Optimization of Protein−Ligand Electrostatic Interactions Using an Alchemical Free-Energy Method

Alexander D. Wade[†] and David J. Huggins*[,†,‡,§]

[†]TCM Group, Cavendish Laboratory, University of Cambridge, 19 J J Thomson Avenue, Cambridge CB3 0HE, U.K.

[‡]Tri-Institutional Therapeutics Discovery Institute, Belfer Research Building, 413 East 69th Street, 16th Floor, Box 300, New York 10021, United States

[§]Department of Physiology and Biophysics, Weill Cornell Medicine, 1300 York Avenue, New York, New York 10065, United States

Ⓢ *Supporting Information*

ABSTRACT: We present an explicit solvent alchemical free-energy method for optimizing the partial charges of a ligand to maximize the binding affinity with a receptor. This methodology can be applied to known ligand−protein complexes to determine an optimized set of ligand partial atomic changes. Three protein−ligand complexes have been optimized in this work: FXa, P38, and the androgen receptor. The sets of optimized charges can be used to identify design principles for chemical changes to the ligands which improve the binding affinity for all three systems. In this work, beneficial chemical mutations are generated from these principles and the resulting molecules tested using free-energy perturbation calculations. We show that three quarters of our chemical changes are predicted to improve the binding affinity, with an average improvement for the beneficial mutations of approximately 1 kcal/mol. In the cases where experimental data are available, the agreement between prediction and experiment is also good. The results demonstrate that charge optimization in explicit solvent is a useful tool for predicting beneficial chemical changes such as pyridinations, fluorinations, and oxygen to sulfur mutations.
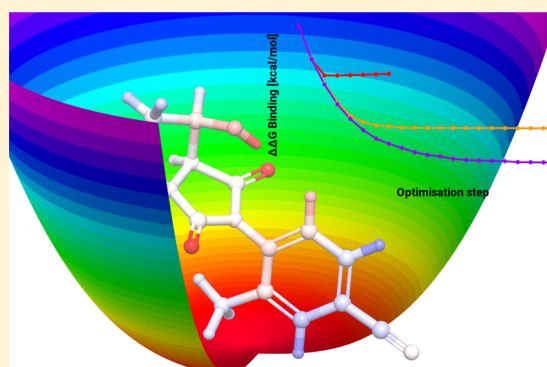
■ INTRODUCTION

In recent years, alchemical methods have garnered increasing attention in drug design.[1−6] In particular, free-energy perturbation (FEP) is now commonly used by pharmaceutical companies because of improvements in efficiency,[7] more accurate force fields,[8,9] and increases in computational power. Based on the Zwanzig equation,[10,11] it is common to use FEP calculations in drug design to calculate the relative binding affinity of two molecules.[11−14] This relative free energy, ΔG_{AB}, can be defined using the Zwanzig equation, as shown in eq 1, as the free-energy difference between thermodynamic states A and B.

$$\Delta G_{AB} = -kT \ln \langle e^{-((E_B - E_A)/kT)} \rangle_A \qquad (1)$$

with k as Boltzmann's constant, T as temperature, E_A and E_B as the potential energies of the system calculated using the Hamiltonian of state A and B, and $\langle . \rangle_A$ as a state average over system A. Typically, in free-energy calculations, eq 1 is expanded upon to include sampling from both states A and B[15] or sampling from intermediate states.[7] This is done to improve the sampling overlap between end states. The drawback here is that for every relative binding affinity calculation, lengthy molecular dynamics (MDs) simulations must be performed for all the end and intermediate states. If, however, the

perturbation between end states remains small enough, such that the overlap between end states is large, eq 1 is sufficient without any intermediate states. Applying eq 1 with no intermediate states is referred to as single-step perturbation (SSP), and this is the primary free-energy method used in this work. Numerous studies have used SSP,[16−19] demonstrating that it is applicable to relative free-energy calculations[20,21] and can be significantly faster than standard FEP.[22] Most recently, the authors of this article have used SSP to perform computational fluorine scanning.[23] We now apply SSP with the goal of optimizing ligand partial charges.

Charge optimization was developed by Tidor and co-workers using an implicit water treatment of electrostatics.[24,25] Poisson−Boltzmann calculations are performed on the bound and unbound states in order to find the optimal partial charges of a given molecule.[26−29] This approach has since been used by other academic groups,[30,31] employed in industries,[32] and been extended to consider induced fit effects.[33] However, the approach suffers from the deficiencies of all implicit water approaches: it is unable to deal effectively with interfacial water molecules. These occur commonly and are very difficult to treat effectively with implicit solvent approaches. The accuracy
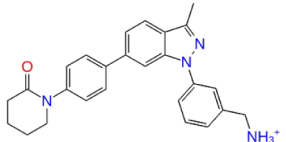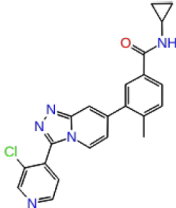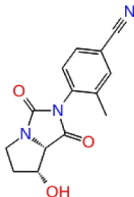
of implicit water models is lacking for many types of free-energy calculations,[34−36] particularly relevant here are the implicit model's shortcomings relative to explicit models in terms of binding free energies.[37] Additionally, while previous work has considered flexibility in the ligand, the receptor and complex were assumed to be rigid. It is known that this may play a significant role in binding free energies.[33] Because of advances in available computing power, explicit water approaches to charge optimization are now possible. We propose to exploit these computational advances by applying SSP to the bound and unbound states of small molecule inhibitors to develop a method for electrostatic charge optimization in explicit solvent.

Combining SSP with explicit water MD calculations and flexible receptors and complexes has the potential to develop a more accurate charge optimizer. To carry out these charge optimizations, we developed a tool to automate their execution. This tool is freely available at https://github.com/adw62/Ligand_Charge_Optimiser. Our ligand charge optimizer uses OpenMM[38] as both an MD engine and a tool to create the modified alchemical systems. The software will generate all of the required mutant ligands from an input wild-type ligand. These mutants are automatically parameterized, built into the complex systems, simulated, and optimized.

■ METHODS

We optimize the ligand partial charges for three protein test cases: FXa, P38 kinase, and the androgen receptor. The chemical structures of the ligands studied are shown in Table 1. The ligands were built from highly related molecules in the Protein Databank[39] (PDB): 2RA0[40] (FXa), 3S3I[41] (P38), and 2NW4[42] (androgen receptor). These small modifications are made from the PDB to allow comparison with experimental

Table 1. Name of the Target for Each System with Ligands' 2D Chemical Structure and the PDB IDs of the Target Ligand Complex

Target	Ligand	PDBIDs
Factor Xa		2RA0[40]
P38		3S3I[41]
Androgen Receptor		2NW4 [42]

data.[40−42] Superpositions of the molecules in the original PDB and the modified molecules are shown in Figures S2−S4.

