EDGE ARTICLE

Check for updates

The role of loop dynamics in the prediction of ligand–protein binding enthalpy†

Süleyman Selim Çınaroğlu [ID] and Philip C. Biggin [ID] *

The enthalpic and entropic components of ligand–protein binding free energy reflect the interactions and dynamics between ligand and protein. Despite decades of study, our understanding and hence our ability to predict these individual components remains poor. In recent years, there has been substantial effort and success in the prediction of relative and absolute binding free energies, but the prediction of the enthalpic (and entropic) contributions in biomolecular systems remains challenging. Indeed, it is not even clear what kind of performance in terms of accuracy could currently be obtained for such systems. It is, however, relatively straight-forward to compute the enthalpy of binding. We thus evaluated the performance of absolute enthalpy of binding calculations using molecular dynamics simulation for ten inhibitors against a member of the bromodomain family, BRD4-1, against isothermal titration calorimetry data. Initial calculations, with the AMBER force-field showed good agreement with experiment (R^2 = 0.60) and surprisingly good accuracy with an average of root-mean-square error (RMSE) = 2.49 kcal mol^{-1}. Of the ten predictions, three were obvious outliers that were all over-predicted compared to experiment. Analysis of various simulation factors, including parameterization, buffer concentration and conformational dynamics, revealed that the behaviour of a loop (the ZA loop on the periphery of the binding site) strongly dictates the enthalpic prediction. Consistent with previous observations, the loop exists in two distinct conformational states and by considering one or the other or both states, the prediction for the three outliers can be improved dramatically to the point where the R^2 = 0.95 and the accuracy in terms of RMSE improves to 0.90 kcal mol^{-1}. However, performance across force-fields is not consistent: if OPLS and CHARMM are used, different outliers are observed and the correlation with the ZA loop behaviour is not recapitulated, likely reflecting parameterization as a confounding problem. The results provide a benchmark standard for future study and comparison.

Introduction

In recent years, there has been significant progress in predicting the binding free energy of small ligands for protein receptors.[1–3] In contrast there has been relatively little progress in the accurate computation of the underlying thermodynamic components, namely the enthalpy (ΔH) and entropy ($T\Delta S$). Accurate computation of enthalpy has historically been viewed as particularly challenging[4–7] due to the large fluctuations in potential energy that systems tend to undergo and thus any estimates of the mean value would likely require vast amounts of sampling. Nevertheless, having a reasonably reliable estimation of the enthalpy, and more importantly the error estimate, would be useful in understanding the role of the underlying contributions, especially in the context of drug-

design.[8,9] From a medicinal chemistry point of view, enthalpic contributions are perhaps intuitively easier to understand and conceptualize than the entropic components. During a fragment or lead compound elaboration, where ΔG is often being optimized, it would be extremely valuable to know during those steps that the changes made to the compound were indeed giving the expected improvement to the ΔG *via* the designs that the medicinal chemist has suggested. Often, such designs are focused around improving interactions between chemical moieties with the expectation that there is a gain in favourable enthalpy. The ability to compute and confirm this as part of the optimization process would be extremely useful. Moreover, being able to compute enthalpy reliably should provide quantitative insight into the phenomenon of entropy-enthalpy compensation.[10]

The increase in computational power over the past decade has meant that the accurate calculation of enthalpies of binding might soon be realized and indeed work, in particular from the Gilson group on small host–guest and other systems[11–14] has showed strong potential. However, despite the promising results obtained for small model systems, the calculation of

Structural Bioinformatics and Computational Biochemistry, Department of Biochemistry, University of Oxford, South Parks Road, Oxford, OX1 3QU, UK. E-mail: philip.biggin@bioch.ox.ac.uk; Fax: +44 (0)1865 613238; Tel: +44 (0)1865 613305

enthalpy contributions for larger proteins has remained challenging and it is not even known what level of performance could be obtained even for well-characterized protein–ligand systems, such as bromodomains.[15,16]

Thus, in order to evaluate the performance of enthalpy calculations for protein–ligand binding, we assembled a data set based on bromodomains, in a similar vein to that which we had previously done for absolute binding free energy (ABFE) calculations.[15] Bromodomains (BRDs) are protein–protein interaction modules that selectively recognize acetylated lysine (K_{ac}) residues as a key event in the epigenetic reading process. A total of 61 human BRDs has been identified in 46 different proteins, which consist of eight protein families.[17] Despite these many different families, all BRDs have a conserved structure that contains a left-handed bundle of four α helices (α_Z, α_A, α_B, α_C), linked by loop regions (ZA and BC loops), which surround the K_{ac} binding site (Fig. 1a). Among eight families present in human proteome, the bromodomain and extraterminal (BET) family is characterized by two tandem N-terminal BRDs and an extraterminal (ET) domain, and is composed of BRD2, BRD3, BRD4, and BRDT.[18] BRD4 is thus a representative member of the BET family and has roles in the activation of critical genes involved in cell growth and cell cycle progression.[19]

Given the strong in cell-cycle, it is perhaps not surprising that BRD4 has also been implicated in inflammation and cancer progression resulting in many inhibitor and probe molecules being developed[17] and indeed this is an ongoing activity. Given the wealth of biophysical and structural data that already exists and the desire to develop yet more probes with improved selectivity, bromodomains represent an ideal test system for computational studies.[15,20–30] In the vast majority of complexes, the binding pocket does not show major differences in conformation, and this probably contributed to the success of rigorous free energy methods. However, it has recently been suggested the bromodomain fold is quite dynamic, including the relatively recent suggestion of cryptic pockets.[24,29,31]

The first bromodomain of BRD4, BRD4-1 has extensive biophysical and structural data associated with it (Table S1†),

and we reasoned this would provide an ideal test case for investigating our current ability to predict enthalpy and what additional insight into binding thermodynamics we might also gain. Here, we perform binding enthalpy calculations using the direct method for a non-redundant set of BRD4-1 and ligand complexes (Fig. 1b).[6,11,14] Our results show that in the very best case scenario the absolute enthalpy can be calculated for such systems to an error of about 1 kcal mol^{-1}. Obtaining results consistent across force-fields remains difficult. However, in some cases, we were able to identify the source of errors in initial outliers and that these can give insight into the well-known problem of enthalpy-entropy compensation (Fig. S1†). In particular, the role of a key loop (the ZA loop) near the binding site is discussed.

Methods

Building the benchmark

Initially, all ITC data was collected from the literature, and ITC entries with PDB structures were filtered for the further steps (Table S1†). Ligands in the PDB structures were clustered using binning clustering with default 0.4 similarity cut-off in the ChemMine Tools server (https://chemminetools.ucr.edu/).[32,33] Representative PDB structures with the best resolution were selected from each cluster for the final benchmark. We use the PDB codes as representing names for all ligands throughout the manuscript for brevity and ease of referral to structures.

System setup

The initial conformations were taken from crystal structures (3MXF, 3U5L, 4LZR, 4QB3, 4XY9, 5D0C, 5D3S, 5DW2, 5FBX, 5IGK, 2OSS). Missing atoms in the crystals were modeled with the DockPrep tool in UCSF Chimera[34] and all heteroatoms were removed from the system except the ligand of interest and all crystallographic waters. The N-terminal tail was deleted up to residue Asn54 to reduce computational cost and complexity of the simulations. Terminal residues were patched with acetylated N-terminus and amidated C-terminus using PDB Reader of CHARMM-GUI.[35,36] Ligand molecules were parameterized with the general AMBER General Force Field for organic molecules (Version 2.11, May 2016)[37] and AM1-BCC[38,39] charges using AmberTools19. We used the Amber ff14SB force field for the protein and the TIP3P water model for water molecules.[40,41] A periodic cubic water box was used for all systems with 20 000 water molecules for the complex and receptor only simulations while 2000 water molecules for used for ligand and solvent-only simulations. The exact same setup was used for the comparison of AMBER to the CHARMM and OPLS Force Fields. Protein topologies were created using the CHARMM36 (ref. 42) and OPLS-AA/M[43] force fields, while ligands topologies and parameters were generated using the CHARMM General Force Field (CGenFF)[44–46] and the OPLS/CM1A Parameter Generator for Organic Ligands (LigParGen).[47,48]
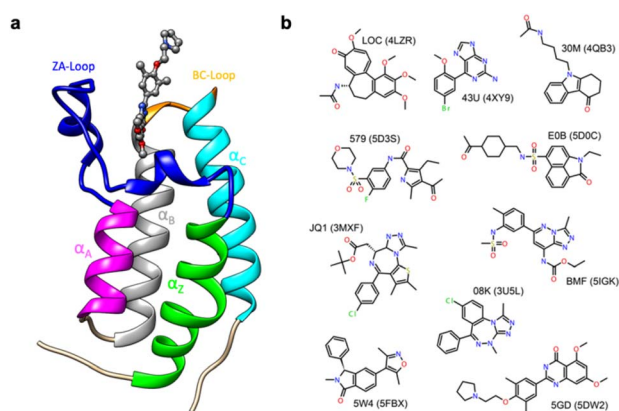

Fig. 1 (a) Cartoon of BRD4-1 with ligand complex (PDB:5DW2). (b) Chemical structures of compounds with their three-letter identifiers and corresponding PDB IDs for the complexes.

Quantum-optimized parameters

We optimized GAFF2 parameters for 08K(3U5L), 30M(4QB3), BMF(5IGK) and three different ionization states of HEPES using the Psi4 *ab initio* quantum engine at the HF/6-31G* level of theory.[49] Atomic charges were fitted to reproduce the electrostatic potential (ESP). All steps for getting optimized parameters and charges were performed using the parameterize parameterization tool (**https://software.acellera.com/docs/latest/parameterize**), which attempts to improve the quality of parameters.[50] Parameter file are available at doi: **https://doi.org/10.5281/zenodo.7534582**.

