

# Computed Protein–Protein Enthalpy Signatures as a Tool for Identifying Conformation Sampling Problems

Süleyman Selim Çınaroğlu and Philip C. Biggin\*



Cite This: *J. Chem. Inf. Model.* 2023, 63, 6095–6108



Read Online

ACCESS |



Metrics & More

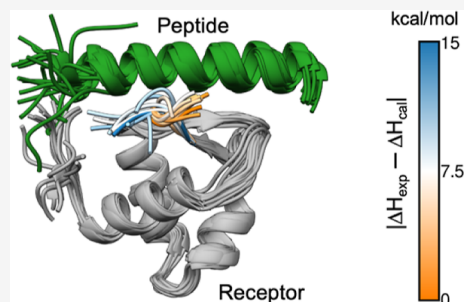


Article Recommendations



Supporting Information

**ABSTRACT:** Understanding the thermodynamic signature of protein–peptide binding events is a major challenge in computational chemistry. The complexity generated by both components possessing many degrees of freedom poses a significant issue for methods that attempt to directly compute the enthalpic contribution to binding. Indeed, the prevailing assumption has been that the errors associated with such approaches would be too large for them to be meaningful. Nevertheless, we currently have no indication of how well the present methods would perform in terms of predicting the enthalpy of binding for protein–peptide complexes. To that end, we carefully assembled and curated a set of 11 protein–peptide complexes where there is structural and isothermal titration calorimetry data available and then computed the absolute enthalpy of binding. The initial “out of the box” calculations were, as expected, very modest in terms of agreement with the experiment. However, careful inspection of the outliers allows for the identification of key sampling problems such as distinct conformations of peptide termini not being sampled or suboptimal cofactor parameters. Additional simulations guided by these aspects can lead to a respectable correlation with isothermal titration calorimetry (ITC) experiments ( $R^2$  of 0.88 and an RMSE of 1.48 kcal/mol overall). Although one cannot know prospectively whether computed ITC values will be correct or not, this work shows that if experimental ITC data are available, then this in conjunction with computed ITC, can be used as a tool to know if the ensemble being simulated is representative of the true ensemble or not. That is important for allowing the correct interpretation of the detailed dynamics of the system with respect to the measured enthalpy. The results also suggest that computational calorimetry is becoming increasingly feasible. We provide the data set as a resource for the community, which could be used as a benchmark to help further progress in this area.



## INTRODUCTION

Protein–protein interactions play a vital role in many biological processes and underpin cellular signal transduction events within (and beyond) the cell.<sup>1</sup> A deeper understanding of the interactome will be necessary to fully appreciate the role of protein–protein interactions in pathology<sup>2</sup> and will aid in the identification of novel targets for therapeutic intervention. The need for safer, more effective medicines is more urgent than ever.<sup>3</sup> Protein–protein interactions are a promising and rapidly growing area of drug discovery, with important implications for treating many diseases.<sup>4</sup>

Understanding the principles of protein–protein association requires a comprehensive description of the thermodynamic and kinetic characteristics of binding.<sup>5</sup> In this work, we focus solely on the thermodynamic aspects. Identifying the driving forces that stabilize the interaction has been actively pursued for decades.<sup>6,7</sup> Protein–protein binding is a complicated process, including hydrophobic, van der Waals, and electrostatic interactions. Three-dimensional structures of protein–protein complexes can be obtained using X-ray crystallography, NMR spectroscopy, and cryo-electron microscopy, which provide valuable information on critical interactions between binding site residues for an atomic-level understanding of binding mechanisms.<sup>8</sup> Alongside the structural efforts,

calorimetric approaches that provide thermodynamic signatures of binding give complementary insight into the forces that stabilize protein–protein complexes.<sup>9</sup>

While there are many techniques for obtaining binding affinity, isothermal titration calorimetry (ITC) can sensitively measure the enthalpic ( $\Delta H$ ) component, and thus, if  $\Delta G$  is known, then the entropy ( $\Delta S$ ) can be also derived to provide a full thermodynamic “signature” for a binding event. However, the quantitative prediction of binding affinities (and indeed the underlying ensembles of structures that give rise to it) remains a significant challenge for computational chemistry.<sup>10,11</sup> Surmounting this challenge would aid our understanding of regulation in biological systems and also improve our prospects for drug design and development.

The development of computing architecture and new algorithms has led to a surge in the use of molecular dynamics