System Setup. To prepare the FXa, P38, and androgen receptor systems, the nonstandard residues were converted to their standard equivalents with pdbfixer.[43] Selenomethionines were changed to methionines and missing sidechains were added using Schrödinger's Preparation Wizard,[44] which was also used to assign the protonation state of all ionizable residues. All buffer solvents and ions were removed. The hydrogen atom positions were then built using tleap and force field parameters, and partial charges were assigned from the AMBER ff14SB force field.[9] Parameters for the inhibitors were generated using Antechamber[45] with AMBER GAFF 2[46] and AM1-BCC.[47] These structures and parameters were then passed to YANK's[48] 0.23.7 automatic setup pipeline to build solvated ligand−protein and ligand systems. For solvation, TIP4P-EW[49] was used. A salt concentration of 150 mM and any required counterion were added using sodium and chloride ions. In every case, the edge of the solvation box was set to be 15 Å from any atom of the receptor and ligand.

Molecular Dynamics. MD sampling in this work is collected to compute SSP and FEP relative free energies, and the amount of MD sampling varies for each application in this work and so is stated explicitly for each case in the relevant section. All simulations were performed with OpenMM 7.3.0[43] as follows. First, OpenMM's default minimizer was used to minimize all structures. Then, equilibration was performed in the *NPT* ensemble at 300 K and 1 atm using a Langevin integrator and Monte Carlo barostat for 250 ps. MD simulations were performed in the *NPT* ensemble using a time step of 2 fs. van der Waals interactions were truncated at 11.0 Å with switching at 9.0 Å. Electrostatics were modeled using the particle mesh Ewald method with a cutoff of 11.0 Å. All other simulation parameters were left as default. Snapshots were collected every 5 ps and all simulations were performed with constrained hydrogens.

Charge Optimization. The objective function of this optimization was $\Delta G_{\text{binding}}(q_i) - \Delta G_{\text{original}}$. $\Delta G_{\text{binding}}(q_i)$ was defined as the difference in free energy between the bound and unbound states of the ligand and target for a ligand with charges q_i. $\Delta G_{\text{original}}$ was defined as the difference in free energy between the bound and unbound states of the ligand and target for a ligand with the charges of the original unoptimized ligand. $\Delta G_{\text{original}}$ is thus a constant. We therefore constructed the optimization problem as follows.

$$\min_{q_i} \Delta G_{\text{binding}}(q_i) - \Delta G_{\text{original}} \tag{2}$$

$$\text{s. t. } \sum_{i=1}^{m} q_i = \text{net charge} \tag{2.1}$$

$$\sqrt{\frac{\sum_{i=1}^{m} (q_{i,0} - q_{i,n+1})^2}{m}} \leq \text{rmsd limit} \tag{2.2}$$

$$|(q_{i,n} - q_{i,n+1})| \leq 0.01e \tag{2.3}$$

where q_i denotes the charges of the ligand and $q_{i,n}$ denotes the charges for iteration n of the optimization, m is the number of atoms in the ligand, and net charge is the total net charge of the ligand. Equation 2.1 constrains the net charge of the ligand to be constant. Equation 2.2 constrains the root mean square difference between the ligands' original charges, $q_{i,0}$, and the

$q_{i,n+1}$ charges to be less than some value of the root-mean-square deviation (rmsd) limit. These limits were chosen to limit the change in $\Delta\Delta G_{total}$ to a sensible range <10 kcal/mol. Without this limit, the optimization continued to very large unphysical values of $\Delta\Delta G_{total}$ because atomic partial charges can reach unphysical values. The results of an unconstrained optimization can be seen in Figures S17−S19. Equation 2.3 constrains the perturbation to each atom to be less than $0.01e$ per iteration, where e is the elementary charge. With this limit of $0.01e$, a determination of how much sampling was required to give converged calculation of $\Delta G_{binding}$ with a perturbation to each atom of $0.01e$ is shown in Figure S14. The amount of sampling needed was determined to be 2.5 ns, and this was then the amount of sampling used in this work to calculate the objective and gradient for each optimization step. The algorithm which was used to find a local minimum in this objective function was the SciPy 1.1.0[50] implementation of the sequential least squares programming algorithm.[51]

The primary calculations in this work were those of $\Delta\Delta G_{bindings}$, calculated as follows: first we applied SSP theory to calculate $\Delta G_{unperturbed \rightarrow perturbed}$ in the bound and unbound states as shown in eq 3.

$$\Delta G_{unperturbed \rightarrow perturbed}$$
$$= -kT \ln \langle e^{-((E_{perturbed} - E_{unperturbed})/kT)} \rangle_{unperturbed} \qquad (3)$$

$E_{perturbed}$ and $E_{unperturbed}$ are the potential energies of the system calculated using the Hamiltonian of the perturbed system and the unperturbed system, respectively. To change Hamiltonians, the charges were switched from unperturbed to perturbed values; however, Lennard-Jones, bonded, angle, and torsion parameters did not change. $\Delta\Delta G_{binding}$ was then constructed as shown in eq 4, with $\Delta G_{unperturbed \rightarrow perturbed}^{bound}$ and $\Delta G_{unperturbed \rightarrow perturbed}^{unbound}$ as $\Delta G_{unperturbed \rightarrow perturbed}$ calculated in the bound and unbound states of the ligand and target.

$$\Delta\Delta G_{binding} = \Delta G_{unperturbed \rightarrow perturbed}^{bound}$$
$$- \Delta G_{unperturbed \rightarrow perturbed}^{unbound} \qquad (4)$$

Note that $\Delta\Delta G_{binding}$ is equal to the objective function in eq 2, if we take eqs 3 and 4 and set the unperturbed state as the ligand with the original unoptimized charges and the perturbed state as the ligand with charges q_i.

To calculate the gradient of the objective function in each direction in charge space, a finite difference was calculated as shown in eq 5.

$$\nabla(\Delta G_{binding}(q_i) - \Delta G_{original})$$
$$= \frac{\Delta G_{binding}(q_i + h) - \Delta G_{binding}(q_i)}{h} \qquad (5)$$

where h is a finite difference of $0.00015e$. The numerator of the rhs of eq 5 is a $\Delta\Delta G_{binding}$ and can be calculated using an SSP approach as detailed in eqs 3 and 4 where q_i are the unperturbed charges and $q_i + h$ are the perturbed charges. This calculation of the gradient shows the advantage of SSP, as numerous (10−100 s) evaluations of $\Delta\Delta G_{binding}$ are required, one for each direction in charge space. This $\Delta\Delta G_{binding}$ is between molecules that are extremely similar, differing only by $0.00015e$ in one atom's partial charge. There is therefore likely to be a large sampling overlap between these states allowing SSP to be applied. Of note is that for each finite difference

calculation, the charge of the simulation box has been changed by $0.00015e$. The potential for finite size effects[52] caused by this change was investigated and the padding of the simulation with solvent was chosen to negate these effects, see Figure S15.

