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

Alexander Wade, David Huggins

Submitted date: 08/07/2019 • Posted date: 09/07/2019
Licence: CC BY-NC-ND 4.0

Citation information: Wade, Alexander; Huggins, David (2019): Optimization of Protein-Ligand Electrostatic Interactions Using an Alchemical Free-Energy Method. ChemRxiv. Preprint.

We present an 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 androgen receptor. The optimization of the ligand charges yielded improvements to binding affinity for all three systems. The sets of optimized charges can be used to identify design principles for chemical changes to the ligand which improve the binding affinity. 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 of approximately 1 kcal/mol. 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 sulphur mutations.

File list (2)

| File Name                                      | Size       | Download                     | View               |
|-----------------------------------------------|------------|------------------------------|--------------------|
| 2019.06.19_ChargeOpt_Paper.pdf                | 1.05 MiB   | download file                | view on ChemRxiv   |
| 2019.06.19_ChargeOpt_Supporting Information.pdf | 573.78 KiB | view on ChemRxiv             | download file      |
Optimization of Protein-Ligand Electrostatic Interactions Using An Alchemical Free-Energy Method

Alexander D. Wade\textsuperscript{a}, David J. Huggins \textsuperscript{a, b, c}

Affiliations: \textsuperscript{a}TCM Group, Cavendish Laboratory, University of Cambridge, 19 J J Thomson Avenue, Cambridge CB3 0HE, United Kingdom
\textsuperscript{b}Tri-Institutional Therapeutics Discovery Institute, Belfer Research Building, 413 East 69th Street, 16th Floor, Box 300, New York, USA
\textsuperscript{c}Department of Physiology and Biophysics, Weill Cornell Medicine, 1300 York Avenue, New York, NY, 10065, USA.

Corresponding Author: David J. Huggins
Email: dhuggins@tritdi.org
Updated: 20th June 2019
ORCIDs: Alex Wade - 0000-0003-1500-3733
David Huggins 0000-0003-1579-2496
Abstract
We present an 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 androgen receptor. The optimization of the ligand charges yielded improvements to binding affinity for all three systems. The sets of optimized charges can be used to identify design principles for chemical changes to the ligand which improve the binding affinity. 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 of approximately 1 kcal/mol. 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 sulphur mutations.
Introduction

In recent years, alchemical methods have garnered increasing attention in drug design. In particular, free energy perturbation (FEP) is now commonly used by pharmaceutical companies due to improvements in efficiency, more accurate force fields and increases in computational power. Based on the Zwanzig equation, it is common to use FEP calculations in drug design to calculate the relative binding affinity of two molecules. A drawback of FEP is that for every relative binding affinity calculation, lengthy molecular dynamics simulations must be performed for that pair of molecules and any alchemical intermediates. A more efficient alternative is to calculate binding affinities for a range of analogues that are all similar to one reference molecule using the method of single step perturbation (SSP). SSP only requires that dynamics are collected for the central reference molecule. Numerous studies have used SSP demonstrating that it is applicable to relative free energy calculations and can be significantly faster than standard FEP. Most recently, the authors of this paper have used SSP to perform computational fluorine scanning. We now apply SSP to the task of optimizing ligand partial charges.

Charge optimization was developed by Tidor and co-workers using an implicit water treatment of electrostatics. Poisson-Boltzmann calculations are performed on the bound and unbound states in order to find the optimal partial charges of a given molecule. This approach has since been used by other academic groups, employed in industry, and been extended to consider induced fit effects. However, the approach suffers from the deficiency 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. Due to 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.

Combining SSP with explicit water molecular dynamics calculations 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 as both a molecular dynamics engine and 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 Figure (1). The ligands were built from highly related molecules in the Protein Databank: 2RA0 (FXa), 3S3I (P38), and 2NW4 (androgen receptor). Superpositions of the molecules in the original PDB and the modified molecules in Figure (1) are shown in Figures (S1-S3).
Figure (1): 3D Chemical structures depicting the wild type ligands in each system. (a) FXa, (b) P38, (c) androgen receptor. With labels showing all atom names. Note the binding pose of FXa has been altered to better show all atoms.

System Setup

The co-crystal structure for the three systems examined here are taken from the Protein Data Bank FXa (PDB ID 2RA0), P38 (PDBID 3S3I) and androgen receptor (PDBID 2NW4). To prepare these systems non-standard residues were converted to their standard equivalents with pdbfixer. Selenomethionines were changed to methionines and missing sidechains were added using Schrödinger’s Preparation Wizard, which was also used to assign protonation state of all ionizable residues. All buffer solvents and ions were removed. The hydrogen atom positions were then built using tleap and forcefield parameters and partial charges were assigned from the AMBER ff14SB force field. Parameters for the inhibitors were generated using Antechamber with AMBER GAFF 2 and AM1-BCC. These structures and parameters are then passed to YANK’s 0.23.7 automatic setup pipeline to build solvated ligand-protein and ligand systems. For solvation, TIP4P-EW is used. A salt concentration of 150mM and any required counter-ions are added using sodium and chloride ions. In every case, the edge of the solvation box is set to be 15 Å from any atom of the receptor and ligand.
Molecular dynamics

All simulations were performed with OpenMM 7.3.0. As follows. First, OpenMM’s default minimizer was used to minimize all structures. Then equilibration was performed in the NPT ensemble for 500 ps at 300 k and 1 atm using a Langevin integrator and Monte Carlo barostat. 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 particle mesh Ewald method with a cutoff of 11.0 Å. All other simulation parameters were left as default. We ran tripllicate simulations of all compounds with the ligand in complex and in solution for 50 ns. Snapshots were collected every 5 ps. All simulations are performed with constrained hydrogens.

Charge optimization

The objective function of this optimization is \( \Delta \Delta G_{binding} \). This is calculated by combining \( \Delta G_{mut->wt} \) from the bound and unbound states. These \( \Delta G \)s in the bound and unbound states are calculated using the Zwanzig equation as shown Equation (1).

\[
\Delta G_{wt->mut} = -kT \ln (< \exp(- (E_{mut} - E_{wt})/kT) >_{wt}) \tag{1}
\]

\( E_{wt} \) and \( E_{mut} \) are the energies of the system using the Hamiltonians, \( H \), of the “wild type” unperturbed system (wt) and the “mutant” perturbed system (mut). The mut Hamiltonian is identical to the wt hamiltonian with the only difference being that the wild type charges have been swapped for the mutant charges. Lennard-Jones, bonded, angle and torsion parameters do not change. The wt subscript on the average of the exponential indicates that it is taken for samples in the wild type ensemble. We then construct an objective function as in Equation (2).

\[
\Delta \Delta G_{binding} = -kT \ln (< \exp(- (E_{mut}^{bound} - E_{wt}^{bound})/kT) >_{wt}) + kT \ln (< \exp(- (E_{mut}^{unbound} - E_{wt}^{unbound})/kT) >_{wt}) \tag{2}
\]

The optimization involves identifying the set of mutant charges that give \( E_{mut}^{bound} \) and \( E_{mut}^{unbound} \) to minimise \( \Delta \Delta G_{binding} \). Importantly, the sum of charges is constrained to remain constant,
preventing the net charge of the ligand from changing. To avoid poorly converged calculations arising from insufficient overlap between the end states of the perturbation, we perform the optimization in small steps. To maximize this overlap, wt charges and the wt ensemble refer to the charges and ensemble of the previous optimization step (not the original charges and ensemble). More explicitly, after every optimization step we set the wt charges as the mutant charges of the previous optimization step, resample the dynamics with the new wt charges then move to the next optimization step. Conveniently, sampling the new wt ensemble is the same as sampling the old mutant ensemble and this allows the calculation of the backwards alchemical step at very little additional computational cost. Therefore, \( \Delta \Delta G_{\text{binding}} \) for the backwards alchemical transformation is calculated for every step and the \( \Delta \Delta G_{\text{binding}} \) is reported as an average of the forwards and backwards transformations. The total change in binding free energy going from the original wild type ligand to the optimized ligand is then the sum of \( \Delta \Delta G_{\text{binding}} \) for every step of the optimization.

