

Computational Fluorine Scanning Using Free-Energy Perturbation

Alexander David Wade, Andrea Rizzi, Yuanqing Wang, and David John Huggins

J. Chem. Inf. Model., **Just Accepted Manuscript** • DOI: 10.1021/acs.jcim.9b00228 • Publication Date (Web): 02 May 2019

Downloaded from <http://pubs.acs.org> on May 7, 2019

Just Accepted

“Just Accepted” manuscripts have been peer-reviewed and accepted for publication. They are posted online prior to technical editing, formatting for publication and author proofing. The American Chemical Society provides “Just Accepted” as a service to the research community to expedite the dissemination of scientific material as soon as possible after acceptance. “Just Accepted” manuscripts appear in full in PDF format accompanied by an HTML abstract. “Just Accepted” manuscripts have been fully peer reviewed, but should not be considered the official version of record. They are citable by the Digital Object Identifier (DOI®). “Just Accepted” is an optional service offered to authors. Therefore, the “Just Accepted” Web site may not include all articles that will be published in the journal. After a manuscript is technically edited and formatted, it will be removed from the “Just Accepted” Web site and published as an ASAP article. Note that technical editing may introduce minor changes to the manuscript text and/or graphics which could affect content, and all legal disclaimers and ethical guidelines that apply to the journal pertain. ACS cannot be held responsible for errors or consequences arising from the use of information contained in these “Just Accepted” manuscripts.

Computational Fluorine Scanning Using Free-Energy Perturbation

Alexander D. Wade^a, Andrea Rizzi^{b,c}, Yuanqing Wang^d, David J. Huggins^{a, e, f}

Affiliations:

^aTCM Group, Cavendish Laboratory, University of Cambridge,
19 J J Thomson Avenue, Cambridge CB3 0HE, United Kingdom

^bTri-Institutional Training Program in Computational Biology and
Medicine, New York, NY 10065, USA

^cComputational and Systems Biology Program, Sloan Kettering
Institute, Memorial Sloan-Kettering Cancer Center, New York,
NY 10065, USA

^d Physiology, Biophysics, and System Biology Program, Weill
Cornell Medicine, 1300 York Avenue, New York, NY 10065,
USA

^eTri-Institutional Therapeutics Discovery Institute, Belfer
Research Building, 413 East 69th Street, 16th Floor, Box 300,
New York, USA

^fDepartment 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: 29th April 2019

ORCID: Alex Wade - 0000-0003-1500-3733

Andrea Rizzi 0000-0001-7693-2013

Yuanqing Wang 0000-0003-4403-2015

David Huggins 0000-0003-1579-2496

Abstract

We present perturbative fluorine scanning, a computational fluorine scanning approach using free-energy perturbation. This method can be applied to molecular dynamics simulations of a single compound and make predictions for the best binders out of numerous fluorinated analogues. We tested the method on nine test systems: Renin, DPP4, Menin, P38, Factor Xa, CDK2, AKT, JAK2, and Androgen Receptor. The predictions were in excellent agreement with more rigorous alchemical free-energy calculations and in good agreement with experimental data for most of the test systems. However, the agreement with experiment was very poor in some of the test systems and this highlights the need for improved force fields in addition to accurate treatment of tautomeric and protonation states. The method is of particular interest due to the wide use of fluorine in medicinal chemistry to improve binding affinity and ADME properties. The promising results on this test case suggest that perturbative fluorine scanning will be a useful addition to the available arsenal of free-energy methods.

Introduction

Fluorine scanning is a common technique in medicinal chemistry and involves systematic replacement of hydrogen with fluorine.¹⁻⁶ It can improve binding affinity as well as ADME properties.^{7,8} A striking example of the potential is given by the application of fluorine scanning on factor Xa.⁹ The authors found that a modification from a hydrogen in compound **1** to a fluorine in compound **2** improved the binding affinity by approximately 55-fold (Figure 1).

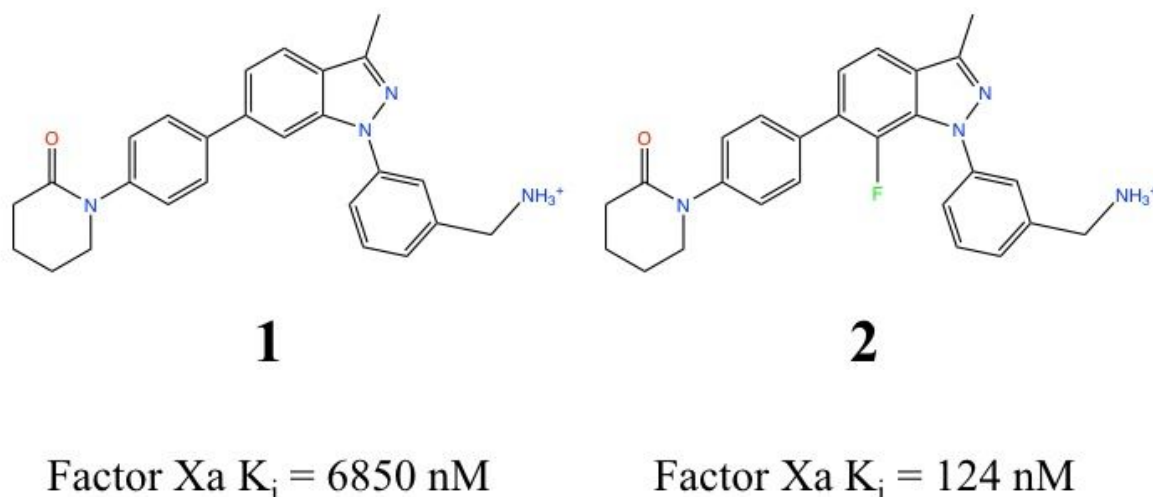


Figure (1): Two factor Xa inhibitors and their K_i values.

One of the drawbacks of fluorine scanning is the requirement to test each hydrogen-to-fluorine mutation individually. For example, testing the aromatic hydrogen-to-fluorine mutations in compound **1** requires synthesizing and assaying 9 compounds and testing all of the hydrogen-to-fluorine mutations requires synthesizing and assaying 16 compounds. Testing combinations of two hydrogen-to-fluorine mutations is an order of magnitude more challenging. Computational fluorine scanning using a molecular mechanics-Poisson-Boltzmann/surface area (MMPBSA) method has been suggested in the past as a way to design molecules with improved binding affinity *in silico*.¹⁰ However, MMPBSA calculations use a simplified implicit solvent model and alchemical free-energy methods with explicit solvent are increasingly used in place of MMPBSA calculations. Alchemical methods are a class of computational methods that use molecular simulation to calculate free energy differences.¹¹⁻¹³ They have garnered a lot of attention in recent years¹⁴⁻¹⁹ and some benchmarks have shown that they outperform other computational methods.²⁰ Free-energy perturbation (FEP)²¹ is one of the most commonly used methods to perform alchemical binding free energy calculations and is based on the Zwanzig equation.^{22,23} FEP can be used to calculate the absolute binding affinity of one molecule²⁴ or the relative binding free energy of two molecules.^{11,23,25,26} Commonly, the system of interest is studied using molecular dynamics (MD) simulation^{27,28} and treated using a forcefield.^{29,30} One of the major drawbacks of alchemical methods is the significant computing resource required for each calculation. Thus, performing a computational fluorine scan for all possible single or pair hydrogen-to-fluorine mutations in compound **1** is computationally demanding. In this study, we develop a method

1
2
3 for computational fluorine scanning using the Zwanzig equation which we term perturbative
4 fluorine scanning (PFS). A number of studies have considered the possibility of using
5 unidirectional transformations from a central reference molecule to calculate relative binding
6 free energies for a number of related molecules – this method has been termed single-step
7 perturbation (SSP). In early work, van Gunsteren and co-workers developed and applied the
8 method to hydration free energies.³¹ They also studied the limits of the method and found that
9 it could be applied to relative binding affinities.^{32,33} Other studies suggest that such an
10 approach is reliable for modification of up to three atoms using thermodynamic integration.³⁴
11 Mackerell and co-workers have also used this approach, which they term single-step free
12 energy perturbation.³⁵ Follow-up studies showed that combining SSP with a ligand grid free
13 energy approach offers a 1000-fold computational savings over traditional FEP for calculating
14 the relative binding affinities once some upfront pre-computations are complete.³⁶ More
15 recently, the method has been studied for bias³⁷ and applied to the efficient optimization of
16 forcefields.³⁸ We expand on this SSP approach to perform computational fluorine scanning by
17 comparing predictions from PFS for nine test systems to experimental data and to traditional
18 alchemical calculations using FEP.
19
20
21
22