To delineate between the free-energy change for individual optimization steps and the free-energy change between the original and final optimized ligands, these will be defined as $\Delta\Delta G_{step}$ and $\Delta\Delta G_{total}$, respectively. Because an SSP method is being used, efforts were made to avoid poor overlap between the end states of any perturbations. To achieve this, the system was resampled after every optimization step. With the new sampling, the perturbed system could be redefined as a new unperturbed system after each optimization step. If this was not done, then the difference between the perturbed and unperturbed systems would grow over the course of the optimization, reducing the overlap in sampling and so reducing the applicability of SSP. Resampling had an additional advantage as it allowed for a calculation of a reverse alchemical step. Therefore, $\Delta\Delta G_{step}$ for both the forward and backward alchemical transformation was calculated for every step, and the $\Delta\Delta G_{step}$ in the results are reported as an average of the forward and backward transformations.

FEP Calculations. The optimization in this work gave a set of optimized charges and an associated $\Delta\Delta G_{binding}$ for going from the original set of ligand charges to the optimal set, named above as $\Delta\Delta G_{total}$. To validate the $\Delta\Delta G_{total}$ given by the optimization, we compared it against standard alchemical relative binding free-energy calculations using the MBAR[7] estimator applied to the system with the original and optimized set of charges as end states. These calculations were performed using the Ligand Charge Optimiser package as follows. For these calculations, 12 equally spaced lambda windows (except when explicitly stated otherwise) were used. Between these windows, the charge parameters were interpolated simultaneously from the original to optimized charges. All windows were sampled independently with 2 ns of Langevin dynamics totaling 24 ns of sampling. All simulation conditions were identical to the optimization MD calculations described above. The samples collected in each intermediate state were decorrelated based on an estimate of the statistical inefficiency of the reduced potential time series before carrying out the MBAR analysis with PyMBAR 3.0.1.[7]

In addition to testing the $\Delta\Delta G_{total}$, we also wished to calculate the $\Delta\Delta G_{binding}$ for a set of chemical changes informed by the optimal charges. This $\Delta\Delta G_{binding}$ is defined as $\Delta\Delta G_{designed}$ and is the change in binding free energy between the original ligand and a designed ligand with some chemical mutation. To perform these calculations, the protocol was the same for the full FEP calculation of $\Delta\Delta G_{total}$ but with the following additional considerations. Now, in addition to the charges, the van der Waals and bonded terms were all interpolated simultaneously from the original to the designed state. In the case of hydrogen to fluorine mutations, the original hydrogen was constrained; therefore, its associated C−H bond could not be interpolated to a C−F bond. When neglecting the interpolation of this bond, the fluorine appeared at the position of the hydrogen, instead of the true physical position of the fluorine. To avoid this issue, for fluorinations, we used a hybrid topology approach where a massless interaction site at the position of the mutation was added. This virtual site represents the fluorine and its position was defined relative to the position of the parent hydrogen such that the C−F distance is always 1.24 times the C−H

distance.[53] When mutating a hydrogen to fluorine, the relevant hydrogen was turned off and fluorine Lennard-Jones and charge parameters were applied to the additional site. When simulating these systems, all bonds to hydrogen were constrained. Because the position of the fluorine was defined relative to the position of its parent hydrogen, it was also implicitly constrained. We therefore made the assumption that the C−F bond length oscillations were negligible. To prevent the hybrid topologies from interacting, the additional sites were excluded from interacting with their parent hydrogens.

■ RESULTS

Using the optimization methodology mentioned above, the partial atomistic charges of three ligand−target systems are optimized. For each optimization, we would like to inspect the convergence over the number of steps. A good metric to analyze the results of the optimization is the set of optimized charges taken as a vector. In the methodology, a mention was made to limiting the rmsd between the original and optimized charges to some value of the rmsd limit. The values which are chosen for the rmsd limit are 0.01, 0.03, and 0.05e. The optimization is therefore repeated with the rmsd bound of these three values. With an rmsd bound of 0.01e, the optimizer is limited to seven steps as adequate convergence is seen at this point. For a larger rmsd, the convergence is slower and therefore for optimizations with rmsd bounds of 0.03 and 0.05e, the optimizer is limited to 20 steps. To assess convergence across simulation steps, we take the dot product of the normalized vector of new charges with the normalized vector of original charges for each step of the optimization (see Figure 1).

Figure 1 shows that the direction of the charge vectors over all systems and rmsds are well converged. The direction of these charge vectors represents where the charge is being applied on the molecule and this is the information that will be used in the following section to make chemical mutations to improve $\Delta\Delta G_{binding}$. It can also be seen that the dot product between the original and optimized charges is different for different rmsds. To quantify this difference, the dot product between the set of optimal charges obtained for rmsd 0.01 with 0.03 and 0.05e can be taken and the results of these projections can be seen in Table 2. Here, we see that sets of charges for the same system are pointing in the same direction. Thus, only the value of the charge changes are dependant on the rmsd, while the direction and relative magnitude of the charge changes are completely consistent. This is an important result because it shows that the design principles identified by the approach will not depend on the arbitrary choice of the rmsd. The invariance in where the charge is being applied can also be seen by the eye if the atoms are colored by change in charge. Figures illustrating this are presented in Figures S5−S13 for all sets of optimized charge.