Absolute binding enthalpy calculations

The binding enthalpy (ΔH) is calculated by computing the terms in eqn (1), where $\langle E \rangle_{complex}$, $\langle E \rangle_{solvent}$, $\langle E \rangle_{receptor}$, and $\langle E \rangle_{ligand}$ are the averaged potential energies of the system as computed *via* four separate simulations (Fig. 2). In this method, the number of atoms between the bound and unbound state of the complex should exactly balance. To achieve this, we first solvate the simulation system and then delete the excess water molecules to balance the bound and unbound states. Note that the pressure–volume contribution for the binding enthalpy is negligible.[11]

$$\Delta H = \langle E \rangle_{complex} + \langle E \rangle_{solvent} - \langle E \rangle_{receptor} - \langle E \rangle_{ligand} \qquad (1)$$

All simulations were performed using the GROMACS v2020.3 software package.[51–54] A 3-step steepest descent energy minimization with a maximum force of 10 kJ mol^{-1} nm^{-2} was applied to all systems. In the first step, position restraints with an harmonic potential with a force constant of 1000 kJ (mol^{-1} nm^{-2}) were applied for all heavy atoms, then there remove for solute heavy atoms, and the final step removed all restraints. NVT and NPT ensemble simulations for 1 ns were performed to equilibrate all systems with position restraints with the harmonic potential at a force constant of 1000 kJ (mol^{-1} nm^{-2}) on heavy atoms of protein and ligand. Additionally, another NPT ensemble simulations for 1 ns was performed without restraints before the production run for data collection. The V-rescale and Parrinello–Rahman algorithms equilibrated the temperature at 300 K and the pressure at 1.0 bar, respectively.
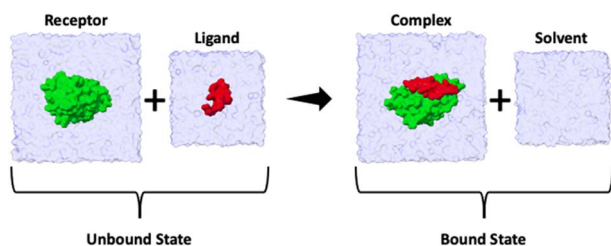

Fig. 2 Summary cartoon of the four simulations required to compute the enthalpy of binding for each ligand. Two simulations correspond to the unbound state (where $\langle E \rangle_{receptor}$ and $\langle E \rangle_{ligand}$ are computed) and two simulations to the bound state (where $\langle E \rangle_{complex}$ and $\langle E \rangle_{solvent}$ are computed). The overall enthalpy can then be computed *via* eqn (1).

The leap-frog algorithm was used to run 20 independent 100 ns MD simulations with 2 fs time step. All input files for all simulations using AMBER, CHARMM and OPLS FF are available at doi: **https://doi.org/10.5281/zenodo.7534582**.

The average of the potential energy and the estimate of the standard error were calculated by performing re-blocking analysis using the pyblock tool (**https://pyblock.readthedocs.io**)[55] for all individual calculations. In all cases blocks were used that led to the maximum standard error of the mean (SEM).

Absolute binding free energy calculations

Absolute binding free energy calculations were performed using conformations obtained from the binding enthalpy simulations. For six complexes (3U5L, 4LZR, 4QB3, 4XY9, 5DW2 and 5IGK), we performed two sets of simulations reflecting the different conformations of the ZA loop. MDRestraintsGenerator, which is a framework for generating restraints for MD simulations was used to provide the optimal Boresch restraints (1 distance, 2 angles and 3 dihedral harmonic restraints) (**https://github.com/IAlibay/MDRestraintsGenerator**). The ligand nonbounded interactions were decoupled using a linear alchemical pathway for the van der Waals and the coulombic transformations with $\Delta\lambda = 0.05$ and 0.1, respectively. The ligand restraints transformation had 12 non-equally distributed λ values (0.0, 0.01, 0.025, 0.05, 0.075, 0.1, 0.15, 0.2, 0.35, 0.5, 0.75, 1.0). Each calculation for absolute binding free energy comprised a total of 44 windows for the complex simulations and 32 windows for the ligand simulations. Each window was completed with 5 step runs. Firstly, energy minimization was carried out using the steepest descent algorithm with 10 000 steps. Then, 10 ps NVT ensemble was performed using leap-frog stochastic dynamics integrator with harmonic position restraints on the solute heavy atoms with a force constant of 1000 kJ mol^{-1} nm^{-2}. After that, 100 ps isotropic ensemble using the Berendsen coupling algorithm was run with the same position restraints. Moreover, another NPT ensemble with the Parrinello–Rahman coupling algorithm was performed for 100 ps without position restraints. Finally, 10 ns production runs were performed for data collection.

Constructing the unit cell

A crystal unit cell for 2OSS was built to investigate the effect of crystal packing on the ZA-loop conformation. The Cell Unit tool in UCSF Chimera was used for constructing the unit cell for 2OSS with the $P2_12_12_1$ space group. Missing atoms in the crystals were modeled with the DockPrep tool in UCSF Chimera and 1,2-ethanediol molecules were removed from the system while crystallographic waters were kept. The cell unit contains 4 chains with lengths 37.418, 44.139, 78.413 on the xyz dimensions. We used three different force field including Amber FF14SB,[40] CHARMM36,[56] and OPLS-AA/M[57] force field. Simulations were performed with 3 replicates for the crystal lattice and 12 replicates for the single chains.

Preparing figures

All plots were prepared by using ggplot2 which is an open-source data visualization package for the R programming language.[58] Visualization and analysis of molecular structures were made with UCSF Chimera.[34]

Results

Collating a non-redundant BRD4-1 benchmark data set

For this study, we first collected all binding data with thermodynamic components derived from ITC for BRD4-1 from the literature. Then only data sets that could be linked to high-resolution (better than 2.0 Å) were retained. Given the interest in generating probes and drug-molecules against BRD4-1, many studies report lead-optimization attempts[59,60] and this leads to many similar ligand molecules within the data set and a significant issue of redundancy. To obtain a non-redundant dataset, ligand molecules were clustered using binning with a 0.4 similarity cut-off with the ChemMine Tools server.[32,33] After clustering, we had 14 different clusters with cluster sizes ranging from 1 to 11 (Table S1†). Four clusters, each with a single protein (4OGI, 4OGJ, 5EGU, 5BT4), were filtered out since interactions between two protein chains in the crystal lattice were bridged by the ligand and this was deemed artificial. Then, representative PDB structures with the best resolution were selected from each remaining clusters for the final benchmark (Table 1 and Fig. 1b).

Absolute binding enthalpy calculations

Relative binding enthalpies have been reported for protein–ligand systems using the direct method.[6,14] Absolute calculation of binding enthalpy using the direct method requires a set of four simulations including bound and unbound states of the system (see Methods). In assessing binding enthalpies, sampling all conformational space is the key factor to achieve sufficient convergence of the potential energy. In previous studies, Roy et al. performed 40 independent 10 ns simulations to get sufficient sampling,[6] moreover, Li and Gilson reached over 250 μs simulations by seeding every 200 ns block with a new random number for the relative binding enthalpy calculation of a protein–ligand system.[14] Here, we performed 20

completely independent repeats of 100 ns simulations for each system. To assess convergence, we employed the blocking method[55] where the enthalpy is averaged over successively larger blocks and for each block size the standard error of the mean is computed. As discussed by Henriksen et al.,[13] in an ideal case the SEM will display a plateau, but this is not always the case and it is not easy to automate detection of such a plateau either. Therefore, again following the work of Henriksen et al.[13] we took the maximum SEM value (Fig. S2 and Table S2†) to err on the side of caution. For most complexes the maximum SEM is ∼0.6 kcal mol^{-1} but even the maximum (for 5FBX) is ∼1.1 kcal mol^{-1}. As expected, the ligand and solvent only profiles converge earlier than the complex.

From these initial simulations with AMBER, the correlation with experiment was moderately good with an $R^2 = 0.60$, and an average of Kendall's $\tau = 0.42$. The accuracy of the calculation is perhaps surprisingly good with an average of root-mean-square error (RMSE) = 2.49 kcal mol^{-1}. Most of the binding enthalpies are within 2 kcal mol^{-1} absolute difference of experimental values (Fig. 3 & Table S2†). The best binding enthalpy predictions were obtained for 5D3S with the XD44 (4-acetyl-3-ethyl-N-[4-fluoro-3-(morpholin-4-ylsulfonyl)phenyl]-5-methyl-1H-pyrrole-2-carboxamide) ligand (579) and 4LZR bound to colchicine (LOC).

Conversely, 3U5L in complex with a benzo-triazepine ligand (08K), 4QB3 with olinone (30M) and 5IGK with bromosporine (BMF) provided the worst binding enthalpy predictions as outliers (Fig. 3 & Table S2†).

Although some of the computed enthalpy values are in excellent agreement with the experimental data, there are some clear outliers (08K, 30M and BMF), all of which are over-estimates. We thus then sought to investigate these further.

We first considered improved parameterization. We optimized ligand charges for the outliers using the Psi4 ab initio quantum engine with the HF/6-31G* level of theory. After optimization, there was good agreement of the energy profile of GAFF2 parameters with the reference quantum mechanics (QM) calculations (Fig. S3†). Then, all simulations with 20 replicates, with a total of 120 simulations, were rerun using the optimized parameters for 3UL5, 4BQ3, and 5IGK. The binding enthalpy calculation for 4QB3 was improved at −5.13 ± 0.72 kcal mol^{-1}, which is closer to the experimental value and within