23 To carry out these fluorine scanning calculation with ease we have developed a tool, Fluorify,
24 for executing the pipeline for these calculation automatically. This tool is freely available at
25 <https://github.com/adw62/Fluorify>. Fluorify uses OpenMM³⁹ as both a molecular dynamics
26 engine and library to create the modified alchemical systems. Fluorify will generate all of the
27 required mutant ligands from an input wild type ligand; these mutants are automatically
28 parameterized, built into complex systems, simulated and analysed.
29
30
31

32 **Methods**

33
34 We consider the effect of hydrogen to fluorine mutations for the hydrogens attached to
35 aromatic carbons of nine different protein ligand binding systems Table (1). The chemical
36 structures of these ligands are shown in Figures (2), all mutated hydrogens are shown
37 explicitly.
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

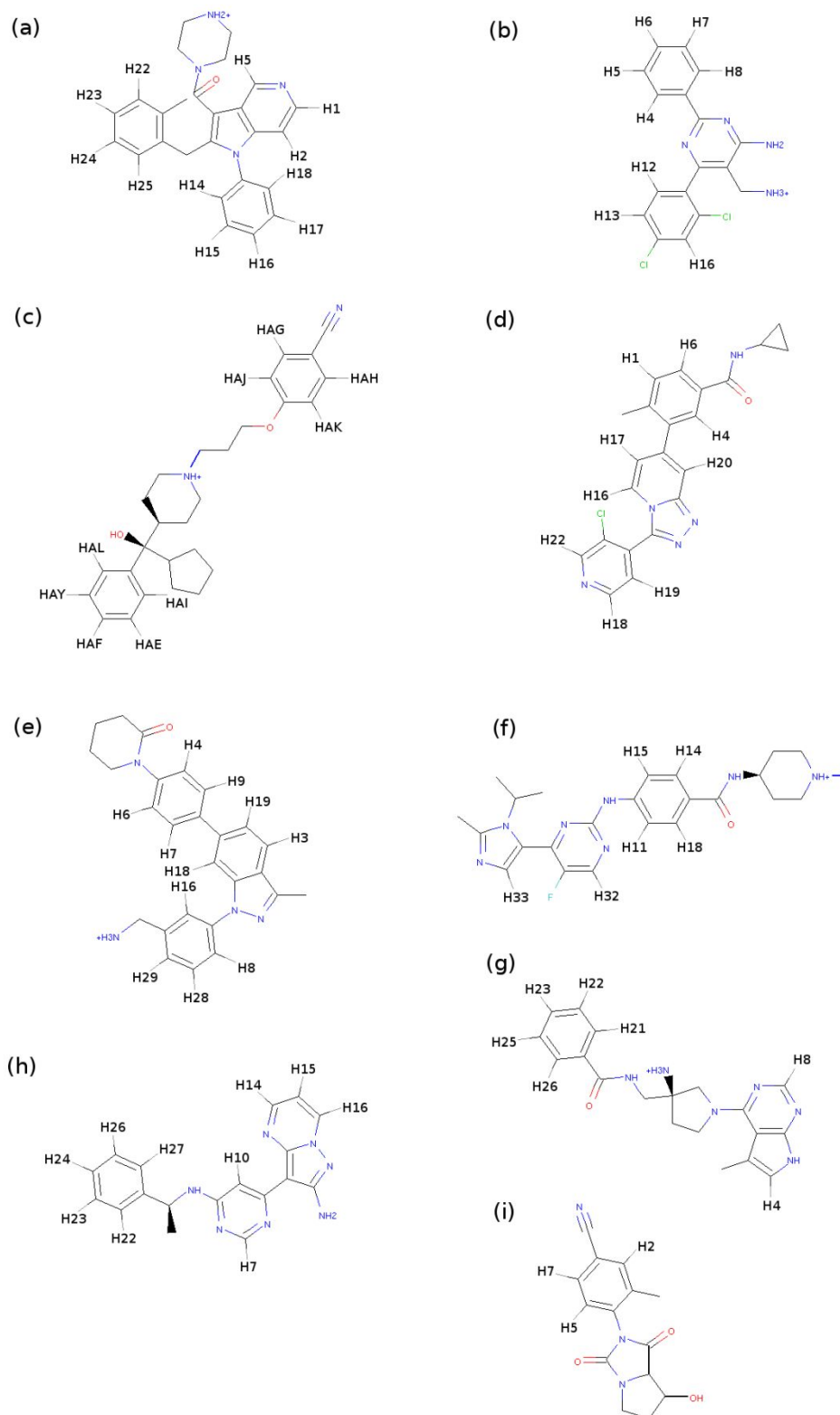


Figure (2): Chemical structures depicting the wild type ligands in each system. (a) Renin ⁴⁰, (b) DPP4 ⁴¹, (c) Menin ⁴², (d) P38 ⁴³, (e) FXa ⁹, (f) CDK2 ⁴⁴, (g) AKT ⁴⁵, (h) JAK2 ⁴⁶, (i) Androgen receptor ⁴⁷.

Examining all of the systems we are considering here, Table (1) shows the experimental data that is available for hydrogen to fluorine mutations taken from the respective papers where these systems were investigated^{9,40–47}. $\Delta\Delta G$ values in Table (1) are calculated from K_i or IC50 values in references, see references for experimental methodologies. These experimental works do not provide estimates of uncertainties. It should be noted that manual preparation of the ligands in the complex PDBs was required. This manual preparation involved changing the ligand structure from that provided in the PDB to the structure for the start point of the experimental fluorine mutation examined. These changes are reflected in the chemical structure shown in Figure (2). In addition, DPP4 was modelled as a monomer rather than the dimer in the crystal structure.

Table (1): $\Delta\Delta G$ for hydrogen to fluorine mutations obtained experimentally. PDB and PDBIDs taken from the Protein Data Bank.⁴⁸ The second column denotes the system as specified by Figure (2). Third column shows hydrogen being mutated as specified by Figure (2).

PDBID	System	Hydrogen	Experimental $\Delta\Delta G$ (kcal/mol)
3OOT ⁴⁰	Renin (a)	H22	-2.47
1RWQ ⁴¹	DPP4 (b)	H5	-2.31
	DPP4 (b)	H6	-2.31
	DPP4 (b)	H7	-2.31
	DPP4 (b)	H4	0.91
	DPP4 (b)	H8	0.91
4OG6 ⁴²	Menin (c)	HAY	-0.40
3S3I ⁴³	P38 (d)	H1	-2.26
2RA0 ⁹	FXa (e)	H18	-2.37
2W17 ⁴⁴	CDK2 (f)	H18	-2.12
3MVH ⁴⁵	AKT (g)	H23	-1.26
	AKT (g)	H22	-0.29
	AKT (g)	H25	-0.29
	AKT (g)	H26	-0.20
3IOK ⁴⁶	JAK2 (h)	H24	-1.15
2NW4 ⁴⁷	Androgen receptor (i)	H2	-1.11

It is worth noting that there are some symmetry considerations to take into account with PFS. Many of the hydrogens in the chemical structures shown in figures (2) are equivalent but a single fluorine substitution will break this symmetry and equivalence. However, for pairs of perturbations (or larger) there are additional symmetry considerations that, whilst not altering the predictions, may reduce the number of calculations needed.