In Table 2, we can see the dot product of the optimized charges from the optimization with an rmsd of 0.01e with themselves returns 1.00 as expected. The dot product of the vector of charges with rmsd = 0.01e and with rmsd = 0.03e also returns 1.00 as these vectors are extremely similar in direction. The dot product of the vector of charges with rmsd = 0.01e and with rmsd = 0.05e returns approximately 1.00 as these vectors are extremely similar in direction but not as close as 0.01e with 0.03e. To summarize, each optimization for a system added the charge in approximately the same place. It is only the magnitude of this charge that varies with rmsd. We can also
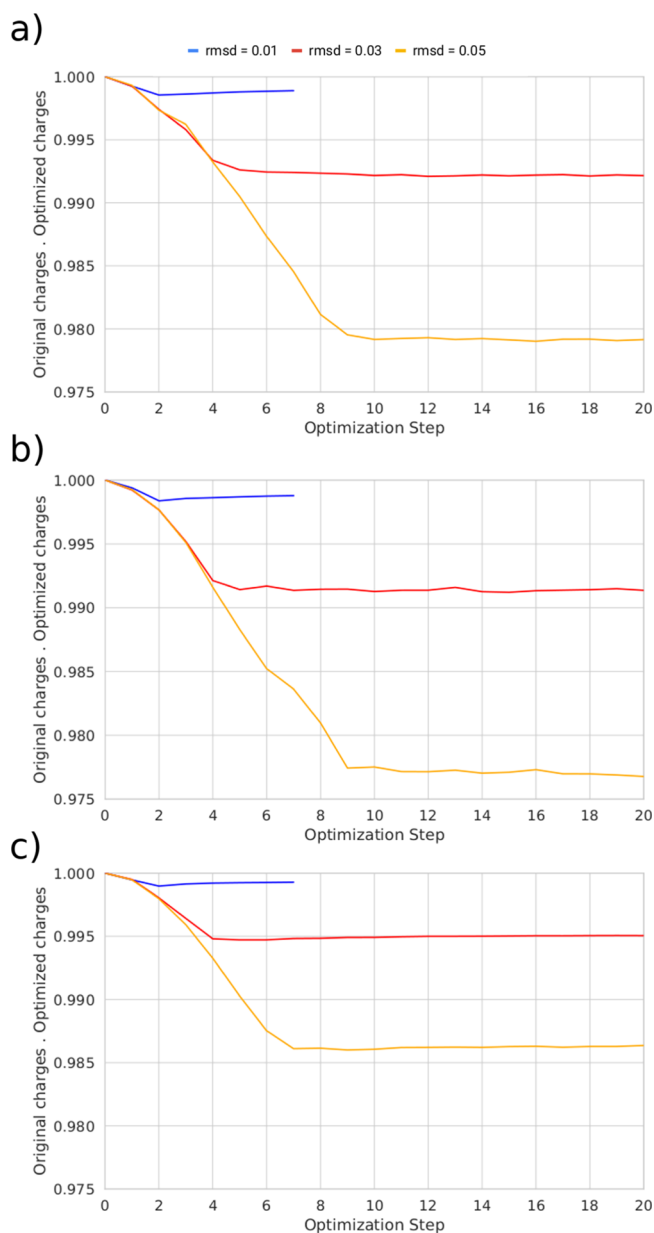

Figure 1. Dot product of the normalized optimized charges with the normalized original charges showing variation of charge vector direction with step. Results are shown for rmsd limits 0.01, 0.03, and 0.05e with the FXa, P38, and androgen receptor systems labeled (a−c), respectively.

Table 2. Dot Products of the Normalized Vector of Optimal Charges Using an rmsd of 0.01 q_e with the Normalized Vector of Optimal Charges Using Different rmsds

rmsd (e)	FXa	P38	androgen receptor
0.01	1.00	1.00	1.00
0.03	1.00	1.00	1.00
0.05	0.99	0.98	0.99

look at the convergence of $\Delta\Delta G_{step}$ with the optimization step, and this can be seen in Figure 2.

With an rmsd of 0.01e, Figure 2 demonstrates that $\Delta\Delta G_{step}$ is well converged for all systems. For an rmsd of 0.03 or 0.05e, the results are only well converged for the androgen receptor system, Figure 2c. This suggests that $\Delta\Delta G_{step}$ for the optimized

a)

b)

c)
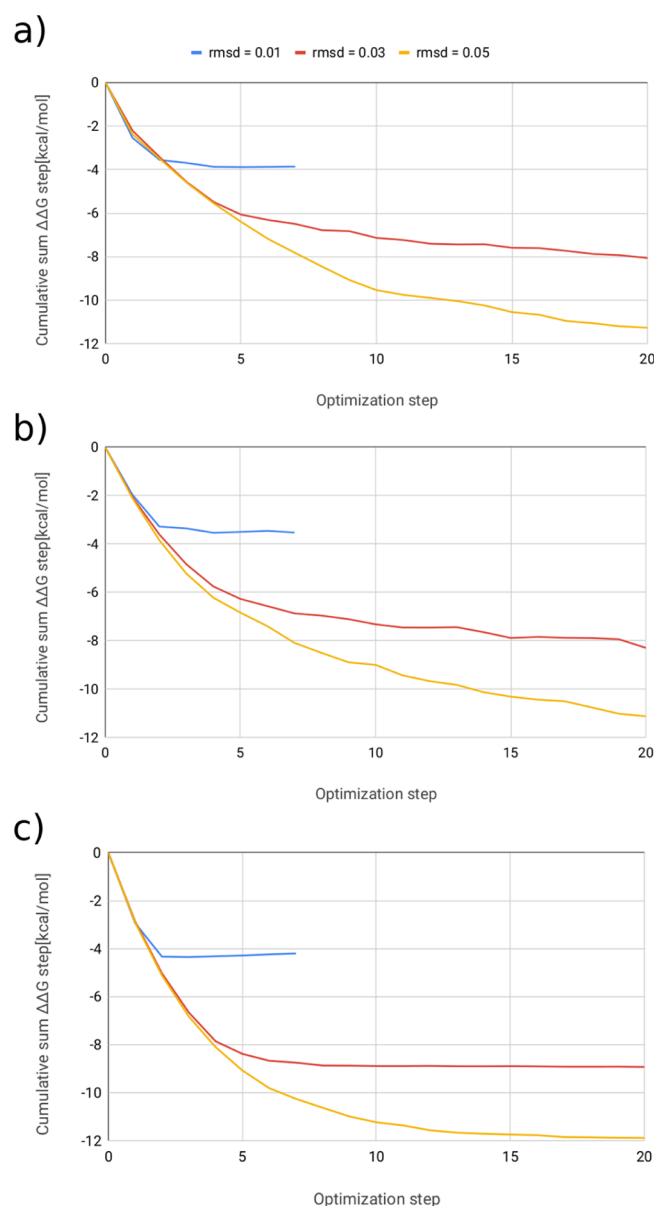
Figure 2. Cumulative sum of $\Delta\Delta G_{step}$ averaged over three replicates for each step of the optimizer. Three optimizations are shown with rmsd bounds of 0.01, 0.03, and 0.05e, with the FXa, P38, and androgen receptor systems labeled (a−c), respectively.

set of charges is dependent on the rmsd and $\Delta\Delta G_{step}$ is slow to converge for larger ligands such as those in the P38 and FXa test cases. $\Delta\Delta G_{total}$ can be calculated between the original and the optimized charges with the cumulative sum of the SSP $\Delta\Delta G_{step}$ for all optimization steps. This gives a $\Delta\Delta G_{total}$ that will be termed $\Delta\Delta G_{total}^{SSP}$. We compare this SSP $\Delta\Delta G_{total}$ for each set of optimized charges with full FEP calculations, see Table 3. These FEP calculations use the original and optimized charges as the two end states and the resulting $\Delta\Delta G_{total}$ will be termed $\Delta\Delta G_{total}^{FEP}$.