**Received:** July 10, 2023

**Published:** September 28, 2023



**Table 1. Nonredundant Data Set for Testing Binding Enthalpies of Protein–Peptide Systems**

PDB	cluster ID <sup>a</sup>	receptor <sup>b</sup>	peptide <sup>b</sup>	$\Delta H$ (kcal/mol) <sup>c</sup>	refs
1DPU	7(1)	P15927 <sub>(201–272)</sub>	P13051 <sub>(66–91)</sub>	$-16.80 \pm 0.28$	Xie et al <sup>32</sup>
1RST	4(1)	P22629 <sub>(39–163)</sub>	AWRHPQFGG	$-12.56 \pm 0.09$	Schmidt et al <sup>33</sup>
2LQC	8(5)	P0DP23 <sub>(1–77)</sub>	Q13936 <sub>(45–68)</sub>	$-6.91 \pm 0.07$	Liu and Vogel <sup>34</sup>
2MNU	8(1)	P02751 <sub>(5–83)</sub>	APT <sub>(201–226)</sub>	$-4.60 \pm 0.10$	Yu et al <sup>35</sup>
2MWY	5(6)	O15151 <sub>(23–111)</sub>	P04637 <sub>(15–29)</sub>	$-17.50 \pm 0.30$	Grace et al <sup>36</sup>
4F14	9(3)	O76041 <sub>(955–1014)</sub>	A4UGR9 <sub>(2245–2258)</sub>	$-8.46 \pm 0.48$	Eulitz et al <sup>37</sup>
4Q6F	11(4)	Q9UIF9 <sub>(1673–1728)</sub>	P68431 <sub>(1–9)</sub>	$-9.81 \pm 0.04$	Tallant et al <sup>38</sup>
5E0M	10(8)	Q24117 <sub>(410–478)</sub>	Q9H4H8 <sub>(434–446)</sub>	$-8.20 \pm 0.10$	Clark et al <sup>39</sup>
SOVC	2(2)	Q9JLU4 <sub>(570–664)</sub>	P97836 <sub>(986–992)</sub>	$-6.80 \pm 0.04$	Ponna et al <sup>40</sup>
6EVO	3(9)	O15460 <sub>(142–236)</sub>	PPGPRGPPG	$-8.70 \pm 0.80$	Murthy et al <sup>41</sup>
6H8C	6(7)	P60520 <sub>(3–117)</sub>	Q9GZZ9 <sub>(337–350)</sub>	$-5.91 \pm 0.09$	Huber et al <sup>42</sup>

<sup>a</sup>Cluster IDs from sequence-based clusters with structure-based cluster IDs in parentheses. As cluster 1 from sequence-based clustering was a single structure, this was omitted. <sup>b</sup>UniProt IDs for receptors and peptides; some peptides are shown as their amino acid residues. <sup>c</sup>Experimental binding enthalpy values.

(MD) simulation tools over recent decades.<sup>12</sup> MD simulations can provide unique insight into the behavior of molecules and their interactions at atomic resolution. Using these simulations, researchers can also estimate thermodynamic quantities to gain knowledge about microscopic details and relate thermodynamic measurements to physical interactions. The advancement of computational methods for estimating binding free energy is well documented.<sup>13</sup> However, the prediction of the  $\Delta S$  and  $\Delta H$  components of binding remains more complicated. Estimating the  $\Delta S$  and  $\Delta H$  components is generally less accurate than estimating the binding free energy itself. The direct method of estimating enthalpy is the most straightforward approach and allows a direct explanation of the physical forces involved.<sup>14</sup> However, the direct method relies on the difference between the energies of the bound and unbound states calculated using four separate simulations.<sup>15</sup> Furthermore, the reliability of the direct method depends on sufficient sampling that can realistically only be obtained with a huge amount of simulation data.<sup>11</sup>

In this work, we set out to explore two aspects of protein–protein enthalpy. First, we wanted to know what kind of performance one could expect with a typical force field. Second, and in anticipation of outliers, what conclusions could be drawn from them about the sampling of the ensemble? To do this, we compute the absolute binding enthalpy from MD simulations of a diverse set of protein–protein complexes using the direct method with the multibox approach. Although the theory has been around for some time, there are very few reports of its application to anything more than host–guest systems, presumably because researchers have assumed that the sampling and consequent error would make the results too difficult to interpret. We show here that although it is still extremely challenging to prospectively predict enthalpy, the (lack of) agreement between experiment and prediction can be used as an indicator of when conformational sampling may not be reflective of the correct ensemble—perhaps as a result of being stuck in a metastable state, which by many metrics could give the illusion of complete sampling. Finally, the data presented here could serve as a benchmark for future studies.

## METHODS

**Building a Nonredundant Data Set.** To investigate how well protein–protein enthalpy can be predicted, we first obtained all protein–protein complexes from the PDBbind v2020.<sup>16</sup> In all cases, the chain with the longest sequence

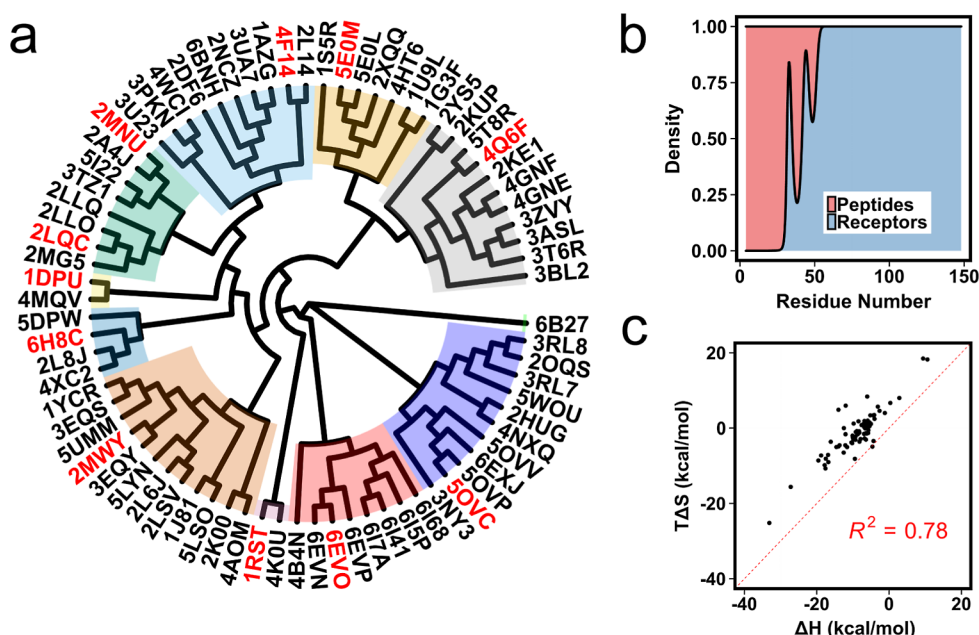
length was defined as the receptor, while the short chain was defined as the peptide for each PDB entry. PDB entries having receptors with sequences longer than 150 residues were eliminated to reduce the computational cost in MD simulations. The remaining PDB entries (736) were then investigated as to whether they had corresponding binding enthalpy data obtained by ITC associated with them in the literature. This resulted in 76 PDBs.

