Evaluating the use of Absolute Binding Free Energy in the fragment optimization process

Irfan Alibay[a], Aniket Magarkar[b], Daniel Seeliger[b], Philip C. Biggin[a],

[a]The University of Oxford, Department of Biochemistry, South Parks Road, Oxford, UK, OX1 3QU

[b]Boehringer Ingelheim Pharma GmbH & Co. KG, Birkendorfer Str. 65, 88397 Biberach an de Riß, Germany

*Correspondence to: philip.biggin@bioch.ox.ac.uk

ORCIDs

Irfan Alibay – 0000-0001-5787-9130

Aniket Magarkar – 0000-0003-3385-964X

Daniel Seeliger – 0000-0002-4119-6068

Philip C. Biggin – 0000-0001-5100-8836

Abstract

Key to the fragment optimization process is the need to accurately capture the changes in affinity that are associated with a given set of chemical modifications. Due to the weakly binding nature of fragments, this has proven to be a challenging task, despite recent advancements in leveraging experimental and computational methods. In this work, we evaluate the use of Absolute Binding Free Energy (ABFE) calculations in guiding fragment optimization decisions, retrospectively calculating binding free energies for 59 ligands across 4 fragment elaboration campaigns. We first demonstrate that ABFEs can be used to accurately rank fragment-sized binders with an overall Spearman's r of 0.89 and a Kendall τ of 0.67, although often deviating from experiment in absolute free energy values with an RMSE of 2.75 kcal/mol. We then also show that in several cases, retrospective fragment optimization decisions can be supported by the ABFE calculations. Cases that were not supported were often limited by large uncertainties in the free energy estimates. Comparing against cheaper endpoint methods, namely N_{wat}-MM/GBSA, we find that ABFEs offer better ranking power and correlation metrics. Our results indicate that ABFE calculations can usefully guide fragment elaborations to maximize affinity.

Introduction

Over the last few decades, the fragment-based drug design (FBDD) process has matured into a popular and effective approach to designing novel binders[1]. Indeed during the 2015 to 2019 period, over 131 successful fragment-to-lead campaigns were published[2-6]. With an intuitive structure-based approach, and ability to more easily sample a large chemical space, FBDD has become a strong contender to more traditional high-throughput screening methods.

In the FBDD process, a library of fragments is first screened against a given protein target in order to identify potential binders. These fragments usually adhere to the so-called "rule of three" [7], having a molecular weight ≤ 300 Da, ClogP ≤ 3, a number of hydrogen bond donors ≤ 3 and a number of hydrogen bond acceptors ≤ 3. From this initial screen, binders are then identified through structural (e.g. X-ray crystallography, nuclear magnetic resonance (NMR)), biochemical, or biophysical (e.g. NMR, surface plasmon resonance (SPR), isothermal titration calorimetry (ITC)) characterisation[8-10]. Confirmed binders are then optimised, heavily relying on structure activity relationships to create bespoke high affinity binders. Optimising strategies either concentrate on improving the affinity of a single binder through fragment *growing*, or by combining multiple fragments occupying distinct binding sites through *linking* and *merging* decisions[11,12].

Accurately characterising fragment-protein interactions is therefore central to the FBDD process. Despite substantial improvements in leveraging *in vitro* methods for fragment screening[1], this remains a challenging task. By their nature fragments tend to be low affinity binders, somewhere in the millimolar to micromolar range, with a propensity to access multiple binding sites in a protein target. Not only do highly sensitive affinity measurements need to be employed, something that cannot always be readily used for large fragment screens, but multiple orthogonal methods are often required to validate low affinity hits[1,13].

As a consequence, *in silico* methods have become increasingly popular in helping guide and support FBDD decisions[14,15]. A variety of approaches have been employed for this task, which can roughly be separated into two categories; i) methods to identify fragment interaction sites, and ii) methods to characterise fragment binding affinities. For the former, both simple methods such as molecular docking[16], and more complex molecular dynamics-based approaches such as hotspot mapping[17 18 19 20 21] or unbiased molecular dynamics (MD) with Markov-state modelling[22] have shown success in identifying potential fragment binding sites on protein targets. In terms of characterising affinity, alchemical relative binding free energy (RBFE) methods have been particularly successful in ranking fragment affinities [15,23-26]. Of

particular note is a 2015 study by Steinbrecher *et al.*[26], which demonstrated that the FEP+ RBFE tool could be used to successfully rank fragment affinities, achieving an RMSE of 1.14 kcal/mol for 96 ligands across 8 fragment optimisation campaigns. Whilst RBFEs are a powerful and relatively cheap tool for this purpose, there are some disadvantages that limit its applicability in FBDD. Firstly, RBFEs do not give a direct measure of ligand binding affinities on an absolute scale, instead the method relies on other methods to normalise output free energy values. Often this would be done via an *in vitro* experimental measurement of a few select compounds, however as discussed above it can often be difficult to do so accurately for fragments and therefore is not always tractable in FBDD. The second major disadvantage is that RBFE protocols are generally developed to investigate small chemical perturbations on a given common chemical scaffold.   This can have limited applicability in some FBDD campaigns, which will often investigate several fragments with different chemical scaffolds, sometimes even in different binding site locations.   Nevertheless, we note some recent success in using RBFE methods for fragment linking purposes, although requiring several extra intermediate steps to achieve good results[27,28].

As an alternative to RBFE methods, we propose that absolute binding free energies (ABFE) [29-33] could instead be used to directly investigate fragment affinities. Despite larger computational costs, ABFEs offer a direct solution to the above described limitations of RBFEs by directly calculating the absolute free energy of binding for each individual ligand, and not requiring a transformation to another chemical entity. Previous works have shown ABFEs to offer highly accurate estimates of binding free energies across a variety of target systems, although often at increased computational costs[31].   Indeed the idea of using ABFEs in fragment binding is not a novel one, and we note several other investigations employing such methods for fragment-sized molecules[15,30,34,35].   That being said these have been mostly limited in scope and to our knowledge there have yet to be any large scale analyses of the applicability of ABFEs to the FBDD process.

Here we specifically look at evaluating the use of ABFEs in the fragment optimisation process. Retrospectively calculating the binding free energies for 59 ligands across four FBDD campaigns, we aim to investigate whether; a) ABFEs offer comparable results to *in vitro* affinity measures, and b) ABFEs could be used to achieve similar synthetic decisions in fragment optimisation. We also look at how ABFEs compare against cheaper methods, namely N_{wat}-MM/GBSA[36,37].

Methods

System selection

Four previously published FBDD campaigns[38-41] (**Fig. 1**), each with a different protein target, were selected for this benchmark. The systems were selected based on the availability of high quality experimental measures, with the presence of both crystallographic and affinity data. We also looked to cover a wide activity range, spanning from millimolar to nanomolar affinities, and various different fragment optimisation cases, such as the presence of multiple distinct binding sites, changes in net charge, fragment growing and fragment merging (but no fragment linking cases). Note that some cases were specifically avoided to ensure this study remained tractable. For example, large multimeric protein targets were not included to reduce computational costs. Similarly, membrane proteins and systems with binding site metals or known protonation issues were not chosen to avoid introducing additional complexity in this initial dataset.
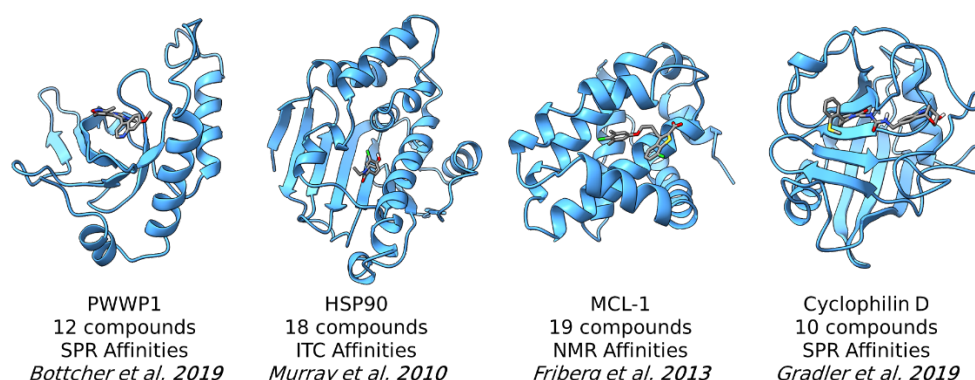

PWWP1	HSP90	MCL-1	Cyclophilin D
12 compounds	18 compounds	19 compounds	10 compounds
SPR Affinities	ITC Affinities	NMR Affinities	SPR Affinities
Bottcher et al. 2019	*Murray et al. 2010*	*Friberg et al. 2013*	*Gradler et al. 2019*

Fig. 1. *Overview of the fragment elaboration datasets. A total of 59 ligands from four different elaboration studies[38-41] with affinities spanning the millimolar to micromolar range are investigated here.*

System preparation and simulation details

As structural data was not available for all protein-ligand complexes, the initial configurations were generated by modifying a chemically close protein-ligand crystal structure using open source PyMOL[42] (see **SI Text S1** for full details). If alternate states were available in the starting crystal, state A was always retained. Where appropriate, acetyl and N-methyl caps were also added to protein structures using PyMOL. Protons were assigned using protoss[43,44], as made available within the ProteinPlus web server[45,46]. In the case of the PWWP1 domain, missing loop residues were modelled using modeller 9v21[47] and the DOPE-HR method. Of

the 500 generated models, the top 10 models were rescored using the QMEAN[48] scoring function as implemented in SWISS-MODEL[49], with the final model selected as the best QMEAN-scored model. The PWWP1 N-termini was also truncated at isoleucine 393 to avoid long timescale interactions between residues in the disordered N-termini region and the ligand binding site (see **SI Text S2**).