System Setup

1
2
3 All FEP and PFS calculations performed in this work were made with Fluorify the details for
4 each stage of these calculations are as follows. The co-crystal structure for the nine systems
5 examined here are taken from the Protein Data Bank ⁴⁸ with PDBIDs shown in Table (1). To
6 prepare these systems non-standard residues were converted to their standard equivalents
7 with pdbfixer ⁴⁹. Selenomethionines were changed to methionines and missing sidechains
8 were added using Schrödinger's Preparation Wizard ⁵⁰, which was also used to assign
9 protonation state of all ionizable residues. All buffer solvents and ions were removed. The
10 hydrogen atom positions were then built using tleap and forcefield parameters and partial
11 charges were assigned from the AMBER ff14SB force field ⁵¹. Parameters for the inhibitors
12 were generated using Antechamber ⁵² with AMBER GAFF 2 ⁵³ and AM1-BCC ⁵⁴. These
13 structures and parameters are then passed to YANK's ⁵⁵ 0.23.7 automatic setup pipeline to
14 build solvated ligand-protein and ligand systems. For solvation, TIP4P-EW ⁵⁶ is used; at this
15 stage a salt concentration of 150mM and any required counter-ions are added. In every case,
16 the edge of the solvation box is 15 Å from any atom of the receptor and ligand.
17
18
19
20

21 *Molecular dynamics*

22
23 All simulations were performed with OpenMM 7.3.0. ⁴⁹ as follows. First OpenMM's default
24 minimizer was used to minimize all structures. Then equilibration was performed in the NPT
25 ensemble for 500ps at 300k and 1atm using a Langevin integrator and Monte Carlo barostat.
26 MD simulations were performed in the NPT ensemble using a time step of 2fs. Van der Waals
27 interactions were truncated at 11.0 Å with switching at 9.0 Å. Electrostatics were modeled
28 using particle mesh Ewald method with a cutoff of 11.0 Å. All other simulation parameters
29 were left as default. We ran triplicate simulations of the non-fluorinated compound with the
30 ligand in complex and in solution, for 50 ns. Snapshots were collected every 5 ps.
31
32
33
34

35 *Perturbative Fluorine Scanning*

36
37 During fluorine scanning, van der Waals, charge, bond and torsion parameters are assumed
38 to change. Since this is a post analysis and the dynamics were collected from the system with
39 wild-type bond parameters, the change in bonded parameters has no effect on dynamics of
40 the molecules geometry. Whilst this should be negligible when considering the change in
41 geometry of non-perturbed atoms this may not be true for the atoms perturbed from hydrogen
42 to fluorine where the C-F bond should be longer than C-H. To aid convergence of the free
43 energy calculation, we would like to include this change in bond length. We therefore use a
44 hybrid topology approach where massless interaction sites at the position of all possible
45 fluorine mutations are added. The positions of these fluorines is defined relative to the position
46 of their parent hydrogen such that the C-F distance is always 1.24 times the C-H distance ⁵⁷.
47 During the simulation the Lennard-Jones, charge, bond and torsions parameters of these
48 additional fluorine sites are turned off. When mutating to a fluorinated system the relevant
49 hydrogen is turned off and fluorine Lennard-Jones and charge parameters are applied to the
50 additional site, this is demonstrated in figure (3). The torsion parameters are mutated from the
51 hydrogenated system to fluorinated system, but the torsions remain on the parent hydrogen
52 and are not transferred to the virtual fluorine. This will have no effect on the energy as the
53 angles remain the same. When simulating these systems, all hydrogen bonds are constrained,
54 since the position of the fluorine is defined relative to the position of its parent hydrogen it is
55 also implicitly constrained. We therefore make the assumption that the C-F bond oscillations
56
57
58
59
60

are negligible. To prevent the hybrid topologies from interacting, the additional fluorine is excluded from interacting with their parent hydrogens.

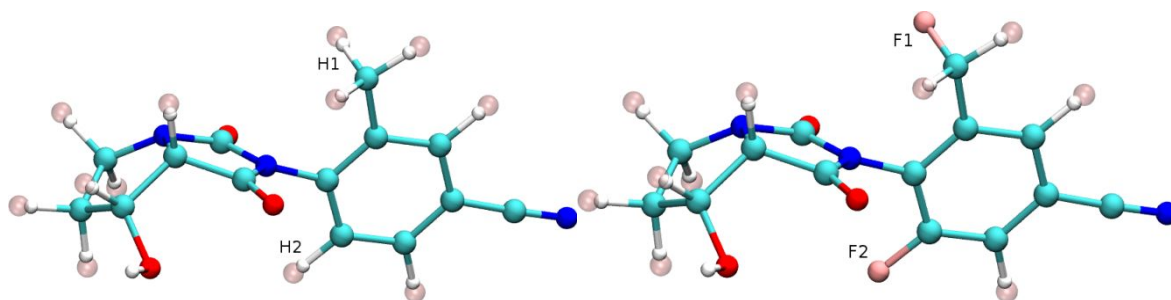


Figure (3) : Left panel Androgen inhibitor with all fluorines turned off. Right panel Androgen inhibitor with H1 and H2 transformed to F1 and F2.

The following demonstrates how the free energy change is calculated. With Equation (1) showing the Zwanzig equation ²².

$$\Delta G_{Zwanzig} = -kT \ln(\langle \exp(-(E_{mut} - E_{wt})/kT) \rangle_{wt}) \quad (1)$$

E_{wt} and E_{mut} are the energies of the system using the Hamiltonians of the “wild type” unperturbed system (wt) and the “mutant” perturbed system (mut). The wt subscript on the average of the exponential indicates that it is taken for samples in the wild type ensemble. E_{wt} and E_{mut} are a sum of the non-bonded, bonded, and torsion energies as shown in Equations (2) and (3).

$$E_{mut} = E_{mut}^{non-bonded} + E_{mut}^{bonded} + E_{mut}^{torsion} \quad (2)$$

$$E_{wt} = E_{wt}^{non-bonded} + E_{wt}^{bonded} + E_{wt}^{torsion} \quad (3)$$

Computationally $E_{wt}^{non-bonded}$ is the energy coming from the Lennard-Jones potential and electrostatics, calculated with PME, for the the wild type system and dynamics. To get $E_{mut}^{non-bonded}$, the wild type system is modified such that the Lennard-Jones and charge parameters of the mutant system are assigned to the atoms of the wild type system. E_{wt}^{bonded} is the energy coming from the harmonic bonds of the wild type system and dynamics. Here to get E_{mut}^{bonded} the bonds in the wild type system are assigned the parameters of the mutant harmonic bonds. The modification of the bonds in this work makes no change to the total energy as both alchemical hydrogen and fluorine are constrained. The modification of the bond energy is included for generality, however, as it would be required in some mutations, an example of which would be pyridination. $E_{wt}^{torsion}$ is the energy coming from the periodic torsions of the wild type system and dynamics. Again, to get $E_{wt}^{torsion}$ the wild type system is modified and the mutant torsion parameters are assigned to the wild type periodic torsion.

Equation (1) can then be applied to simulations of the unperturbed system in the bound and unbound states. Combining $\Delta G_{Zwanzig}$ for the bound and unbound states yields an estimate of the $\Delta\Delta G_{binding}$. This effectively involves performing a unidirectional alchemical transformation to each target molecule.

FEP

To validate the PFS result, we compare it against standard alchemical relative binding free energy calculations using the MBAR⁵⁸ estimator. These FEP simulations used the same hybrid topology described previously for PFS. The code to perform these relative free energy calculation is included as part of the Fluorify package. We used a total of 12 equally spaced lambda windows in which Lennard-Jones, charge, and torsion parameters were interpolated simultaneously from the wild-type to the mutated state. All windows were sampled independently with 2 ns of Langevin dynamics. Giving a total of 24ns of sampling per mutant half of the 50ns used for all mutants of a ligand in PFS. All simulation conditions were identical to the PFS molecular dynamics calculation 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 3.0.1⁵⁸. This FEP protocol is run automatically as part of the Fluorify package to check the $\Delta\Delta G$ for the top-ranked mutants as determined by PFS.

Results

We first analysed the convergence of PFS predictions as the simulation time increased. Figure (4) shows the $\Delta\Delta G$ predictions for the Factor Xa test case.