We then employed two different clustering approaches to reduce the redundancy in the data. First, we extracted receptor amino acid sequences from RCSB PDB,<sup>17</sup> then the amino acid sequences were aligned using Clustal omega<sup>18</sup> and clusters were defined by the result of their phylogenetic relationships of the sequences. Second, a sequence-independent clustering based on the local backbone similarity matching was performed using the MaxCluster tool for receptors.<sup>19</sup> The final data set was selected by choosing the lowest number of total residues of the protein–peptide system from each cluster, with a view to reducing computational cost as far as possible. We also paid attention to the existence of complete ITC details and the same temperature (25 °C) when choosing the representatives. The final data set is composed of 11 complexes.

**MD Simulation Setup.** The PDB structures were obtained from the RCSB PDB database for the data set (Table 1). Missing atoms and loops were modeled with the Modeler implemented in UCSF Chimera,<sup>20</sup> and all heteroatoms were removed from the system except all crystallographic waters. We used the AMBER ff14SB force field for the protein and the TIP3P water model for water molecules.<sup>21</sup> The zinc AMBER force field (ZAFF) was additionally used for Zn<sup>2+</sup> atoms parameterization in the PHD zinc finger of BAZ2A (PDB: 4Q6F),<sup>22,23</sup> while default parameters in ff14SB were used for Ca<sup>2+</sup> atoms in calmodulin (PDB: 2LQC).

All simulations were run using the Gromacs v2020 software package.<sup>24</sup> A 3-step steepest descent energy minimization with a maximum force of 10 kJ/mol/nm<sup>2</sup> was applied to all systems.<sup>25</sup> In the first step, position restraints with a harmonic potential with a force constant of 1000 kJ/(mol/nm<sup>2</sup>) were applied to all heavy atoms, followed by a second step with position restraints only on the solute heavy atoms and then a final step with no restraints.

NVT and NPT ensemble simulations for 1 ns were performed to equilibrate all systems with positional restraints with the harmonic potential at a force constant of 1000 kJ/



**Figure 1.** (a) Phylogenetic tree of receptor sequences from 76 PDBs. Red PDB IDs are the ones used for the final benchmark. (b) Density distribution for sequence lengths of peptides and receptors. (c) Entropy–enthalpy scatter plot for the 76 protein–protein complexes.

(mol/nm<sup>2</sup>) on heavy atoms of protein and ligand. Additionally, another NPT ensemble simulation for 1 ns was performed without restraints before the production run for data collection. The V-rescale<sup>26</sup> and Parrinello and Rahman<sup>27</sup> algorithms equilibrated the temperature at 300 K and the pressure at 1.0 bar, respectively. Unbonded interactions were calculated up to a cutoff of 1.0 nm with a potential shift. A dispersion correction was applied to the energy and pressure. All H-bond lengths were constrained with a LINear Constraint Solver (LINCS) algorithm.<sup>28</sup> Coulomb interactions were evaluated with the fast smooth particle-mesh Ewald electrostatics method with an initial short-range cutoff of 1.0 nm.<sup>29</sup> The leapfrog algorithm was used to run 20 independent 100 ns MD simulations with a 2 fs time step. Snapshots from the production runs were taken every 100 ps for 3D coordinates, while snapshots were taken every 100 fs for energy data.

**Absolute Binding Enthalpy Calculations.** The binding enthalpy ( $\Delta H$ ) is calculated by eq 1, where  $\langle E \rangle_{\text{complex}}$ ,  $\langle E \rangle_{\text{solvent}}$ ,  $\langle E \rangle_{\text{receptor}}$ , and  $\langle E \rangle_{\text{ligand}}$  are the averaged potential energies of the system from four separate simulations. In this method, the number of atoms between the bound and unbound states of the complex should exactly balance. Note that the pressure–volume contribution for the binding enthalpy is negligible.<sup>15</sup>

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

where  $\langle \dots \rangle$  is the time-averaged value obtained from multiple independent MD trajectories. The use of multiple simulations can provide a broader sampling of conformational space than a single long simulation. Thus, the trajectories allow the determination of mean values distributed across the potential energy surface.

The mean energy for each trajectory is estimated by averaging over the time period corresponding to the  $N$  number of snapshots, while the expected energy value for the set of trajectories,  $\langle E \rangle$ , is determined by calculating the means of the individual trajectories. For  $K$  number of trajectories, the ensemble mean is

$$\langle E \rangle = \frac{1}{KN} \sum_{t=1}^K \sum_{n=1}^N E_{t,n} \quad (2)$$

where  $E_{t,n}$  is the energy value of the  $n$ 'th snapshot of the  $t$ 'th trajectory.  $K$  is the set of 20 trajectories for complex and apo-receptor simulations, while 10 trajectories for peptide in water and water-only simulations, and  $N$  (1,000,000) is defined as the number of snapshots taken from a 100 ns trajectory. Convergence of the potential energy was evaluated by plotting the cumulative average of the simulation data while the uncertainty [standard error of the mean (SEM)] in the energy values obtained from  $K$  independent trajectories was estimated using reblocking analysis,<sup>30</sup> as implemented in the pyblock tool (<https://pyblock.readthedocs.io>). The SEM of  $\Delta H$  is given by