Table 3 shows that the SSP and FEP calculations well agree with an rmsd of 0.01e (differing by less than 1.0 kcal/mol in all cases). For an rmsd of 0.03 and 0.05e, SSP and FEP are less well agreed (differing by more than 1.0 kcal/mol in some cases). Table 3 also shows clearly that changing the rmsd changes the calculated $\Delta\Delta G_{total}$. The reason is that increasing the rmsd bound increases how much the charges can be

Table 3. Calculated $\Delta\Delta G_{total}$ for the Set of Optimal Charges[a]

FXa			
rmsd (e)	0.05	0.03	0.01
$\Delta\Delta G_{total}^{FEP}$ [kcal/mol]	−8.7, [−9.2, −8.2]	−6.3, [−6.5, −6.1]	−3.1, [−3.4, −2.9]
$\Delta\Delta G_{total}^{SSP}$ [kcal/mol]	−11.3, [−12.4, −10.1]	−8.1, [−9.1, −7.0]	−3.9, [−4.4, −3.3]

P38			
rmsd (e)	0.05	0.03	0.01
$\Delta\Delta G_{total}^{FEP}$ [kcal/mol]	−9.4, [−10.8, −8.0]	−6.6, [−7.2, −6.0]	−3.2, [−3.4, −3.1]
$\Delta\Delta G_{total}^{SSP}$ [kcal/mol]	−11.1, [−11.5, −10.8]	−8.3, [−8.7, −7.9]	−3.5, [−4.0, −3.1]

Androgen Receptor			
rmsd (e)	0.05	0.03	0.01
$\Delta\Delta G_{total}^{FEP}$ [kcal/mol]	−11.5, [−12.0, −10.9]	−8.8, [−8.8, −8.8]	−4.2, [−4.2, −4.1]
$\Delta\Delta G_{total}^{SSP}$ [kcal/mol]	−11.9, [−12.3, −11.5]	−8.9, [−9.2, −8.7]	−4.2, [−4.4, −4.0]

[a]SSP $\Delta\Delta G_{total}$ values are calculated by summing the average of forward and backward SSP calculations made for each step of the optimizer. FEP $\Delta\Delta G_{total}$ values are calculated from an alchemical transformation from the original charges to the optimal charges. SSP and FEP predictions are reported as the mean of three replicates with 95% confidence interval reported between square brackets computed as mean $\pm$ $t2 \cdot$SEM, where $t2$ is the t-distribution statistic with two degrees of freedom, and SEM is the standard error of the mean computed from the sample standard deviation of the three independent replicate predictions.

changed and so increases the change in $\Delta\Delta G_{total}$. However, as discussed above, the convergence of $\Delta\Delta G_{total}$ is an unnecessary condition, providing no additional information. It is only critical that the direction of the charge vectors are well converged and consistent for all rmsd values for all test cases, which has been shown in Figure 1 and Table 2, as it is this information that will inform what chemical mutations are proposed for the ligands. As such, we then use where the optimizer has placed the charge as a design tool to generate ideas for beneficial chemical mutations and this is presented in Figure 3.

We developed specific design ideas to improve $\Delta\Delta G_{binding}$ based on the changes in charge. First, analyzing the FXa ligand, three options are selected:

● Replacing the hydrogen with a fluorine at position 1a.
● Replacing the nitrogen with a carbon at position 2a.
● Replacing one or more hydrogen atoms with a fluorine atom on the methyl group at position 3a.

Analyzing the P38 ligand, four options are selected:

● Replacing the hydrogen with a fluorine at position 1b or 4b.
● Replacing the nitrogen with a carbon at position 1b or 4b.
● Replacing the oxygen with a sulfur at position 2b.
● Replacing one or more hydrogen atoms with a fluorine atom on the methyl group at position 3b.

The final set of changes applies to the ligand of the androgen receptor with three options selected:

● Replacing the oxygen with a sulfur at position 1c and 2c.
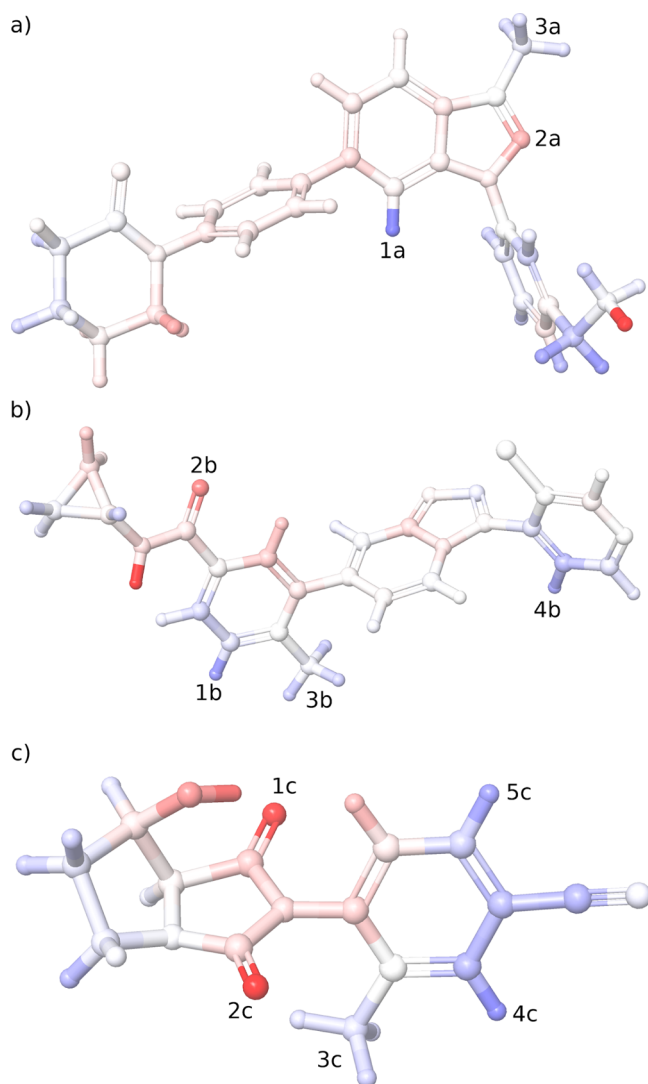
Figure 3. Panels (a−c) show the FXa, P38, and androgen receptor ligands with atoms colored by change in charge relative to the original partial charges. The optimized charge is taken from the optimization with an rmsd bound of 0.03e. Blue represents atoms which are more negative, and red represents atoms which are more positive. Selected sites for chemical modification are numbered.

- Replacing the hydrogen with a fluorine at position 3c, 4c, or 5c.
- Replacing the bonded carbon with a nitrogen at positions 4c or 5c.