Table 1 Non-redundant BRD4(1) benchmark. All values are in kcal mol^{-1}

| PDB ID	Resolution (Å)	Space group	Ligand ID	ΔH	$T\Delta S$	ΔG	Reference
3MXF	1.60	$P2_12_12_1$	JQ1	−8.42	1.22	−9.64	Filippakopoulos et al.[61]
3U5L	1.39	$P2_12_12_1$	08K	−6.16	2.00	−8.16	Filippakopoulos et al.[17]
4LZR	1.85	$P2_12_12_1$	LOC	−9.00	−2.60	−6.40	Lucas et al.[62]
4QB3	0.94	$P2_12_12_1$	30M	−6.62	0.93	−7.55	Gacias et al.[63]
4XY9	1.83	$P2_12_12_1$	43U	−6.09	0.94	−7.03	Picaud et al.[64]
5D0C	1.49	$P2_12_12_1$	E0B	−10.20	−2.52	−7.68	Xue et al.[65]
5D3S	1.75	$P2_12_12_1$	579	−9.77	−1.73	−8.04	Hügle et al.[66]
5DW2	1.12	$P2_12_12_1$	5GD	−10.10	2.10	−8.00	Raux et al.[67]
5FBX	1.85	$P2_12_12_1$	5W4	−15.57	−4.66	−10.90	Montenegro et al.[68]
5IGK	1.70	$P2_12_12_1$	BMF	−11.09	−1.36	−9.73	Picaud et al.[69]
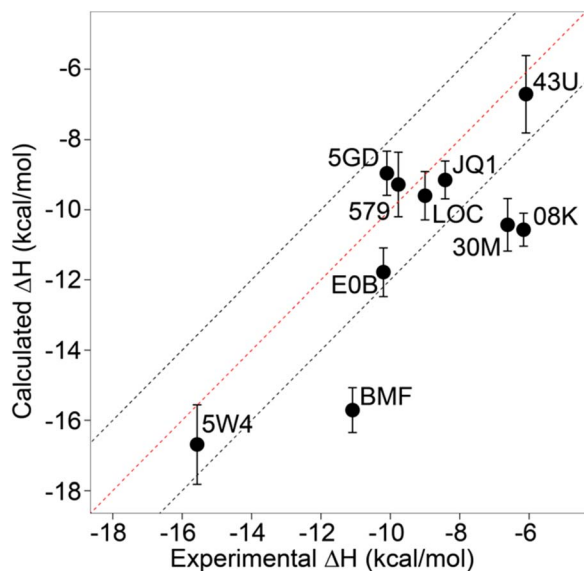
Fig. 3 Comparison of calculated binding enthalpies from experimental values. Error bars show SEM of the mean. The line of equivalence is shown in red and the black dashed lines indicate the 2 kcal mol^{-1} error limit. Three letter codes are the ligand codes as in Table 1 and Fig. 1.

2 kcal mol^{-1}. However, both 3UL5 and 5IGK still remained as outliers. The predicted binding enthalpy for 5IGK with bromosporine was slightly improved, but for 3UL5 the values actually got worse and increased the absolute difference to 7.69 kcal mol^{-1} from the experimental value (Table S3†). Thus, re-parameterization could not account for all of the poor performance.

We next considered the role of the buffer. Thus far, we have performed all simulations with pure water to reduce complexity, whereas the ITC experiments were mostly performed in 50 mM HEPES and 150 mM NaCl solution (Table S1†). However, binding thermodynamics may be sensitive to the solvent composition for both experimental[70,71] and computational[12,16] studies. To investigate the role of the buffer and in particular to what extent replicating the conditions of the experiments influenced the calculations, we set up simulations having three different ionization states of HEPES (Fig. S4†) and NaCl for the apo-receptor (2OSS) and the 3U5L and 5IGK complexes. Force field parameters of HEPES were obtained using Psi4 with the HF/6-31G* level of theory while Na$^+$ and Cl$^-$ parameters were used as provided in the Amber ff14SB force field. We then performed further simulations with 20 replicates with a total of 60 simulations of the apo-receptor (2OSS), 3U5L and 5IGK in 50 mM HEPES and 150 mM NaCl solution. Although we obtained sufficient convergence of potential energy for these simulations (data not shown), the uncertainty goes up, as expected, because HEPES and NaCl make the system more complex, and this requires longer simulations or more replicas. However, the accuracy of enthalpy prediction itself remained poor. The enthalpy for 3U5L was slightly improved but 5IGK gave a worse result than previous calculations (Table S3†). Thus, we concluded that explicit treatment of buffer in the

calculations was not the main reason for large deviations from experimental data.

The ZA-loop adopts an alternative conformation that strongly affects binding enthalpy

Simple observation of trajectories revealed a significant structural deviation in the ZA loop of some simulations, especially the apo structure, 2OSS (Fig. 4a–c and 1a). Whilst some ligands appear to stabilize the ZA-loop in the crystal-like conformation (5IGK, 5FBX, 5D3S, 5D0C, and 3U5L simulations), it is clear that the others afford the ZA loop a greater level of dynamics as evidenced by simple root-mean squared fluctuations (RMSF – Fig. 4b). Closer inspection revealed that in fact the ZA-loop can move to a distinct and alternative conformation, which in the case of the apo (2OSS) state, exists for approximately 75% of the 2 μs simulation time (Fig. 4c). In this alternative conformation, the ZA-loop moves outwards away from the acetyl-lysine binding pocket and induces a short helix (residue 88–91) within the ZA-loop. This outward movement makes the binding pocket open and more accessible. To investigate the role of this loop behaviour on the enthalpy we extended the number of repeats of the apo state to 100, to ensure that we obtained sufficient sampling of the crystal-like conformation of the ZA-loop. For the remainder of the discussion, we refer to the crystal-like conformation of the ZA-loop as ZA1 while the alternative ZA-loop conformation as ZA2.

After completing 100 replicates of the apo-receptor (a total of 10 μs of simulation), we checked the difference in potential energy between the ZA1 and ZA2 conformational states, which was 0.84 ± 0.2 kcal mol^{-1}. Previous work has reported the barrier between these states as being of the order of ~2 kcal mol^{-1}.[26] In addition to the apo-receptor (2OSS) displaying this alternative ZA2 loop conformation, so did some
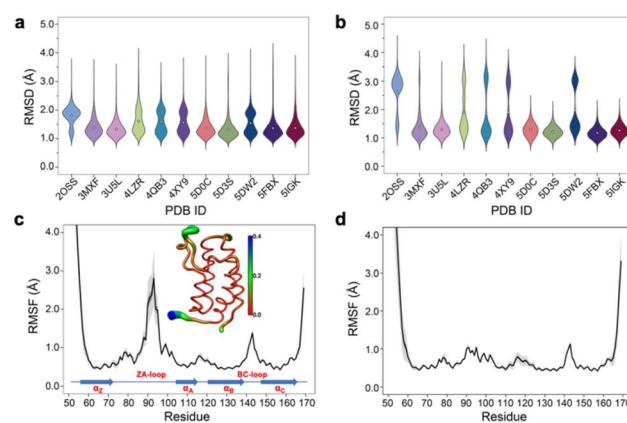

Fig. 4 (a) RMSD violin plots for backbone atoms of 2OSS apo-receptor and complex simulations, white circle represents overall mean value of all 20 repeats. (b) RMSD violin plots for ZA-loop (76–106 residues – see Fig. 1) backbone atoms of 2OSS apo-receptor and complex simulations after fitting whole protein backbone. (c) RMSF for backbone atoms of 2OSS apo-receptor. The inset figure shows regional fluctuations on the BRD4-1. (d) RMSF for backbone atoms of 5IGK complex simulations. Black line gives average RMSF of 20 repeats while grey shade represents standard deviation.

ligand-bound simulations including 4LZR, 4QB3, 4XY9, and 5DW2 (Fig. 4b and S5†). We then also checked the difference of the potential energy between ZA1 and ZA2 (Table S4†) for these complexes. The 4LZR complex gave the biggest difference of the potential energy with 5.77 ± 1.5 kcal mol^{-1} while 4XY9, at 1.83 ± 1.93 kcal mol^{-1}, had the lowest difference amongst these four complexes. Of these four complexes, only 5DW2 exhibited lower potential energy for the ZA2 conformation.

The existence of these significant differences of potential energy thus raised the question of how these two conformations affect binding enthalpy calculations. To explore this, we
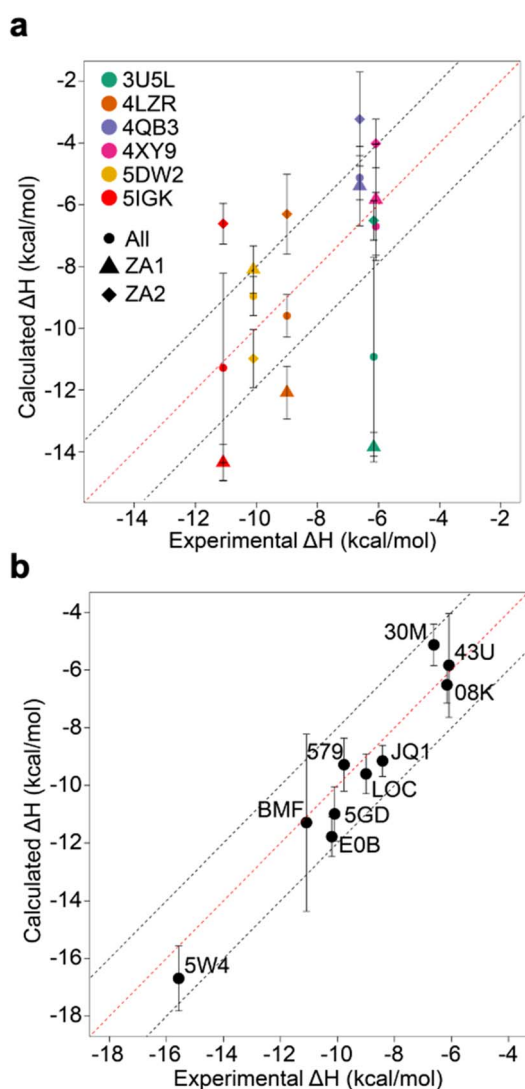

Fig. 5 Comparison of calculated binding enthalpies to experimental values. (a) Values obtained by considering only ZA1 (triangles) or ZA2 (diamonds) or all (circles) ZA loop conformations for outliers and complexes observed to adopt alternative conformations of the ZA loop. (b) The best calculated binding enthalpy values that can be obtained for the whole dataset. Note that for 4X9Y the "All" data point does not sit between the ZA1 and ZA2 points as might be intuitively expected – this is due to the nature of the blocking analysis where we use blocks that give the largest SEM. Error bars show the maximum standard error the mean estimate.

calculated binding enthalpies using ZA1 or ZA2 conformations exclusively for the 4LZR, 4QB3, 4XY9, and 5DW2 complexes (Fig. 5 & Tables S4 and S5† – note strictly speaking the apo states of ZA1 and ZA2 should be considered but in practice the 0.84 kcal mol^{-1} difference in ΔH between these two conformations in the apo state was within error). Using exclusively ZA1 or ZA2 conformations for 4LZR and 4XY9 gave less accurate predictions compared to use of both conformations (Fig. 5). 4QB3 gives a more accurate enthalpy prediction with ZA1 than with ZA2 alone or use of all simulation data combined. This also explains why the simulations with QM-refined charges (Table S3 and Fig. S3†) and gave more precise results than the initial runs, since occupation of the ZA1 conformation in the second simulation set was more than the first one (Table S5†). Conversely, 5DW2 interestingly gave more accurate binding enthalpy estimates when only the ZA2 conformation was used.