Solvation and force field assignment was achieved through a combination of AmberTools 18[50] and GROMACS 2019[51]. Ligand parameters and partial charges were assigned using the GAFF2 force field and the AM1-BCC partial charge model. The ff99SB-ILDN force field was used for protein parameters and the complexes were solvated in TIP3P[52] cubic boxes with a minimum distance of 12Å from the solute to the box edge using GROMACS' *solvate* module. Sodium and chloride ions were added to neutralise the systems and achieve a concentration of 150 mM using the *genion* module of GROMACS. ParmEd version 3.2.0 (https://github.com/ParmEd/ParmEd) was used to convert input topologies and coordinates between AMBER and GROMACS file formats.

Unless otherwise mentioned, a hydrogen mass repartitioning scheme (HMR)[53,54] was used to achieve a 4 fs integration timestep for our simulations. Due to differences in how MD engines handle hydrogens with analytical constraints[54], hydrogen masses, except those of waters, were increased to 3 and 4 atomic mass units for AMBER and GROMACS respectively. Whilst HMR has been employed before in alchemical free energy calculations[55-57], there is limited data on its use in absolute binding free energies in GROMACS. To this end, a small validation using the Cyclophilin D dataset is shown in Supplementary Information (see **SI Text S3**). Water hydrogen motions were constrained using SETTLE [58] and either SHAKE[59] or LINCS[60,61] for other constraints in AMBER or GROMACS respectively. In all cases, simulation temperature was maintained at 298.15 K though Langevin dynamics with a collision frequency of 2 ps^{-1}. A simulation pressure of 1 atmosphere was maintained using various barostats. For AMBER simulations, a Monte Carlo barostat[62] was employed with a volume exchange attempt frequency of 100 ps. For GROMACS, the initial equilibration steps used the Berendsen barostat[63] with a time constant of 1 ps, followed by the Parrinello-Rahman barostat[64] with a time constant of 2.0 ps for all follow-on equilibration and production simulations. In all cases, a cut-off of 1 nm was used for short range interactions, and long range electrostatics are handled via PME[65,66]. Input topologies, coordinates, and simulation control files are provided as Supplementary Information (zenodo: 10.5281/zenodo.5913469 and https://github.com/bigginlab/fragment-opt-abfe-benchmark).

System equilibration

A two-step equilibration procedure is followed here. First, binding site waters were equilibrated using the AMBER MC/MD procedure[67] as implemented in AMBER18's pmemd.cuda engine[68,69] . The system first underwent a short initial equilibration phase consisting of 10,000 steps of minimisation, followed by a 500 ps NVT phase, and then 5 ns of NPT. During this equilibration, protein backbone atoms and non-hydrogen ligand atoms were restrained using a 5 kcal/mol/Å^2 force constant. The system then underwent a 5 ns MC/MD with the same positional restraints with 25 000 MC attempts every 1 000 MD steps. This was followed by a further 10 ns of MC/MD water exchange with position restraints on the ligand removed. In all MC/MD simulations, the NVT ensemble was sampled and the MC swap region box was trimmed to ensure as many exchanges with binding site waters as possible.

Once complete, the final frame from the MC/MD procedure was converted to GROMACS using ParmEd. The system was then further equilibrated using the *mdrun* engine in GROMACS 2019[51]. This included an initial 10,000 step minimisation, followed by 1 ns of restrained (2.39 kcal/mol/Å^2 applied to the protein backbone and ligand non-hydrogen atoms) NVT and NPT equilibration. The latter step employed the Berendsen barostat[63] as detailed above. The system was then relaxed using 5 ns of unrestrained NPT simulation using the Parrinello-Rahman barostat[64]. Finally, a further 20 ns of NPT simulation was generated. This final 20 ns simulation was used to both derive Boresch-style[70] orientational restraint parameters and for N_{wat}-MM/GBSA analysis.

Absolute binding free energy calculations

Here we employed an ABFE protocol similar to the one previously described by Aldeghi *et al.*[32,33]. Following the equilibration phase, a partial decoupling scheme is employed to trace the alchemical path from a fully interacting protein ligand complex to a ligand in solution as shown in **Fig. 2**. This partial decoupling scheme involves annihilating ligand partial charges through 11 windows spaced at λ intervals of 0.1 from each other. This is then followed by 21 Van der Waals decoupling windows spaced with the following λ schedule [0.0, 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95, 1.0]. A soft-core potential for decoupled Van der Waals interactions was used[71]. Additionally, to restrict ligand motion in the complex phase an orientational restraint, as defined by Boresch *et al.*[70], was employed. This restraint was applied over 12 windows in the complex decoupling phase

with the following schedule [0.0 0.01, 0.025, 0.05, 0.075, 0.1, 0.15, 0.2, 0.35, 0.5, 0.75, 1.0].
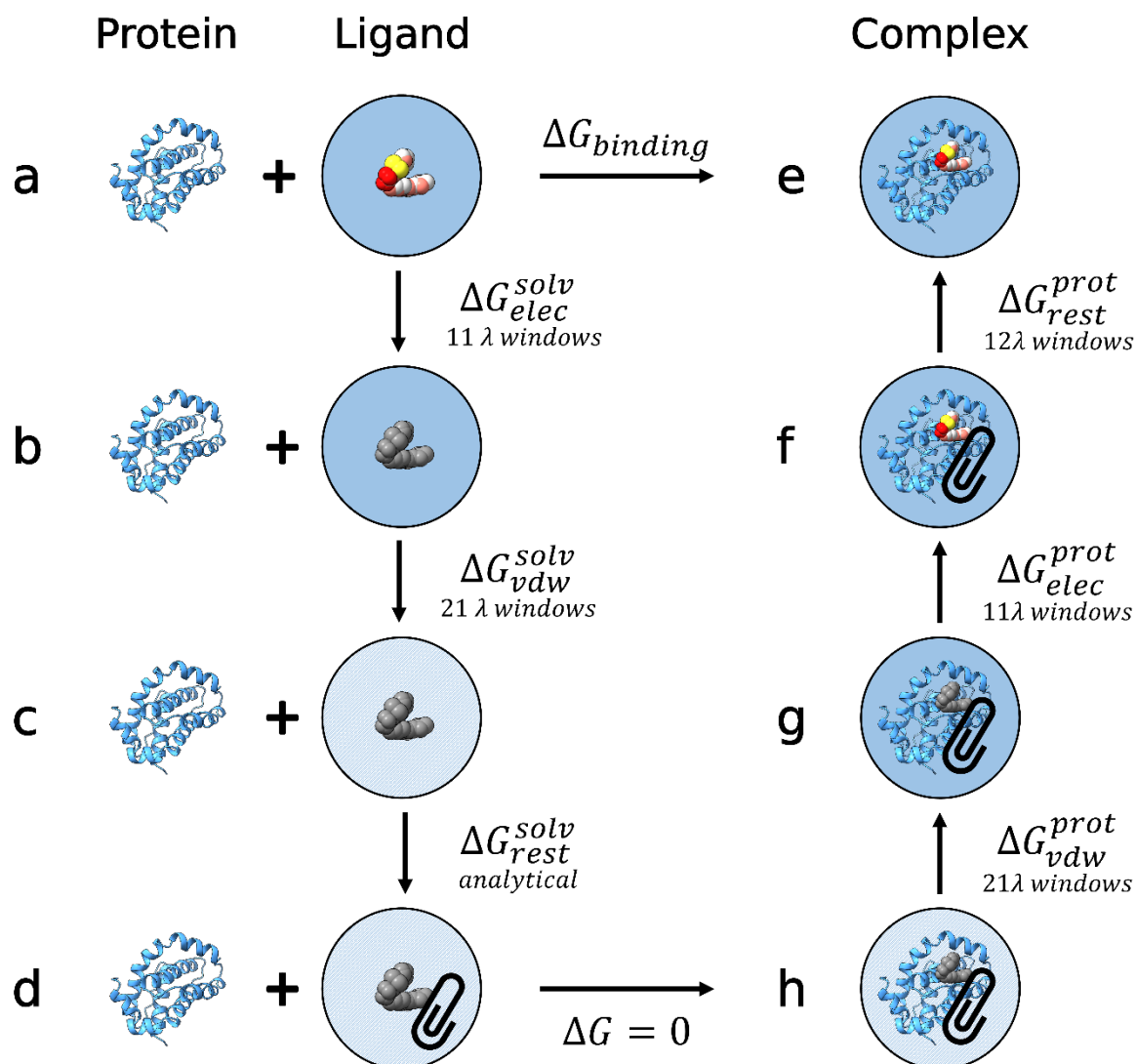In the solvent phase, the influence of this restraint was accounted for analytically.