$\Delta\Delta G_{\text{designed}}$ for all these changes was calculated using the FEP protocol discussed in the Methods section. Each FEP calculation was performed in triplicate, and the averaged results of these calculations can be seen in Table 4.

The atoms indicated by the optimization to beneficially be more negative in Figure 3 line up with experimental work on these test cases.[40−42] Mutants 1, 6, and 19 are predicted by FEP to be beneficial (−2.2, −2.2, and −2.5 kcal/mol, respectively), and this is in good agreement with experimental data (−2.1, −2.3, and −1.1 kcal/mol, respectively). Experimental data do not exist for the remaining proposed mutations. However, 73% of the mutations in Table 4 are predicted to be favorable by FEP. Both the FXa and androgen systems have a higher success rate with 80 and 89% of ideas, respectively, from charge optimization being beneficial as

assessed by FEP. P38 has a lower (though still promising) success rate with 50% of mutations being beneficial as assessed by FEP.

■ CONCLUSIONS

We have demonstrated ligand charge optimization in explicit solvent to be a useful tool to rationally design ligands with improved binding affinities. The electrostatics of three ligand–receptor systems were systematically optimized using the alchemical SSP method, yielding sets of optimal ligand charges. These sets of optimal charges were used to generate design principles for chemical mutations to the ligands that would yield improved receptor binding affinity. These chemical mutations were assessed with a more rigorous FEP method. Using FEP, 73% of the predicted chemical mutations were found to be beneficial. The average improvement of the beneficial mutations was approximately 1 kcal/mol. In three of these cases, experimental data exist and are in excellent agreement with calculations, with mutants 1, 6, and 19 in Table 4 predicted by FEP to be beneficial (−2.2, −2.2, and −2.5 kcal/mol, respectively) compared to the experimental data (−2.1, −2.3, and −1.1 kcal/mol, respectively).

The major advantage of SSP shown in this work is the calculation of the gradient. SSP allows for many highly related mutations to be assessed quickly, as is required to calculate the gradient via a finite difference method. For comparison, the collection of 2.5 ns of sampling for the FXa system with 99 000 and 13 000 atoms in the complex and solvent systems, respectively, takes 29 min. To calculate a gradient from this sampling takes 15 min and so, including sampling, this totals to 44 min per gradient. The calculation of a perturbation of 0.00015e with full FEP (assuming 1.0 ns of sampling converges $\Delta\Delta G_{\text{binding}}$, see Figure S16 for convergence plot) takes 14 min. With 58 charges in the FXa ligand, which must all be perturbed in the complex and solvent systems, this gives approximately 27 h to calculate one gradient. This advantage would only be compounded if a more complex optimization scheme, which required a calculation of the Hessian, was used. Both these calculations of the gradient are run in parallel (see the Supporting Information for parallelization strategy) across 4 NVIDIA P100 GPUs using OpenMM 7.3.0 and CUDA 10.0.

This ligand charge optimization methodology could be easily extended by considering the optimization of any other parameters of the force field with respect to the binding free energy. For example, the van der Waals parameters could be "optimized". Additionally, the calculation for the gradient of force field parameters with respect to potential energies could be used in the systematic optimization of a small molecule force field. Because here we have demonstrated a methodology for quickly calculating gradients of force field parameters with respect to free energy, this could lead to some interesting studies of force field optimization using experimental ligand–receptor binding energies as a target data set. This method is relatively unique in providing drug design principles from alchemical free-energy calculations along with a rationale for increase or decrease in binding because of specific changes to the ligand.[54]

In summary, we have presented a novel technique for identifying partial charges that optimize protein–ligand binding affinities and highlighted ways in which these predictions can be turned into design principles for drug discovery. The method is fast compared to traditional FEP,

Table 4. $\Delta\Delta G_{\text{designed}}$ for Proposed Chemical Mutations to the FXa, P38, and Androgen Receptor Ligands Calculated with FEP[a]

	FXa		
Position and mutation		**Mutant**	**FEP $\Delta\Delta G_{designed}$ [kcal/mol]**
		1.	
(1a) Hydrogen to fluorine			-2.2 [-2.3, -2.1]‡
		2.	
(2a) Nitrogen to carbon			3.0 [1.9, 4.1]
		3.	
(3a) Hydrogen to fluorine			-0.1 [-0.2, 0.0]*
		4.	
(3a) Double hydrogen to fluorine			-0.6 [-0.7, -0.5]*
		5.	
(3a) Triple hydrogen to fluorine			-0.8 [-1.2, -0.5]

	P38		
Position		**Mutant**	**FEP $\Delta\Delta G_{designed}$ [kcal/mol]**
		6.	
(1b) Hydrogen to fluorine			-2.2 [-2.7, -1.6]‡
		7.	
(1b) Carbon to nitrogen			2.0 [1.2, 2.7]
		8..	
(2b) Oxygen to sulphur			0.0 [-2.7, 2.7]†
		9.	
(3b) Hydrogen to fluorine			0.3 [0.1, 0.4]*
		10.	
(3b) Double hydrogen to fluorine			-0.5 [-0.9, -0.1]*
		11.	
(3b) Triple hydrogen to fluorine			1.0 [-0.1, 2.1]

	P38		
Position		**Mutant**	**FEP $\Delta\Delta G_{designed}$ [kcal/mol]**
		12.	
(4b) Hydrogen to fluorine			-1.6 [-1.7, -1.4]‡
		13.	
(4b) Carbon to nitrogen			-0.4 [-1.7, 1.0]

	Androgen Receptor		
Position		**Mutant**	**FEP $\Delta\Delta G_{designed}$ [kcal/mol]**
		14.	
(1c) Oxygen to sulphur			-2.1 [-3.3, -0.9]
		15.	
(2c) Oxygen to sulphur			-2.0 [-3.2, -0.8]
		16.	
(3c) Hydrogen to fluorine			-0.6 [-0.8, -0.4]*
		17.	
(3c) Double hydrogen to fluorine			-1.6 [-1.8, -1.4]*
		18.	
(3c) Triple hydrogen to fluorine			-0.1 [-0.9, 1.0]
		19.	
(4c) Hydrogen to fluorine			-2.5 [-2.8, -2.1]‡
		20.	
(4c) Carbon to nitrogen			-0.9 [-1.4, -0.5]
		21.	
(5c) Hydrogen to fluorine			-0.3 [-0.3, -0.2]‡
		22.	
(5c) Carbon to nitrogen			1.1 [0.0, 2.2]

[a]The positions denoted numerically correspond to numerical positions in Figure 3. 2D structures of mutation are presented in the column-labeled mutant. FEP predictions are reported as the mean value of three replicates with 95% confidence interval reported between square brackets computed as mean $\pm$ $t2\cdot$SEM, where $t2$ is the t-distribution statistic with two degrees of freedom, and SEM is the standard error of the mean computed from the sample standard deviation of the three independent replicate predictions. The asterisk label * indicates single or double

Table 4. continued

fluorinations of a methyl group. These are averaged over every hydrogen or pair of hydrogen in the methyl group and as such, these data represent the averaging of nine replicates with the confidence interval reported such that t2 is now the t-distribution statistic with eight degrees of freedom. The obelisk label † denotes calculations that were slow to converge and run with 24 lambda windows of 2 ns. The diesis ‡ label denotes data taken from previous work of the authors.[23]

easy to interpret, and simple to test by using more thorough approaches such as MBAR.