To ensure good overlap between the wt and mut ensembles the change in charge per atom per iteration is bounded to be +/- 0.01 \( q_e \). In an effort to avoid unphysical predictions, the total change per atom is constrained to be +/- 0.50 \( q_e \) across the whole optimization. To limit \( \Delta \Delta G_{\text{binding}} \) to a sensible range, a constraint is added to the RMSD between original and optimized charge sets. To assess the influence of this on the results we perform three optimizations, limiting the RMSD of the charges to be 0.01, 0.03 or 0.05 \( q_e \). The optimization is performed using Scipy’s 1.1.0 \(^{45} \) implementation of the Sequential Least Squares Programming algorithm. For this, the optimizer needs access to an objective function and the gradient of the change in \( \Delta \Delta G_{\text{binding}} \) with respect to every dimension. In this case, the dimensions are the charge of all the ligand atoms. The calculation of the objective has been discussed above and the gradient is calculated as follows.

To construct the gradient of \( \Delta \Delta G_{\text{binding}} \) with respect to the charges, SSP is used in combination with a finite difference method, \( Q \) is a vector of all \( i \) charges in the ligand. We construct \( i \) Hamiltonians \( H(q_i) \) with \( q_i \) shown in Equation (3).

\[
q_i = Q + e_i h
\]  

Equation (3)
Where $h$ is the finite difference. In this work, we take $h = 0.00015 q_e$ and $e_i$ is the $i$th unit vector. We then have as many Hamiltonians as charges in the ligand. Evaluating the energy for all $i$ mutant Hamiltonian gives $i$ values of $E_{\text{mut}}$ for the bound and unbound systems. $E_{\text{wt}}$ can be evaluated from $H(Q)$. Following Equation (2), combining $E_{\text{wt}}$ with the $i$ values of $E_{\text{mut}}$ will give $i$ free energy differences which are the required gradients. Here, the advantage of SSP combined with finite difference is clear, as numerous (10s-100s) of evaluations of $\Delta \Delta G$ are required for molecules which are extremely similar, differing only by 0.00015 $q_e$ in one atom’s partial charge. This gradient calculation would be intractable with full FEP. Of note is that for each finite difference calculation the charge of the simulation box has been changed by 0.00015 $q_e$. The potential for finite size effects caused by this change will be investigated in the results section.

**FEP Calculations**

To validate the optimization result, we compare it against standard alchemical relative binding free energy calculations using the MBAR estimator. We used a total of 12 equally spaced lambda windows (except when explicitly stated otherwise) in which the charge parameters were interpolated simultaneously from the wild-type to the mutated state. All windows were sampled independently with 2 ns of Langevin dynamics. All simulation conditions were identical to the optimization molecular dynamics calculations described above. The samples collected in each intermediate state were decorrelated based on an estimate of the statistical inefficiency of the reduced potentials time series before carrying out the MBAR analysis with the PyMBAR 3.0.1. This FEP protocol is run automatically as part of the Ligand Charge Optimiser package in order to check the $\Delta \Delta G_{\text{binding}}$ for final set of optimized charges.

In addition to testing the $\Delta \Delta G_{\text{binding}}$ of the optimal set of charges, we also wish to calculate the $\Delta \Delta G_{\text{binding}}$ for a set of chemical changes informed by the optimal charges. To perform these calculations, the protocol is the same as above but with the following additional considerations. Now, in addition to the charges, the van der Waals and bonded terms are all interpolated simultaneously from the wild-type to the mutated state. In the case of hydrogen to fluorine mutations the original hydrogen is constrained, therefore its associated C-H bond cannot be interpolated to a C-F bond. Without interpolating, this bond the fluorine would appear at the position of the hydrogen, instead of the true physical position of the fluorine. To avoid this issue for fluorinations we use a hybrid topology approach where a massless interaction sites at
the position of the mutation is added. This virtual site represents the fluorine and its position is defined relative to the position of the parent hydrogen such that the C-F distance is always 1.24 times the C-H distance \(^{47}\). When mutating a hydrogen to fluorine, the relevant hydrogen is turned off and fluorine Lennard-Jones and charge parameters are applied to the additional site. When simulating these systems, all bonds to hydrogen are constrained. Since the position of the fluorine is defined relative to the position of its parent hydrogen it is also implicitly constrained. We therefore make the assumption that the C-F bond length oscillations are negligible. To prevent the hybrid topologies from interacting, the additional sites are excluded from interacting with their parent hydrogens.
Results

We first analyzed the convergence of the calculation of $\Delta \Delta G$ using SSP for a small perturbation ($0.01 \ q_e$). To perform this calculation, we use the test system FXa and we apply a perturbation of $+0.01 \ q_e$ to half the atoms and $-0.01 \ q_e$ to the other half, maintaining the simulations net charge. Calculations for the SSP $\Delta \Delta G$ of this mutant are performed in triplicate scanning across simulation length and presented in Figure (2).

Figure (2): Convergence of the $\Delta \Delta G_{\text{binding}}$ predictions in the Factor Xa test case for a perturbation of $+0.01 \ q_e$ to half the atoms and $-0.01 \ q_e$ to the other half (maintaining the net charge) as the simulation time is increased. Calculations were performed from at 0.01 ns and then from 0.05 ns to 2.5 ns in 0.05 ns increments. The values of $\Delta \Delta G_{\text{binding}}$ are reported as mean of three replicates with the shaded area showing the 95% confidence interval computed as mean $\pm t_2 \cdot \text{SEM}$, where $t_2$ 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.

Figure (2) shows that when the maximum step of the optimizer is bound to $0.01 \ q_e$ per atom per iteration, 2.5 ns of sampling provides satisfactorily converged $\Delta \Delta G_{\text{binding}}$ calculations. The optimiser is therefore limited to take a maximum step of $0.01 \ q_e$ per atom per iteration.
It was discussed in the methods that the charge perturbation of 0.00015 \( q_e \) used to calculate the gradient is unbalanced, in that no counter charge is added to keep the total charge of the system neutral. This may lead to finite size effects if the periodic images of these charges interact with each other. To investigate this we calculate the \( \Delta G \) values resulting from a perturbation of 0.00015 \( q_e \) whilst varying the size of the simulation box. See Figure (3).

**Figure (3):** \( \Delta \Delta G_{\text{binding}} \) values for a 0.00015 \( q_e \) perturbation to one atom against \( r \) (where \( r \) is the minimum padding of solvent added between the protein and edge of the box). \( \Delta G \) values are calculated using SSP and 2.5 ns of sampling. \( \Delta G \)s are reported as the mean of six replicates with shaded area showing 95% confidence interval computed as mean ± \( t_2 \cdot \text{SEM} \), where \( t_2 \) is the \( t \)-distribution statistic with five degrees of freedom, and \( \text{SEM} \) is the standard error of the mean computed from the sample standard deviation of the six independent replicate predictions.

Figure (3) shows that the calculation of \( \Delta G \) bound and unbound are not dependant on the size of the simulation box (for the size of the simulation box considered in this work). For each optimization we would also like to inspect the convergence over the number of steps. A good metric to analyse the results of the optimization is the set of optimized charges taken as a vector. In the methodology, mention was made to limiting the RMSD between original and optimized charges to be 0.01, 0.03 or 0.05 \( q_e \). The optimization is therefore repeated with the RMSD bound to these three values. With an RMSD bound of 0.01 \( q_e \), 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 0.05 $q_e$ the optimizer is limited to 20 steps. To assess convergence across simulation steps, we take the dot product of the normalized new charges with the normalized original set of charges for each step of the optimization (see Figures 4-6).