Fig. 2. *Absolute binding free energy thermodynamic cycle employed. The free energy of binding, i.e.
going from a ligand in solution (state **a**) to a protein-ligand complex (state **e**), is captured through a non-
physical path. First, the ligand electrostatic are annihilated to zero (state **b**) over 11 λ windows. This is
followed by a further 21 λ windows which decouple the ligand van der Waals interaction, from the solvent
(state **c**). The decoupled ligand is then analytically restrained as defined by Boresch et al.[70] (state **d**).
By accounting for this restraint the ligand state is then equivalent to a non-interacting ligand in a protein-
ligand complex (state **h**). The ligand interactions with the environment are then turned back on, first re-
coupling the van der Waals interactions over 21 λ windows (state **g**), followed by a further 11 λ windows
to add back electrostatics (state **f**). Finally the orientational restraints are turned off over 12 λ windows
resulting in a fully interacting protein-ligand complex (state **e**).*

Appropriately choosing which 6 atoms to involve in Boresch-style[70] orientational restraints can be a complex issue, here we use the procedure as implemented in MDRestraintsGenerator (10.5281/zenodo.4570555). Briefly, we pick out the least mobile ligand atoms from our final 20 ns equilibration simulation (see *System Equilibration* above) as potential anchor points for our orientational restraint. We then analyse the trajectory to select all available alpha carbon protein anchor atoms within an 8 Å cut-off of the ligand anchor atoms, generating a list of potential orientational restraints (where the nearest bonded ligand heavy atoms and protein backbone atoms are selected as the remaining atoms involved in the orientational restraint). Bond, angle and dihedral timeseries for all identified restraints are obtained and the restraint with the lowest standard deviation across all values is picked as our orientational restraint of choice. The frame closest to the mean bond, angle and dihedral values of the restraint over the 20 ns simulation is then used as the starting point for the ABFE cycle.

Each ABFE window consists of a short equilibration similar in protocol to the previously detailed equilibration cycles, although using 10 ps for the restrained NVT equilibration, 100 ps for the restrained NPT Berendsen barostat equilibration, and 500 ps for the unrestrained NPT Parinnello-Rahman barostat equilibration. This is followed by 20 ns of production NPT simulation. A total of five independent replicas of the ABFE cycle are simulated. For the protein-ligand complexes, each replica is independently equilibrated and uses different orientational restraints, allowing the replicas to more representatively sample the available conformational space within the binding site. GROMACS 2021 was used for all ABFE calculations. For charged ligands, to reproduce the work of Rocklin *et al.*[72] , the ABFE simulations were carried out with a net system charge by adding or removing a counterion as necessary. Files containing the sampled ΔH and ΔH/Δλ energies have been made publicly available at zenodo : https://zenodo.org/record/5906262, https://zenodo.org/record/5906110, https://zenodo.org/record/5904110, https://zenodo.org/record/5906019 and https://zenodo.org/record/5906805)

N_{wat}-MM/GBSA calculations

The last 20 ns of our equilibration procedure was analysed using the N_{wat}-MM/GBSA method[36,37]. Here the *N* = 20 nearest waters to the ligand in the binding site are included as part of the protein in the MM/GBSA calculation. GROMACS XTC files were converted to AMBER NETCDF format using MDAnalysis 1.1.1[73,74], and the trajectory and topologies manipulated to remove excess waters using cpptraj v5.1.0[75]. The mmbondi2 radii with GB

model 2 parameters[76] were employed with a GB salt concentration of 150 mM. Frames were sampled every 25 ps from the 20 ns trajectories. The AmberTools21 versions of MMPBSA.py and *sander* were used for the MM/GBSA calculations[77].

Analysis

Analysis of the alchemical simulations was achieved using the alchemlyb v0.3.0 library[78] and the MBAR estimator as implemented in pyMBAR v3.0.3[79]. The first 1 ns of each production window was discarded as extra equilibration time and samples were decorrelated based on the derivative of the potential with respect to λ up to a maximum frequency of 100 ps^{-1}. Sampling error is shown as the standard deviation of the mean free energy estimates across all five repeats. Where appropriate, an analytical correction was included to account for finite size errors as defined by Rocklin et al.[72] and implemented in rocklinc (https://github.com/xiki-tempula/rocklinc), using ABPS v3.0[80,81] for Poisson-Botlzmann calculations[81]. It should be noted that no additional long range dispersion corrections (e.g. EXP-LR)[82] were employed here.

All other simulation analyses were carried out using MDAnalysis 1.1.1[73,74], spyrmsd for symmetry corrected RMSDs[83], and the numpy[84], scipy[85], and scikit-learn[86] libraries. Plotting was done using the matplotlib library[87], and images of atomic coordinates through ChimeraX[88]. The correlation between the calculated and experimental affinities are analysed via Spearman *r*, Kendall τ, and RMSE, with error bars obtained as the standard deviation of the means generated through bootstrap resampling (100,000 iterations).

Evaluation of synthetic decisions

As part of our analysis we attempt to retrospectively evaluate the influence our predicted values would have had on the original synthetic decisions undertaken in the fragment optimisations. Overall this is a rather difficult task with many caveats, including; synthetic decisions not always being well defined, steps involving testing affinity changes between more than one ligand pair, and inequalities in the difficulty between synthetic steps. For example; for PWWP1 Böttcher *et al*[39] first detail investigating the impact of adding a 3,5-dimethyl-1,2-oxazole moiety on ligand 8 into order to form ligand 9. The authors then detail their next step as testing 2-methyl- and 2,6-dimethyl-phenyl substituents instead of the dimethyl-1,2-oxazole, showing moderately improve potency (ligands 10 and 11). Given the context of the tested substituents, one could look at these as either two or three optimisation decisions, based on whether one considers the synthesis of ligands 10 and 11 as a combined test of affinity change

in the optimisation process. Here we have, to the best of our ability, attempted to group these syntheses into decisions based on the text outlined in the source reference (see **SI Tables S2-3, S5-S6, S8-S9, and S11-12**) . This was easily done for HSP90[41] and PWWP1[39] but was more complex for MCL-1[40] and Cyclophilin D[38]. Indeed, for MCL-1 we opted to group the decisions into three sets of various substitutions, each for a specific fragment core, and one fragment merging decision. For Cyclophilin D, five clear fragment merging cases could be identified from the simulated ligands and these were used as the synthetic decisions. It is worth noting that the non-equal number of syntheses per decision, combined with our assessment that a failed decision occurs when in that set has an insufficiently estimated $\Delta\Delta G$, makes it such that some decisions have a much higher predictive difficulty than others.

Results and Discussion

ABFE results across all system

The absolute binding free energies of 59 ligands for fragment optimisation campaigns of the PWWP1[89], HSP90[90], MCL-1[91] and Cyclophillin D[92] receptors were computed (see Methods). As shown in **Fig. 3a**, we see very good agreement between the calculated and experimental affinities, with a Pearson r of 0.89 ± 0.03, and a Kendall τ of 0.67 ± 0.05. However, we observe a RMSE of 2.75 ± 0.20 kcal/mol. As outlined in **Fig. 3b-e**, each dataset deviates to varying degrees from experiment, with only PWWP1 showing an RMSE close to 1 kcal/mol. The correlation shown here is on par, if not better, than other comparable alchemical free energy studies[26,93]. Whilst the RMSE is larger than the 1 kcal/mol limit shown by some other ABFE studies[33,93], similar system dependent shifts have been reported previously[94].
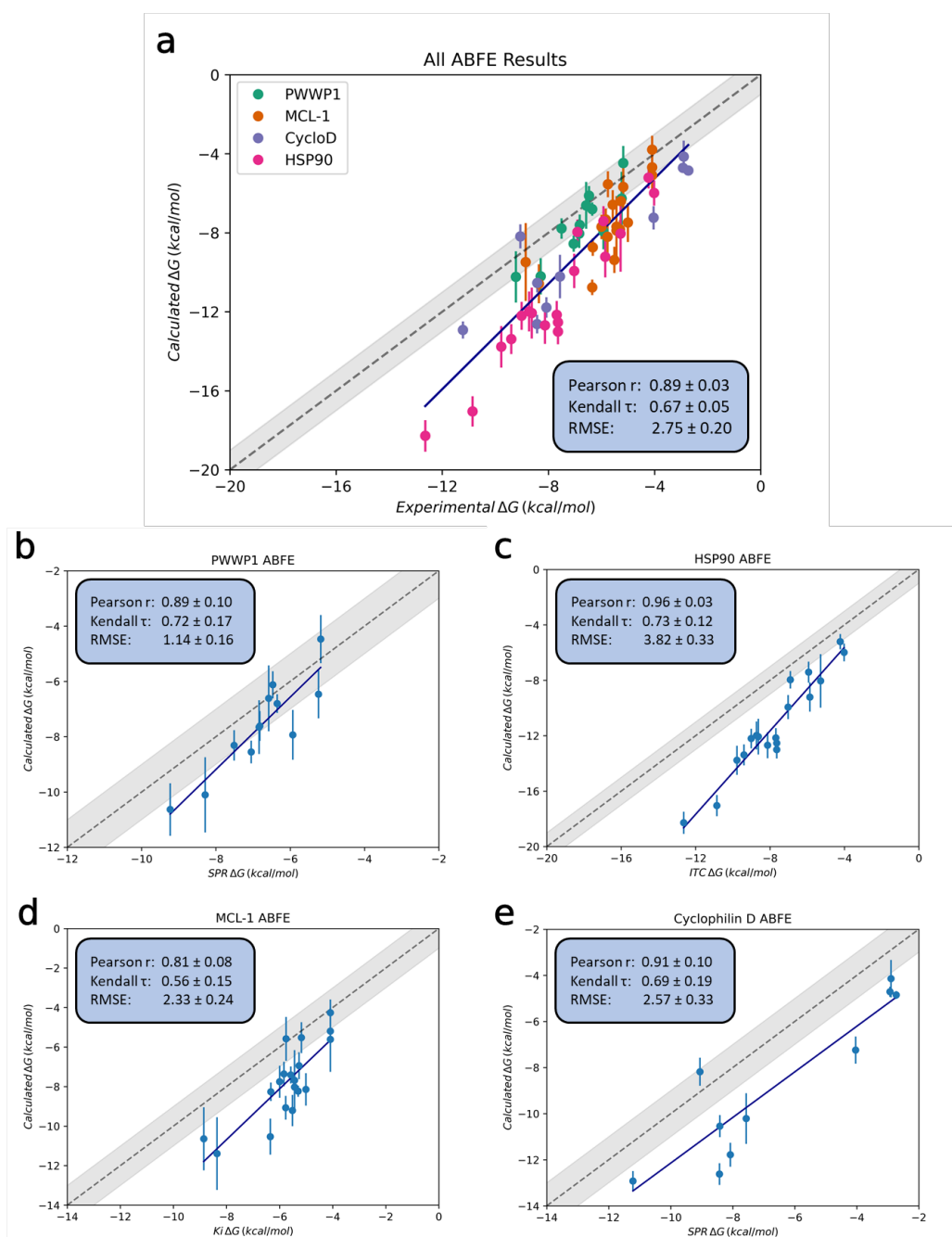
Fig. 3. *ABFE calculations results for; **a**) all four datasets, **b**) PWWP1, **c**) HSP90, **d**) MCL-1, **e**) Cyclophilin D. All free energy results, including RMSE values have units of kcal/mol.*

