw, D. E. Desmond/GPU Performance as of October 2015; *D. E. Shaw Research Technical Report DESRES/TR--2015-01*, 2015.

(4) Eastman, P.; Friedrichs, M. S.; Chodera, J. D.; Radmer, R. J.; Bruns, C. M.; Ku, J. P.; Beauchamp, K. A.; Lane, T. J.; Wang, L.-P.; Shukla, D.; Tye, T.; Houston, M.; Stich, T.; Klein, C.; Shirts, M. R.; Pande, V. S. OpenMM 4: A Reusable, Extensible, Hardware Independent Library for High Performance Molecular Simulation. *J. Chem. Theory Comput.* **2013**, *9*, 461−469.

(5) Wang, L.; Wu, Y.; Deng, Y.; Kim, B.; Pierce, L.; Krilov, G.; Lupyan, D.; Robinson, S.; Dahlgren, M. K.; Greenwood, J.; Romero, D. L.; Masse, C.; Knight, J. L.; Steinbrecher, T.; Beuming, T.; Damm, W.; Harder, E.; Sherman, W.; Brewer, M.; Wester, R.; Murcko, M.; Frye, L.; Farid, R.; Lin, T.; Mobley, D. L.; Jorgensen, W. L.; Berne, B. J.; Friesner, R. A.; Abel, R. Accurate and Reliable Prediction of Relative Ligand Binding Potency in Prospective Drug Discovery by Way of a Modern Free-Energy Calculation Protocol and Force Field. *J. Am. Chem. Soc.* **2015**, *137*, 2695−2703.

(6) Brown, S. P.; Muchmore, S. W.; Hajduk, P. J. Healthy skepticism: assessing realistic model performance. *Drug Discovery Today* **2009**, *14*, 420−427.

(7) Harder, E.; Damm, W.; Maple, J.; Wu, C.; Reboul, M.; Xiang, J. Y.; Wang, L.; Lupyan, D.; Dahlgren, M. K.; Knight, J. L.; Kaus, J. W.; Cerutti, D. S.; Krilov, G.; Jorgensen, W. L.; Abel, R.; Friesner, R. A. OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *J. Chem. Theory Comput.* **2016**, *12*, 281−296.

(8) Steinbrecher, T. B.; Dahlgren, M.; Cappel, D.; Lin, T.; Wang, L.; Krilov, G.; Abel, R.; Friesner, R.; Sherman, W. Accurate Binding Free Energy Predictions in Fragment Optimization. *J. Chem. Inf. Model.* **2015**, *55*, 2411−2420.

(9) Mikulskis, P.; Genheden, S.; Ryde, U. A Large-Scale Test of Free-Energy Simulation Estimates of Protein−Ligand Binding Affinities. *J. Chem. Inf. Model.* **2014**, *54*, 2794−2806.

(10) Lovering, F.; Aevazelis, C.; Chang, J.; Dehnhardt, C.; Fitz, L.; Han, S.; Janz, K.; Lee, J.; Kaila, N.; McDonald, J.; Moore, W.; Moretto, A.; Papaioannou, N.; Richard, D.; Ryan, M. S.; Wan, Z.-K.; Thorarensen, A. Imidazotriazines: Spleen Tyrosine Kinase (Syk) Inhibitors Identified by Free-Energy Perturbation (FEP). *ChemMedChem* **2016**, *11*, 217−233.

(11) Rombouts, F. J. R.; Tresadern, G.; Buijnsters, P.; Langlois, X.; Tovar, F.; Steinbrecher, T. B.; Vanhoof, G.; Somers, M.; Andrés, J.-I.; Trabanco, A. A. Pyrido[4,3-e][1,2,4]triazolo[4,3-a]pyrazines as Selective, Brain Penetrant Phosphodiesterase 2 (PDE2) Inhibitors. *ACS Med. Chem. Lett.* **2015**, *6*, 282−286.