**Figure (4)**: Dot product of the normalized optimized charges with the normalized original charges for increasing optimization steps using the FXa test case. Results are shown for RMSD limits 0.01, 0.03 and 0.05 $q_e$. 
Figure (5): Dot product of the normalized optimized charges with the normalized original charges for increasing optimization steps using the p38 test case. Results are shown for RMSD limits 0.01, 0.03 and 0.05 $q_e$.

![Graph showing dot product with RMSD limits 0.01, 0.03, and 0.05](image)

Figure (6): Dot product of the normalized optimized charges with the normalized original charges for increasing optimization steps using the androgen receptor test case. Results are shown for RMSD limits 0.01, 0.03 and 0.05 $q_e$.

Figure (4-6) show 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 $\Delta\Delta G_{\text{binding}}$. 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 0.05 $q_e$ can be taken and the result of these projections can be seen in Table (1). Here we see that sets of charges for the same system are pointing in the same direction, it can therefore be said that it is only the value of charges which are strongly dependant on RMSD and not where the charges are applied. 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 eye if the atoms are colored by change in charge. Figures illustrating this are presented in the Figures (S4-S12) for all sets of optimized charge.
Table (1): 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 ($q_e$) | 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             |

In Table (1) we can see the dot product of the optimized charges from the optimization with an RMSD of 0.01 $q_e$ with themselves returns 1.00 as expected. The dot product of the vector of charges with RMSD = 0.01 $q_e$ with RMSD = 0.03 $q_e$ also returns 1.00 as these vectors are extremely similar in direction. The dot product of the vector of charges with RMSD = 0.01 $q_e$ with RMSD = 0.05 $q_e$ returns approximately 1.00 as these vectors are extremely similar in direction but not as close as 0.01 $q_e$ with 0.03 $q_e$. To summarize, each optimization for a system approximately added the charge in the same place it is only the magnitude of this charge that varies with RMSD. We can also look at the convergence of $\Delta\Delta G_{\text{binding}}$ with optimisation step and this can be seen in Figures (7-9).
**Figure (7):** \( \Delta \Delta G_{\text{binding}} \) averaged over three replicates for each step the optimizer in the FXa system. Three optimizations are shown with RMSD bound to 0.01, 0.03 and 0.05 \( q_e \).

**Figure (8):** \( \Delta \Delta G_{\text{binding}} \) averaged over three replicates for each step the optimizer in the p38 system. Three optimizations are shown with RMSD bound to 0.01, 0.03 and 0.05 \( q_e \).
**Figure (9):** $\Delta\Delta G_{\text{binding}}$ averaged over three replicates for each step the optimizer in the Androgen receptor system. Three optimizations are shown with RMSD bound to 0.01, 0.03 and 0.05 $q_e$.

With an RMSD of 0.01 $q_e$ these figures demonstrate that $\Delta\Delta G_{\text{binding}}$ is well converged for all systems. For a RMSDs of 0.03 or 0.05 $q_e$ the results are not converged with the exception of the androgen receptor system (Figure 9) which remains well converged. This suggests that $\Delta\Delta G_{\text{binding}}$ for the optimized set of charges is dependent on the RMSD and $\Delta\Delta G_{\text{binding}}$ is slow to converge for larger ligands such as those in the p38 and FXa test cases. $\Delta\Delta G_{\text{binding}}$ can be calculated between the original and the optimized charges by summing the $\Delta\Delta G_{\text{binding}}$ of each optimization step. This gives a $\Delta\Delta G_{\text{binding}}$ that will be termed as the “SSP total $\Delta\Delta G_{\text{binding}}$”. We compare this SSP total $\Delta\Delta G_{\text{binding}}$ for each set of optimized charges with three FEP calculations per set of charges, see Table (2). These FEP calculations use the original and optimized charges as the two end states and the resulting $\Delta\Delta G_{\text{binding}}$ will be termed as the “FEP total $\Delta\Delta G_{\text{binding}}$”

**Table (2):** Calculated $\Delta\Delta G_{\text{binding}}$ for the set of optimal charges. SSP $\Delta\Delta G_{\text{binding}}$ values are calculated by summing the average of forward and backwards SSP calculations made for each step of the optimizer. FEP $\Delta\Delta G_{\text{binding}}$ values are calculated from an alchemical transformation from the wild type 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-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.

|       | FXa       |
|-------|-----------|
|       | RMSD ($q_e$) | 0.05       | 0.03       | 0.01       |
| FEP Total $\Delta\Delta G_{\text{binding}}$ [kcal/mol] | -8.7       | -6.3       | -3.1       |
|       | [-9.2, -8.2] | [-6.5, -6.1] | [-3.4, -2.9] |
| SSP Total $\Delta\Delta G_{\text{binding}}$ | -11.3      | -8.1       | -3.9       |
|       | [-12.4, -10.1] | [-9.1, -7.0] | [-4.4, -3.3] |
Table (2) shows that the SSP and FEP calculations are well agreed with an RMSD of 0.01 $q_e$ (differing by less than 1.0 kcal/mol in all cases). For an RMSD of 0.03 and 0.05 $q_e$, SSP and FEP are less well agreed (differing by more than 1.0 kcal/mol in some cases). Table (2) also shows clearly that changing the RMSD changes the calculated $\Delta \Delta G_{\text{binding}}$. The relation here is that increasing the RMSD bound increases how much the charges can be changed and so increases the change in $\Delta \Delta G_{\text{binding}}$. However, as discussed above, the convergence of $\Delta \Delta G_{\text{binding}}$ 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 as been shown in Figures(4-6) and Table (1), as it is this information that will
inform what chemical mutations are proposed for the ligands. As such we then use where the optimiser has placed the charge (Figures (10-12)) as a design tool to generate ideas for beneficial chemical mutations.

**Figure (10):** The FXa ligand with atoms coloured by change in charge relative to the original partial charges. The optimised charge taken from the optimisation with RMSD bound to 0.03 $q_e$. Blue represents atoms which are more negative and red represents atoms which are more positive. Selected sites for chemical modification are highlighted and numbered.

**Figure (11):** The p38 ligand with atoms coloured by change in charge relative to the original partial charges. The optimised charge taken from the optimisation with RMSD bound to 0.03 $q_e$. Blue represents atoms which are more negative and red represents atoms which are more positive. Selected sites for chemical modification are highlighted and numbered.
$q_e$. Blue represents atoms which are more negative and red represents atoms which are more positive. Selected sites for chemical modification are highlighted and numbered.

**Figure (12):** Androgen receptor ligand with atoms coloured by change in charge relative to the original partial charges. The optimised charge taken from the optimisation with RMSD bound to 0.03 $q_e$. Blue represents atoms which are more negative and red represents atoms which are more positive. Selected sites for chemical modification are highlighted and numbered.

We developed specific design ideas to improve $\Delta \Delta G_{\text{binding}}$ based on the changes in charge. Analyzing first the Fxa ligand in Figure (10) we selected three options:

- Replacing the hydrogen with a fluorine at position (1).
- Replacing the nitrogen with a carbon at position (2).
- Replacing one or more of the hydrogens with a fluorine on the methyl group at position (3).

Turning to Figure (11), depicting the ligand in the P38 system:

- Replacing the hydrogen with a fluorine at positions (1) and (4)
- Replacing the nitrogen with a carbon at positions (1) and (4).
- Replacing the oxygen with a sulphur at position (2).
Replacing one or more of the hydrogens with a fluorine on the methyl group at position (3).