Sampling of free energies agrees well between replicas with a mean error of 0.79 ± 0.38 kcal/mol, although approximately 22% of the estimates show an error bar greater than 1 kcal/mol, with the highest value being 1.93 kcal/mol. This is generally on a par with sampling errors shown by other absolute alchemical free energy studies[33,94,95] and matching a similar level of uncertainty as the cycle closure errors shown in the Steinbrecher *et al*. 2015 fragment FEP+ study[26]. Whilst the error in experimental measurements was not provided for all our

systems, our sampling errors are for the most part approximately twice as large as the ~0.5 kcal/mol limit which we might expect from experimental measurements[96].

ABFEs of fragments for the PWWP1 domain of NSD3

Specifically looking at each individual case, we start with a fragment elaboration for the PWWP1 domain of NSD3 by Böttcher *et al.*[89]. Here 11 ligands were elaborated from an initial 160 µM fragment hit (ligand 8), eventually resulting in the 170 nM ligand BI-9321 (**SI Fig. S1, Table S1**). As shown in **Fig. 3b**, the ABFE results correlate very well with experiment, with an RMSE of 1.14 ± 0.16 kcal/mol, the lowest of all four optimisation campaigns investigated. In the original elaboration study, seven specific elaboration decisions were outlined (**SI Table S2**). Of these, only two would have been definitively supported by our ABFE results, that is to say - there is a greater-than-error difference between the calculated ligand binding energies in each decision. Whilst the ABFE results accurately predict the direction of the affinity change for six of the seven decisions, the majority of the changes in free energy remained within the range of the errors of the estimates. We note that the one decision where the wrong sign is predicted involves experimental ΔΔG values of -0.12 and -0.23 kcal/mol, which is well within the limit of both experimental[97,98] and force field[99] accuracy.

ABFEs of a fragment optimisation for HSP90

Next, we look at a fragment elaboration study by Murray *et al.*[90] elaborating 17 ligands from an initial fragment hit (ligand 3, **SI Fig. S2**, **SI Table S4**). Whilst the original study does outline two separate fragment elaboration campaigns, we specifically looked at the second elaboration set which involved a larger number of ligands. As shown in **Fig. 3c**, the free energy estimates correlate very well with the experimental values with a Pearson *r* of 0.96 ± 0.03 and a Kendall *τ* of 0.73 ± 0.12. However, we do see a progressive deviation from experiment as the affinity of the ligands increases leading to a RMSE of 3.82 ± 0.33 kcal/mol.

This difference between calculated and experimental values has been observed by other free energy studies of HSP90 and as detailed by Baumann *et al.*[100], is possibly explained by several slow degrees of freedom associated with HSP90 binding. These include; the presence of varying waters in the binding site, ligand re-orientation, and side-chain motions. Interestingly, Baumann *et al.*[100] identify that the presence of waters in the binding site worsened free energy estimates for their HSP90 test case, although it offered improved convergence in results. It is therefore possible that our attempts to optimise binding site waters

through MC/MD in fully interacting protein-ligand complexes may have influenced the deviations in absolute values seen in our results.

One such slow motion which may have had an impact here is the re-arrangement of one of the binding site loops in the region of residues ASN106 through to SER113. As shown in **Fig. 4**, the helicity of this loop region changes between the initial models used for ligands 3-20, 24, 28 and 31 (PDB IDs: 2XDL[41] and 2XAB[41] to the one used for ligands 21-23, 25-27, 29-30 (PDB ID: 2XHT[41]. The re-arrangement of this loop from a helix-loop-helix, akin to the 2XHT conformation, to a continuous helix is well documented and previous work has shown it to be variably induced depending on the binding ligand [101-103]. In their work Murray *et al.* only crystalised a small subset of their ligands in either one or the other crystal forms[41]. As a result, for this study initial models for each ligand were purely chosen based on chemical similarity with ligands from one of the resolved structures. Given the subtle differences between the ligands investigated, it is unclear if the correct loop conformer was chosen, especially given the simulation timescales employed here. Validation would require much longer simulations than are attempted here, likely coupled with experimental methods (e.g. ATR-FTIR spectroscopy [103]).
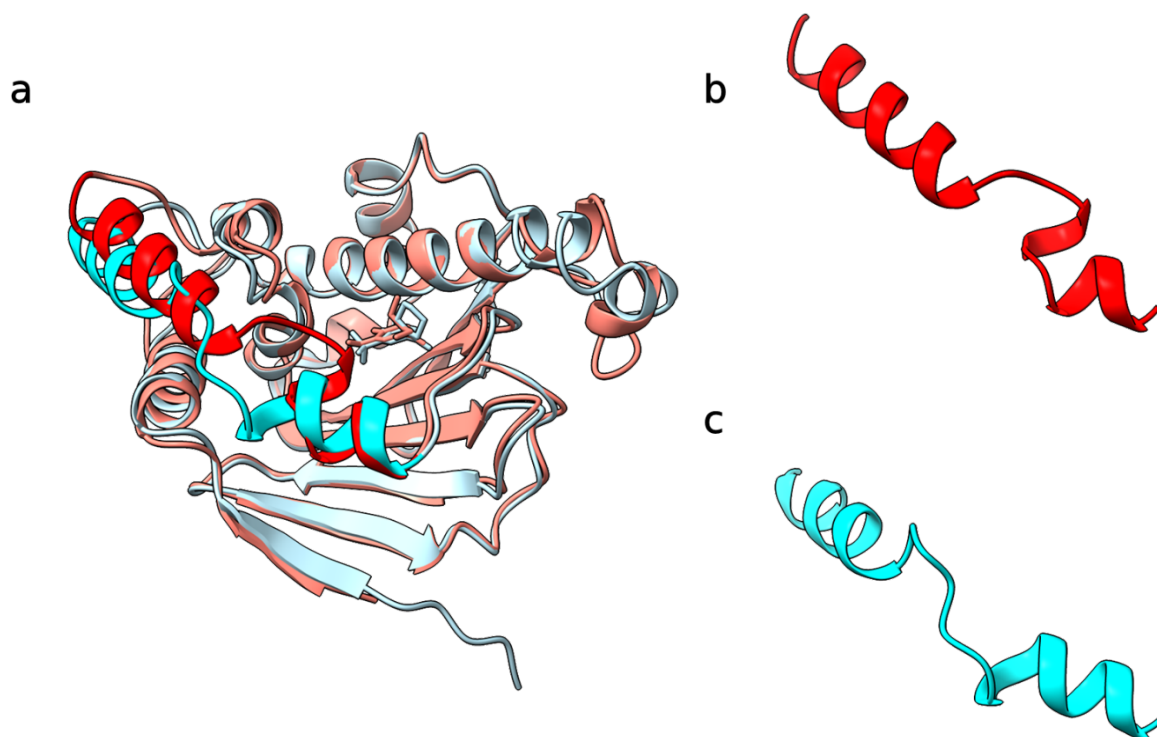

Fig. 4. *Comparison of the structures of PDB IDs 2XHT (blue) and 2XDL (red) demonstrating the difference in helicity between the two models, with a) showing the overlap of the two structures, and b-c) the a zoomed view of the affected helix. Residues 100 to 124 have been highlighted in a darker colour to aid in visualisation. PDB ID 2XAB, which is similar in structure to 2XDL is not shown.*

In this study six elaboration decisions were outlined, of these three would have been clearly supported by our ABFE results (**SI Table S5**). Of the three failed decisions only one (decision 2), would have been fully miscalculated. For the other two, one involved small changes in experimental affinity (decision 5) with ΔΔG values ~ 0.6-0.8 kcal/mol. This likely still falls within the accuracy of ITC measurement, but is too small a change to be properly captured by ABFE given the size of sampling errors. The other elaboration decision clearly identified improvements in affinity but failed to identify ligand 24 as a better binder than ligands 25 and 26 (decision 4).