$$\sigma_{\Delta H} = \sqrt{\sum_{i=1}^4 \sigma_i^2} \quad (3)$$

where  $\sigma_i$  obtained from the reblocking analysis is the SEM of the overall potential energy of each system (complex, water, receptor, and peptide) because  $\Delta H$  is an additive combination of mean energies. The main obstacle in interpreting the reblocking analysis is the choice of an optimal block size, giving the true SEM. A clear plateau in the reblock plot of the SEM versus block size ideally represents the SEM. However, it is neither objective nor efficient to individually analyze the reblock plot for each calculation to make decisions. It is possible to select an optimal block size via the use of eq 4

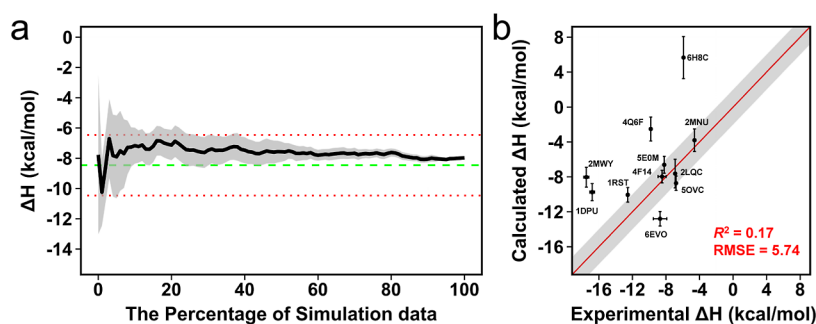
$$B_{\text{opt}} = \sqrt[3]{2nn_{\text{corr}}^2} \quad (4)$$

where  $n$  is the number of data points and  $n_{\text{corr}}$  is the correlation length described elsewhere.<sup>31</sup> However, to err on the side of caution, in this study, we took the maximum SEM value from each reblocking analysis.

## RESULTS

**Obtaining a Nonredundant Data Set for Protein–Protein Enthalpy Calculations.** We extracted 76 PDB





**Figure 2.** (a) Convergence pattern of the calculated  $\Delta H$  by using 60 trajectories for the protein–peptide complex in PDB:4F14, giving good agreement with the experimental  $\Delta H$  value. The green dashed line is the experimental  $\Delta H$  and the red dotted lines indicate the 2 kcal/mol error limit. (b) Comparison of calculated binding enthalpies from experimental values. Error bars were drawn by using the highest SEM values from the reblocking analysis. The line of equivalence is shown in red, and the black shadow indicates the 2 kcal/mol error limit.

entries where the larger of the protein–peptides in the complex was smaller than 150 residues and where binding enthalpy data were available in the literature (Figure 1a and Supporting Information). The data set contains 25 NMR and 51 X-ray crystal structures with a resolution ranging from 1.11 to 2.6 Å. Receptor proteins come from 49 different genes of 12 organisms including two virus genomes. The receptor size ranges from 54 to 148 residues, while the peptides range from 4 to 49 residues (Figure 1b).

After obtaining PDB entries having binding enthalpies, we performed clustering based on the sequence and structure of the receptors. First, we ran Clustal Omega, a multiple-sequence alignment program, for receptor sequences and obtained the phylogenetic tree of the receptor sequences. The tree-based clustering gave us 11 different clusters with cluster sizes ranging from 1 to 11 (Figure 1a). 6B27 was the only member of cluster 1, however, it is a tandem repeat of SH3 domains, which mostly exist in cluster 9.

Indeed, closer inspection revealed that some domains (e.g., SH3 and EF-Hand) were found in different clusters derived from the sequence-based clustering process, and thus, we decided to use a structure-based clustering via the MaxCluster tool (<http://www.sbg.bio.ic.ac.uk/maxcluster/>) using the local backbone similarity matching method. Structure-based clustering gave us eight different clusters, with cluster sizes ranging from 4 to 12 members. We also had an additional cluster for all other PDBs having unrelated structures that are not similar (cluster ID 1—see Table 1).

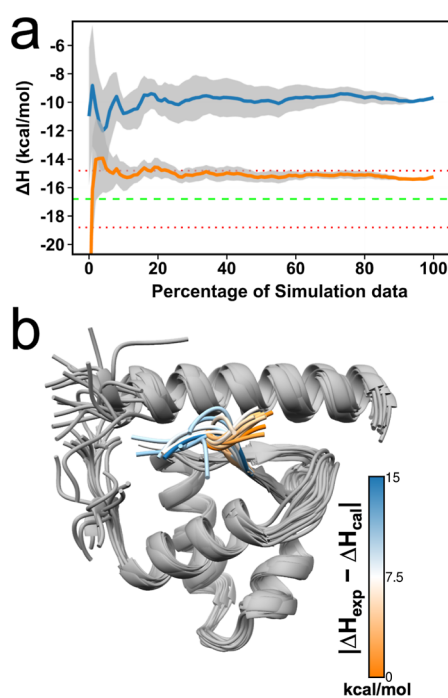
Finally, we picked representative PDBs with the smallest number of total residues and complete ITC that were all performed at the same temperature (25 °C), giving us 11 different complexes in the data set (Table 1), which represent the clusters obtained from both methods. Cluster 1 obtained by structure-based clustering is equivalent to the three different clusters via the sequence-based method.