The final set of changes apply to the ligand of the androgen receptor in Figure (12):

- Replacing the oxygen with a sulphur at position (1)
- Replacing the hydrogen with a fluorine at positions (2), (3) or (4).
- Replacing the bonded carbon with a nitrogen at positions (3) or (4).

$\Delta \Delta G_{\text{binding}}$ for all of 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 (3).

**Table (3):** $\Delta \Delta G_{\text{binding}}$ for proposed chemical mutations to the FXa, P38 and androgen receptor ligands calculated with FEP. The positions denoted numerically corresponds to numerical positions in Figures (10-12). 2D structures of the 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 ± $t_2$·SEM, where $t_2$ 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 fluorinations of a methyl. These are averaged over every hydrogen or pair of hydrogen in the methyl and as such this data represents the averaging of nine replicates with the confidence interval reported such that $t_2$ 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 2ns, so twice the sampling. The diesis ‡ label denotes data taken from previous work of the authors.

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

P38
| No. | Type of Bond          | Image   | Values                |
|-----|-----------------------|---------|-----------------------|
| 6.  | Hydrogen to fluorine  | ![Image](image1) | -2.2 [-2.7, -1.6]‡    |
| 7.  | Carbon to nitrogen    | ![Image](image2) | 2.0 [1.2, 2.7]        |
| 8.  | Oxygen to sulphur     | ![Image](image3) | 0.0 [-2.7, 2.7]†      |
| 9.  | Hydrogen to fluorine  | ![Image](image4) | 0.3 [0.1, 0.4]*       |
| 10. | Double hydrogen to fluorine | ![Image](image5) | -0.5 [-0.9, -0.1]*    |
| Position | Mutant | FEP ΔΔG\textsubscript{binding} [kcal/mol] |
|----------|--------|------------------------------------------|
| (3) Triple hydrogen to fluorine | | 1.0 [-0.1, 2.1] |
| (4) Hydrogen to fluorine | | -1.6 [-1.7, -1.4]‡ |
| (4) Carbon to nitrogen | | -0.4 [-1.7, 1.0] |

**Androgen Receptor**

| Position | Mutant | FEP ΔΔG\textsubscript{binding} [kcal/mol] |
|----------|--------|------------------------------------------|
| (1) Oxygen to sulphur | | -2.1 [-3.3, -0.9] |
|   | Chemical Bond | No. | Value 1 | Value 2 | Value 3 |
|---|---------------|-----|---------|---------|---------|
| 1 | Oxygen to sulphur | 15. | -2.0    | [-3.2, -0.8] |
| 2 | Hydrogen to fluorine | 16. | -0.6    | [-0.8, -0.4]* |
| 2 | Double hydrogen to fluorine | 17. | -1.6    | [-1.8, -1.4]* |
| 2 | Triple hydrogen to fluorine | 18. | -0.1    | [-0.9, 1.0] |
| 3 | Hydrogen to fluorine | 19. | -2.5    | [-2.8, -2.1]‡ |
The atoms indicated by the optimization to beneficially be more negative in Figures (10-12) line up with experimental work on these test cases. Mutants 1, 6, and 19 are predicted by FEP to be beneficial (-2.2 kcal/mol, -2.2 kcal/mol, and -2.5 kcal/mol respectively) and this is in good agreement with experimental data (-2.1 kcal/mol, -2.3 kcal/mol, and -1.1 kcal/mol respectively). Experimental data does not exist for the remaining proposed mutations. However, 73% of the mutations in Table (6) are predicted to be favourable by FEP. Both the FXa and androgen systems have a higher success rate with 80% and 89% of ideas from charge optimization being beneficial as assessed by FEP respectively. 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 beneficial mutations was approximately 1 kcal/mol. In three of these cases, experimental data exists and is in excellent agreement with calculation.

When calculating the binding affinity objective function we are only sampling one end state and as such this a unidirectional FEP calculation. One concern raised about this methodology in the methods section is that the optimizer must be bound not to take large steps to prevent the overlap with the wild type ensemble from becoming too small. Arguably this issue could be avoided and larger steps in charge space could be taken by the optimizer if a full FEP was used to calculate the objective. There most likely exists an optimal ratio of computation per step and step size which allows the evaluation of the objective to remain accurate but reaches the optimal set of charges in the minimum amount of computation. No effort in this work has been made find this ratio nor full FEP trialed in the evaluation of the objective function. The major advantage of the unidirectional FEP calculation in the calculation of the gradient as SSP allow for many highly related mutations to be assessed quickly as is required to calculate the gradient via a finite difference method.

This ligand charge optimization methodology could be easily extended by considering the optimization of any other parameter of the forcefield 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. Since 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 increases or decreases in binding due to specific changes to the ligand.\textsuperscript{48}

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 relatively fast, easy to interpret, and simple to test by using more accurate FEP calculations.
Supporting Information

Figures (S1-S3) showing superposition of ligands in PDBs and manually built ligands used in this work. Table (S1) contains original charges of ligands used in this work. Table (S2) contains optimised charges for all optimisations. Figures(S4-S12) show ligands colored by all set of optimised charges.

Acknowledgments: The authors would like to thank Andrea Rizzi, Yuanqing Wang, and John Chodera for useful discussions. A.D.W. would like to acknowledge the EPSRC Centre for Doctoral Training in Computational Methods for Materials Science for funding under grant number EP/L015552/1. A.D.W would also like to acknowledge Michael Hutcheon and Nick Woods for enlightening discussions and suggestions that are reflected in this manuscript. All calculations were performed using the Darwin Supercomputer of the University of Cambridge High Performance Computing Service (http://www.hpc.cam.ac.uk/) and were funded by the EPSRC under grant EP/P020259/1.

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

(1) Wang, L.; Wu, Y.; Deng, Y.; Kim, B.; Pierce, L.; Krilov, G.; Lupyan, D.; Robinson, S.; Dahlgren, M. K.; Greenwood, J.; et al. 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. *Journal of the American Chemical Society*. 2015, pp 2695–2703. https://doi.org/10.1021/ja512751q.

(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.* 2017, 61 (3), 638–649.

(3) Seeliger, D.; de Groot, B. L. Protein Thermostability Calculations Using Alchemical Free Energy Simulations. *Biophys. J.* 2010, 98 (10), 2309–2316.

(4) Loeffler, H. H.; Michel, J.; Woods, C. FESetup: Automating Setup for Alchemical Free Energy Simulations. *J. Chem. Inf. Model.* 2015, 55 (12), 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 (12), 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. *Biochimica et Biophysica Acta (BBA) - General Subjects* 2015, 1850 (5), 966–971.

(7) Shirts, M. R.; Chodera, J. D. Statistically Optimal Analysis of Samples from Multiple Equilibrium States. *J. Chem. Phys.* 2008, 129 (12), 124105.

(8) Roos, K.; Wu, C.; Damm, W.; Reboul, M.; Stevenson, J. M.; Lu, C.; Dahlgren, M. K.; Mondal, S.; Chen, W.; Wang, L.; et al. OPLS3e: Extending Force Field Coverage for Drug-Like Small Molecules. *J. Chem. Theory Comput.* 2019. https://doi.org/10.1021/acs.jctc.8b01026.

(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 (8), 3696–3713.

(10) Zwanzig, R. W. Erratum : High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *The Journal of Chemical Physics*. 1954, pp 2099–2099. https://doi.org/10.1063/1.1740022.

(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 (7),
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 (2), 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 (4), 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.; et al. Computationally-Guided Optimization of Small-Molecule Inhibitors of the Aurora A kinase–TPX2 Protein–protein Interaction. *Chemical Communications*. 2017, pp 9372–9375. https://doi.org/10.1039/c7cc05379g.