ABFEs of a fragment optimisation for MCL-1

The third system presented here is an elaboration of fragments for MCL-1 by Friberg *et al.*[91]. In this study, two fragment series are grown in parallel, with an eventual merge to form a nanomolar compound (**SI Fig S3**, **Table S7**). Here we have one of the lowest correlations between calculated and experimental ΔG values with a Pearson r of 0.81 ± 0.08 and a Kendall τ of 0.56 ± 0.15. We also see a shift from experiment in the absolute values, with an RMSE of 2.33 ± 0.24 kcal/mol. It should also be noted that in this series only an upper bound K_i of > 1000 μM was assigned for three of the ligands (ligands 1, 6, and 12, **SI Table S7**). Disregarding these worsens correlation with experiment, yielding a Pearson r of 0.72 ± 0.16, Kendall τ of 0.43 ± 0.19, and a RMSE of 2.50 ± 0.24 kcal/mol. This rather poor correlation by comparison to the other datasets may in part be explained by the narrower activity range covered by the ligands in this series (**Fig. 3d**). Except from two (ligands 60 and 65), the ligands cover an approximately 2 kcal/mol activity range. This is possibly too narrow a range to distinguish between binding, especially given the size of the uncertainties in our calculated estimates.

In their paper Friberg *et al.*[40] did not specifically outline a set of synthetic decisions for these ligands, but instead demonstrated four activity cliffs (**SI Table S8**) based on the growing of > 1 mM class I fragments to micromolar ligands (decisions 1 through 3), and the merging of the class I and II ligands occupying distinct parts of the MCL-1 binding site (decision 4). The majority of fragment growing activity cliffs are clearly captured by the ABFE results, demonstrating greater than error ΔΔG values. However both decisions 1 and 3 have one ligand pair (ligands **1** to **5**, and **3** to **13**) which are within uncertainty of each other. Additionally, the predicted affinities for the optimised ligands are at times predicted to be several kcal/mol away from each other, even though they are experimentally determined to be within a ~ 1 kcal/mol range. The fragment merging case (decision 4) could be clearly captured by our calculated

results, despite large errors in the binding free energy estimates of the merged scaffold ligands 60 and 65. Indeed, the MCL-1 dataset has on average the highest standard deviations between ABFE replicates. These large sampling errors are reflect the charged nature of these ligands and the large conformational space available for the ligand in the binding site. As demonstrated for ligand 60 in **Figure 5**, the ligand can effectively roll within the binding site, resulting in a pose with an RMSD > 4 Å from the others. The flexibility of ligands binding to MCL-1 was also noted by Steinbrecher *et al.*[26], and indicates a need for either longer simulation times or enhanced sampling schemes to reach sufficient convergence for systems such as these.
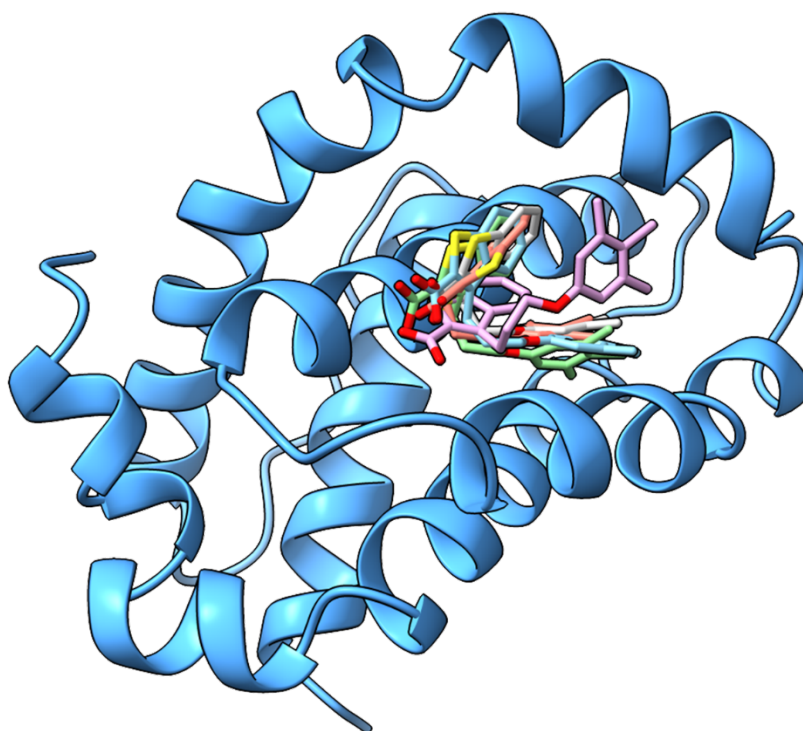

Fig. 5. *Overlay of the starting configurations of each replica of the ABFE calculations demonstrating the rearrangement (pink coloured ligand) of the merged ligand 60 in the MCL-1 binding site.*

ABFEs of a fragment optimisation for Cyclophillin D

Our final elaboration dataset is a set of fragment merging decisions for a fragment screen of Cyclophilin D by Gradler *et al.*[92]. Here we simulate a subset of the ligands investigated in the original study, covering a total of 5 merging decisions (**SI Fig. S4**, **SI Table S10**). We again find good correlation between the calculated and experimental free energies, with a Pearson r of 0.91 ± 0.10 and a Kendall τ 0.69 ± 0.19 albeit with a relatively large RMSE of 2.57 ± 0.33

kcal/mol. For the most part (with ligand 2 being an exception) this seems to represent a systematic shift in the free energy by approximately 2.5 kcal/mol (**Fig. 3e**).

Of the five merging decisions captured by the simulated ligands, four would have been clearly supported by the ABFE results (**SI Table S11**). The large activity cliffs seen in these merging decisions, often exceeding 5 kcal/mol, are easily captured by the ABFE simulations even in cases where uncertainty exceeds 1 kcal/mol. The only miscalculated merging decision involves a relatively small $\Delta\Delta G$ change of 0.64 kcal/mol in experimental affinities (ligand 2 to ligand 16), which is miscalculated through ABFE by 3 kcal/mol in the opposite direction. Beyond merging decisions, whilst the estimates can differentiate between low affinity fragments (within the limit of error), the higher affinity merged ligands cannot be clearly distinguished from each other. For example, the 6 nM ligand 16, is estimated as within 0.3 kcal/mol of the 660 nM ligand 39.

Comparison to other methods

The computational cost of ABFEs is not insignificant, being easily orders of magnitude higher than cheaper endpoint methods. Previous works[26,104,105] have shown that endpoint methods such as MM/PBSA and MM/GBSA can often be competitive with alchemical methods, especially given the much-reduced computational costs. Here we compare our ABFE results with those calculated using N_{wat}-MM/GBSA. As shown in **Fig 3 and 6**, we find that overall ABFE calculations yield improved free energy estimates compared to N_{wat}-MM/GBSA, with up to 0.2 improvements in average correlation metrics (Pearson r and Kendall τ) for all datasets except MCL-1 where N_{wat}-MM/GBSA shows a higher Pearson r of 0.91 ± 0.11. As is usual for methods like N_{wat}-MM/GBSA[104], as a consequence of not directly accounting for entropy[106], the absolute free energies are vastly overestimated, leading to RMSE values in the range of ~ 30 to 50 kcal/mol. Nevertheless, the N_{wat}-MM/GBSA method shows a reasonably good accuracy in ranking fragment optimisations, being able to clearly predict several of the elaboration decisions (**SI Tables S3, S6, S9, S12**). Surprisingly, a similar amount of decisions to the ABFE calculations are predicted here. Whilst the N_{wat}-MM/GBSA results predict one fewer decision for the HSP90 dataset, two more decisions are predicted for PWWP1. Additionally, for MCL-1 whilst the fragment optimisation activity cliffs for decision 1 cannot be clearly distinguished due to large uncertainties, the optimisation of ligand 12 into ligand 13 (decision 3), which could not be supported by the ABFE results can be clearly identified using N_{wat}-MM/GBSA.
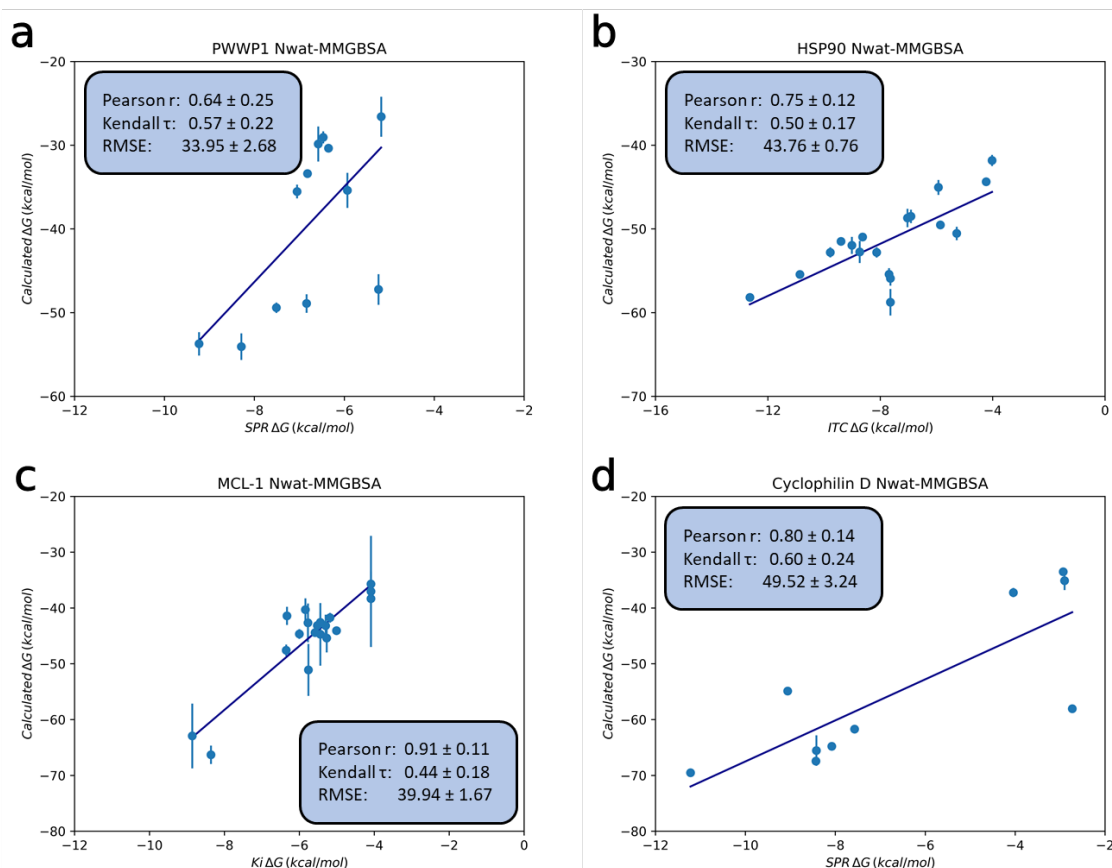
Fig. 6. Binding free energies estimated by N_{wat}-MM/GBSA for the a) PWWP1, b) HSP90, c) MCL-1, and d) Cyclophilin D datasets. Energies, including RMSE values, in units of kcal/mol.