(12) Mey, A. S. J. S.; Juárez-Jiménez, J.; Hennessy, A.; Michel, J. Blinded predictions of binding modes and energies of HSP90-α ligands for the 2015 D3R grand challenge. *Bioorg. Med. Chem.* **2016**, *24*, 4890−4899.

(13) Rocklin, G. J.; Boyce, S. E.; Fischer, M.; Fish, I.; Mobley, D. L.; Shoichet, B. K.; Dill, K. A. Blind Prediction of Charged Ligand Binding Affinities in a Model Binding Site. *J. Mol. Biol.* **2013**, *425*, 4569−4583.

(14) Jarzynski, C. Comparison of far-from-equilibrium work relations. *C. R. Phys.* **2007**, *8*, 495−506.

(15) Bennett, C. H. Efficient estimation of free energy differences from Monte Carlo data. *J. Comput. Phys.* **1976**, *22*, 245−268.

(16) Zwanzig, R. W. High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *J. Chem. Phys.* **1954**, *22*, 1420−1426.

(17) Shirts, M. R.; Chodera, J. D. Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.* **2008**, *129*, 124105.

DOI: 10.1021/acs.accounts.7b00083
Acc. Chem. Res. 2017, 50, 1625−1632

(18) Kirkwood, J. G. Statistical Mechanics of Fluid Mixtures. *J. Chem. Phys.* **1935**, *3*, 300−313.
(19) Kollman, P. Free energy calculations: Applications to chemical and biochemical phenomena. *Chem. Rev.* **1993**, *93*, 2395−2417.
(20) Perez, A.; Morrone, J. A.; Simmerling, C.; Dill, K. A. Advances in free-energy-based simulations of protein folding and ligand binding. *Curr. Opin. Struct. Biol.* **2016**, *36*, 25−31.
(21) Pohorille, A.; Jarzynski, C.; Chipot, C. Good Practices in Free-Energy Calculations. *J. Phys. Chem. B* **2010**, *114*, 10235.
(22) McCammon, J. A.; Gelin, B. R.; Karplus, M. Dynamics of folded proteins. *Nature* **1977**, *267*, 585−590.
(23) Levitt, M.; Warshel, A. Computer simulation of protein folding. *Nature* **1975**, *253*, 694−698.
(24) Jorgensen, W. L.; Tirado-Rives, J. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* **1988**, *110*, 1657−1666.
(25) Robertson, M. J.; Tirado-Rives, J.; Jorgensen, W. L. Improved Peptide and Protein Torsional Energetics with the OPLS-AA Force Field. *J. Chem. Theory Comput.* **2015**, *11*, 3499−3509.
(26) Jorgensen, W. L.; Schyman, P. Treatment of Halogen Bonding in the OPLS-AA Force Field: Application to Potent Anti-HIV Agents. *J. Chem. Theory Comput.* **2012**, *8*, 3895−3901.
(27) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(28) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157−1174.
(29) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696−3713.
(30) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(31) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; MacKerell, A. D. Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ, ψ and Side-Chain χ1 and χ2 Dihedral Angles. *J. Chem. Theory Comput.* **2012**, *8*, 3257−3273.
(32) Schuler, L. D.; Daura, X.; van Gunsteren, W. F. An improved GROMOS96 force field for aliphatic hydrocarbons in the condensed phase. *J. Comput. Chem.* **2001**, *22*, 1205−1218.
(33) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Struct., Funct., Genet.* **2006**, *65*, 712−725.
(34) Lim, N. M.; Wang, L.; Abel, R.; Mobley, D. L. Sensitivity in Binding Free Energies Due to Protein Reorganization. *J. Chem. Theory Comput.* **2016**, *12*, 4620−4631.