**Initial Results for Absolute Binding Enthalpy Calculations.** To the best of our knowledge, there has been no previous attempt to compute binding enthalpies for protein–protein systems using MD simulations. Thus, in the first instance, we wanted to evaluate just how well a standard approach would perform. Here, the absolute binding enthalpy for 11 protein–peptide complexes was estimated by the direct method (eq 1) using ensemble averaging (eq 2). In obtaining binding enthalpies, sufficient sampling to cover all conformational space is crucial to obtaining a converged pattern of potential energy. In previous studies, Roy et al. performed multiple independent simulations,<sup>14</sup> while Li and Gilson ran

long simulations to get sufficient sampling for the relative binding enthalpy calculation of a protein–ligand system.<sup>11</sup> Here, we opted to perform multiple independent simulations to maximize sampling as we previously obtained good results for bromodomain–ligand complexes using this strategy.<sup>43</sup> Thus, 60 trajectories, totaling 6  $\mu$ s of simulation data, were generated from MD simulations of the bound and unbound states for each protein–peptide complex. We produced one million snapshots per simulation for the potential energy, which means 60 million data points were used for each  $\Delta H$  calculation. Before performing enthalpy calculations, we first checked the overall convergence pattern for each complex. The cumulative convergence in increasing blocks of all simulation data shows a reasonably well-converged pattern of the calculated  $\Delta H$ , as exemplified here for 4F14 (Figure 2a). This pattern was almost the same with most of the protein–peptide complexes (Figure S1). However, 2LQC and 6H8C exhibited a rather uneven profile, indicating that these complexes require more simulation data for convergence. Nevertheless, 2LQC gave a highly accurate result, but 6H8C gave the worst result (Figure 2b). Overall, the correlation with the experiment was poor, with an  $R^2 = 0.17$ , and the accuracy of the calculations was weak with an average of  $RMSE = 5.74$  kcal/mol. The results for only 5 PDB complexes are within the 2 kcal/mol error limit.

**Tail Conformations Affect the Accuracy of Predictions.** Although the initial results were quite poor in terms of correlation with the experiment, about half of the predictions were close to the experimental value, and this encouraged us to identify possible causes of error in the inaccurate results. First, 1DPU, where the receptor is the 32 kDa subunit of replication protein A (RPA32) gave an inaccurate  $\Delta H$  with a difference of 7.12 kcal/mol between calculated and experimental values (Figure 3a, blue line). RPA32 interacts with DNA damage response proteins including SMARCA1, Tipin, UNG2, and XPA.<sup>32</sup> In the 1DPU structure, RPA32<sub>(201–272)</sub> is in complex with a 73–88 residue long peptide from UNG2.

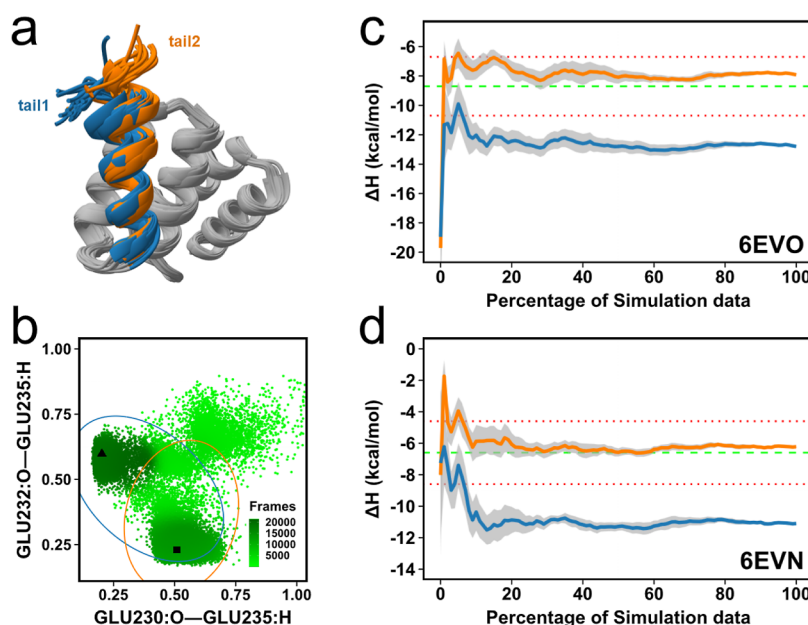
To identify potential problems with the  $\Delta H$  calculation, we first performed clustering on the simulations of the complex using the single-linkage method with a 1 Å cutoff. We extracted the most populated conformations from each simulation. After superposition, the conformations were inspected and we observed highly flexible termini. Coloring conformations by the difference between calculated and experimental  $\Delta H$  (Figure 3b), revealed two distinct conformations for the C-terminal, highly correlative with the experimental  $\Delta H$  (Figure 3b). In one conformation (which we term 1DPU<sub>tail2</sub>), good