■ ASSOCIATED CONTENT

ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.9b00976.

> 3D structures of ligands with all atoms named; superposition of ligands in PDBs and manually built ligands used in this work; original charges of ligands used in this work; optimized charges for all optimizations; ligands colored by all sets of optimized charges; convergence plot for $\Delta\Delta G$ with sampling time calculated using SSP; convergence plot for ΔG with box size calculated using SSP; convergence plot for $\Delta\Delta G$ with sampling time calculated using full FEP; FXa ligand colored by a set of optimized charges from an optimization with no rmsd constraint; $\Delta\Delta G$ plotted against step from an optimization for FXa with no rmsd constraint; dot product of the normalized optimized charges with the normalized original charges showing variation of charge vector direction with step for FXa with no rmsd constraint; and discussion of parallelization strategy for all calculations (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: dhuggins@tritdi.org.

ORCID ⓞ

Alexander D. Wade: 0000-0003-1500-3733

David J. Huggins: 0000-0003-1579-2496

Notes

The authors declare the following competing financial interest(s): D.H. is a founder and shareholder of Integrated Biomedical Solutions Ltd.

■ REFERENCES

(1) 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.

(2) Williams-Noonan, B. J.; Yuriev, E.; Chalmers, D. K. Free Energy Methods in Drug Design: Prospects of "alchemical Perturbation" in Medicinal Chemistry: Miniperspective. *J. Med. Chem.* **2018**, *61*, 638−649.

(3) Seeliger, D.; de Groot, B. L. Protein Thermostability Calculations Using Alchemical Free Energy Simulations. *Biophys. J.* **2010**, *98*, 2309−2316.

(4) Loeffler, H. H.; Michel, J.; Woods, C. FESetup: Automating Setup for Alchemical Free Energy Simulations. *J. Chem. Inf. Model.* **2015**, *55*, 2485−2490.

(5) 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.

(6) Cole, D. J.; Tirado-Rives, J.; Jorgensen, W. L. Molecular Dynamics and Monte Carlo Simulations for Protein−ligand Binding and Inhibitor Design. *Biochim. Biophys. Acta, Gen. Subj.* **2015**, *1850*, 966−971.

(7) Shirts, M. R.; Chodera, J. D. Statistically Optimal Analysis of Samples from Multiple Equilibrium States. *J. Chem. Phys.* **2008**, *129*, 124105.

(8) Roos, K.; Wu, C.; Damm, W.; Reboul, M.; Stevenson, J. M.; Lu, C.; Dahlgren, M. K.; Mondal, S.; Chen, W.; Wang, L.; Abel, R.; Friesner, R. A.; Harder, E. D. OPLS3e: Extending Force Field Coverage for Drug-Like Small Molecules. *J. Chem. Theory Comput.* **2019**, *15*, 1863−1874.

(9) 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.

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

(11) Abel, R.; Wang, L.; Harder, E. D.; Berne, B. J.; Friesner, R. A. Advancing Drug Discovery through Enhanced Free Energy Calculations. *Acc. Chem. Res.* **2017**, *50*, 1625−1632.

(12) 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.

(13) Lybrand, T. P.; McCammon, J. A.; Wipff, G. Theoretical Calculation of Relative Binding Affinity in Host-Guest Systems. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 833−835.

(14) Cole, D. J.; Janecek, M.; Stokes, J. E.; Rossmann, M.; Faver, J. C.; McKenzie, G. J.; Venkitaraman, A. R.; Hyvönen, M.; Spring, D. R.; Huggins, D. J.; Jorgensen, W. L. Computationally-Guided Optimization of Small-Molecule Inhibitors of the Aurora A kinase−TPX2 Protein−protein Interaction. *Chem. Commun.* **2017**, *53*, 9372−9375.

(15) Shirts, M. R.; Pande, V. S. Comparison of Efficiency and Bias of Free Energies Computed by Exponential Averaging, the Bennett Acceptance Ratio, and Thermodynamic Integration. *J. Chem. Phys.* **2005**, *122*, 144107.

(16) Liu, H.; Mark, A. E.; van Gunsteren, W. F. Estimating the Relative Free Energy of Different Molecular States with Respect to a Single Reference State. *J. Phys. Chem.* **1996**, *100*, 9485−9494.

(17) Mordasini, T. Z.; McCammon, J. A. Calculations of Relative Hydration Free Energies: A Comparative Study Using Thermody-

namic Integration and an Extrapolation Method Based on a Single Reference State. *J. Phys. Chem. B* **2000**, *104*, 360−367.

(18) Raman, E. P.; Vanommeslaeghe, K.; MacKerell, A. D.; MacKerell, A. D. Site-Specific Fragment Identification Guided by Single-Step Free Energy Perturbation Calculations. *J. Chem. Theory Comput.* **2012**, *8*, 3513−3525.

(19) Stroet, M.; Koziara, K. B.; Malde, A. K.; Mark, A. E. Optimization of Empirical Force Fields by Parameter Space Mapping: A Single-Step Perturbation Approach. *J. Chem. Theory Comput.* **2017**, *13*, 6201−6212.

(20) Oostenbrink, C.; Van Gunsteren, W. F. Single-Step Perturbations to Calculate Free Energy Differences from Unphysical Reference States: Limits on Size, Flexibility, and Character. *J. Comput. Chem.* **2003**, *24*, 1730−1739.

(21) Oostenbrink, C.; van Gunsteren, W. F. Free Energies of Ligand Binding for Structurally Diverse Compounds. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6750−6754.

(22) Raman, E. P.; Lakkaraju, S. K.; Denny, R. A.; MacKerell, A. D.; MacKerell, A. D. 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.

(23) Wade, A. D.; Rizzi, A.; Wang, Y.; Huggins, D. J. Computational Fluorine Scanning Using Free-Energy Perturbation. *J. Chem. Inf. Model.* **2019**, *59*, 2776−2784.

(24) Lee, L.-P.; Tidor, B. Optimization of Electrostatic Binding Free Energy. *J. Chem. Phys.* **1997**, *106*, 8681−8690.