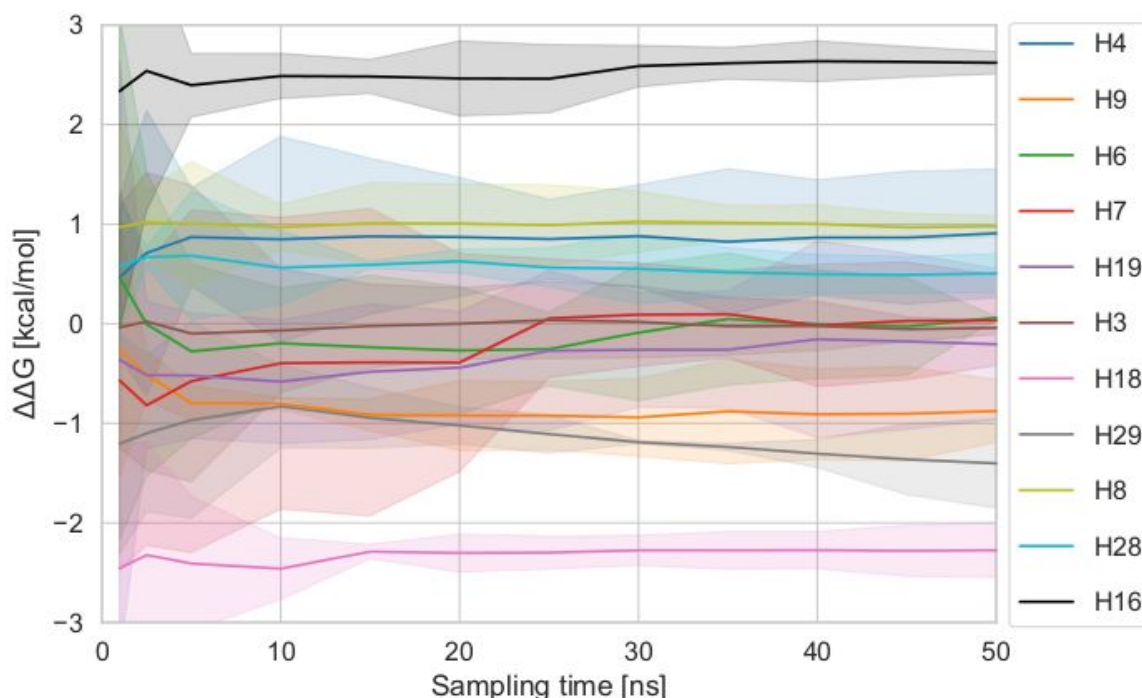


Figure (4) - Convergence of the $\Delta\Delta G$ predictions for the hydrogen to fluorine mutations in the Factor Xa test case as the simulation time is increased, H labels shown in Figure (2). Calculations were performed at 1.0 ns, 2.5 ns and then from 5.0 ns to 50.0 ns in 5.0 ns increments. $\Delta\Delta G$ s reported as mean of three replicates with shaded area showing 95% confidence interval computed as $\text{mean} \pm t_2 \cdot \text{SEM}$, 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.

The results for each system are $\Delta\Delta G$ s calculated by PFS and FEP. $\Delta\Delta G$ is calculated for the top n mutants. n is taken to be either three or the rank (as determined by PFS) of the best experimental mutant. Such that PFS and FEP results are always collected for the best experimental mutant. Table (2) shows the results of these $\Delta\Delta G$ calculations. All computational values in Table (2) are the average of three calculations, unaveraged values for PFS are reported in Table (S1) along with $\Delta\Delta G$ values from PFS calculations for all possible aromatic hydrogen to fluorine mutations. The largest error determined by MBAR for any of the FEP results in Table (2) was 0.17 kcal/mol.

Table (2): PFS, FEP and experimental $\Delta\Delta G$ s for all test cases, in kcal/mol. PFS and FEP predictions are reported as the mean $\Delta\Delta G$ s of three replicates with 95% confidence interval reported between square brackets 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. Symmetry related positions are indicated by †.

System/Hydrogen	PFS $\Delta\Delta G$ [kcal/mol]	FEP $\Delta\Delta G$ [kcal/mol]	EXP $\Delta\Delta G$ [kcal/mol]
(a) Renin			
H22	-1.8 [-2.4, -1.2]	-1.4 [-1.8, -0.9]	-2.47
H25	-1.6 [-2.1, -1.0]	-1.5 [-2.3, -0.8]	
H15	-0.6 [-1.0, -0.1]	-0.4 [-1.1, 0.4]	
(b) DPP4			
H13	-1.2 [-3.6, 1.3]	-0.8 [-2.1, 0.6]	
H16	-0.5 [-1.7, 0.8]	-0.2 [-1.7, 1.2]	
H7	-0.3 [-1.0, 0.4]	-0.2 [-0.8, 0.4]	-2.31†
H5	-0.2 [-0.6, 0.1]	-0.4 [-0.7, 0.0]	-2.31†
H4	-0.2 [-0.6, 0.2]	0.2 [-0.6, 1.1]	0.91†
H6	-0.2 [-0.3, 0.0]	-0.2 [-0.3, -0.1]	-2.31
H8	-0.1 [-1.5, 1.3]	0.0 [-0.7, 0.6]	0.91†
(c) Menin			
HAY	-1.5 [-2.0, -1.0]	-1.4 [-2.3, -0.5]	-0.40
HAI	-1.3 [-1.6, -0.9]	-0.7 [-1.2, -0.3]	
HAL	-0.9 [-1.1, -0.6]	-0.8 [-1.0, -0.5]	
(d) P38			
H1	-2.2 [-2.8, -1.6]	-2.2 [-2.7, -1.6]	-2.26
H19	-1.9 [-2.1, -1.6]	-1.6 [-1.7, -1.4]	
H16	-0.6 [-0.9, -0.3]	-0.3 [-0.5, -0.1]	
(e) Fxa			

H18	-2.3 [-2.5, -2.0]	-2.2 [-2.3, -2.1]	-2.37
H29	-1.4 [-1.9, -1.0]	-0.6 [-1.2, -0.1]	
H9	-0.9 [-1.2, -0.6]	-0.8 [-1.3, -0.4]	
(f) CDK2			
H33	-1.0 [-1.4, -0.5]	-0.6 [-1.4, 0.1]	
H14	-0.3 [-1.2, 0.6]	-0.4 [-1.6, 0.8]	
H18	-0.2 [-0.9, 0.5]	0.1 [-0.5, 0.7]	-2.12
(g) AKT			
H22	-2.2 [-2.6, -1.8]	-0.9 [-2.3, 0.4]	-0.29†
H25	-1.3 [-2.1, -0.5]	-0.8 [-1.0, -0.6]	-0.29†
H26	-1.2 [-2.2, -0.3]	-1.9 [-2.3, -1.6]	-0.20
H23	-0.7 [-1.1, -0.3]	-0.5 [-0.8, -0.1]	-1.26
(h) JAK2			
H24	-2.0 [-2.4, -1.6]	-2.1 [-2.7, -1.5]	-1.15
H27	-1.4 [-2.2, -0.6]	-1.2 [-1.7, -0.6]	
H14	-1.0 [-1.3, -0.7]	-0.8 [-1.4, -0.2]	
(i) Androgen receptor			
H2	-2.5 [-3.7, -1.3]	-2.5 [-2.8, -2.1]	-1.11
H7	-0.3 [-0.4, -0.2]	-0.27 [-0.32, -0.21]	
H5	3.5 [2.9, 4.0]	3.5 [2.9, 4.1]	

Overall PFS demonstrated excellent agreement with FEP with a mean unsigned difference of 0.3 [0.2,0.4] kcal/mol, where the 95% confidence square between square brackets was estimated with the bias-corrected and accelerated bootstrap method. PFS also shows good agreement with experiment in most systems with a mean unsigned error of 1.2 [0.9,1.5] kcal/mol and this is comparable to the mean unsigned difference of FEP and experiment, 1.1 [0.8,1.5]. To examine the agreement of these methods more closely the correlation of their results should be inspected. The correlation between PFS and FEP is good as seen in Figure (5) where fitting a trend line yields a R^2 value of 0.9 [0.6,1.0]. This correlation is favourably impacted by the uppermost top right data point and excluding this data point R^2 would be 0.8 [0.6 1.0]