(15) 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 (22), 9485–9494.

(16) Mordasini, T. Z.; Andrew McCammon, J. Calculations of Relative Hydration Free Energies: A Comparative Study Using Thermodynamic Integration and an Extrapolation Method Based on a Single Reference State. *The Journal of Physical Chemistry B*. 2000, pp 360–367. https://doi.org/10.1021/jp993102o.

(17) Raman, E. P.; Prabhu Raman, E.; Vanommeslaeghe, K.; MacKerell, A. D. Site-Specific Fragment Identification Guided by Single-Step Free Energy Perturbation Calculations. *Journal of Chemical Theory and Computation*. 2012, pp 3513–3525. https://doi.org/10.1021/ct300088r.

(18) 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. *Journal of Chemical Theory and Computation*. 2017, pp 6201–6212. https://doi.org/10.1021/acs.jctc.7b00800.

(19) Oostenbrink, C.; Van Gunsteren, W. F. Single-Step Perturbations to Calculate Free Energy Differences from Unphysical Reference States: Limits on Size, Flexibility, and Character. *Journal of Computational Chemistry*. 2003, pp 1730–1739. https://doi.org/10.1002/jcc.10304.

(20) Oostenbrink, C.; van Gunsteren, W. F. Free Energies of Ligand Binding for Structurally Diverse Compounds. *Proc. Natl. Acad. Sci. U. S. A.* 2005, 102 (19), 6750–6754.
(21) Raman, E. P.; Prabhu Raman, E.; Lakkaraju, S. K.; Denny, R. A.; 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. *Journal of Computational Chemistry*. 2017, pp 1238–1251. https://doi.org/10.1002/jcc.24522.

(22) Wade, A. D.; Rizzi, A.; Wang, Y.; Huggins, D. J. Computational Fluorine Scanning Using Free-Energy Perturbation. *J. Chem. Inf. Model.* 2019. https://doi.org/10.1021/acs.jcim.9b00228.

(23) Lee, L.-P.; Tidor, B. Optimization of Electrostatic Binding Free Energy. *J. Chem. Phys.* 1997, 106 (21), 8681–8690.

(24) Kangas, E.; Tidor, B. Charge Optimization Leads to Favorable Electrostatic Binding Free Energy. *Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics* 1999, 59 (5 Pt B), 5958–5961.

(25) Kangas, E.; Tidor, B. Electrostatic Specificity in Molecular Ligand Design. *J. Chem. Phys.* 2000, 112 (20), 9120–9131.

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

(27) 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.

(28) Kangas, E.; Tidor, B. Electrostatic Complementarity at Ligand Binding Sites: Application to Chorismate Mutase. *J. Phys. Chem. B* 2001, 105 (4), 880–888.

(29) 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 (11), 1416–1429.

(30) Sulea, T.; Purisima, E. O. Optimizing Ligand Charges for Maximum Binding Affinity. A Solvated Interaction Energy Approach. *J. Phys. Chem. B* 2001, 105 (4), 889–899.

(31) Armstrong, K. A.; Tidor, B.; Cheng, A. C. Optimal Charges in Lead Progression: A Structure-Based Neuraminidase Case Study. *J. Med. Chem.* 2006, 49 (8), 2470–2477.

(32) Shen, Y.; Gilson, M. K.; Tidor, B. Charge Optimization Theory for Induced-Fit Ligands. *J. Chem. Theory Comput.* 2012, 8 (11), 4580–4592.

(33) 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.; et al. OpenMM 4: A Reusable,
Extensible, Hardware Independent Library for High Performance Molecular Simulation. *J. Chem. Theory Comput.* **2013**, *9*(1), 461–469.

(34) 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.; et al. 7-Fluoroindazoles as Potent and Selective Inhibitors of Factor Xa†. *Journal of Medicinal Chemistry*. 2008, pp 282–297. https://doi.org/10.1021/jm701217r.

(35) Aiguadé, J.; Balagué, C.; Carranco, I.; Caturla, F.; Domínguez, M.; Eastwood, P.; Esteve, C.; González, J.; Lumeras, W.; Orellana, A.; et al. Novel Triazolopyridylbenzamides as Potent and Selective p38α Inhibitors. *Bioorg. Med. Chem. Lett.* **2012**, *22*(10), 3431–3436.

(36) Hamann, L. G.; al., E. al et. Tandem Optimization of Target Activity and Elimination of Mutagenic Potential in a Potent Series of N-Aryl Bicyclic Hydantoin-Based Selective Androgen Receptor Modulators. *ChemInform*. 2007. https://doi.org/10.1002/chin.200733119.

(37) Berman, H. M.; Westbrook, J.; Feng, Z. The Protein Data Bank. *Nucleic acids* **2000**.

(38) Eastman, P.; Swails, J.; Chodera, J. D. OpenMM 7: Rapid Development of High Performance Algorithms for Molecular Dynamics. *PLoS Comput. Biol.* **2017**.

(39) 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*(3), 221–234.

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

(41) 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*(9), 1157–1174.

(42) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, Efficient Generation of High-quality Atomic Charges. AM1-BCC Model: II. Parameterization and Validation. *Journal of computational* **2002**.

(43) Rizzi, A.; Chodera, J.; Naden, L.; Beauchamp, K.; Albanese, S.; Grinaway, P.; Rustenburg, B.; Saladi, S.; Boehm, K. *Choderalab/yank: Bugfix Release*; 2018. https://doi.org/10.5281/zenodo.1447109.

(44) 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* (20), 9665–9678.

(45) Oliphant, T. E. SciPy: Open Source Scientific Tools for Python. *Computing in Science and Engineering* **2007**, *9* (1), 10–20.

(46) Reif, M. M.; Hünenberger, P. H. Computation of Methodology-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* (14), 144103.

(47) 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**, No. 12, S1–S19.

(48) Irwin, B. W. J.; Huggins, D. J. Estimating Atomic Contributions to Hydration and Binding Using Free Energy Perturbation. *J. Chem. Theory Comput.* **2018**, *14* (6), 3218–3227.
Supporting Information for:

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

Alexander D. Wade\textsuperscript{a}, David J. Huggins \textsuperscript{a, b, c}

Affiliations:

\textsuperscript{a}TCM Group, Cavendish Laboratory, University of Cambridge, 19 JJ Thomson Avenue, Cambridge CB3 0HE, United Kingdom
\textsuperscript{b}Tri-Institutional Therapeutics Discovery Institute, Belfer Research Building, 413 East 69th Street, 16th Floor, Box 300, New York, USA
\textsuperscript{c}Department of Physiology and Biophysics, Weill Cornell Medicine, 1300 York Avenue, New York, NY, 10065, USA.

Corresponding Author: David J. Huggins
Email: dhuggins@tritdi.org
Updated: 20th June 2019
ORCID:s:
Alex Wade - 0000-0003-1500-3733
David Huggins 0000-0003-1579-2496
Complexes for the ligands shown in Figure (1) were built from crystal structures of highly related molecules. Superpositions of the ligands in the PDBs are shown in Figures (S1-S3)

Figure (S1) : Superposition of the 3D structure of the ligand in PDBID 2RA0 shown in red and the manually build 3D structure of the ligand depicted in Figure (1a) shown in blue.

Figure (S2) : Superposition of the 3D structure of the ligand in PDBID 3S3I shown in red and the manually build 3D structure of the ligand depicted in Figure (1b) shown in blue.
Figure (S3): Superposition of the 3D structure of the ligand in PDBID 2NW4 shown in red and the manually build 3D structure of the ligand depicted in Figure (1c) shown in blue.