As the remaining outliers, 5IGK and 3U5L, were comprised solely of ZA1 conformations in all 60 replicates and given the above indication of the importance of the ZA loop behaviour, we decided to initiate replicates for these two complexes starting from the ZA2 conformation. For this purpose, we extracted a snapshot having the ZA2 conformation from a random apo-receptor simulation, then we manually docked BzT-7 (08K) of 3U5L and bromosporine (BMF) of 5IGK to the receptor *via* superimposition and performed 20 repeats of 100 ns. The ZA-loop stayed as the ZA2 conformation in all simulations for 3U5L (Fig. S9†) and almost all simulations for 5IGK (there was a transition from ZA2 to ZA1 in the last 20 ns of only one simulation). We obtained sufficient convergence of potential energy for these simulations, then binding enthalpies were calculated. Surprisingly, 3U5L gave a highly accurate prediction with -6.52 ± 0.64 kcal mol^{-1}, which is 0.36 kcal mol^{-1} away from the experimental value (-6.16). However, we did not obtain an improvement in accuracy of enthalpy for 5IGK with ZA2, indeed the experimental value is nearly in the middle of the predicted binding enthalpies for ZA1 and ZA2, suggesting that both conformational states contribute to the enthalpy. To confirm this, we combined simulation data for both ZA1 and ZA2 from 40 simulations in total and recomputed the binding enthalpy. The combined simulation data gave a prediction of -11.29 ± 3.07 kcal mol^{-1}, which is 0.20 kcal mol^{-1} away from the experimental value (-11.09 kcal mol^{-1}) (Fig. 5a). Taken together these computations reveal a hidden complexity of binding thermodynamics in what might be considered a relatively simple system. When the best predictions are all combined (Fig. 5b), an R^2 of 0.95 and an RMSE of 0.90 kcal mol^{-1} can be obtained, and although this was arrived at retrospectively and in dependent on known observations, it does at least illustrate that "chemical accuracy" (~1 kcal mol^{-1}) of these calculations is within reach.

The relationship to the absolute binding free energy

Given the influence of the ZA loop conformation on the enthalpy predictions, we were interested to see to what extent the ZA loop conformation also affected the absolute binding free energy, ΔG. Thus, for the four complexes that displayed alternative ZA loop conformations (see Fig. 5a); 4LZR, 4QB3,

4XY9 and 5DW2, along with the two outliers 5IGK and 3U5L (where loop stability is greater for ZA1 Fig. S7†), we computed the binding free energies (Table 2). For four of the complexes (4LZR, 4QB3, 4ZY9 and 5DW2) the loop conformation that favours the lower ΔG is mirrored by the enthalpy results (Table S4†). The 3U5L complex however, is a more complicated result. Whereas the enthalpy calculations suggest that the ZA2 conformation leads to better agreement with experiment (Fig. 5a), the free energy calculations give a lower binding free energy for the ZA1 conformation (albeit heavily overestimated compared with experiment). Similar trends for 3U5L have been reported by Heinzelmann et al.[26] using a different approach, the Attach-Pull-Release method, which produced an over-estimated result (-10.61 kcal mol^{-1}) for ZA1 but a closer-to-experiment result (-8.07 kcal mol^{-1}) for ZA2. Furthermore, an almost similar result (-9.9 ± 0.8) for ZA1 has been observed by Aldeghi et al.[15] and also by Bertin[72] (-9.1 ± 0.3) (**https://thesis.unipd.it/handle/20.500.12608/21280**) using the alchemical decoupling free energy method. Together, these results suggest that the 08K ligand in 3U5L tends to give an overestimated binding affinity when the crystallographic conformation (ZA1) is used. Compared with other ligands that bind BRD4(1), the ligand has modest enthalpic contributions to the binding free energy, but one of the most favourable entropic contributions (Table 1). Analysis of the energetic components of the enthalpy (Table S6†) shows that there is a large coulombic contribution in the ZA1 conformation that is almost completely absent in ZA2. In the case of 5IGK (bromosporine complex) the ΔG values is higher than experiment for the ZA1 loop conformation but lower for the ZA2 conformation. The value obtained for the ZA1 conformation in this work is completely consistent with the value we obtained in previous work[21] (and was initiated from a dock to the apo state, thus using the ZA1 conformation). Calculated ΔG and ΔH values for both ZA loop conformations are nearly equidistant to the experimental values for 5IGK.

The dynamics of the ZA loop

Given the clear role of the ZA-loop conformations in enthalpy prediction accuracy, we analysed the 100 apo-receptor simulations in terms of the transition between ZA1 and ZA2. The transition from ZA1 to ZA2 occurs in all simulations with a mean transition time of 22.59 ± 1.84 ns. Interestingly, the transition was in most cases irreversible and the reverse

transition from ZA2 to ZA1 happened in only two simulations. Moreover, the ZA-loop quickly transitioned back again to ZA2 whenever a reverse transition happened. Heinzelmann et al. reported ZA2 to be more favourable than the ZA1 by -2.54 kcal mol^{-1}.[26] Fig. 6a shows an example of the reverse transition from ZA2 to ZA1 around 80–90 ns. A pairwise RMSD analysis also confirmed the reverse transition from ZA2 to ZA1 in the simulation (Fig. S6†) and is clearly obvious by simple observation (Movie M1†).

During the transition, the hydrogen bond profiles of key backbone residues change dramatically (Fig. 6b and c) as well as sidechains. The transition of ZA1 to ZA2 is associated with the backbone torsion angles of ψ (N-CA-C-N) Asp88 and φ (C-N-CA-C) Asp96 (Fig. 7). These two dihedrals behave like a hinge allowing the ZA-loop to transit from ZA1 and ZA2. The ψ Asp88 shuttles from 50 to $-40°$ while the φ Asp96 moves from -150 to $-60°$. The distributions of the other backbone torsions in the ZA-loop do not show such clear-cut modal distributions, except ω (CA-C-N-CA) of Gln84 but is possibly not related to the transition from ZA1 and ZA2 (Fig. S8†). Potentially, ω GLN84 is related to a recently explained hidden transient state of the ZA-loop, where the event includes the breaking of two backbone hydrogen bonds between the ZA-loop and the α_A helix.[29]

Crystal-packing of apo-BRD4-1 explains ZA-loop conformations

Given the ubiquity of the ZA2 conformation in our simulation data, but lack of observation in crystallographic data we next investigated the role of crystal lattice packing. We built a crystal unit cell for the apo BRD4-1 receptor (2OSS) with the $P2_12_12_1$ space group (Fig. 8a). The cell unit contains 4 chains with lengths 37.418, 44.139, 78.413 Å on the x, y and z dimensions. We performed 3 replicates (100 ns) for the crystal unit and 12 replicates (100 ns) for simulations with a single chain (monomer).
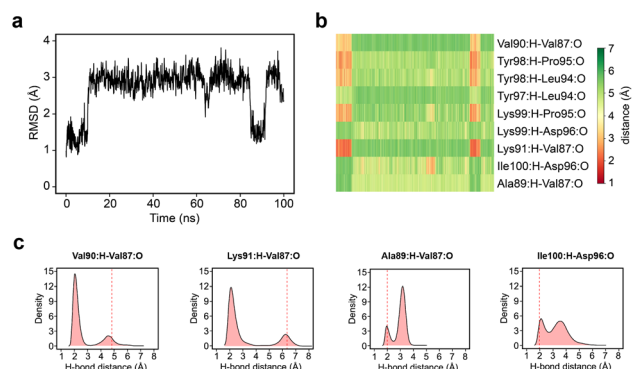

Fig. 6 (a) RMSD plot for ZA-loop (76−106 residues) backbone atoms from a representative case for the transition between ZA1 and ZA2 from one apo-receptor simulation of 100 replicates. (b) Hydrogen bond distance profile of some important backbone hydrogen bonds in the ZA-loop from the representative simulation. (c) Density plots of distance distribution for key hydrogen bonds across 100 replicates of apo-receptor simulations. The red dashed line shows the average distance of starting minimized structures from ZA1 conformations.

Table 2 ABFE results for ZA loop conformations[a]

PDB ID	ΔG_{Exp}	ΔG_{ZA1}	ΔG_{ZA2}
3U5L	-8.16	-11.36 ± 0.28	-6.85 ± 0.28
4LZR	-6.40	-6.52 ± 0.53	-2.50 ± 0.21
4QB3	-7.55	-7.55 ± 0.37	-4.94 ± 0.64
4XY9	-7.03	-4.88 ± 0.48	-2.87 ± 0.95
5DW2	-8.00	-7.69 ± 0.69	-8.10 ± 0.32
5IGK	-9.73	-11.79 ± 0.37	-4.84 ± 0.40

[a] ΔG values were obtained via running 3 independent ABFE calculation using different starting structures.
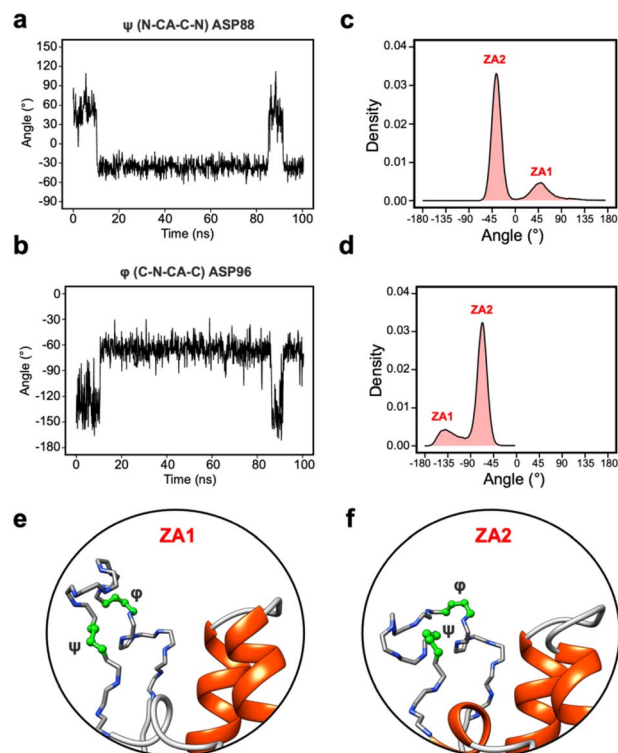
Fig. 7 (a) The angle change of the ψ (N-CA-C-N) Asp88 from the simulation in Fig. 5a. (b) The angle change of the φ (C-N-CA-C) Asp96 from the simulation in Fig. 5a. (c) The angle distribution of the ψ (N-CA-C-N) Asp88 across 100 apo-receptor simulations. (d) The angle distribution of the φ (C-N-CA-C) Asp96 across 100 apo-receptor simulations. (e and f) Main chain representation of ZA1 and ZA2 conformations. Green ball & stick regions show φ and ψ dihedral angles.