(25) Kangas, E.; Tidor, B. Charge Optimization Leads to Favorable Electrostatic Binding Free Energy. *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.* **1999**, *59*, 5958−5961.

(26) Kangas, E.; Tidor, B. Electrostatic Specificity in Molecular Ligand Design. *J. Chem. Phys.* **2000**, *112*, 9120−9131.

(27) Lee, L.-P.; Tidor, B. Optimization of Binding Electrostatics: Charge Complementarity in the Barnase-barstar Protein Complex. *Protein Sci.* **2001**, *10*, 362.

(28) Bardhan, J. P.; Lee, J. H.; Kuo, S. S.; Altman, M. D.; Tidor, B.; White, J. K. *Fast Methods for Biomolecule Charge Optimization*; Modeling and Simulation of Microsystems (Nanotech), 2003.

(29) Kangas, E.; Tidor, B. Electrostatic Complementarity at Ligand Binding Sites: Application to Chorismate Mutase. *J. Phys. Chem. B* **2001**, *105*, 880−888.

(30) Sims, P. A.; Wong, C. F.; McCammon, J. A. Charge Optimization of the Interface between Protein Kinases and Their Ligands. *J. Comput. Chem.* **2004**, *25*, 1416−1429.

(31) Sulea, T.; Purisima, E. O. Optimizing Ligand Charges for Maximum Binding Affinity. A Solvated Interaction Energy Approach. *J. Phys. Chem. B* **2001**, *105*, 889−899.

(32) Armstrong, K. A.; Tidor, B.; Cheng, A. C. Optimal Charges in Lead Progression: A Structure-Based Neuraminidase Case Study. *J. Med. Chem.* **2006**, *49*, 2470−2477.

(33) Shen, Y.; Gilson, M. K.; Tidor, B. Charge Optimization Theory for Induced-Fit Ligands. *J. Chem. Theory Comput.* **2012**, *8*, 4580−4592.

(34) Zhang, J.; Zhang, H.; Wu, T.; Wang, Q.; van der Spoel, D. Comparison of Implicit and Explicit Solvent Models for the Calculation of Solvation Free Energy in Organic Solvents. *J. Chem. Theory Comput.* **2017**, *13*, 1034−1043.

(35) Harris, R. C.; Pettitt, B. M. Examining the Assumptions Underlying Continuum-Solvent Models. *J. Chem. Theory Comput.* **2015**, *11*, 4593−4600.

(36) Zhang, H.; Tan, T.; van der Spoel, D. Generalized Born and Explicit Solvent Models for Free Energy Calculations in Organic Solvents: Cyclodextrin Dimerization. *J. Chem. Theory Comput.* **2015**, *11*, 5103−5113.

(37) Zhang, H.; Yin, C.; Yan, H.; van der Spoel, D. Evaluation of Generalized Born Models for Large Scale Affinity Prediction of Cyclodextrin Host−Guest Complexes. *J. Chem. Inf. Model.* **2016**, *56*, 2080−2092.

(38) 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.

(39) Berman, H. M.; Westbrook, J.; Feng, Z. The Protein Data Bank. *Nucleic Acids* **2000**, *28*, 235−242.

(40) Lee, Y.-K.; Parks, D. J.; Lu, T.; Thieu, T. V.; Markotan, T.; Pan, W.; McComsey, D. F.; Milkiewicz, K. L.; Crysler, C. S.; Ninan, N.; Abad, M. C.; Giardino, E. C.; Maryanoff, B. E.; Damiano, B. P.; Player, M. R. 7-Fluoroindazoles as Potent and Selective Inhibitors of Factor Xa†. *J. Med. Chem.* **2008**, *51*, 282−297.

(41) Aiguadé, J.; Balagué, C.; Carranco, I.; Caturla, F.; Domínguez, M.; Eastwood, P.; Esteve, C.; González, J.; Lumeras, W.; Orellana, A.; Preciado, S.; Roca, R.; Vidal, L.; Vidal, B. Novel Triazolopyridylben-zamides as Potent and Selective p38α Inhibitors. *Bioorg. Med. Chem. Lett.* **2012**, *22*, 3431−3436.

(42) Hamann, L. G.; et al. Tandem Optimization of Target Activity and Elimination of Mutagenic Potential in a Potent Series of N-Aryl Bicyclic Hydantoin-Based Selective Androgen Receptor Modulators. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 1860−1864.

(43) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L.-P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D.; Wiewiora, R. P.; Brooks, B. R.; Pande, V. S. OpenMM 7: Rapid Development of High Performance Algorithms for Molecular Dynamics. *PLoS Comput. Biol.* **2017**, *13*, No. e1005659.

(44) Sastry, G. M.; 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.

(45) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Antechamber: An Accessory Software Package for Molecular Mechanical Calculations. *J. Chem. Inf. Comput. Sci.* **2001**, *222*, U403.

(46) 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.

(47) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, Efficient Generation of High-quality Atomic Charges. AM1-BCC Model: II. Parameter-ization and Validation. *J. Comput. Chem.* **2002**, *23*, 1623−1641.

(48) Rizzi, A.; Chodera, J.; Naden, L.; Beauchamp, K.; Albanese, S.; Grinaway, P.; Rustenburg, B.; Saladi, S.; Boehm, K. *Choderalab/yank*: Bugfix Release, 2018.

(49) Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. Development of an Improved Four-Site Water Model for Biomolecular Simulations: TIP4P-Ew. *J. Chem. Phys.* **2004**, *120*, 9665−9678.

(50) Oliphant, T. E. Python for Scientific Computing. *Comput. Sci. Eng.* **2007**, *9*, 10−20.

(51) Kraft, D. A Software Package for Sequential Quadratic Programming. Forschungsbericht-; Deutsche Forschungs- und Versuchsanstalt fur Luft- und Raumfahrt, 1988.

(52) Reif, M. M.; Hünenberger, P. H. Computation of Method-ology-Independent Single-Ion Solvation Properties from Molecular Simulations. III. Correction Terms for the Solvation Free Energies, Enthalpies, Entropies, Heat Capacities, Volumes, Compressibilities, and Expansivities of Solvated Ion. *J. Chem. Phys.* **2011**, *134*, 144103.

(53) Allen, F. H.; Kennard, O.; Watson, D. G.; Brammer, L.; Orpen, A. G.; Taylor, R. Tables of Bond Lengths Determined by X-Ray and Neutron Diffraction. Part 1. Bond Lengths in Organic Compounds. *J. Chem. Soc., Perkin Trans. 2* **1987**, S1−S19.

(54) Irwin, B. W. J.; Huggins, D. J. Estimating Atomic Contributions to Hydration and Binding Using Free Energy Perturbation. *J. Chem. Theory Comput.* **2018**, *14*, 3218−3227.