Parameters for the inhibitors were generated using Antechamber with AMBER GAFF and AM1-BCC. The partial charges for the wild type inhibitors are shown in Table (S1).

Table (S1): Original charges for fxa, p38 and androgen ligands. Atom names given in main text.

| FXa     | Charge \( q_e \) | P38     | Charge \( q_e \) | Androgen receptor | Charge \( q_e \) |
|---------|-----------------|---------|-----------------|------------------|-----------------|
| Atom names | C1  | -0.215 | C2  | -0.034 | C3  | -0.049 | C4  | -0.064 | C5  | -0.035 | C6  | -0.148 | C7  | -0.058 | C8  | -0.154 | C9  | -0.095 | C10 | 0.089 | C11 | -0.065 | C12 | -0.360 |
|          | C2  | 0.089  | C3  | -0.095 | C4  | 0.771  | C5  | -0.095 | C6  | 0.234  | C7  | 0.049  | C8  | 0.771  | C9  | 0.090  | C10 | -0.095 | C11 | -0.025 | C12 | -0.430 |
|   |   |   |   |   |
|---|---|---|---|---|
| C11 | 0.705 | C10 | -0.066 | N14 | -0.356 |
| C12 | -0.093 | C11 | -0.144 | C15 | -0.085 |
| C13 | 0.098 | N2 | -0.075 | C17 | 0.147 |
| C14 | -0.120 | N3 | -0.312 | C19 | -0.106 |
| C15 | -0.181 | C12 | 0.303 | C22 | 0.040 |
| C16 | -0.144 | C13 | -0.064 | O25 | -0.594 |
| C17 | -0.085 | C14 | 0.675 | C27 | 0.711 |
| C18 | 0.177 | N4 | -0.544 | O28 | -0.575 |
| N1  | -0.521 | C15 | 0.035 | O29 | -0.571 |
| N2  | 0.125 | C16 | -0.131 | C30 | -0.068 |
| C19 | -0.142 | C17 | -0.131 | H25 | 0.425 |
| C20 | 0.026 | O1  | -0.595 | H2  | 0.150 |
| C21 | -0.152 | N5  | -0.646 | H7  | 0.150 |
| N3  | -0.388 | C18 | 0.392 | H5  | 0.154 |
| O1  | -0.620 | C19 | -0.202 | H15 | 0.117 |
| C22 | 0.074 | C20 | -0.042 | H17 | 0.061 |
| C23 | -0.104 | C21 | -0.102 | H19 | 0.076 |
| C24 | -0.113 | C22 | 0.402 | H192| 0.076 |
| N4  | -0.846 | C11 | -0.064 | H221| 0.078 |
| C25 | -0.146 | H7  | 0.053 | H222| 0.078 |
| C26 | -0.083 | H8  | 0.053 | H301| 0.058 |
| H5  | 0.042 | H9  | 0.053 | H302| 0.058 |
| H10 | 0.458 | H10 | 0.315 | H303| 0.058 |
| H11 | 0.066 | H11 | 0.079 |   |   |
| H12 | 0.066 | H12 | 0.081 |   |   |
| H13 | 0.066 | H13 | 0.081 |   |   |
| H14 | 0.112 | H14 | 0.081 |   |   |
To inspect the effect of the RMSD on where the optimiser places the charge optimisations were run for 0.01, 0.03 and 0.05 $q_e$. For each RMSD the optimisation was repeated three times the average of the optimised charges across these three replicates for all RMSDs is shown in Table (S2). It can be seen from the low standard deviation for all sets of optimised charges in Table (S2) that the optimisation agrees well on the optimal set of charges across replicates.
Table (S2) : Sets of optimized charges for each system calculated with an RMSD of 0.01, 0.03 and 0.05 $q_e$ and averaged from 3 replicates. $\sigma$ denotes standard deviation.