We also explored potential force-field influence, by examining two additional force fields: CHARMM36 and OPLS-AA/M. As a result, in total, we performed a total of 9 simulations for a complete unit cell and 36 simulations for a single chain. We were first interested in investigating the dynamics of ZA-loop in both simulation setups. As expected, all chains retained their crystal-like conformation in the unit cell simulations (Fig. 8b). In contrast, in all single-chain simulations in all three different force-fields, the ZA-loop exhibited much higher flexibility (Fig. 8b), thus supporting the notion that crystal-packing artefacts likely constrain the ZA conformation in the apo state. Crystal-packing effects are likely to be present in the complexes as well – simulations of similar unit-cell simulations of the complexes (Fig. S9†) reveal the ZA-loop does not move away from its lattice conformation. Almost all complex PDBs except 5D0C share same space group with the apo BRD4-1 receptor (2OSS) (see Table 1). Interestingly, simulations of the 5D0C lattice appear to allow more flexibility of the ZA-loop.

Role of force fields in general

In the previous section, we investigated the potential influence of force fields on the flexible nature of the ZA-loop and showed the force-fields behaved similarly. We thus decided to extend
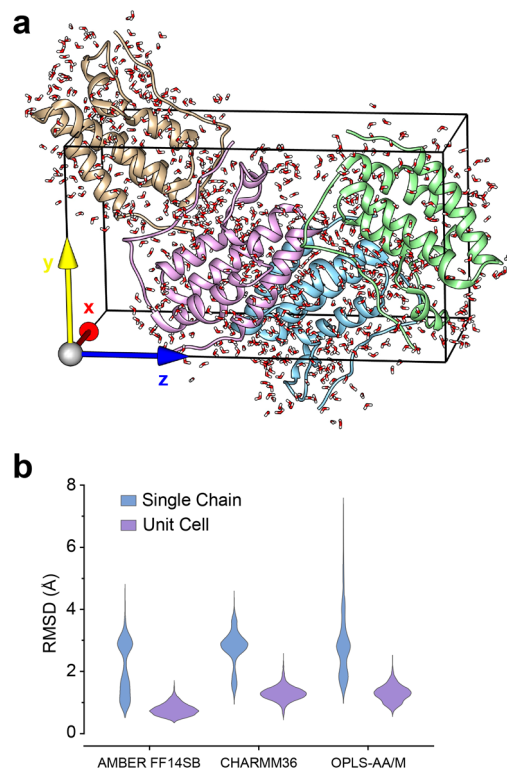

Fig. 8 (a) 2OSS crystal cell unit containing 4 identical chains. (b) RMSD violin plots for ZA-loop (76–106 residues) backbone atoms. Blue violin shows ZA-loop RMSD from 12 independent single-chain simulations while purple violins show for the unit cell simulations.

this approach further and evaluated the performance of different force-fields for computing the absolute binding enthalpies. Our calculations revealed that OPLS and AMBER showed comparable accuracy, whilst the RMSE of the CHARMM calculations was 8.07 kcal mol^{-1} (Fig. 9a). However, it is worth noting that while AMBER produced overestimated values (more negative), it had a higher correlation with experimental values than OPLS and CHARMM. We did not consider the QM optimization and the ZA loop effect in AMBER simulations to ensure a fair comparison in Fig. 9. Among the PDBs evaluated, only 4XY9 showed good agreement with experimental values for all force field (Fig. 9b and Table S7†).

The large error associated with the CHARMM calculations can be attributed to various things. For example, our results indicate that 3MXF and 5D0C with AMBER and OPLS simulations with stable ZA loops in the crystal-like conformation (ZA1) provided more accurate binding enthalpy (Fig. S11 and Table S7†). In contrast, within the CHARMM simulations, the ZA loop exhibited greater flexibility resulting in inaccurate binding enthalpy for 3MXF and 5D0C. We identified 10 and 9 simulations for 3MXF and 5D0C respectively, where the ligand completely left the binding pocket in these simulations, suggesting that default ligand parameterization within the CHARMM system may require additional optimization. Indeed, all but one ligand likely require some addition parameterization as suggested by the CGENFF server (Table S8†).
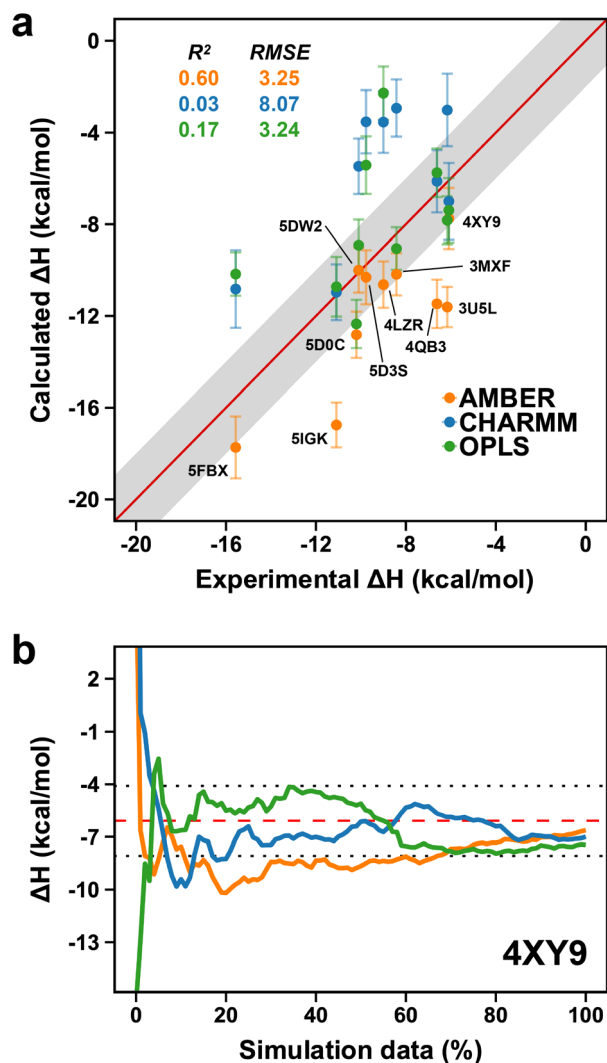
Fig. 9 (a) Comparison of AMBER (orange points) to CHARMM (blue dots) and OPLS (green dots). Labels are only shown on the AMBER data points. Error bars show the maximum standard error the mean estimate. (b) Convergence pattern of the calculated ΔH for 4XY9 with the three force fields (AMBER, orange; CHARMM, blue and OPLS, green). The red dashed-line is the experimental ΔH and the dotted-lines indicate the 2 kcal mol^{-1} error limit.

3U5L with AMBER provided an accurate enthalpy estimate with only the ZA2 conformation, while 5IGK required a combination of ZA1 and ZA2 to obtain a good prediction of enthalpy in AMBER simulations (Fig. 5a). To validate these results with CHARMM and OPLS, we also ran simulations using starting structures with both ZA1 and ZA2 for 3U5L and 5IGK (Fig. S12 and S13†). The ZA loop remained very stable when simulations were started from ZA2 in all cases (Fig. S12†). We observed that the ZA loop of 3U5L remained stable in the ZA1 conformation with AMBER and CHARMM, while it also interestingly transitioned to the ZA2 conformation with OPLS, yielding an accurate result for the binding enthalpy (Fig. S12 and Table S9†).

As CHARMM gave many outliers, we restricted further outlier analysis to OPLS. The OPLS simulations had three distinct outliers (4LZR, 5D3S, and 5FBX) that were not only different

from AMBER but also all underpredictions. Whereas for the AMBER outliers there appeared to be a link between the ZA loop conformation and the quality of the prediction, for the OPLS outliers we could not find such a correspondence.

Firstly, 4LZR yielded accurate results with the AMBER forcefield, and the simulations displayed a transition to the ZA2 conformation. In the AMBER simulations, the ZA2 conformation had an occupancy of 26.94%, while in OPLS, it was only 4.41%. Given our previous observations, we suspected that the lower occupancy of ZA2 in OPLS simulations might be responsible for the inaccurate binding enthalpy predictions. We thus conducted six additional simulations explicitly starting with the ZA2 conformation. However, these additional simulations did not result in any improvement in the outcome. We did, however, observe that the ligand in the OPLS simulations appeared to have more flexibility and indeed tended to move away from the binding pocket. Thus, in this case some refinement of the OPLS ligand parameters might well be necessary.

For 5D3S, both AMBER and OPLS displayed a similar occupancy profile for the ZA2 conformation. However, OPLS introduced greater flexibility to the ZA loop compared to AMBER. To investigate the relationship between the ZA loop dynamics and the binding enthalpy, we calculated the binding enthalpies individually for each complex simulation. Surprisingly, seven simulations provided accurate results within a 2 kcal mol^{-1} error limit. These seven simulations exhibited a much more stable ZA loop compared to the remaining simulations and yielded binding enthalpy results similar to those obtained with AMBER (Fig. S14†). Thus, although it is difficult to address why the ZA loop exhibits greater dynamics with OPLS, it does at least support the overall theme that the ZA loop behaviour is important in enthalpy predictions.