(35) Wang, L.; Berne, B. J.; Friesner, R. A. On achieving high accuracy and reliability in the calculation of relative protein-ligand binding affinities. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 1937−1942.
(36) Wang, L.; Friesner, R. A.; Berne, B. J. Replica exchange with solute scaling: a more efficient version of replica exchange with solute tempering (REST2). *J. Phys. Chem. B* **2011**, *115*, 9431−9438.
(37) Wang, L.; Deng, Y.; Knight, J. L.; Wu, Y.; Kim, B.; Sherman, W.; Shelley, J. C.; Lin, T.; Abel, R. Modeling Local Structural Rearrangements Using FEP/REST: Application to Relative Binding Affinity Predictions of CDK2 Inhibitors. *J. Chem. Theory Comput.* **2013**, *9*, 1282−1293.
(38) Mobley, D. L.; Chodera, J. D.; Dill, K. A. Confine-and-Release Method: Obtaining Correct Binding Free Energies in the Presence of Protein Conformational Change. *J. Chem. Theory Comput.* **2007**, *3*, 1231−1235.
(39) Kaus, J. W.; Harder, E.; Lin, T.; Abel, R.; McCammon, J. A.; Wang, L. How To Deal with Multiple Binding Poses in Alchemical Relative Protein−Ligand Binding Free Energy Calculations. *J. Chem. Theory Comput.* **2015**, *11*, 2670−2679.
(40) Cole, D. J.; Tirado-Rives, J.; Jorgensen, W. L. Enhanced Monte Carlo Sampling through Replica Exchange with Solute Tempering. *J. Chem. Theory Comput.* **2014**, *10*, 565−571.
(41) Shivakumar, D.; Harder, E.; Damm, W.; Friesner, R. A.; Sherman, W. Improving the Prediction of Absolute Solvation Free Energies Using the Next Generation OPLS Force Field. *J. Chem. Theory Comput.* **2012**, *8*, 2553−2558.
(42) Wang, L.; Deng, Y.; Wu, Y.; Kim, B.; LeBard, D. N.; Wandschneider, D.; Beachy, M.; Friesner, R. A.; Abel, R. Accurate Modeling of Scaffold Hopping Transformations in Drug Discovery. *J. Chem. Theory Comput.* **2017**, *13*, 42−54.
(43) Liu, S.; Wu, Y.; Lin, T.; Abel, R.; Redmann, J. P.; Summa, C. M.; Jaber, V. R.; Lim, N. M.; Mobley, D. L. Lead optimization mapper: automating free energy calculations for lead optimization. *J. Comput.-Aided Mol. Des.* **2013**, *27*, 755−770.
(44) Wang, L.; Lin, T.; Abel, R. Cycle Closure Estimation of Relative Binding Affinities and Errors. EP Patent EP 2972311 A2, 2014.
(45) Cappel, D.; Hall, M. L.; Lenselink, E. B.; Beuming, T.; Qi, J.; Bradner, J.; Sherman, W. Relative Binding Free Energy Calculations Applied to Protein Homology Models. *J. Chem. Inf. Model.* **2016**, *56*, 2388−2400.
(46) Sherborne, B.; Shanmugasundaram, V.; Cheng, A. C.; Christ, C. D.; DesJarlais, R. L.; Duca, J. S.; Lewis, R. A.; Loughney, D. A.; Manas, E. S.; McGaughey, G. B.; Peishoff, C. E.; van Vlijmen, H. Collaborating to improve the use of free-energy and other quantitative methods in drug discovery. *J. Comput.-Aided Mol. Des.* **2016**, *30*, 1139−1141.
(47) Mondal, J.; Tiwary, P.; Berne, B. J. How a Kinase Inhibitor Withstands Gatekeeper Residue Mutations. *J. Am. Chem. Soc.* **2016**, *138*, 4608−4615.
(48) Abel, R.; Mondal, S.; Masse, C.; Greenwood, J.; Harriman, G.; Ashwell, M. A.; Bhat, S.; Wester, R.; Frye, L.; Kapeller, R.; Friesner, R. A. Accelerating drug discovery through tight integration of expert molecular design and predictive scoring. *Curr. Opin. Struct. Biol.* **2017**, *43*, 38−44.

DOI: 10.1021/acs.accounts.7b00083
Acc. Chem. Res. 2017, 50, 1625−1632