| FXa | RMSD | $\sigma$ | 0.03 $\sigma$ | 0.01 $\sigma$ |
|-----|------|----------|---------------|---------------|
| C1  | 0.405 | 0.008 | 0.409 | 0.010 | 0.414 | 0.003 |
| C2  | -0.190 | 0.002 | -0.196 | 0.001 | -0.211 | 0.003 |
| C3  | -0.021 | 0.015 | -0.043 | 0.005 | -0.049 | 0.002 |
| C4  | -0.001 | 0.008 | -0.022 | 0.004 | -0.031 | 0.003 |
| C5  | 0.106 | 0.008 | 0.109 | 0.002 | 0.099 | 0.002 |
| C6  | -0.138 | 0.007 | -0.139 | 0.005 | -0.145 | 0.001 |
| C7  | -0.086 | 0.012 | -0.078 | 0.006 | -0.085 | 0.001 |
| C8  | -0.057 | 0.010 | -0.062 | 0.000 | -0.074 | 0.003 |
| C9  | -0.083 | 0.009 | -0.082 | 0.007 | -0.091 | 0.002 |
| C10 | -0.157 | 0.004 | -0.140 | 0.005 | -0.147 | 0.002 |
| C11 | 0.690 | 0.010 | 0.701 | 0.005 | 0.705 | 0.005 |
| C12 | -0.075 | 0.007 | -0.084 | 0.004 | -0.088 | 0.002 |
| C13 | 0.128 | 0.005 | 0.122 | 0.004 | 0.108 | 0.002 |
| C14 | -0.148 | 0.005 | -0.131 | 0.004 | -0.124 | 0.004 |
| C15 | -0.138 | 0.007 | -0.173 | 0.003 | -0.188 | 0.002 |
| C16 | -0.180 | 0.003 | -0.172 | 0.001 | -0.157 | 0.003 |
| C17 | -0.118 | 0.003 | -0.102 | 0.003 | -0.091 | 0.001 |
| C18 | 0.141 | 0.007 | 0.136 | 0.008 | 0.159 | 0.002 |
| N1  | -0.433 | 0.006 | -0.467 | 0.008 | -0.499 | 0.002 |
| N2  | 0.153 | 0.006 | 0.139 | 0.004 | 0.132 | 0.004 |
| C19 | -0.091 | 0.013 | -0.126 | 0.009 | -0.136 | 0.001 |
| C20 | 0.079 | 0.010 | 0.053 | 0.005 | 0.032 | 0.001 |
|   |   |   |   |   |   |
|---|---|---|---|---|---|
| C21 | -0.151 | 0.013 | -0.142 | 0.006 | -0.149 | 0.003 |
| N3  | -0.393 | 0.014 | -0.380 | 0.003 | -0.385 | 0.002 |
| O1  | -0.599 | 0.011 | -0.610 | 0.002 | -0.618 | 0.004 |
| C22 | 0.078  | 0.007 | 0.073  | 0.009 | 0.071  | 0.003 |
| C23 | -0.100 | 0.007 | -0.117 | 0.005 | -0.112 | 0.002 |
| C24 | -0.069 | 0.013 | -0.098 | 0.006 | -0.120 | 0.003 |
| N4  | -0.845 | 0.008 | -0.849 | 0.004 | -0.852 | 0.002 |
| C25 | -0.172 | 0.004 | -0.152 | 0.005 | -0.149 | 0.002 |
| C26 | -0.127 | 0.016 | -0.101 | 0.007 | -0.088 | 0.003 |
| H5  | 0.095  | 0.012 | 0.097  | 0.006 | 0.063  | 0.003 |
| H10 | 0.638  | 0.004 | 0.579  | 0.005 | 0.497  | 0.006 |
| H11 | 0.020  | 0.004 | 0.038  | 0.013 | 0.059  | 0.002 |
| H12 | 0.023  | 0.007 | 0.044  | 0.000 | 0.060  | 0.002 |
| H13 | 0.026  | 0.008 | 0.043  | 0.012 | 0.061  | 0.002 |
| H14 | 0.019  | 0.013 | 0.053  | 0.004 | 0.088  | 0.003 |
| H15 | 0.006  | 0.007 | 0.048  | 0.004 | 0.082  | 0.001 |
| H17 | 0.445  | 0.009 | 0.442  | 0.006 | 0.455  | 0.006 |
| H20 | 0.097  | 0.008 | 0.079  | 0.005 | 0.057  | 0.002 |
| H21 | 0.014  | 0.008 | 0.053  | 0.002 | 0.078  | 0.006 |
| H22 | 0.106  | 0.005 | 0.090  | 0.006 | 0.085  | 0.004 |
| H23 | 0.112  | 0.009 | 0.079  | 0.008 | 0.067  | 0.002 |
| H24 | 0.095  | 0.011 | 0.077  | 0.003 | 0.067  | 0.002 |
| H25 | 0.030  | 0.006 | 0.049  | 0.007 | 0.053  | 0.003 |
| H26 | -0.008 | 0.002 | 0.025  | 0.007 | 0.047  | 0.005 |
| H27 | 0.416  | 0.008 | 0.429  | 0.015 | 0.447  | 0.005 |
| H4  | 0.193  | 0.010 | 0.161  | 0.009 | 0.160  | 0.002 |
| H9  | 0.181  | 0.007 | 0.140  | 0.002 | 0.132  | 0.001 |
|    |    |    |    |    |    |    |
|----|----|----|----|----|----|----|
| H6 | 0.160 | 0.001 | 0.158 | 0.006 | 0.157 | 0.002 |
| H7 | 0.129 | 0.018 | 0.144 | 0.006 | 0.141 | 0.002 |
| H19 | 0.207 | 0.007 | 0.171 | 0.006 | 0.160 | 0.001 |
| H3 | 0.152 | 0.004 | 0.158 | 0.004 | 0.150 | 0.002 |
| H18 | 0.000 | 0.002 | 0.053 | 0.003 | 0.110 | 0.002 |
| H29 | 0.094 | 0.011 | 0.097 | 0.009 | 0.124 | 0.004 |
| H8 | 0.117 | 0.017 | 0.151 | 0.002 | 0.171 | 0.002 |
| H28 | 0.117 | 0.007 | 0.134 | 0.005 | 0.154 | 0.002 |
| H16 | 0.110 | 0.004 | 0.128 | 0.005 | 0.138 | 0.002 |
| P38 |    |    |    |    |    |    |
| RMSD | 0.05 | σ | 0.03 | σ | 0.01 | σ |
| C1 | -0.175 | 0.009 | -0.162 | 0.008 | -0.156 | 0.002 |
| C2 | 0.011 | 0.022 | -0.034 | 0.006 | -0.037 | 0.001 |
| C3 | -0.030 | 0.023 | -0.044 | 0.013 | -0.060 | 0.001 |
| C4 | -0.020 | 0.016 | -0.024 | 0.005 | -0.046 | 0.004 |
| C5 | -0.143 | 0.010 | -0.144 | 0.005 | -0.145 | 0.006 |
| C6 | -0.154 | 0.008 | -0.122 | 0.005 | -0.105 | 0.001 |
| C7 | -0.019 | 0.022 | -0.019 | 0.005 | -0.023 | 0.002 |
| C8 | -0.057 | 0.017 | -0.066 | 0.008 | -0.063 | 0.002 |
| C9 | -0.009 | 0.013 | -0.050 | 0.005 | -0.071 | 0.004 |
| N1 | -0.007 | 0.004 | -0.030 | 0.004 | -0.045 | 0.001 |
| C10 | -0.067 | 0.010 | -0.070 | 0.005 | -0.065 | 0.001 |
| C11 | -0.122 | 0.012 | -0.140 | 0.005 | -0.143 | 0.003 |
| N2 | -0.067 | 0.024 | -0.068 | 0.014 | -0.072 | 0.002 |
| N3 | -0.326 | 0.003 | -0.328 | 0.001 | -0.316 | 0.001 |
| C12 | 0.286 | 0.009 | 0.307 | 0.009 | 0.299 | 0.003 |
| C13 | -0.085 | 0.011 | -0.070 | 0.005 | -0.077 | 0.003 |
|   | C14     | 0.695   | 0.009 | 0.697 | 0.006 | 0.691 | 0.003 |
|---|---------|---------|-------|-------|-------|-------|-------|
| N4| -0.485  | 0.008   | -0.520| 0.005 | -0.530| 0.005 |
| C15| 0.038   | 0.010   | 0.043 | 0.010 | 0.039 | 0.001 |
| C16| -0.078  | 0.012   | -0.109| 0.004 | -0.117| 0.001 |
| C17| -0.152  | 0.009   | -0.140| 0.002 | -0.132| 0.002 |
| O1 | -0.503  | 0.005   | -0.519| 0.005 | -0.558| 0.002 |
| N5 | -0.624  | 0.007   | -0.645| 0.006 | -0.645| 0.003 |
| C18| 0.366   | 0.008   | 0.374 | 0.004 | 0.383 | 0.005 |
| C19| -0.258  | 0.016   | -0.243| 0.012 | -0.217| 0.003 |
| C20| -0.060  | 0.004   | -0.061| 0.004 | -0.051| 0.004 |
| C21| -0.126  | 0.011   | -0.100| 0.008 | -0.104| 0.007 |
| C22| 0.405   | 0.002   | 0.410 | 0.004 | 0.403 | 0.003 |
| Cl1| -0.047  | 0.010   | -0.064| 0.007 | -0.062| 0.003 |
| H7 | 0.006   | 0.010   | 0.023 | 0.007 | 0.036 | 0.005 |
| H8 | 0.006   | 0.008   | 0.023 | 0.008 | 0.037 | 0.001 |
| H9 | 0.007   | 0.010   | 0.019 | 0.007 | 0.038 | 0.002 |
| H10| 0.506   | 0.004   | 0.433 | 0.000 | 0.341 | 0.002 |
| H11| -0.002  | 0.009   | 0.047 | 0.006 | 0.074 | 0.001 |
| H12| 0.077   | 0.001   | 0.096 | 0.007 | 0.090 | 0.002 |
| H13| 0.163   | 0.010   | 0.118 | 0.008 | 0.094 | 0.001 |
| H14| 0.030   | 0.013   | 0.053 | 0.016 | 0.076 | 0.001 |
| H15| 0.070   | 0.014   | 0.065 | 0.004 | 0.082 | 0.003 |
| H6 | 0.095   | 0.009   | 0.120 | 0.007 | 0.132 | 0.001 |
| H4 | 0.191   | 0.008   | 0.204 | 0.007 | 0.180 | 0.001 |
| H1 | 0.048   | 0.009   | 0.079 | 0.007 | 0.115 | 0.001 |
| H22| 0.063   | 0.004   | 0.038 | 0.006 | 0.042 | 0.001 |
| H18| 0.003   | 0.009   | 0.017 | 0.007 | 0.027 | 0.002 |
|     | H19 | H16 | H17 | H20 | Androgen recp. | RMSD | σ  | 0.03 | σ  | 0.01 | σ  |
|-----|-----|-----|-----|-----|----------------|------|----|------|----|------|----|
|     | 0.088 | 0.168 | 0.158 | 0.140 |               | 0.05 |    | 0.03 |    | 0.01 |    |
| C2  | -0.137 | -0.034 | 0.104 | -0.099 |               | -0.137 | 0.009 | -0.121 | 0.002 | -0.104 | 0.000 |
| C3  | -0.034 | 0.003 | -0.056 | 0.006 |               | -0.034 | 0.003 | -0.056 | 0.004 | -0.059 | 0.001 |
| C4  | 0.104 | 0.009 | 0.092 | 0.006 |               | 0.104 | 0.009 | 0.092 | 0.002 | 0.077 | 0.002 |
| C5  | -0.099 | 0.006 | -0.112 | 0.003 |               | -0.099 | 0.006 | -0.112 | 0.002 | -0.119 | 0.001 |
| C7  | -0.116 | 0.006 | -0.110 | 0.003 |               | -0.116 | 0.006 | -0.110 | 0.003 | -0.098 | 0.001 |
| C9  | -0.044 | 0.002 | -0.039 | 0.003 |               | -0.044 | 0.002 | -0.039 | 0.003 | -0.024 | 0.001 |
| C10 | 0.201 | 0.010 | 0.204 | 0.002 |               | 0.201 | 0.010 | 0.204 | 0.002 | 0.220 | 0.001 |
| N11 | -0.349 | 0.004 | -0.366 | 0.001 |               | -0.349 | 0.004 | -0.366 | 0.001 | -0.368 | 0.001 |
| N12 | -0.414 | 0.004 | -0.411 | 0.003 |               | -0.414 | 0.004 | -0.411 | 0.003 | -0.421 | 0.000 |
| C13 | 0.781 | 0.006 | 0.791 | 0.002 |               | 0.781 | 0.006 | 0.791 | 0.002 | 0.781 | 0.001 |
| N14 | -0.370 | 0.010 | -0.357 | 0.004 |               | -0.370 | 0.010 | -0.357 | 0.004 | -0.354 | 0.001 |
| C15 | -0.083 | 0.010 | -0.079 | 0.001 |               | -0.083 | 0.010 | -0.079 | 0.001 | -0.081 | 0.002 |
| C17 | 0.176 | 0.006 | 0.161 | 0.007 |               | 0.176 | 0.006 | 0.161 | 0.007 | 0.153 | 0.001 |
| C19 | -0.122 | 0.004 | -0.116 | 0.002 |               | -0.122 | 0.004 | -0.116 | 0.002 | -0.110 | 0.000 |
| C22 | 0.019 | 0.010 | 0.029 | 0.005 |               | 0.019 | 0.010 | 0.029 | 0.005 | 0.038 | 0.002 |
| O25 | -0.537 | 0.008 | -0.552 | 0.003 |               | -0.537 | 0.008 | -0.552 | 0.003 | -0.576 | 0.001 |
| C27 | 0.722 | 0.006 | 0.734 | 0.000 |               | 0.722 | 0.006 | 0.734 | 0.000 | 0.723 | 0.001 |
| O28 | -0.452 | 0.001 | -0.501 | 0.001 |               | -0.452 | 0.001 | -0.501 | 0.001 | -0.545 | 0.001 |
| O29 | -0.429 | 0.003 | -0.490 | 0.001 |               | -0.429 | 0.003 | -0.490 | 0.001 | -0.543 | 0.001 |
| C30 | -0.092 | 0.004 | -0.084 | 0.002 |               | -0.092 | 0.004 | -0.084 | 0.002 | -0.073 | 0.001 |
| Symbol | Charge 1 | Error 1 | Charge 2 | Error 2 | Charge 3 | Error 3 |
|--------|----------|---------|----------|---------|----------|---------|
| HO25   | 0.526    | 0.010   | 0.483    | 0.001   | 0.444    | 0.002   |
| H2     | 0.075    | 0.003   | 0.103    | 0.003   | 0.131    | 0.001   |
| H7     | 0.070    | 0.002   | 0.110    | 0.003   | 0.135    | 0.001   |
| H5     | 0.209    | 0.006   | 0.184    | 0.004   | 0.163    | 0.001   |
| H15    | 0.108    | 0.009   | 0.110    | 0.002   | 0.116    | 0.002   |
| H17    | 0.030    | 0.002   | 0.046    | 0.001   | 0.061    | 0.001   |
| H191   | 0.032    | 0.004   | 0.053    | 0.003   | 0.069    | 0.002   |
| H192   | 0.035    | 0.007   | 0.056    | 0.002   | 0.071    | 0.002   |
| H221   | 0.074    | 0.005   | 0.074    | 0.003   | 0.078    | 0.001   |
| H222   | 0.028    | 0.011   | 0.046    | 0.000   | 0.068    | 0.001   |
| H301   | 0.028    | 0.001   | 0.041    | 0.004   | 0.050    | 0.001   |
| H302   | 0.029    | 0.008   | 0.042    | 0.004   | 0.049    | 0.000   |
| H303   | 0.032    | 0.004   | 0.038    | 0.001   | 0.049    | 0.001   |