Finally, for 5FBX, although the ZA loop exhibited similar behaviour in OPLS as in AMBER, the loop was highly flexible in all simulations. We explored this by additional simulations with dihedral restraints on the ψ (N-CA-C-N) of Asp88 and φ (C-N-CA-C) of Asp96 to maintain stability in the ZA loop. However, this approach did not yield any improvement in enthalpy prediction and thus other factors must be at play here.

Discussion

Although much work has been done on model systems[12,73,74] the increase in the number of degrees of freedom associated with protein–ligand systems that are the size of typical drug-targets has led to an apparent reticence to explore enthalpy predictions and their potential utility. Given the growth in available computational power, we reasoned it would be useful to investigate the level of accuracy (and precision) one might obtain for a well-characterized protein–ligand system. To that end we focussed on bromodomains, a system for which we have previously shown that accurate ΔG predictions could be obtained *via* absolute binding free energies both for different compounds[15] at the same protein and for the same compound at different proteins.[21]

We focussed on the BRD4-1 system as it is well characterized both in terms of structure and biophysical characterizations.

We first asked the question of how accurate the calculation of enthalpy for this system would be assuming standard procedures reported in the literature, similar to our previous approach for ΔG predictions.[15] For the AMBER forcefield, the results were surprisingly accurate and precise (Fig. 3) with only three obvious outliers from a 2 kcal mol^{-1} error boundary, thus suggesting that it is indeed possible to make reasonably accurate predictions of enthalpy. Such performance is remarkably encouraging. Outliers provide an opportunity to gain more insight into what factors are important in the prediction of enthalpy. Parameterization is certainly one aspect that can offer improvement as indeed was found for some systems here when charges were optimized. This aspect is also highlighted by the direct comparison between force-fields (Fig. 9a), where it is notable that AMBER tends to overpredict and both OPLS and CHARMM tend to underpredict ΔH. It is also the case the outliers predicted for AMBER are not the same as those for OPLS and CHARMM.

Comparison of the force-fields suggests that OPLS and AMBER were of comparable (similar RMSE), while CHARMM tended to give worse predictions, likely as a consequence of ligand parameters in CGenFF requiring further optimization. This is consistent with our previous observations on cucurbit[7] uril–guest systems where CGenFF performed worse compared to other force fields.[75] The observations here reinforce the need for caution and care to be taken when using new parameters.

However, in some instances force-field was clearly not the major influence, and this led us ultimately to the identification of the influence of the ZA loop on the enthalpy prediction.

The dynamics and flexibility of the ZA loop has been reported across many different bromodomains, including for example ATAD2,[76,77] BRD2-2,[78] BRD4-1,[79,80] BRPF1,[81] BAZ2B,[82,83] BRD9,[84] BRG1 (ref. 85) and CBP.[81] Thus, it has been proposed that the dynamic nature of the ZA-loop plays an important role in selectivity due to its flexibility and sequence variation between bromodomains.[17,82,86] However, experimental 3D structures in the PDB databank mostly show no significant changes in the secondary structure despite the apparent plasticity. A recent work reported an investigation of 297 crystal structures of BRD4-1 and concluded that there is a high level of similarity in the binding pocket region, regardless of the bound ligand.[87] Nevertheless, there are many studies highlighting flexibility. Eron et al., using hydrogen–deuterium exchange mass spectrometry (HDX-MS), reported significant deuterium uptake on the ZA-loop of the apo state, but in contrast, the solvent shielding data pointed out a high degree of stabilization of the ZA-loop with the CFT-1297 degrader ligand.[88] Yu et al. reported significant chemical shift around the ZA-loop for BRD4-1 upon ligand binding even though they showed that there were no significant differences between the solution structure of BRD4-1 and its crystal structure in the backbone resonance assignment data.[89] Furthermore, ATAD2 displays "open" or "closed" conformational states of the ZA-loop.[76,90–92] Further examples from BRG1,[85] BRD7 (ref. 93) and BRD2-2,[78] show large structural deviations in the ZA-loop when compared to the other regions according to NMR experiments.

Moreover, computational studies have provided detailed analysis of the dynamic nature of the ZA-loop.[26,31,94–96] Tumdam, R., et al. showed that the ZA-loop in the apo-BRD4-1 can adopt a similar conformation to that which we observe here.[96] The nature of the transitions of the ZA loop we observe appear very similar to that reported by Heinzelmann et al.[26] who also computed that the free energy difference was 2.54 kcal mol^{-1} (with TIP3P) more favourable for their "open state" (equivalent to our ZA2 state here). Additionally, Cheng et al. showed an "in/out" transition of the ZA-loop in BRD4-1 using QM/MM, explaining the differential binding of RVX-208 & 297.[97] More recently, a hidden state across all bromodomain families was proposed via used MD simulations and Markov state modelling in which important backbone hydrogen bonds are broken and the ZA-loop displaces away from the α_A helix.[29]

Seven of the ten complexes here gave excellent predictions of enthalpy of binding. For six of those complexes, most of the simulation time is spent in the ZA1 (close to crystal) conformation. The 5DW2 complex gives an excellent prediction of enthalpy, but rather interesting, readily transitions to the ZA2 loop conformation and indeed using only these conformations gives much more accurate predictions. Analysis of the 5DW2 simulations with the ZA2 conformations reveals that the ligand makes more close contacts (<0.4 A) with Asp88 and Tyr139 in ZA2 than ZA1 (Fig. S10†).

Of the three initial outliers with AMBER, simulations of the olinone (30M) complex (4QB3) gave some improvement when charges were refined with QM calculations. However, in these simulations, the complex occupied more time in the ZA1 conformation and it appears this conformation gives a more accurate estimate of free energy (with AMBER). Interestingly, olinone appears to make interactions with the BC loop in both conformations.

Accurate enthalpy predictions for bromosporine (BMF) in complex with BRD4-1 (5IGK) were only possible when combining predictions from both ZA1 and ZA2 conformations. Whilst simulations of the BzT-7 (08K) complex (3U5L) demonstrate that more accurate predictions can be obtained with just the ZA2 conformation, even though this conformation was never transitioned to in the initial set of 20 simulations. It may be the BzT-7 creates an energy barrier for the loop transition, though this is likely an indirect effect as BzT-7 is one of the smaller ligands and does not interact directly with the ZA loop. Further work would be necessary to explore this in more detail.

Of course, enthalpy is only one component of the binding free energy, ΔG. The results above demonstrate that the interaction and behaviour of a small, but crucial loop, near the binding site plays a key role in shaping that component. The calculations performed here were all retrospective, but such calculations in the future will only be useful if they can be useful prospectively (i.e. where we do not know/have the ITC measurements). What do these results mean in that context? Firstly, the trends in the initial data are reasonably good. Secondly, if one has confidence in the parameters, then sampling of key conformation states is a key issue (as would be expected). If one has knowledge of important (for ligand-binding) dynamics up-front, then strategies can be

incorporated to mitigate this. Thirdly, if one has ITC data and ΔG calculations are in agreement but ΔH are not, that could be indicative of hidden states/cryptic sites and thus open up new avenues for design against existing targets. The ΔG would hide the presence of these states through entropy-enthalpy compensation. At the very least this would enable one to approach the predicted values with the necessary caution for sensible interpretations. Thus, in a prospective scenario, if one observes conformational transitions (in for example loops as here) and obtains enthalpy values that differ by several kcal mol^{-1}, that should at least suggest that great care should be taken over any future interpretation of the thermodynamics.

Conclusions

We have demonstrated that absolute ligand-binding enthalpy calculations for a well-characterized drug-target, BRD4-1, can give reasonably accurate results. Our results clearly show a strong dependence on the behaviour of the ZA loop to the predicted enthalpy for AMBER but this is less apparent in the OPLS simulations. We have also demonstrated that this alternative loop conformation is likely readily accessible, if not dominant, in the apo state in solution and that crystal lattice packing likely constrains the conformation. Indeed, it may well be the case that the ZA1 conformation of the loop is particularly amenable to lattice formation and thus the reason why many complexes exhibit this conformation. This observation highlights the need to take particular care when using apo state for docking studies and in the subsequent processes of rational drug design, like FEP calculations.

A key question that remains very open at this stage is just how generalizable is this approach in giving accurate predictions of ΔH. Can we expect this approach to become prospective? Prediction of ΔH may well be useful in the context of trying to optimize enthalpic contributions during a drug-discovery campaign. However, to do that with confidence in a prospective fashion will be dependent on more studies showing the approach can deliver across a variety of different systems. The work here should be taken as evidence and encouragement that it is feasible, at least for some systems. Studies on additional systems would also allow us to begin to understand how strong entropic contributions (by inference) might be linked to particular moieties or water molecules.

An alternative way to use this approach however, might be to draw researchers attention to hidden conformational states that from the initial structural biology work may not be immediately apparent. Outliers could be a way to identify such behaviour. The predicted contribution of different states to the enthalpic signature may provide a useful metric with which to gauge the importance of different states and how valuable they might be in terms of targeting. The ability to make accurate enthalpy predictions alongside accurate ΔG free energy predictions moves us considerably closer to being able to design ligands with desired thermodynamic properties, something that has long been sought after.[98] Indeed, for some systems like membrane proteins these approaches will be very important as

experimentally they are often more difficult to work with. This will be significantly easier if one has a good understanding of the dynamics of the protein before commencing such studies.

Data availability

Input coordinate and parameters files used in this study have been deposited at: **https://doi.org/10.5281/zenodo.7534582**.

Author contributions

PCB formulated the project. SSC performed all simulations. SSC and PCB analysed the data. SSC wrote the first draft and both authors contributed to revising and editing the manuscript.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

Notes and references

1 C. Chipot and A. Pohorille, *Free energy calculations: theory and applications in chemistry and biology*, Springer, 2007.

2 M. R. Shirts, D. L. Mobley and J. D. Chodera, in *Annual Reports in Computational Chemistry*, ed. D. C. Spellmeyer and R. Wheeler, Elsevier, 2007, ch. 4, vol. 3, pp. 41–59.