We can also compare some of the results with previously published RBFE data. Specifically, a subset of the MCL-1 dataset investigated here was also calculated using FEP+ in Steinbrecher *et al.*'s 2015 fragment optimisation study[26]. Whilst only offering a limited comparison, we find the two methods to give comparable results (**Fig. 7**), with the two methods offering within error values for Pearson r and Kendall τ. Given the narrow activity range and the relatively large uncertainties in the estimates, further comparison of correlation between calculated and experimental values is unlikely to yield much insight. We do note that the reported RMSE for FEP+ is much lower at 0.93 kcal/mol, however the absolute numbers in this FEP+ study were normalised based on the experimental affinity and therefore cannot be directly compared to the ABFE absolute results which required no *a priori* experimental affinity data.
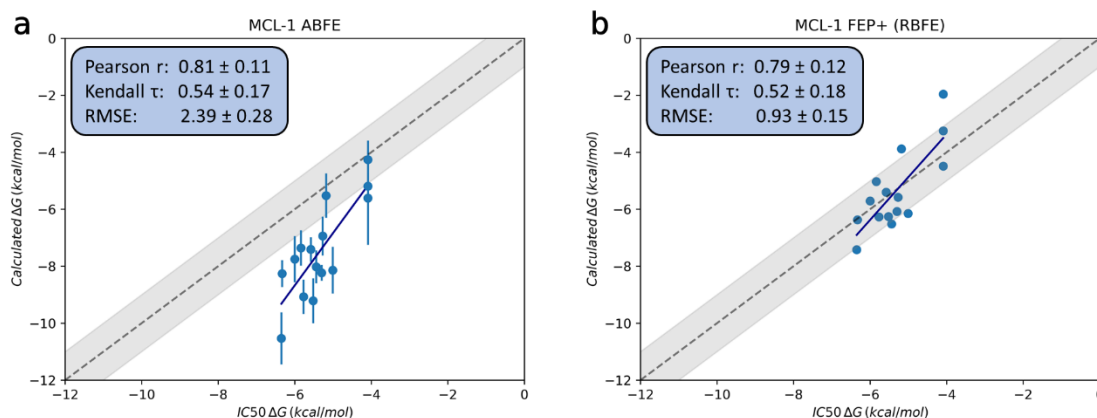
Fig. 7. *Comparison of a subset of the MCL-1 dataset as calculated by **a)** ABFE simulations and **b)** the 2015 FEP+ study by Steinbrecher et al.[26] Free energy units, including RMSE, in kcal/mol.*

Conclusions

In this study, we demonstrate that absolute binding free energy calculations can successfully estimate affinity changes in the fragment optimisation process. Whilst not always being able to clearly support elaboration decisions, the right direction in the affinity change was captured for the vast majority of cases (46 of the 52 ligand pairs involved in the 22 elaboration decisions). Despite occasional challenges in accurately matching experimentally derived binding free energies, the predictive power displayed here shows that ABFEs can be suitably used to not only act to orthogonally validate fragment hits but also to guide fragment optimisation.

We also show that ABFE calculations outcompete cheaper endpoint methods such as N_{wat}-MM/GBSA and, for one system of our benchmark, offer comparable results to relative binding free energy methods. Although ABFEs do have a several times larger computational cost to such methods, advances in GPU utilisation have been making the high throughput use of such a technique more tractable. That being said the N_{wat}-MM/GBSA method employed here showed reasonably good predictive ability, particularly in supporting retrospective elaboration decisions, and could easily be used as a low computational cost prefilter to more expensive ABFE calculations.

Our results did demonstrate cases of sampling limitations. This was either seen through poor inter-replica convergence (as shown in some of the MCL-1 ligands), or large deviations from experimental affinities (as shown in HSP90). We note that the ABFE protocol used here is reasonably simple and could easily be improved to address some of these sampling limitations [107]. Indeed, recent work by Khalak *et al.*[93] demonstrates that non-equilibrium absolute binding

free energy can be used to bridge differences between the *apo* and *holo* states of a protein, and possibly could alleviate some of the issues shown here.

It is important to also note that *a priori* system specific knowledge remains crucial in ensuring the performance of alchemical methods. Whilst one of the main advantages of ABFE calculations being of directly estimating the affinity of a binder, having access to experimental affinities for a subset of the calculated binders to the same target or a structurally similar one, would significantly help in identifying and adjusting issues in a model. For example, we were able to confidently use an N-terminal truncated model of the PWWP1 system (see **SI Text 2**) after observing that initial calculated free energies were close to those obtained from experiment. Similarly the overestimations in the free energy estimates seen here are sufficiently large that they could mislead synthetic efforts, but also could easily be adjusted through prior affinity data on a few of the data points.

Despite the positive results displayed here, it is important to make clear that in many ways this benchmark only tackles a simplified set of tasks in the FBDD process and that several challenges remain ahead in validating and using absolute binding free energies. Of particular note, this dataset solely looked at cases where most binding poses were known and as a consequence were likely to not include ligands occupying more than one binding pose significantly contributing to the total free energy. Thus, our use of orientational restraints in this study could be justified. However there are FBDD cases where accounting for multiple binding sites either through less restrictive restraints, such as center of mass restraints[100,108-111] or through enhanced sampling means (e.g. NCMC[112] or metadynamics[113]), will be required. Additionally, as mentioned in our methods we also specifically attempted to avoid known cases where issues may arise in ABFE calculations, such as the presence of structural ions, membrane proteins, or ambiguous protonation states. Future work will look at tackling some of these more complex issues, particularly within the context of FBDD and see what type of solutions can be leveraged to ensure the accurate calculation of absolute binding free energies.

Acknowledgements

References

1   Lamoree, B. & Hubbard, R. Current perspectives in fragment-based lead discovery (FBLD). *Essays Biochem.* **61**, 453-464, (2017).

2   Johnson, C. N., Erlanson, D. A., Jahnke, W., Mortenson, P. N. & Rees, D. C. Fragment-to-Lead Medicinal Chemistry Publications in 2016. *Journal of Medicinal Chemistry* **61**, 1774-1784, (2018).

3   Mortenson, P. N., Erlanson, D. A., de Esch, I. J. P., Jahnke, W. & Johnson, C. N. Fragment-to-Lead Medicinal Chemistry Publications in 2017. *Journal of Medicinal Chemistry* **62**, 3857-3872, (2019).

4   Erlanson, D. A., de Esch, I. J. P., Jahnke, W., Johnson, C. N. & Mortenson, P. N. Fragment-to-Lead Medicinal Chemistry Publications in 2018. *Journal of Medicinal Chemistry* **63**, 4430-4444, (2020).

5   Jahnke, W. *et al.* Fragment-to-Lead Medicinal Chemistry Publications in 2019. *Journal of Medicinal Chemistry* **63**, 15494-15507, (2020).

6   Johnson, C. N., Erlanson, D. A., Murray, C. W. & Rees, D. C. Fragment-to-Lead Medicinal Chemistry Publications in 2015. *Journal of Medicinal Chemistry* **60**, 89-99, (2017).

7   Congreve, M., Carr, R., Murray, C. & Jhoti, H. A 'Rule of Three' for fragment-based lead discovery? *Drug Discovery Today* **8**, 876-877, (2003).