To visually assert that the charge is being placed in the same places independent of RMSD Figures (S4-S12) show all inhibitors colored by change in charge across all RMSDs. Figures (S4-S12) show that the information about which atoms could be beneficially changed to be more positive or negative is largely invariant as the RMSD is changed.
Figure (S4): Fxa ligand with atoms colored by charge such that more negative atoms are blue and more positive atoms are red. Optimal set of charges computed with and RMSD of 0.05 $q_e$. 
Figure (S5): Fxa ligand with atoms colored by charge such that more negative atoms are blue and more positive atoms are red. Optimal set of charges computed with an RMSD of 0.03 $q_e$. 
Figure (S6): Fxa ligand with atoms colored by charge such that more negative atoms are blue and more positive atoms are red. Optimal set of charges computed with and RMSD of 0.01 $q_e$.

Figure (S7): P38 ligand with atoms colored by charge such that more negative atoms are blue and more positive atoms are red. Optimal set of charges computed with and RMSD of 0.05 $q_e$.

Figure (S8): P38 ligand with atoms colored by charge such that more negative atoms are blue and more positive atoms are red. Optimal set of charges computed with and RMSD of 0.03 $q_e$. 

Figure (S9): P38 ligand with atoms colored by charge such that more negative atoms are blue and more positive atoms are red. Optimal set of charges computed with and RMSD of 0.01 $q_e$.

Figure (S10): Androgen receptor with atoms colored by charge such that more negative atoms are blue and more positive atoms are red. Optimal set of charges computed with and RMSD of 0.05 $q_e$. 
Figure (S11): Androgen receptor with atoms colored by charge such that more negative atoms are blue and more positive atoms are red. Optimal set of charges computed with and RMSD of $0.03 \, q_e$.

Figure (S12): Androgen receptor with atoms colored by charge such that more negative atoms are blue and more positive atoms are red. Optimal set of charges computed with and RMSD of $0.01 \, q_e$. 
References

(1) 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.; et al. 7-Fluoroindazoles as Potent and Selective Inhibitors of Factor Xa†. *Journal of Medicinal Chemistry*. 2008, pp 282–297. https://doi.org/10.1021/jm701217r.

(2) Aiguadé, J.; Balagué, C.; Carranco, I.; Caturla, F.; Domínguez, M.; Eastwood, P.; Esteve, C.; González, J.; Lumeras, W.; Orellana, A.; et al. Novel Triazolopyridylbenzamides as Potent and Selective p38α Inhibitors. *Bioorg. Med. Chem. Lett.* 2012, 22 (10), 3431–3436.

(3) Hamann, L. G.; al., E. al et. Tandem Optimization of Target Activity and Elimination of Mutagenic Potential in a Potent Series of N-Aryl Bicyclic Hydantoin-Based Selective Androgen Receptor Modulators. *ChemInform*. 2007. https://doi.org/10.1002/chin.200733119.

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

(5) 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 (9), 1157–1174.

(6) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, Efficient Generation of High-quality Atomic Charges. AM1-BCC Model: II. Parameterization and Validation. *Journal of computational 2002.*