3 A. S. J. S. Mey, B. K. Allen, H. E. Bruce Macdonald, J. D. Chodera, D. F. Hahn, M. Kuhn, J. Michel, D. L. Mobley, L. N. Naden, S. Prasad, A. Rizzi, J. Scheen, M. R. Shirts, G. Tresadern and H. Xu, *Living J. Comput. Mol. Sci.*, 2021, **1**, 18378.

4 R. M. Levy and E. Gallicchio, *Annu. Rev. Phys. Chem.*, 1998, **49**, 531–567.

5 N. Lu, D. A. Kofke and T. B. Woolf, *J. Phys. Chem. B*, 2003, **107**, 5598–5611.

6 A. Roy, D. P. Hua, J. M. Ward and C. B. Post, *J. Chem. Theory Comput.*, 2014, **10**, 2759–2768.

7 M. A. Wyczalkowski, A. Vitalis and R. V. Pappu, *J. Phys. Chem. B*, 2010, **114**, 8166–8180.

8 E. Freire, *Drug Discovery Today*, 2008, **13**, 869–874.

9 J. D. Chodera and D. L. Mobley, *Annu. Rev. Biophys.*, 2013, **42**, 121–142.

10 J. M. Fox, M. Zhao, M. J. Fink, K. Kang and G. M. Whitesides, *Annu. Rev. Biophys.*, 2018, **47**, 223–250.

11 A. T. Fenley, N. M. Henriksen, H. S. Muddana and M. K. Gilson, *J. Chem. Theory Comput.*, 2014, **10**, 4069–4078.

12 K. Gao, J. Yin, N. M. Henriksen, A. T. Fenley and M. K. Gilson, *J. Chem. Theory Comput.*, 2015, **11**, 4555–4564.

13 N. M. Henriksen, A. T. Fenley and M. K. Gilson, *J. Chem. Theory Comput.*, 2015, **11**, 4377–4394.

14 A. Li and M. K. Gilson, *J. Chem. Phys.*, 2018, **149**, 072311.

15 M. Aldeghi, A. Heifetz, M. J. Bodkin, S. Knapp and P. C. Biggin, *Chem. Sci.*, 2016, **7**, 207–218.

16 D. L. Mobley and M. K. Gilson, *Annu. Rev. Biophys.*, 2017, **46**, 531–558.

17 P. Filippakopoulos and S. Knapp, *FEBS Lett.*, 2012, **586**, 2692–2704.

18 S. Shu and K. Polyak, *Cold Spring Harbor Symp. Quant. Biol.*, 2016, **81**, 123–129.

19 C.-Y. Wang and P. Filippakopoulos, *Trends Biochem. Sci.*, 2015, **40**, 468–479.

20 M. Aldeghi, M. J. Bodkin, S. Knapp and P. C. Biggin, *J. Chem. Inf. Model.*, 2017, **57**, 2203–2221.

21 M. Aldeghi, A. Heifetz, M. J. Bodkin, S. Knapp and P. C. Biggin, *J. Am. Chem. Soc.*, 2017, **139**, 946–957.

22 M. Aldeghi, G. A. Ross, M. J. Bodkin, J. W. Essex, S. Knapp and P. C. Biggin, *Commun. Chem.*, 2018, **1**, 19.

23 M. Brand, J. Clayton, M. Moroglu, M. Schiedel, S. Picaud, J. P. Bluck, A. Skwarska, H. Bolland, A. K. N. Chan, C. M. C. Laurin, A. R. Scorah, L. See, T. P. C. Rooney, K. H. Andrews, O. Fedorov, G. Perell, P. Kalra, K. B. Vinh, W. A. Cortopassi, P. Heitel, K. E. Christensen, R. I. Cooper, R. S. Paton, W. C. K. Pomerantz, P. C. Biggin, E. M. Hammond, P. Filippakopoulos and S. J. Conway, *J. Med. Chem.*, 2021, **64**, 10102–10123.

24 A. Dickson, *Biophys. J.*, 2018, **115**, 1707–1719.

25 E. E. Guest, L. F. Cervantes, S. D. Pickett, C. L. Brooks and J. D. Hirst, *J. Chem. Inf. Model.*, 2022, **62**, 1458–1470.

26 G. Heinzelmann, N. M. Henriksen and M. K. Gilson, *J. Chem. Theory Comput.*, 2017, **13**, 3260–3275.

27 L. E. Jennings, M. Schiedel, D. S. Hewings, S. Picaud, C. M. C. Laurin, P. A. Bruno, J. P. Bluck, A. R. Scorah, L. See, J. K. Reynolds, M. Moroglu, I. N. Mistry, A. Hicks, P. Guzanov, J. Clayton, C. Evans, G. Stazi, P. C. Biggin, A. K. Mapp, E. M. Hammond, P. G. Humphreys, P. Filippakopoulos and S. J. Conway, *Bioorg. Med. Chem.*, 2018, **26**, 2937–2957.

28 C. M. C. Laurin, J. P. Bluck, A. K. N. Chan, M. Keller, A. Boczek, A. R. Scorah, K. F. L. See, L. E. Jennings, D. S. Hewings, F. Woodhouse, J. K. Reynolds, M. Schiedel, P. G. Humphreys, P. C. Biggin and S. J. Conway, *ACS Infect. Dis.*, 2021, **7**, 2238–2249.

29 L. Raich, K. Meier, J. Günther, C. D. Christ, F. Noe and S. Olsson, *bioRxiv*, 2020, preprint, DOI: 10.1101/2020.04.01.019547.

30 S. Wan, A. P. Bhati, S. J. Zasada, I. Wall, D. Green, P. Bamborough and P. V. Coveney, *J. Chem. Theory Comput.*, 2017, **13**, 784–795.

31 X. Zhang, K. Chen, Y.-D. Wu and O. Wiest, *PLoS One*, 2017, **12**, e0186570.

32 T. W. H. Backman, Y. Cao and T. Girke, *Nucleic Acids Res.*, 2011, **39**, W486–W491.

33 Y. Cao, A. Charisi, L.-C. Cheng, T. Jiang and T. Girke, *Bioinformatics*, 2008, **24**, 1733–1734.

34 E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, *J. Comput. Chem.*, 2004, **25**, 1605–1612.

35 S. Jo, X. Cheng, J. Lee, S. Kim, S.-J. Park, D. S. Patel, A. H. Beaven, K. I. Lee, H. Rui, S. Park, H. S. Lee, B. Roux, A. D. MacKerell Jr, J. B. Klauda, Y. Qi and W. Im, *J. Comput. Chem.*, 2017, **38**, 1114–1124.

36 S. Jo, T. Kim, V. G. Iyer and W. Im, *J. Comput. Chem.*, 2008, **29**, 1859–1865.

37 X. He, V. H. Man, W. Yang, T.-S. Lee and J. Wang, *J. Chem. Phys.*, 2020, **153**, 114502.

38 A. Jakalian, D. B. Jack and C. I. Bayly, *J. Comput. Chem.*, 2002, **23**, 1623–1641.

39 A. Jakalian, B. L. Bush, D. B. Jack and C. I. Bayly, *J. Comput. Chem.*, 2000, **21**, 132–146.

40 J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser and C. Simmerling, *J. Chem. Theory Comput.*, 2015, **11**, 3696–3713.

41 P. Mark and L. Nilsson, *J. Phys. Chem. A*, 2001, **105**(43), 9954–9960.

42 J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B. L. de Groot, H. Grubmueller and A. D. MacKerell Jr, *Nat. Methods*, 2017, **14**, 71–73.

43 M. J. Robertson, J. Tirado-Rives and W. L. Jorgensen, *J. Chem. Theory Comput.*, 2015, **11**, 3499–3509.

44 K. Vanommeslaeghe and A. D. MacKerell Jr, *J. Chem. Inf. Model.*, 2012, **52**, 3144–3154.

45 K. Vanommeslaeghe, E. P. Raman and A. D. MacKerell Jr, *J. Chem. Inf. Model.*, 2012, **52**, 3155–3168.

46 K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov and A. D. MacKerell Jr, *J. Comput. Chem.*, 2010, **31**, 671–690.

47 L. S. Dodda, I. C. de Vaca, J. Tirado-Rives and W. L. Jorgensen, *Nucleic Acids Res.*, 2017, **45**, W331–W336.

48 L. S. Dodda, J. Z. Vilseck, J. Tirado-Rives and W. L. Jorgensen, *J. Phys. Chem. B*, 2017, **121**, 3864–3870.

49 R. M. Parrish, L. A. Burns, D. G. A. Smith, A. C. Simmonett, A. E. DePrince, E. G. Hohenstein, U. Bozkaya, A. Y. Sokolov, R. Di Remigio, R. M. Richard, J. F. Gonthier, A. M. James, H. R. McAlexander, A. Kumar, M. Saitow, X. Wang, B. P. Pritchard, P. Verma, H. F. Schaefer, K. Patkowski, R. A. King, E. F. Valeev, F. A. Evangelista, J. M. Turney, T. D. Crawford and C. D. Sherrill, *J. Chem. Theory Comput.*, 2017, **13**, 3185–3197.

50 R. Galvelis, S. Doerr, J. M. Damas, M. J. Harvey and G. De Fabritiis, *J. Chem. Inf. Model.*, 2019, **59**, 3485–3493.

51 S. Páll, A. Zhmurov, P. Bauer, M. Abraham, M. Lundborg, A. Gray, B. Hess and E. Lindahl, *J. Chem. Phys.*, 2020, **153**, 134110.

52 M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindahl, *SoftwareX*, 2015, **1–2**, 19–25.

53 H. J. C. Berendsen, D. van der Spoel and R. van Drunen, *Comput. Phys. Commun.*, 1995, **95**, 43–56.

54 D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark and H. J. C. Berendsen, *J. Comput. Chem.*, 2005, **26**, 1701–1718.

55 H. Flyvbjerg and H. G. Petersen, *J. Chem. Phys.*, 1989, **91**, 461–466.

56 J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B. L. de Groot, H. Grubmüller and A. D. MacKerell, *Nat. Methods*, 2017, **14**, 71–73.