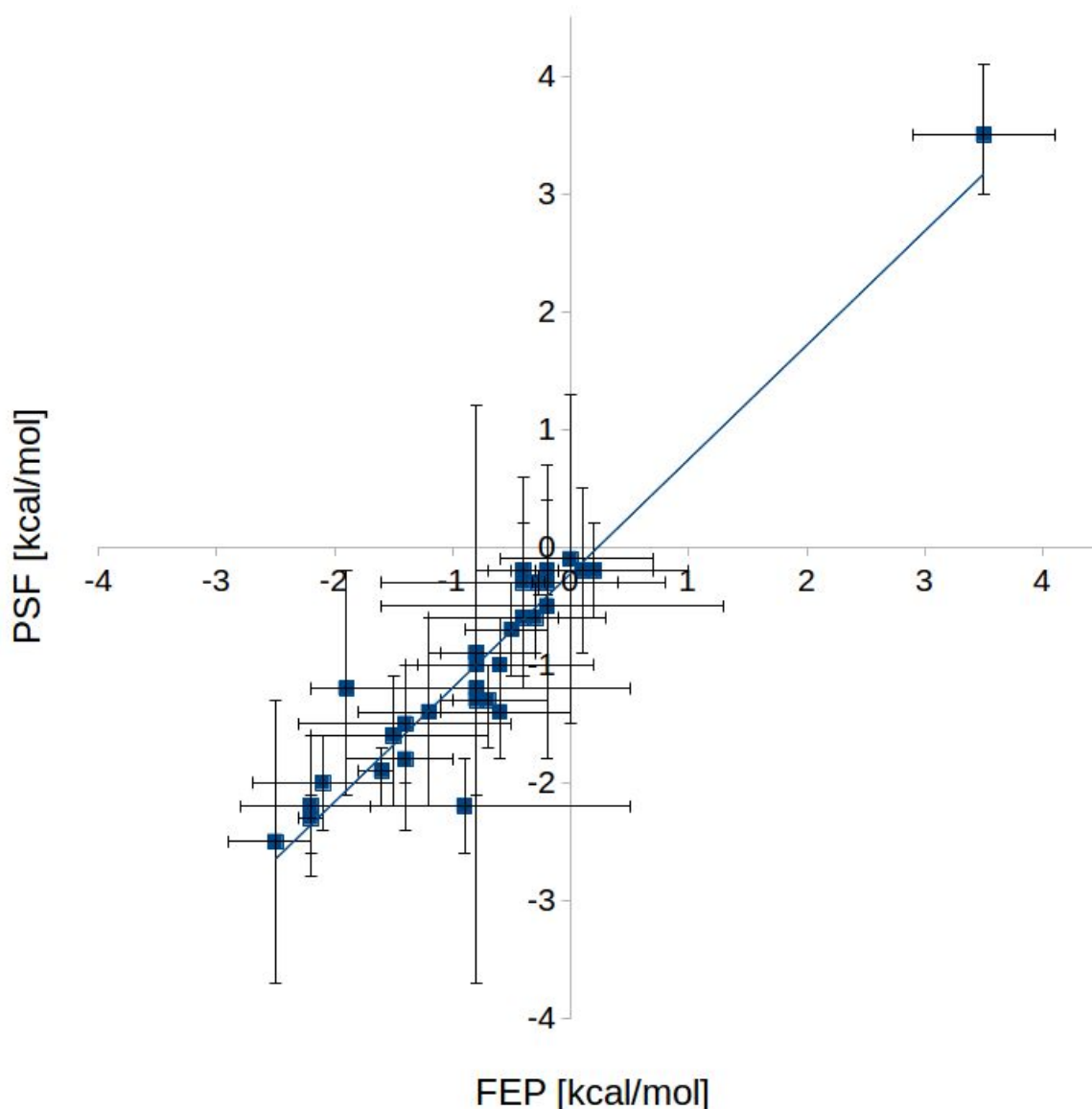


Figure (5): Plot for $\Delta\Delta G$ for fluorine mutants for calculations from perturbative fluorine scanning (PFS) and FEP for all data points in Table (2). $\Delta\Delta G$ s are reported as the means with error bars as 95% t-based confidence interval computed from the three independent replicate calculations.

In terms of agreement with experimental data, errors are greater than 2.0 kcal/mol in only three cases (all for DPP4). Overall poor correlation is seen, however, because the experimental range for $\Delta\Delta G$, 0.4 kcal/mol - 2.47kcal/mol, is small. Table (3) shows the correlation R^2 , mean unsigned difference and RMSD between the PFS, FEP and experimental data.

Table (3) - Correlation R^2 , mean unsigned difference, and RMSD between the PFS, FEP and experimental data for all data points in Table (2). 95% confidence intervals were estimated with the bias-corrected and accelerated bootstrap method and are reported between square brackets.

Comparison	R^2	Mean Unsigned Difference [kcal/mol]	RMSD [kcal/mol]
PFS vs FEP	0.9 [0.6,1.0]	0.3 [0.2,0.4]	0.4 [0.2,0.5]
FEP vs EXP	0.0 [0.0,0.3]	1.1 [0.8,1.5]	1.3 [1.0,1.6]
PFS vs EXP	0.0 [0.0,0.2]	1.2 [0.9,1.5]	1.3 [1.0,1.6]

Despite these poor correlations, both PFS and FEP have a reasonable accuracy in terms of mean unsigned difference.

Looking at each test case individually we see that PFS is a reasonably good predictor of the the mutant highlighted by experiment work. For systems Renin, Menin, P38, FXa, JAK2 and Androgen receptor PFS correctly predicts the mutant highlighted by experiment work. System DPP4 was more challenging, the top mutants, H5, H6 and H7 all have the same free energy. PFS ranks one of the best mutants, H7, as 3rd and incorrectly calculated the best mutant as H13 and second best as H16. FEP does better, again incorrectly ranking the best mutant as H13 but ranking two of the best experimental mutations H5 and H7 as second and third respectively. Whilst PFS and FEP are well agreed (within 1 kcal/mol) for this test case neither of these methods predict the best experimental mutant correctly. This may be due to the system preparation, modelling the DPP4 monomer rather than the dimer. Additionally it can be seen from Figure (2b) that H13 and H16 are on a phenyl already selected as favourable for chlorination and this may explain why PFS indicates these positions over the best position determined by experiment. For system CDK2, PFS fails to predict the top mutant however this failure is mirrored in FEP. The predictions made by PFS and FEP for the $\Delta\Delta G$ of the top experimental mutant agree within 1 kcal/mol. However neither are within 1 kcal/mol of the experimental $\Delta\Delta G$. PFS and FEP select H33 as the best position for fluorination. It can be seen in Figure (2(f)) that H33 is close to a position already selected as favourable for fluorination and this might explain why it selected over the position highlighted in the experimental work.

Conclusion

We have developed a new methodology for calculating relative binding affinities, which we term perturbative fluorine scanning. For a typical small molecule inhibitor, PFS applied to molecular dynamics simulations of a single molecule has the potential to combinatorially assess all possible fluorination sites yielding millions of predictions. These predictions can then be further assessed using more rigorous approaches and would be particularly useful in medicinal chemistry, providing insight for which analogs to synthesize. The results in Figure (4) suggest that the molecular dynamics simulations need to be run for at least 30.0 ns (or with multiple replicates) to reach converged predictions. This is likely due to the poor overlap between the two end states. This could be resolved by identifying mutants with poor overlap and running additional MD simulations in the mutant ensemble, such that a bi-directional estimator⁵⁹ could be applied. This would, however, incur additional computational cost. In its current state, using only one directional calculations, PFS consumes far less computational resources compared to traditional FEP approaches. For example, the FEP calculations in this

1
2
3 work used 24.0 ns of sampling for a single mutant whereas PFS used 50.0 ns of sampling for
4 all possible mutants (For the FXa test case this is 11 single hydrogens, 55 pairs of hydrogens
5 and increasingly more for additional mutations). As a point of reference the FXa case has
6 99,000 and 13,000 atoms in the complex and solvent systems respectively. Run in parallel
7 (see SI for parallelization strategy) across 4 NVIDIA P100 GPUs using OpenMM 7.3.0⁴⁹ and
8 CUDA 8.0 it takes approximately 8.5 hours to collect 50ns for both the complex and solvent
9 systems. Using the methods outlined above PFS analysis then takes 1 hour to calculate $\Delta\Delta G$
10 for all 11 mutant ligands. Comparatively with the same hardware and software full FEP takes
11 4 hours to compute $\Delta\Delta G$ for one mutant ligand. This improved computational speed allows for
12 more sampling in the PFS method, additionally all of the sampling in PFS is taken from
13 physical states, rather than alchemical intermediates in FEP. More sampling and longer
14 classical trajectories should provide a more thorough exploration of phase space compared to
15 FEP, particularly with reference to moderate conformational events on the order of 10's ns
16 which may be missed in short FEP lambda windows. The PFS method is simple and could
17 easily be improved by enhanced sampling techniques such as replica exchange⁶⁰ or solute
18 tempering.⁶¹ It represents one of few alchemical free energy methods⁶² that can be used
19 as a design tool, informing the user of which are the best mutations to make, rather than a tool
20 to assess specific ideas.
21
22
23
24
25