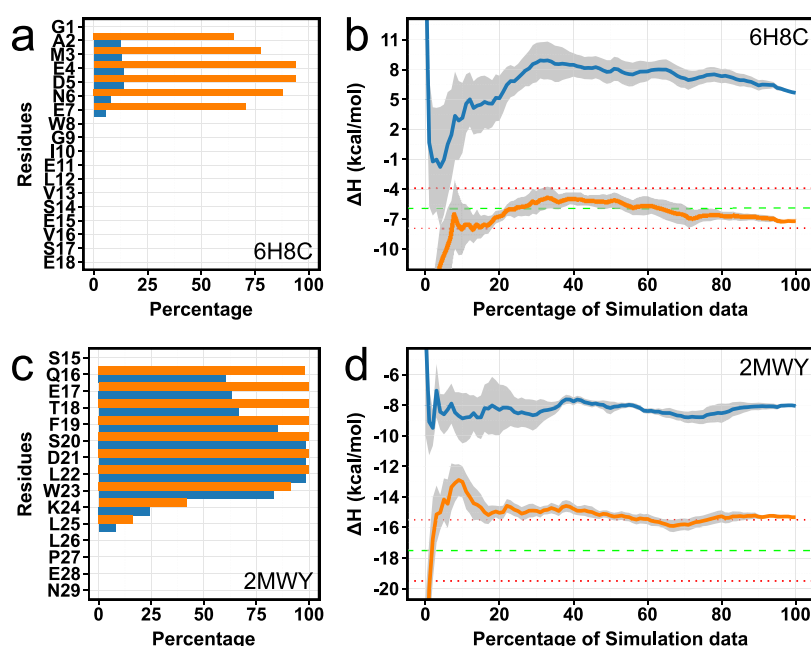
**Figure 3.** (a) Convergence pattern of the calculated  $\Delta H$  by using 60 trajectories for two simulation setups for the RPA32–UNG2 complex. The blue line shows calculated  $\Delta H$  for the first simulation set, while the orange line shows for the second simulation set. The green dashed line is the experimental  $\Delta H$  while the red dotted lines indicate the 2 kcal/mol error limit. (b) Most populated conformations from each simulation for the 1DPU complex. The peptide is the helix at the top of the diagram. The blue–orange color scale represents the absolute difference between calculated and experimental  $\Delta H$  in kcal/mol.

agreement is found with the experiment, and here, the terminal residue E270 interacts with K265 and K78 of UNG2. On the other hand, in the other conformation ( $1\text{DPU}_{\text{tail1}}$ ) (Figure S2a,b), E270 interacts with K217 and R88 of UNG2. This apparent correlation between conformations and binding enthalpies led us to perform additional simulations to increase the sampling and check the stability of the interaction between the terminal residue and K265. We then performed further simulations with 20 replicates using the  $1\text{DPU}_{\text{tail2}}$  conformation as the starting structure. We analyzed the distance between the C-terminal C atom and the side-chain N atom of residue K265 to check the stability of the interaction (Figure S2c). The second set of simulations revealed that the  $1\text{DPU}_{\text{tail2}}$  conformation is highly stable across 20 independent simulations. The most striking result to emerge from the binding enthalpy calculation is that the second set of simulations provided a noticeably more accurate  $\Delta H$  (Figure 3a, orange line) than did the first simulation set. Thus, the C-terminal conformation of the receptor is not only flexible with two distinct conformations that interact with the peptide in distinct ways but also that one of these conformations is highly correlated with an accurate prediction of  $\Delta H$ .

To further corroborate this finding, we considered a modified system,  $\text{UNG2}^{\text{RNK}/\text{AAA}}$ , in which the peptide has three alanine mutations. ITC results have also been reported for this mutation.<sup>32</sup> Importantly, this complex contains the K78A mutation, thus disrupting the interaction between E270 and K78 observed above. Nonetheless, the tail was quite stable in the  $1\text{DPU}_{\text{tail2}}$  conformation, as shown in Figure S2b despite the mutation (Figure S2c). To our surprise, the  $\text{UNG2}^{\text{RNK}/\text{AAA}}$  peptide also provided an accurate  $\Delta H$  with the  $1\text{DPU}_{\text{tail2}}$  conformation (Figure S2d)—i.e., despite the removal of the key salt-bridge.



**Figure 4.** (a) Most populated conformations from each simulation for 40 apo-receptor simulations of two setups. (b) Convergence pattern of the calculated  $\Delta H$  for 6EVO (c) distance (nm) distribution of two hydrogen bonds from apo-receptor simulations with two different tail setups. Elliptic circles show a 75% probability distribution for  $\text{tail1}$  (blue) and  $\text{tail2}$  (orange) simulations. The triangle shows hydrogen bond distances in the 6EVO crystal structure, while the square shows distances in 6EVM. (d) Convergence pattern of the calculated  $\Delta H$  for 6EVN. For (b,d), the blue line shows calculated  $\Delta H$  for the first simulation set ( $\text{tail1}$ ), while the orange line shows for the second simulation set ( $\text{tail2}$ ). The green dashed line is the experimental  $\Delta H$ , while the red dotted lines indicate the 2 kcal/mol error limit.



**Figure 5.** (a,c)  $\alpha$ -Helix percentage per residue of the peptide. (b,d) Convergence pattern of the calculated  $\Delta H$  (kcal/mol). The blue line shows calculated  $\Delta H$  for the first simulation set, while the orange line shows for the second simulation set with helix formation. The green dashed line is the experimental  $\Delta H$ , while the red dotted lines indicate the 2 kcal/mol error limit.

In all of these calculations so far, the apo-receptor simulations (see eq 1) were initiated from the  $1\text{DPU}_{\text{tail1}}$  conformation and as outlined above, good agreement with the experiment can be obtained when the complex simulations are initiated from the  $1\text{DPU}_{\text{tail2}}$  conformation (Figure S2). We explored this relationship further and performed additional simulations using only  $1\text{DPU}_{\text{tail2}}$  conformations for the apo-receptor component of the simulations to examine the effects on the  $\Delta H$  prediction. The result was always worse when the receptor was initiated from the  $1\text{DPU}_{\text{tail2}}$  conformation in the apo-receptor simulations (Figure S2). Taken together, these findings suggest that the predominant conformational state of the apo-receptor differs from the preferred conformational state for the complex, at least in terms of the enthalpic component.