57 M. J. Robertson, J. Tirado-Rives and W. L. Jorgensen, *J. Chem. Theory Comput.*, 2015, **11**, 3499–3509.

58 H. Wickham, *ggplot2: Elegant Graphics for Data Analysis*, Springer-Verlag, New York, 2009.

59 M. Hügle, P. Regenass, R. Warstat, M. Hau, K. Schmidtkunz, X. Lucas, D. Wohlwend, O. Einsle, M. Jung, B. Breit and S. Günther, *J. Med. Chem.*, 2020, **63**, 15603–15620.

60 M. Zhang, Y. Zhang, M. Song, X. Xue, J. Wang, C. Wang, C. Zhang, C. Li, Q. Xiang, L. Zou, X. Wu, C. Wu, B. Dong, W. Xue, Y. Zhou, H. Chen, D. Wu, K. Ding and Y. Xu, *J. Med. Chem.*, 2018, **61**, 3037–3058.

61 P. Filippakopoulos, J. Qi, S. Picaud, Y. Shen, W. B. Smith, O. Fedorov, E. M. Morse, T. Keates, T. T. Hickman, I. Felletar, M. Philpott, S. Munro, M. R. McKeown, Y. Wang, A. L. Christie, N. West, M. J. Cameron, B. Schwartz, T. D. Heightman, N. La Thangue, C. A. French, O. Wiest, A. L. Kung, S. Knapp and J. E. Bradner, *Nature*, 2010, **468**, 1067–1073.

62 X. Lucas, D. Wohlwend, M. Hügle, K. Schmidtkunz, S. Gerhardt, R. Schüle, M. Jung, O. Einsle and S. Günther, *Angew. Chem., Int. Ed.*, 2013, **52**, 14055–14059.

63 M. Gacias, G. Gerona-Navarro, A. N. Plotnikov, G. Zhang, L. Zeng, J. Kaur, G. Moy, E. Rusinova, Y. Rodriguez, B. Matikainen, A. Vincek, J. Joshua, P. Casaccia and M.-M. Zhou, *Chem. Biol.*, 2014, **21**, 841–854.

64 S. Picaud, M. Strocchia, S. Terracciano, G. Lauro, J. Mendez, D. L. Daniels, R. Riccio, G. Bifulco, I. Bruno and P. Filippakopoulos, *J. Med. Chem.*, 2015, **58**, 2718–2736.

65 X. Xue, Y. Zhang, Z. Liu, M. Song, Y. Xing, Q. Xiang, Z. Wang, Z. Tu, Y. Zhou, K. Ding and Y. Xu, *J. Med. Chem.*, 2016, **59**, 1565–1579.

66 M. Hügle, X. Lucas, G. Weitzel, D. Ostrovskyi, B. Breit, S. Gerhardt, O. Einsle, S. Günther and D. Wohlwend, *J. Med. Chem.*, 2016, **59**, 1518–1530.

67 B. Raux, Y. Voitovich, C. Derviaux, A. Lugari, E. Rebuffet, S. Milhas, S. Priet, T. Roux, E. Trinquet, J.-C. Guillemot, S. Knapp, J.-M. Brunel, A. Y. Fedorov, Y. Collette, P. Roche, S. Betzi, S. Combes and X. Morelli, *J. Med. Chem.*, 2016, **59**, 1634–1641.

68 R. C. Montenegro, P. G. K. Clark, A. Howarth, X. Wan, A. Ceroni, P. Siejka, G. A. Nunez-Alonso, O. Monteiro, C. Rogers, V. Gamble, R. Burbano, P. E. Brennan, C. Tallant, D. Ebner, O. Fedorov, E. Neill, S. Knapp, D. Dixon and S. Müller, *Oncotarget*, 2016, **7**, 43997–44012.

69 S. Picaud, K. Leonards, J.-P. Lambert, O. Dovey, C. Wells, O. Fedorov, O. Monteiro, T. Fujisawa, C.-Y. Wang, H. Lingard, C. Tallant, N. Nikbin, L. Guetzoyan, R. Ingham, S. V. Ley, P. Brennan, S. Muller, A. Samsonova,

A.-C. Gingras, J. Schwaller, G. Vassiliou, S. Knapp and P. Filippakopoulos, *Sci. Adv.*, 2016, **2**, e1600760.

70 T. Fiala, K. Sleziakova, K. Marsalek, K. Salvadori and V. Sindelar, *J. Org. Chem.*, 2018, **83**, 1903–1912.

71 V. H. Le, M. Yanney, M. McGuire, A. Sygula and E. A. Lewis, *J. Phys. Chem. B*, 2014, **118**, 11956–11964.

72 G. Bertin, https://hdl.handle.net/20.500.12608/21280, Padua, 2020.

73 S. S. Çınaroğlu and P. C. Biggin, *J. Phys. Chem. B*, 2021, **125**, 1558–1567.

74 C. Schönbeck and R. Holm, *J. Phys. Chem. B*, 2019, **123**, 6686–6693.

75 S. S. Cinaroglu and P. C. Biggin, *J. Phys. Chem. B*, 2021, **125**, 1558–1567.

76 G. Poncet-Montange, Y. Zhan, J. P. Bardenhagen, A. Petrocchi, E. Leo, X. Shi, G. R. Lee IV, P. G. Leonard, M. K. Geck Do, M. G. Cardozo, J. N. Andersen, W. S. Palmer, P. Jones and J. E. Ladbury, *Biochem. J.*, 2015, **466**, 337–346.

77 Y. Zhou, M. Hussain, G. Kuang, J. Zhang and Y. Tu, *Physical Chemistry Chemical Physics*, 2018, **20**(36), 23222–23232.

78 H. Huang, J. Zhang, W. Shen, X. Wang, J. Wu, J. Wu and Y. Shi, *BMC Struct. Biol.*, 2007, **7**, 57.

79 C. Cheng, H. Diao, F. Zhang, Y. Wang, K. Wang and R. Wu, *PCCP Phys. Chem. Chem. Phys.*, 2017, **19**, 23934–23941.

80 S. J. Eron, H. Huang, R. V. Agafonov, M. E. Fitzgerald, J. Patel, R. E. Michael, T. D. Lee, A. A. Hart, J. Shaulsky, C. G. Nasveschuk, A. J. Phillips, S. L. Fisher and A. Good, *ACS Chem. Biol.*, 2021, **16**(11), 2228–2243.

81 J. Zhu, C. Zhou and A. Caflisch, *Eur. J. Med. Chem.*, 2018, **155**, 337–352.

82 S. Steiner, A. Magno, D. Huang and A. Caflisch, *FEBS Lett.*, 2013, **587**, 2158–2163.

83 F. M. Ferguson, D. M. Dias, J. P. G. L. M. Rodrigues, H. Wienk, R. Boelens, A. M. J. J. Bonvin, C. Abell and A. Ciulli, *Biochemistry*, 2014, **53**, 6706–6716.

84 N. H. Theodoulou, P. Bamborough, A. J. Bannister, I. Becher, R. A. Bit, K. H. Che, C.-w. Chung, A. Dittmann, G. Drewes, D. H. Drewry, L. Gordon, P. Grandi, M. Leveridge, M. Lindon, A.-M. Michon, J. Molnar, S. C. Robson, N. C. O. Tomkinson, T. Kouzarides, R. K. Prinjha and P. G. Humphreys, *J. Med. Chem.*, 2016, **59**, 1425–1439.

85 W. Shen, C. Xu, W. Huang, J. Zhang, J. E. Carlson, X. Tu, J. Wu and Y. Shi, *Biochemistry*, 2007, **46**, 2100–2110.

86 L. R. Vidler, N. Brown, S. Knapp and S. Hoelder, *J. Med. Chem.*, 2012, **55**, 7346–7359.

87 E. E. Guest, S. D. Pickett and J. D. Hirst, *Org. Biomol. Chem.*, 2021, **19**, 5632–5641.

88 S. J. Eron, H. Huang, R. V. Agafonov, M. E. Fitzgerald, J. Patel, R. E. Michael, T. D. Lee, A. A. Hart, J. Shaulsky, C. G. Nasveschuk, A. J. Phillips, S. L. Fisher and A. Good, *ACS Chem. Biol.*, 2021, **16**, 2228–2243.

89 J.-l. Yu, T.-t. Chen, C. Zhou, F.-l. Lian, X.-l. Tang, Y. Wen, J.-k. Shen, Y.-c. Xu, B. Xiong and N.-x. Zhang, *Acta Pharmacol. Sin.*, 2016, **37**, 984–993.

90 C. Langini, A. Caflisch and A. Vitalis, *J. Biol. Chem.*, 2017, **292**, 16734–16745.

91 J. T. Lloyd, K. McLaughlin, M. Y. Lubula, J. C. Gay, A. Dest, C. Gao, M. Phillips, M. Tonelli, G. Cornilescu, M. R. Marunde, C. M. Evans, S. P. Boyson, S. Carlson, M.-C. Keogh, J. L. Markley, S. Frietze and K. C. Glass, *J. Med. Chem.*, 2020, **63**, 12799–12813.

92 Y. Zhou, M. Hussain, G. Kuang, J. Zhang and Y. Tu, *PCCP Phys. Chem. Chem. Phys.*, 2018, **20**, 23222–23232.

93 H. Sun, J. Liu, J. Zhang, W. Shen, H. Huang, C. Xu, H. Dai, J. Wu and Y. Shi, *Biochem. Biophys. Res. Commun.*, 2007, **358**, 435–441.

94 M. Kuang, J. Zhou, L. Wang, Z. Liu, J. Guo and R. Wu, *J. Chem. Inf. Model.*, 2015, **55**, 1926–1935.

95 J. Su, X. Liu, S. Zhang, F. Yan, Q. Zhang and J. Chen, *J. Biomol. Struct. Dyn.*, 2018, **36**, 1212–1224.

96 R. Tumdam, A. Kumar, N. Subbarao and B. S. Balaji, *SAR QSAR Environ. Res.*, 2018, **29**, 975–996.

97 C. Cheng, H. Diao, F. Zhang, Y. Wang, K. Wang and R. Wu, *Phys. Chem. Chem. Phys.*, 2017, **19**, 23934–23941.

98 G. Klebe, *Drug Discovery Today*, 2019, **24**, 943–948.