8   Li, Q. Application of fragment-based drug discovery to versatile targets. *Front. Mol. Biosci.* **7**, (2020).

9   Coyle, J. & Walser, R. Applied biophysical methods in fragment-based drug discovery. *SLAS Discov.* **25**, 471-490, (2020).

10   Jacquemard, C. & Kellenberger, E. A bright future for fragment-based drug discovery: what does it hold? *Expert Opin. Drug Discov.* **14**, 413-416, (2019).

11   de Souza Neto, L. R. *et al.* In silico strategies to support fragment-to-lead optimization in drug discovery. *Front. Chem.* **8**, (2020).

12   Bian, Y. & Xie, X.-Q. Computational fragment-based drug design: Current trends, strategies, and applications. *AAPS J.* **20**, 59, (2018).

13   Gossert, A. D. & Jahnke, W. NMR in drug discovery: A practical guide to identification and validation of ligands interacting with biological macromolecules. *Prog. Nuc. Mag. Res. Spec.* **97**, 82-125, (2016).

14   Bissaro, M., Sturlese, M. & Moro, S. The rise of molecular simulations in fragment-based drug design (FBDD): an overview. *Drug Discovery Today* **25**, 1693-1701, (2020).

15   Georgiou, C. *et al.* Pushing the limits of detection of weak binding using fragment-based drug discovery: Identification of new cyclophilin binders. *J. Mol. Biol.* **429**, 2556-2570, (2017).

16   Verdonk, M. L. *et al.* Docking performance of fragments and druglike compounds. *J. Med. Chem.* **54**, 5422-5431, (2011).

17   Defelipe, L. A. *et al.* Solvents to fragments to drugs: MD applications in drug design. *Molecules* **23**, (2018).

18   Guvench, O. Computational functional group mapping for drug discovery. *Drug Discov. Today* **21**, 1928-1931, (2016).

19   Bakan, A., Nevins, N., Lakdawala, A. S. & Bahar, I. Druggability assessment of allosteric proteins by dynamics simulations in the presence of probe molecules. *J. Chem. Theory Comput.* **8**, 2435-2447, (2012).

20   Lexa, K. W. & Carlson, H. A. Full protein flexibility is essential for proper hot-spot mapping. *J. Am. Chem. Soc.* **133**, 200-202, (2011).

21   Guvench, O. & MacKerell, A. D., Jr. Computational fragment-based binding site identification by ligand competitive saturation. *PLOS Comput. Biol.* **5**, e1000435, (2009).

22   Linker, S. M., Magarkar, A., Köfinger, J., Hummer, G. & Seeliger, D. Fragment binding pose predictions using unbiased simulations and Markov-state models. *J. Chem. Theory Comput.* **15**, 4974-4981, (2019).

23   Matricon, P. *et al.* Fragment-based design of selective GPCR ligands guided by free energy simulations. *Chem. Comm.* **57**, 12305-12308, (2021).

24   Chen, D., Ranganathan, A., Ijzerman, A. P., Siegal, G. & Carlsson, J. Complementarity between in silico and biophysical screening approaches in fragment-based lead discovery against the A2a adenosine receptor. *J. Chem. Inf. Model.* **53**, 2701-2714, (2013).

25   Matricon, P. *et al.* Fragment optimization for GPCRs by molecular dynamics free energy calculations: Probing druggable subpockets of the A2Aadenosine receptor binding site. *Sci. Reps.* **7**, 6398, (2017).

26   Steinbrecher, T. B. *et al.* Accurate binding free energy predictions in fragment optimization. *J. Chem. Inf. Model.* **55**, 2411-2420, (2015).

27   Zara, L., Efrém, N.-L., van Muijlwijk-Koezen, J. E., de Esch, I. J. P. & Zarzycka, B. Progress in free energy perturbation: Options for evolving fragments. *Drug Discov. Today* **40**, 36-42, (2021).

28   Mobley, D. L. & Gilson, M. K. Predicting binding free energies: Frontiers and benchmarks. *Ann. Rev. Biophys.* **46**, 531-558, (2017).

29   Deng, Y. & Roux, B. Calculation of Standard Binding Free Energies: Aromatic Molecules in the T4 Lysozyme L99A Mutant. *J. Chem. Theory Comput.* **2**, 1255-1273, (2006).

30   Mobley, D. L. *et al.* Predicting absolute ligand binding free energies to a simple model site. *J Mol Biol* **371**, 1118-1134, (2007).

31   Khalak, Y. *et al.* Alchemical absolute protein–ligand binding free energies for drug design. *Chemical Science* **12**, 13958-13971, (2021).

32   Aldeghi, M., Heifetz, A., Bodkin, M. J., Knapp, S. & Biggin, P. C. Predictions of ligand selectivity from absolute binding free energy calculations. *J. Am. Chem. Soc.* **139**, 946-957, (2017).

33   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.* **7**, 207-218, (2016).

34   Boyce, S. E. *et al.* Predicting ligand binding affinity with alchemical free energy methods in a polar model binding site. *J Mol Biol* **394**, 747-763, (2009).

35   Rizzi, A. *et al.* Overview of the SAMPL6 host–guest binding affinity prediction challenge. *J. Comp-aided Mol. Des.* **32**, 937-963, (2018).

36   Maffucci, I. & Contini, A. Improved Computation of Protein–Protein Relative Binding Energies with the Nwat-MMGBSA Method. *J. Chem. Inf. Model.* **56**, 1692-1704, (2016).

37   Maffucci, I. & Contini, A. Explicit ligand hydration shells improve the correlation between MM-PB/GBSA binding energies and experimental activities. *J. Chem. Theor. Comput.* **9**, 2706-2717, (2013).

38   Grädler, U. *et al.* Discovery of novel Cyclophilin D inhibitors starting from three dimensional fragments with millimolar potencies. *Bioorg. Med. Chem. Letts* **29**, 126717, (2019).

39   Böttcher, J. *et al.* Fragment-based discovery of a chemical probe for the PWWP1 domain of NSD3. *Nature Chem. Biol.* **15**, 822-829, (2019).

40   Friberg, A. *et al.* Discovery of potent myeloid cell leukemia 1 (Mcl-1) inhibitors using fragment-based methods and structure-based design. *J. Med. Chem.* **56**, 15-30, (2013).

41   Murray, C. W. *et al.* Fragment-based drug discovery applied to HSP90. Discovery of two lead series with high ligand efficiency. *J. Med. Chem.* **53**, 5942-5955, (2010).

42   The PyMOL molecular graphics system. (2004).

43   Bietz, S., Urbaczek, S., Schulz, B. & Rarey, M. Protoss: a holistic approach to predict tautomers and protonation states in protein-ligand complexes. *Journal of Cheminformatics* **6**, 12, (2014).

44   Lippert, T. & Rarey, M. Fast automated placement of polar hydrogen atoms in protein-ligand complexes. *Journal of Cheminformatics* **1**, 13, (2009).

45   Fährrolfes, R. *et al.* ProteinsPlus: a web portal for structure analysis of macromolecules. *Nucleic Acids Research* **45**, W337-W343, (2017).

46   Schöning-Stierand, K. *et al.* ProteinsPlus: interactive analysis of protein–ligand binding interfaces. *Nucleic Acids Research* **48**, W48-W53, (2020).

47   Webb, B. & Sali, A. Comparative protein structure modeling using modeller. *Curr. Prot. Bioinf.* **5**, 5.61-65.66.32, (2014).

48   Benkert, P., Tosatto, S. C. E. & Schomburg, D. QMEAN: A comprehensive scoring function for model quality assessment. *Proteins: Structure, Function, and Bioinformatics* **71**, 261-277, (2008).

49   Waterhouse, A. *et al.* SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296-W303, (2018).

50   AMBER 2018 (University of California, 2018).

51   Abraham, M. J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19-25, (2015).

52   Jorgensen, W. L. Transferable intermolecular potential functions for water, alcohols, and ethers, application to liquid water. *J. Am. Chem. Soc.* **103**, 335-340, (1981).

53   Feenstra, K. A., Hess, B. & Berendsen, H. J. C. Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. *J. Comp. Chem.* **20**, 786-798, (1999).

54   Hopkins, C. W., Le Grand, S., Walker, R. C. & Roitberg, A. E. Long-time-step molecular dynamics through hydrogen mass repartitioning. *J. Chem. Theory Comput.* **11**, 1864-1874, (2015).

55   Heinzelmann, G. & Gilson, M. K. Automation of absolute protein-ligand binding free energy calculations for docking refinement and compound evaluation. *Scientific Reports* **11**, 1116, (2021).

56   Lee, T.-S. *et al.* Alchemical binding free energy calculations in AMBER20: Advances and best practices for drug discovery. *J. Chem. Inf. Model.* **60**, 5595-5623, (2020).

57   Zhang, H. *et al.* CHARMM-GUI free energy calculator for practical ligand binding free energy simulations with AMBER. *J. Chem. Inf. Model.* **61**, 4145-4151, (2021).

58   Miyamoto, S. & Kollman, P. A. SETTLE: An analytical version of the SHAKE and RATTLE algorithms for rigid water molecules. *J. Comp. Chem.* **18**, 1463-1472, (1992).

59   Ryckaert, J. P., Ciccotti, G. & Berendsen, H. J. C. Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **23**, 327, (1977).