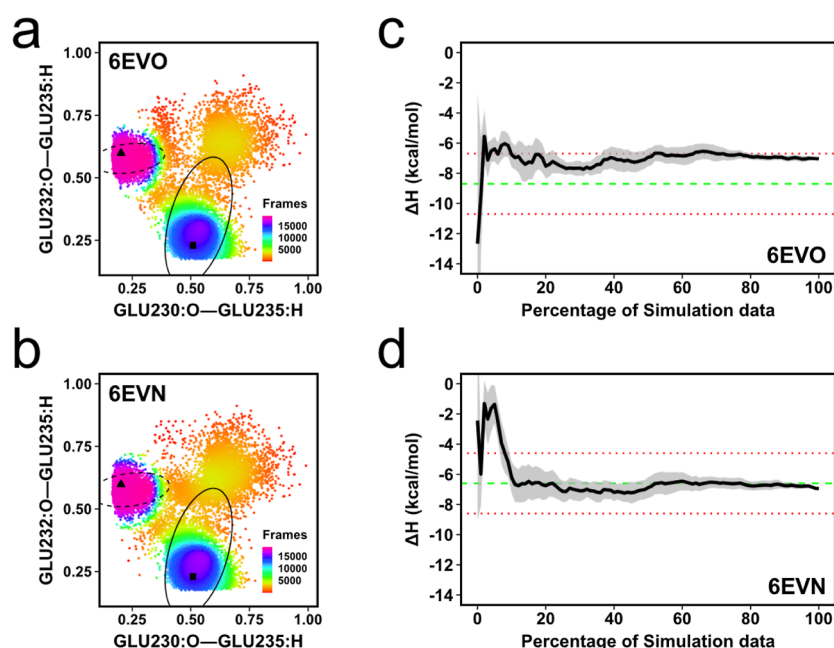
The 6EVO complex also exists in two distinct tail conformations at the C-terminus. 6EVO contains the peptide-substrate-binding (PSB) domain of human type II collagen prolyl-4-hydroxylase (C-P4H-II) complexed with a proline-rich procollagen peptide (PPGPRGPPG).<sup>41</sup> We found two different tail conformations at the C-terminal in both complex and apo-receptor simulations and also in two different crystal structures: 6EVO (**tail1**) and 6EVM (**tail2**) (Figures 4a and S3a). Like 1DPU (Figure S2), the tail contains negatively charged residues and these residues interact with positively charged residues around the tail (Figure S3b,c). In our initial calculations with 6EVO, the difference between the calculated and experimental enthalpy was 4.10 kcal/mol (Figure 4b, blue line). In the initial simulations, the tail mostly stayed in the **tail1** conformation. However, it was not clear how the **tail2** conformation would affect the enthalpy predictions.

Thus, we run an additional 20 independent simulations using the **tail2** conformation for both the complex and apo-receptor. Interestingly, the tail mostly remained stable with whatever conformation was used as the starting conformation (Figures 4c and S3f,g). We used two important backbone hydrogen bonds as a proxy for the tail's conformational state.

E235:H makes a hydrogen bond with E230:O in 6EVO, while it makes a different hydrogen bond with E232:O in 6EVM. This analysis revealed that the **tail1** conformation was more stable in the complex form than the apo-receptor (Figures 4b and S3h,i). We then checked the potential energies for each simulation setup. Complex simulations for protein–peptide did not give a significant difference with 1.41 kcal/mol for both simulation setups; on the other hand, apo-receptor simulations had very different potential energy profiles with 4.88 kcal/mol difference between **tail1** and **tail2** setups (Figure S3d,e). As a result, simulations of apo-receptor with **tail2** conformations improved the binding enthalpy prediction to a difference of 0.83 kcal/mol between calculation and experiment (Figure 4c, orange line).

To further confirm this finding, we performed an additional  $\Delta H$  calculation for 6EVN having the same receptor complex with the PPGAGPPG peptide. This peptide has an RSA mutation and  $-6.6 \pm 0.1$  kcal/mol binding enthalpy, as determined by ITC. Apo-receptor simulations with the **tail2** conformation provided more accurate  $\Delta H$  than simulations starting from the **tail1** conformation (Figure 4d). Thus, overall, the results show that the receptor should have the **tail2** conformation in the apo state, but the **tail1** conformation in the bound state in order to obtain an accurate binding enthalpy calculation for the PSB domain of human C-P4H-II.

**Accurate Binding Enthalpy Hidden by Unknown Helix Formation.** In two other outliers, we observed that a small  $\alpha$ -helix had a tendency to form at the N-termini of the peptides (Figure 5). First, 6H8C is an NMR structure of the human GABARAPL2 protein in complex with the LIR motif that is found within UBA5 (ubiquitin-like modifier activating enzyme 5).<sup>42</sup> Starting from the PDB coordinates of the complex gave us the worst initial  $\Delta H$  prediction with the highest absolute difference of 11.78 kcal/mol between the calculated and experimental values (Table S1 and Figures 1b and 5b). However, only using trajectories where the helix was present in the complex resulted in a greatly improved enthalpy