26 The change in binding affinity for a wide range of hydrogen to fluorine mutations has been
27 investigated computationally. Two computational methods were applied: FEP and PFS. It was
28 demonstrated that FEP and PFS are in excellent agreement. However, the correlation
29 between the computational methods and experiment for the $\Delta\Delta G$ calculations was not good.
30 This poor correlation could come from many potential sources, such as systematic errors in
31 the force fields or differences between computational and experimental systems. For example
32 DPP4 has the worst accuracy, compared to experiment, of any system investigated in this
33 work and this may stem from simulating it as a monomer compared to its dimer biological unit.
34 Another potential source of error could be the fluorination making a significant change to the
35 protonation, tautomeric, or conformational states of the ligand, an effect which we do not
36 account for here. Finally, differences in mass between fluorine and hydrogen are not
37 accounted for in either our FEP or PFS protocols and this may lead to differences in dynamics.
38 However, it will not affect the energies or free energies. This poor correlation does not raise a
39 major concern regarding the PFS method since it is similar for both FEP and PFS. Additionally,
40 the mean unsigned error for both methods remains low at 1.1 [0.8,1.5] kcal/mol and 1.2
41 [0.9,1.5] kcal/mol for FEP and PFS respectively and this is very close to 1.0 kcal/mol which
42 has been suggested as the acceptable error for free energy calculations. Where PFS performs
43 well is in reproducing FEP results (both $\Delta\Delta G$ values and rankings) with good correlation, $R^2 =$
44 0.9 [0.6, 1.0] and high accuracy, mean unsigned difference = 0.3 [0.2,0.4] kcal/mol at a fraction
45 of the computational cost. Combining this method with a more accurate forcefield could yield
46 a very useful tool. With many efforts continually undertaken to improve forcefield accuracy 64-
47 66. The scope could also be expanded significantly by considering additional mutations such
48 as chlorination, aromatic C to N and methylation. These changes, either alone or in
49 combination, can be considered from single simulations of the bound and unbound states.
50 With improved force fields, this method represents a powerful tool to explore chemical space
51 in an efficient and accurate way. In this work we have focused on hydrogen to fluorine changes
52 which are particularly attractive in a medicinal chemistry context due to the potential for fluorine
53 to act as a metabolic block in addition to a source of increased binding affinity⁶³. We envisage
54
55
56
57
58
59
60

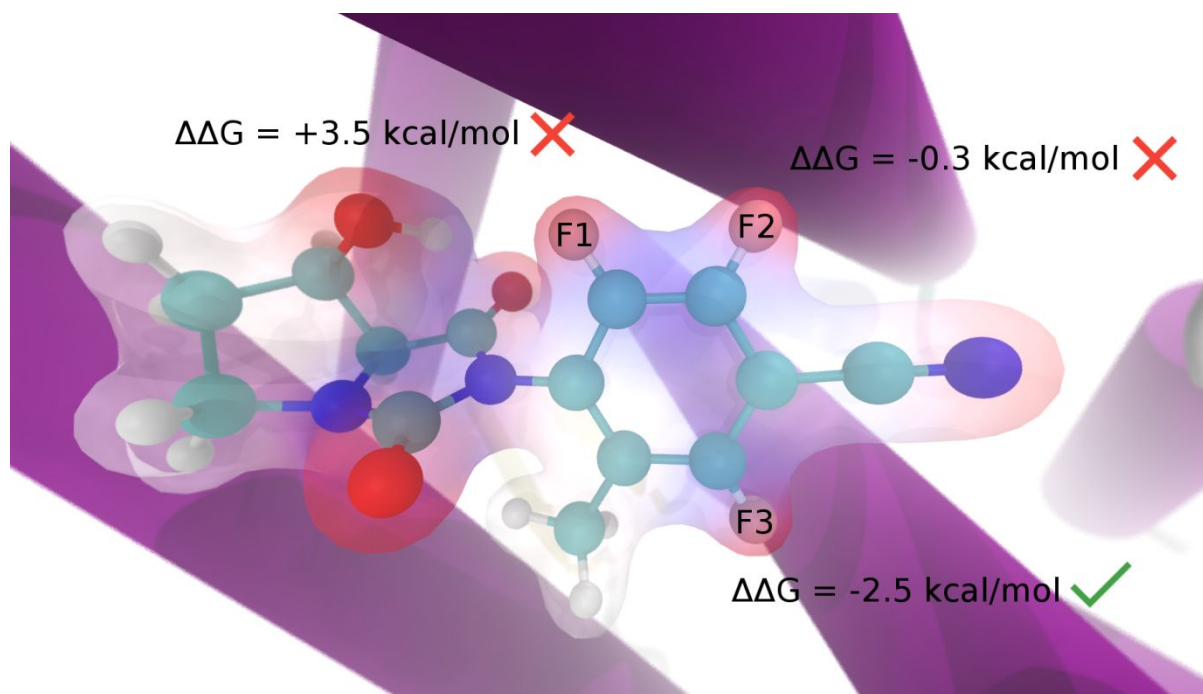
1
2
3 the use of PFS to identify hydrogen to fluorine changes predicted to increase binding affinity
4 in addition to hydrogen to fluorine changes predicted to improve ADME properties whilst
5 maintaining binding affinity.
6

7
8 Supporting Information Available: Table (S1) containing $\Delta\Delta G$ s for all hydrogen to fluorine
9 mutations highlighted in Figure (2), calculated by PFS. Figures (S1 - S9) presenting RMSD
10 plots for all complex systems. Details of computational parallelization strategies used in this
11 work are also provided.
12
13

14 **Acknowledgments**

15
16 Work in the D.J.H. laboratory was supported by the Medical Research Council under grant
17 ML/L007266/1. A.D.W. would like to acknowledge the EPSRC Centre for Doctoral Training in
18 Computational Methods for Materials Science for funding under grant number EP/L015552/1.
19 A.R. would like to acknowledge John Chodera (ORCID 0000-0003-0542-119X) for support
20 and enlightening discussions and suggestions that are reflected in this manuscript. A.R. also
21 acknowledges partial support from the Sloan Kettering Institute and the Tri-Institutional
22 Program in Computational Biology and Medicine. All calculations were performed using the
23 Darwin Supercomputer of the University of Cambridge High Performance Computing Service
24 ([http:// www.hpc.cam.ac.uk/](http://www.hpc.cam.ac.uk/)) and were funded by the EPSRC under grant EP/P020259/1.
25
26
27
28
29
30
31
32
33
34
35
36
37
38

39 **Table of contents graphic:**
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



Notes

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

References

- (1) Hyohdoh, I.; Furuichi, N.; Aoki, T.; Itezono, Y.; Shirai, H.; Ozawa, S.; Watanabe, F.; Matsushita, M.; Sakaitani, M.; Ho, P.-S.; Takanashi, K.. Fluorine Scanning by Nonselective Fluorination: Enhancing Raf/MEK Inhibition While Keeping Physicochemical Properties. *ACS Medicinal Chemistry Letters*. **2013**, *4*, pp 1059–1063. <https://doi.org/10.1021/ml4002419>.
- (2) Lou, Y.; Sweeney, Z. K.; Kuglstatter, A.; Davis, D.; Goldstein, D. M.; Han, X.; Hong, J.; Kocer, B.; Kondru, R. K.; Litman, R.; McIntosh, J. Finding the Perfect Spot for Fluorine: Improving Potency up to 40-Fold during a Rational Fluorine Scan of a Bruton's Tyrosine Kinase (BTK) Inhibitor Scaffold. *Bioorganic & Medicinal Chemistry Letters*. **2015**, *25*, pp 367–371. <https://doi.org/10.1016/j.bmcl.2014.11.030>.
- (3) Morgenthaler, M.; Aebi, J. D.; Grüninger, F.; Mona, D.; Wagner, B.; Kansy, M.; Diederich, F. A Fluorine Scan of Non-Peptidic Inhibitors of Neprilysin: Fluorophobic and Fluorophilic Regions in an Enzyme Active Site. *J. Fluor. Chem.* **2008**, *129*, 852–865.
- (4) Naret, T.; Bignon, J.; Bernadat, G.; Benchekroun, M.; Levaique, H.; Lenoir, C.; Dubois, J.; Pruvost, A.; Saller, F.; Borgel, D.; Manoury, B. A Fluorine Scan of a Tubulin Polymerization Inhibitor Isocombretastatin A-4: Design, Synthesis, Molecular Modelling, and Biological Evaluation. *Eur. J. Med. Chem.* **2018**, *143*, 473–490.
- (5) Olsen, J. A.; Banner, D. W.; Seiler, P.; Sander, U. O.; D'Arcy, A.; Stihle, M.; Müller, K.; Diederich, F. A Fluorine Scan of Thrombin Inhibitors to Map the Fluorophilicity/Fluorophobicity of an Enzyme Active Site: Evidence for C-F...C-O Interactions. *Angewandte Chemie International Edition*. **2003**, *42*, pp 2507–2511. <https://doi.org/10.1002/anie.200351268>.
- (6) Schweizer, E.; Hoffmann-Röder, A.; Schäfer, K.; Olsen, J. A.; Fäh, C.; Seiler, P.; Obst-Sander, U.; Wagner, B.; Kansy, M.; Diederich, F. A Fluorine Scan at the Catalytic Center of Thrombin: C-F, C-OH, and C-OMe Bioisosterism and Fluorine Effects on pKa