60   Hess, B. P-lincs: A parallel linear constraint solver for molecular simulation. *J. Chem. Theor. Comput.* **4**, 116-122, (2008).

61   Hess, B., Bekker, J., Berendsen, H. J. C. & Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comp. Chem.* **18**, 1463-1472, (1997).

62   Allen, M. P. & Tildesley, D. J. *Computer Simulation of Liquids*. (Oxford University Press, 1987).

63   Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684-3690, (1984).

64   Parinello, M. & Rahman, A. Polymorphic transitions in single crystals - a new molecular dynamics method. *J. Appl. Phys.* **52**, 7182-7190, (1981).

65   Essman, U. *et al.* A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577-8593, (1995).

66   Darden, T., Pearlman, D. & Pedersen, L. G. Ionic charging free energies: Spherical versus periodic boundary conditions. *J. Chem. Phys.* **109**, 10921-10935, (1998).

67   Ben-Shalom, I. Y., Lin, C., Kurtzman, T., Walker, R. C. & Gilson, M. K. Simulating Water Exchange to Buried Binding Sites. *Journal of Chemical Theory and Computation* **15**, 2684-2691, (2019).

68   Salomon-Ferrer, R., Götz, A. W., Poole, D., Le Grand, S. & Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *Journal of Chemical Theory and Computation* **9**, 3878-3888, (2013).

69   Le Grand, S., Götz, A. W. & Walker, R. C. SPFP: Speed without compromise—A mixed precision model for GPU accelerated molecular dynamics simulations. *computer physics communications* **184**, 374-380, (2013).

70   Boresch, S., Tettinger, F., Leitgeb, M. & Karplus, M. Absolute binding free energies: A quantitative approach for their calculation. *J. Phys. Chem. B.* **107**, 9535-9551, (2003).

71   Beutler, T. C., Mark, A. E., van Schaik, R. C., Gerber, P. R. & van Gunsteren, W. F. Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations. *Chem. Phys. Letts.* **222**, 529-539, (1994).

72   Rocklin, G. J., Mobley, D. L., Dill, K. A. & Hünenberger, P. H. Calculating the binding free energies of charged species based on explicit-solvent simulations employing lattice-sum methods: An accurate correction scheme for electrostatic finite-size effects. *J. Chem. Phys.* **139**, 184103, (2013).

73   Gowers, R. J. *et al.* in *Python in Science.* (eds S. Benthall & S. Rostrup).

74   Michaud-Agrawal, N., Denning, E. J., Woolf, T. & Beckstein, O. MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem.* **32**, 2319-2327, (2011).

75   Roe, D. R. & Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *Journal of Chemical Theory and Computation* **9**, 3084-3095, (2013).

76   Onufriev, A., Bashford, D. & Case, D. A. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins: Structure, Function, and Bioinformatics* **55**, 383-394, (2004).

77   AMBER 2021 (University of California, 2020).

78   alchemistry/alchemlyb (2021).

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

80   Jurrus, E. *et al.* Improvements to the APBS biomolecular solvation software suite. *Prot. Sci.* **27**, 112-128, (2018).

81   Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. USA* **98**, 10037-10041, (2001).

82   Shirts, M. R., Mobley, D. L., Chodera, J. D. & Pande, V. S. Accurate and efficient corrections for missing dispersion interactions in molecular simulations. *J. Phys. Chem. B.* **111**, 13052-13063, (2007).

83   Meli, R. & Biggin, P. C. spyrmsd: symmetry-corrected RMSD calculations in Python. *J. Cheminf.* **12**, 49, (2020).

84   Harris, C. R. *et al.* Array programming with NumPy. *Nature* **585**, 357-362, (2020).

85   Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Meth.* **17**, 261-272, (2020).

86   Pedregosa, F. *et al.* Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825-2830, (2011).

87   Hunter, J. D. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **9**, 90-95, (2007).

88   Pettersen, E. F. *et al.* UCSF Chimera—A visualization system for exploratory research and analysis. *Journal of Computational Chemistry* **25**, 1605-1612, (2004).

89   Böttcher, J. *et al.* Fragment-based discovery of a chemical probe for the PWWP1 domain of NSD3. *Nature Chemical Biology* **15**, 822-829, (2019).

90   Murray, C. W. *et al.* Fragment-Based Drug Discovery Applied to Hsp90. Discovery of Two Lead Series with High Ligand Efficiency. *Journal of Medicinal Chemistry* **53**, 5942-5955, (2010).

91   Friberg, A. *et al.* Discovery of Potent Myeloid Cell Leukemia 1 (Mcl-1) Inhibitors Using Fragment-Based Methods and Structure-Based Design. *Journal of Medicinal Chemistry* **56**, 15-30, (2013).

92   Grädler, U. *et al.* Discovery of novel Cyclophilin D inhibitors starting from three dimensional fragments with millimolar potencies. *Bioorganic & Medicinal Chemistry Letters* **29**, 126717, (2019).

93   Khalak, Y., Tresadern, G., de Groot, B. L. & Gapsys, V. Non-equilibrium approach for binding free energies in cyclodextrins in SAMPL7: force fields and software. *J. Comput.-Aided Mol. Des.* **35**, 49-61, (2021).

94   Gapsys, V. *et al.* Accurate absolute free energies for ligand–protein binding based on non-equilibrium approaches. *Comm. Chem.* **4**, 61, (2021).

95   Gapsys, V. *et al.* Large scale relative protein ligand binding affinities using non-equilibrium alchemy. *Chem. Sci.* **11**, 1140-1152, (2020).

96   Brown, S. P., Muchmore, S. W. & Hajduk, P. J. Healthy skepticism: assessing realistic model performance. *Drug Discov. Today* **14**, 420-427, (2009).

97   Myszka, D. G. *et al.* The ABRF-MIRG'02 study: Assembly state, thermodynamic, and kinetic analysis of an enzyme/inhibitor interaction. *J. Biomol. Tech.* **14**, 247-269, (2003).

98   Tellinghuisen, J. & Chodera, J. D. Systematic errors in isothermal titration calorimetry: Concentrations and baselines. *Analytical Biochemistry* **414**, 297-299, (2011).

99   Henriksen, N. M. & Gilson, M. K. Evaluating force field performance in thermodynamic calculations of cyclodextrin host–guest binding: Water models, partial charges, and host force field parameters. *J. Chem. Theory Comput.* **13**, 4253-4269, (2017).

100   Baumann, H. M., Gapsys, V., de Groot, B. L. & Mobley, D. L. Challenges encountered applying equilibrium and nonequilibrium binding free energy calculations. *J. Phys. Chem. B.*, (2021).

101   Wright, L. *et al.* Structure-activity relationships in purine-based inhibitor binding to HSP90 isoforms. *Chem. Biol.* **11**, 775-785, (2004).

102   Immormino, R. M., Kang, Y., Chiosis, G. & Gewirth, D. T. Structural and quantum chemical studies of 8-aryl-sulfanyl adenine class hsp90 inhibitors. *J. Med. Chem.* **49**, 4953-4960, (2006).

103   Güldenhaupt, J. *et al.* Ligand-induced conformational changes in HSP90 monitored time resolved and label free—towards a conformational activity screening for drug discovery. *Ang.Chem. Int. Ed.* **57**, 9955-9960, (2018).

104   Aldeghi, M., Bodkin, M. J., Knapp, S. & Biggin, P. C. A statistical analysis on the performance of MMPBSA versus absolute binding free energy calculations: bromodomains as a case study. *J. Chem. Inf. Model.* **57**, 2203-2221, (2017).

105   Genheden, S. & Ryde, U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin. Drug Discov.* **10**, 449-461, (2015).

106   Tuccinardi, T. What is the current value of MM/PBSA and MM/GBSA methods in drug discovery? *Expert Opin. Drug Discov.* **16**, 1233-1237, (2021).

107   Procacci, P. Methodological uncertainties in drug-receptor binding free energy predictions based on classical molecular dynamics. *Current Opinion in Structural Biology* **67**, 127-134, (2021).

108   Cabeza de Vaca, I., Qian, Y., Vilseck, J. Z., Tirado-Rives, J. & Jorgensen, W. L. Enhanced Monte Carlo methods for modeling proteins including computation of absolute free energies of binding. *J. Chem. Theory Comput.* **14**, 3279-3288, (2018).

109   Duboué-Dijon, E. & Hénin, J. Building intuition for binding free energy calculations: Bound state definition, restraints, and symmetry. *J. Chem. Phys.* **154**, 204101, (2021).

110   Mobley, D. L., Chodera, J. D. & Dill, K. A. On the use of orientational restraints and symmetry corrections in alchemical free energy calculations. *J. Chem. Phys.* **125**, 084902, (2006).

111   Shi, Y., Laury, M. L., Wang, Z. & Ponder, J. W. AMOEBA binding free energies for the SAMPL7 TrimerTrip host–guest challenge. *J. Comput-Aid Mol. Des.* **35**, 79-93, (2021).

112   Gill, S. C. *et al.* Binding Modes of Ligands Using Enhanced Sampling (BLUES): Rapid Decorrelation of Ligand Binding Modes via Nonequilibrium Candidate Monte Carlo. *J. Phys. Chem. B*, (2018).

113   Tanida, Y. & Matsuura, A. Alchemical free energy calculations via metadynamics: Application to the theophylline-RNA aptamer complex. *Journal of Computational Chemistry* **41**, 1804-1819, (2020).