- and logD Values. *ChemMedChem*. **2006**, 1, pp 611–621. <https://doi.org/10.1002/cmdc.200600015>.
- (7) Gillis, E. P.; Eastman, K. J.; Hill, M. D.; Donnelly, D. J.; Meanwell, N. A. Applications of Fluorine in Medicinal Chemistry. *Journal of Medicinal Chemistry*. **2015**, 58, pp 8315–8359. <https://doi.org/10.1021/acs.jmedchem.5b00258>.
- (8) Meanwell, N. A. Fluorine and Fluorinated Motifs in the Design and Application of Bioisosteres for Drug Design. *J. Med. Chem.* **2018**, 61, pp 5822–5880. <https://doi.org/10.1021/acs.jmedchem.7b01788>.
- (9) 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. 7-Fluorindazoles as Potent and Selective Inhibitors of Factor Xa†. *Journal of Medicinal Chemistry*. **2008**, 51, pp 282–297. <https://doi.org/10.1021/jm701217r>.
- (10) Kuhn, B.; Kollman, P. A. A Ligand That Is Predicted to Bind Better to Avidin than Biotin: Insights from Computational Fluorine Scanning. *J. Am. Chem. Soc.* **2000**, 122, 3909–3916.
- (11) 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.
- (12) Jorgensen, W. L. Free Energy Calculations: A Breakthrough for Modeling Organic Chemistry in Solution. *Acc. Chem. Res.* **1989**, 22, 184–189.
- (13) Michel, J.; Foloppe, N.; Essex, J. W. Rigorous Free Energy Calculations in Structure-Based Drug Design. *Mol. Inform.* **2010**, 29, 570–578.
- (14) Wang, L.; Wu, Y.; Deng, Y.; Kim, B.; Pierce, L.; Krilov, G.; Lupyan, D.; Robinson, S.; Dahlgren, M. K.; Greenwood, J.; Romero, D.L. 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**, 137, pp 2695–2703. <https://doi.org/10.1021/ja512751q>.
- (15) 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, 638–649.
- (16) Seeliger, D.; de Groot, B. L. Protein Thermostability Calculations Using Alchemical Free Energy Simulations. *Biophys. J.* **2010**, 98, 2309–2316.
- (17) Loeffler, H. H.; Michel, J.; Woods, C. FESetup: Automating Setup for Alchemical Free Energy Simulations. *J. Chem. Inf. Model.* **2015**, 55, 2485–2490.
- (18) 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.
- (19) 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, 966–971.
- (20) Aldeghi, M.; Heifetz, A.; Bodkin, M. J.; Knapp, S.; Biggin, P. C. Accurate Calculation of the Absolute Free Energy of Binding for Drug Molecules. *Chem. Sci.* **2016**, 7, 207–218.
- (21) Kollman, P. Free Energy Calculations: Applications to Chemical and Biochemical Phenomena. *Chem. Rev.* **1993**, 93, 2395–2417.
- (22) Zwanzig, R. W. Erratum : High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *The Journal of Chemical Physics*. **1954**, 22, pp 2099–2099. <https://doi.org/10.1063/1.1740022>.
- (23) 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.
- (24) Woo, H.-J.; Woo, H.; Roux, B. Calculation of Absolute Protein-Ligand Binding Free Energy from Computer Simulations. *Proceedings of the National Academy of Sciences*. **2005**, 102, pp 6825–6830. <https://doi.org/10.1073/pnas.0409005102>.
- (25) Lybrand, T. P.; McCammon, J. A.; Wipff, G. Theoretical Calculation of Relative

- 1
2
3 Binding Affinity in Host-Guest Systems. *Proc. Natl. Acad. Sci. U. S. A.* **1986**, *83*, 833–
4 835.
- 5 (26) Cole, D. J.; Janecek, M.; Stokes, J. E.; Rossmann, M.; Faver, J. C.; McKenzie, G. J.;
6 Venkitaraman, A. R.; Hyvönen, M.; Spring, D. R.; Huggins, D. J.; Jorgensen, W.L.
7 Computationally-Guided Optimization of Small-Molecule Inhibitors of the Aurora A
8 kinase–TPX2 Protein–protein Interaction. *Chemical Communications*. **2017**, *53*, pp
9 9372–9375. <https://doi.org/10.1039/c7cc05379g>.
- 10 (27) Alder, B. J.; Wainwright, T. E. Studies in Molecular Dynamics. I. General Method. *J.*
11 *Chem. Phys.* **1959**, *31*, 459–466.
- 12 (28) McCammon, J. A.; Gelin, B. R.; Karplus, M. Dynamics of Folded Proteins. *Nature*
13 **1977**, *267*, 585–590.
- 14 (29) Levitt, M.; Lifson, S. Refinement of Protein Conformations Using a Macromolecular
15 Energy Minimization Procedure. *J. Mol. Biol.* **1969**, *46*, 269–279.
- 16 (30) Lifson, S.; Warshel, A. Consistent Force Field for Calculations of Conformations,
17 Vibrational Spectra, and Enthalpies of Cycloalkane and n-Alkane Molecules. *J. Chem.*
18 *Phys.* **1968**, *49*, 5116–5129.
- 19 (31) Liu, H.; Mark, A. E.; van Gunsteren, W. F. Estimating the Relative Free Energy of
20 Different Molecular States with Respect to a Single Reference State. *J. Phys. Chem.*
21 **1996**, *100*, 9485–9494.
- 22 (32) Oostenbrink, C.; Van Gunsteren, W. F. Single-Step Perturbations to Calculate Free
23 Energy Differences from Unphysical Reference States: Limits on Size, Flexibility, and
24 Character. *Journal of Computational Chemistry*. **2003**, *24*, pp 1730–1739.
25 <https://doi.org/10.1002/jcc.10304>.
- 26 (33) Oostenbrink, C.; van Gunsteren, W. F. Free Energies of Ligand Binding for
27 Structurally Diverse Compounds. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 6750–6754.
- 28 (34) Mordasini, T. Z.; Andrew McCammon, J. Calculations of Relative Hydration Free
29 Energies: A Comparative Study Using Thermodynamic Integration and an Extrapolation
30 Method Based on a Single Reference State. *The Journal of Physical Chemistry B*. **2000**,
31 *104*, pp 360–367. <https://doi.org/10.1021/jp993102o>.
- 32 (35) Raman, E. P.; Prabhu Raman, E.; Vanommeslaeghe, K.; MacKerell, A. D. Site-
33 Specific Fragment Identification Guided by Single-Step Free Energy Perturbation
34 Calculations. *Journal of Chemical Theory and Computation*. **2012**, *8*, pp 3513–3525.
35 <https://doi.org/10.1021/ct300088r>.
- 36 (36) Raman, E. P.; Prabhu Raman, E.; Lakkaraju, S. K.; Denny, R. A.; MacKerell, A. D.
37 Estimation of Relative Free Energies of Binding Using Pre-Computed Ensembles Based
38 on the Single-Step Free Energy Perturbation and the Site-Identification by Ligand
39 Competitive Saturation Approaches. *Journal of Computational Chemistry*. **2017**, *38*, pp
40 1238–1251. <https://doi.org/10.1002/jcc.24522>.
- 41 (37) Boresch, S.; Woodcock, H. L. Convergence of Single-Step Free Energy Perturbation.
42 *Mol. Phys.* **2017**, *115*, 1200–1213.
- 43 (38) Stroet, M.; Koziara, K. B.; Malde, A. K.; Mark, A. E. Optimization of Empirical Force
44 Fields by Parameter Space Mapping: A Single-Step Perturbation Approach. *Journal of*
45 *Chemical Theory and Computation*. **2017**, *13*, pp 6201–6212.
46 <https://doi.org/10.1021/acs.jctc.7b00800>.
- 47 (39) Eastman, P.; Friedrichs, M. S.; Chodera, J. D.; Radmer, R. J.; Bruns, C. M.; Ku, J. P.;
48 Beauchamp, K. A.; Lane, T. J.; Wang, L.-P.; Shukla, D.; Tye, T. OpenMM 4: A
49 Reusable, Extensible, Hardware Independent Library for High Performance Molecular
50 Simulation. *J. Chem. Theory Comput.* **2013**, *9*, 461–469.
- 51 (40) Matter, H.; Scheiper, B.; Steinhagen, H.; Böcskei, Z.; Fleury, V.; McCort, G.
52 Structure-Based Design and Optimization of Potent Renin Inhibitors on 5- or 7-
53 Azaindole-Scaffolds. *Bioorg. Med. Chem. Lett.* **2011**, *21*, 5487–5492.
- 54 (41) Peters, J.-U.; Weber, S.; Ritter, S.; Weiss, P.; Wallier, A.; Boehringer, M.; Hennig,
55 M.; Kuhn, B.; Loeffler, B.-M. Aminomethylpyrimidines as Novel DPP-IV Inhibitors: A
56 10(5)-Fold Activity Increase by Optimization of Aromatic Substituents. *Bioorg. Med.*
57 *Chem. Lett.* **2004**, *14*, 1491–1493.

- 1
2
3 (42) He, S.; Senter, T. J.; Pollock, J.; Han, C.; Upadhyay, S. K.; Purohit, T.; Gogliotti, R.
4 D.; Lindsley, C. W.; Cierpicki, T.; Stauffer, S. R.; Grembecka, J. High-Affinity Small-
5 Molecule Inhibitors of the Menin-Mixed Lineage Leukemia (MLL) Interaction Closely
6 Mimic a Natural Protein–Protein Interaction. *J. Med. Chem.* **2014**, *57*, 1543–1556.
- 7 (43) Aiguadé, J.; Balagué, C.; Carranco, I.; Caturla, F.; Domínguez, M.; Eastwood, P.;
8 Esteve, C.; González, J.; Lumeras, W.; Orellana, A.; Preciado, S. Novel
9 Triazolopyridylbenzamides as Potent and Selective p38 α Inhibitors. *Bioorg. Med. Chem.*
10 *Lett.* **2012**, *22*, 3431–3436.
- 11 (44) Jones, C. D.; Andrews, D. M.; Barker, A. J.; Blades, K.; Daunt, P.; East, S.; Geh, C.;
12 Graham, M. A.; Johnson, K. M.; Loddick, S. A.; McFarland, H.M. The Discovery of
13 AZD5597, a Potent Imidazole Pyrimidine Amide CDK Inhibitor Suitable for Intravenous
14 Dosing. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 6369–6373.
- 15 (45) Freeman-Cook, K. D.; Autry, C.; Borzillo, G.; Gordon, D.; Barbacci-Tobin, E.;
16 Bernardo, V.; Briere, D.; Clark, T.; Corbett, M.; Jakubczak, J.; McFarland, H.M. Design
17 of Selective, ATP-Competitive Inhibitors of Akt. *J. Med. Chem.* **2010**, *53*, 4615–4622.
- 18 (46) Ledebøer, M. W.; Pierce, A. C.; Duffy, J. P.; Gao, H.; Messersmith, D.; Salituro, F.
19 G.; Nanthakumar, S.; Come, J.; Zuccola, H. J.; Swenson, L.; Shlyakter, D. 2-
20 Aminopyrazolo[1,5-A]pyrimidines as Potent and Selective Inhibitors of JAK2. *Bioorg.*
21 *Med. Chem. Lett.* **2009**, *19*, 6529–6533.
- 22 (47) Hamann, L. G.; al., E. al et. Tandem Optimization of Target Activity and Elimination
23 of Mutagenic Potential in a Potent Series of N-Aryl Bicyclic Hydantoin-Based Selective
24 Androgen Receptor Modulators. *ChemInform.* **2007**, *17*, 1860-1864.
25 <https://doi.org/10.1002/chin.200733119>.
- 26 (48) Berman, H. M.; Westbrook, J.; Feng, Z. The Protein Data Bank. *Nucleic acids* **2000**.
- 27 (49) Eastman, P.; Swails, J.; Chodera, J. D. OpenMM 7: Rapid Development of High
28 Performance Algorithms for Molecular Dynamics. *PLoS Comput. Biol.* **2017**.
- 29 (50) Sastry, G. M.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W. Protein and
30 Ligand Preparation: Parameters, Protocols, and Influence on Virtual Screening
31 Enrichments. *J. Comput. Aided Mol. Des.* **2013**, *27*, 221–234.
- 32 (51) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling,
33 C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters
34 from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713.
- 35 (52) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Antechamber: An Accessory
36 Software Package for Molecular Mechanical Calculations. *J. Am. Chem. Soc.* **2001**,
37 *222*, U403.
- 38 (53) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and
39 Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.
- 40 (54) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, Efficient Generation of High-quality Atomic
41 Charges. AM1-BCC Model: II. Parameterization and Validation. *Journal of*
42 *computational* **2002**, *23*, 1623-1641.
- 43 (55) Rizzi, A.; Chodera, J.; Naden, L.; Beauchamp, K.; Albanese, S.; Grinaway, P.;
44 Rustenburg, B.; Saladi, S.; Boehm, K. *Choderalab/yank: Bugfix Release*; **2018**.
45 <https://doi.org/10.5281/zenodo.1447109>.
- 46 (56) Horn, H. W.; Swope, W. C.; Pitara, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.;
47 Head-Gordon, T. Development of an Improved Four-Site Water Model for Biomolecular
48 Simulations: TIP4P-Ew. *J. Chem. Phys.* **2004**, *120*, 9665–9678.
- 49 (57) Allen, F. H.; Kennard, O.; Watson, D. G.; Brammer, L.; Orpen, A. G.; Taylor, R.
50 Tables of Bond Lengths Determined by X-Ray and Neutron Diffraction. Part 1. Bond
51 Lengths in Organic Compounds. *J. Chem. Soc. Perkin Trans. 2* **1987**, , S1–S19.
- 52 (58) Shirts, M. R.; Chodera, J. D. Statistically Optimal Analysis of Samples from Multiple
53 Equilibrium States. *J. Chem. Phys.* **2008**, *129*, 124105.
- 54 (59) Bennett, C. H. Efficient Estimation of Free Energy Differences from Monte Carlo
55 Data. *J. Comput. Phys.* **1976**, *22*, 245–268.
- 56 (60) Rao, F.; Caflisch, A. Replica Exchange Molecular Dynamics Simulations of
57 Reversible Folding. *J. Chem. Phys.* **2003**, *119*, 4035–4042.
- 58
59
60

- 1
2
3 (61) Liu, P.; Kim, B.; Friesner, R. A.; Berne, B. J. Replica Exchange with Solute
4 Tempering: A Method for Sampling Biological Systems in Explicit Water. *Proceedings of*
5 *the National Academy of Sciences*. **2005**, 102, pp 13749–13754.
6 <https://doi.org/10.1073/pnas.0506346102>.
7
8 (62) Irwin, B. W. J.; Huggins, D. J. Estimating Atomic Contributions to Hydration and
9 Binding Using Free Energy Perturbation. *J. Chem. Theory Comput.* **2018**, 14, 3218–
10 3227.
11 (63) Shah, P.; Westwell, A. D. The Role of Fluorine in Medicinal Chemistry. *J. Enzyme*
12 *Inhib. Med. Chem.* **2007**, 22, 527–540.
13 (64) Roos, K.; Wu, C.; Damm, W.; Reboul, M.; Stevenson, J. M.; Lu, C.; Dahlgren, M. K.;
14 Mondal, S.; Chen, W.; Wang, L.; Abel, R. OPLS3e: Extending Force Field Coverage for
15 Drug-Like Small Molecules. *J. Chem. Theory Comput.* **2019**.
16 <https://doi.org/10.1021/acs.jctc.8b01026>.
17
18 (65) Vanommeslaeghe, K.; MacKerell Jr, A.D. Automation of the CHARMM General Force
19 Field (CGenFF) I: bond perception and atom typing. *J. Chem. Inf. Model.* 2012, 52,
20 3144-3154.
21
22 (66) Wang, L.P.; McKiernan, K.A.; Gomes, J.; Beauchamp, K.A.; Head-Gordon, T.; Rice,
23 J.E.; Swope, W.C.; Martínez, T.J.; Pande, V.S. Building a more predictive protein force
24 field: a systematic and reproducible route to AMBER-FB15. *J. Phys. Chem. B.* 2017,
25 121, 4023-4039.
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60