**Figure 6.** (a,b) Distance distribution of two hydrogen bonds from apo-receptor and complex simulations. Elliptic circles show a 75% probability distribution for complex (dashed) and apo-receptor simulations. The triangle shows hydrogen bond distances in the 6EVO crystal structure, while the square shows distances in 6EVM. (c,d) Convergence pattern of the calculated  $\Delta H$  for 6EVO and 6EVN in experimental buffer conditions. The green dashed line is the experimental  $\Delta H$ , while the red dotted lines indicate the 2 kcal/mol error limit.

prediction (Figure S4a). This finding prompted us to conduct further simulations to increase the sampling of the helical conformation and test the stability of the helix. Thus, we ran additional 20 independent simulations using the complex with the helix present from the start. The secondary structure was computed for each simulation, and analysis revealed that the helix was relatively stable across almost all simulations (Figure 5a). Consequently, the second simulation set with the helix present dramatically improved the  $\Delta H$  prediction (Figure 5b). However, we should be cautious in the interpretation here because the peptide used for the ITC measurements excluded the first 4 residues used in the NMR experiments (and also included 3 extra residues at the C-terminus)<sup>42</sup> (Figure S4b). Thus, it is difficult to make a true direct comparison here.

The other complex where helix formation was observed was the MDMX-p53 complex from 2MWY. In this complex, the N-terminal domain of MDMX<sub>(23–111)</sub> binds the first trans-activation domain (TAD1) of p53<sub>(15–29)</sub>.<sup>44</sup> When we looked at the experimental structures (Figure S5), the TAD1 peptide has a small helical structure, but an increase in the amount of helix of the TAD1 region was noticeable in our simulations. Interestingly, this same helix formation at the same region is also found in the AlphaFold model (<https://alphafold.ebi.ac.uk/entry/P04637>). This helix increase was clearly visible in >50% of the simulations (Figure 5c). We identified 10 simulations with the highest helix formation and examined their contribution to the binding enthalpy. While the simulations with the highest amount of helix improved the agreement with experimentally obtained enthalpy (−9.68 kcal/mol), the simulations with the least amount made the result even worse (−6.40 kcal/mol). Overall, the  $\Delta H$  for 2MWY was estimated at  $-8.04 \pm 0.08$  kcal/mol for all simulations (Figures 2b and 5d). This suggested to us that once again, the presence of an additional helix in the peptide would improve the enthalpy prediction. Thus, we ran an additional 20 independent simulations for MDMX-p53 with the additional

helical content. The secondary structure analysis revealed that the helix was highly stable across the simulations (Figure 5c), and the second simulation set with the helix formation significantly improved the  $\Delta H$  prediction (Figure 5d).

**Metal Cation Parameters in Metalloproteins.** In this data set, we had two metalloproteins; calmodulin having two  $\text{Ca}^{2+}$  ions (PDB: 2LQC) and the PHD zinc finger domain of BAZ2A that has two  $\text{Zn}^{2+}$  ions (PDB: 4Q6F) (Figure S6a). Calmodulin makes a complex with the “NSCaTE” peptide from the N-terminal cytoplasmic domain of the L-type voltage-gated calcium channel  $\alpha 1C$  subunit.<sup>34</sup> This complex gave us an accurate  $\Delta H$  with the absolute difference of 0.72 kcal/mol between calculated and experimental values (Figure S1 and Table S1) when we used the default parameters for  $\text{Ca}^{2+}$  within AMBER ff14SB. Given the known problems with calcium parametrization,<sup>45</sup> this was a surprising result. On the other hand, the BAZ2A PHD zinc finger in complex with an unmodified H3K4 histone peptide in the 4Q6F structure provided an incorrect  $\Delta H$  (Table S1 and Figure S6c) when using the default parameters for  $\text{Zn}^{2+}$  in AMBER ff14SB. Therefore, we investigated other parameters for  $\text{Zn}^{2+}$  and we ran an additional 40 simulations using the zinc AMBER force field (ZAFF)<sup>46</sup> for the complex and apo-receptor simulations. The new setup improved the  $\Delta H$  prediction with a 5.16 kcal/mol improvement for 4Q6F (Figure S6c) and placed just around the 2 kcal/mol error level.

**Considering Experimental Conditions.** We have performed all simulations with pure water to reduce complexity since we recently obtained good results for bromodomain–ligand complexes in pure water.<sup>43</sup> The ITC experiments were mostly conducted under diverse buffer conditions (Table S2). Nevertheless, we wanted to see how accurate this approach is given an explicit buffer composition typical of that used in ITC experiments. We picked the PSB domain of human C-P4H-II (6EVN and 6EVO) since it required two different sampling approaches for complex and apo-receptor simulations. Thus,



we set up additional simulations having 20 mM TRIS (tris(hydroxymethyl)aminomethane), 50 mM NaCl, and 50 mM glycine for the receptor–peptide complex and the apo-receptor. We used the **tail1** conformation in 6EVO for complex simulations as a starting conformation while the **tail2** conformation found in 6EVM was used for the apo-receptor simulations (Figure S3a) since this setup provided the most accurate results (Figure 4c,d). The same hydrogen bonds (Figure 4b) were used as a proxy to check the stability of the tail conformation (Figure 6a,b). The simulations show that an accurate  $\Delta H$  prediction can be obtained whether simulating in pure water or a complete buffer condition for the PSB domain of human C-P4H-II (Figure 6c,d). Perhaps surprisingly, 6EVN gives a more accurate  $\Delta H$  prediction in pure water than in buffer conditions (Figures 4c and 6d).

**Decompositions of Binding Enthalpies into Physical Components.** Enthalpies obtained in this way from MD simulations can be further analyzed in terms of the energetic subcomponents.<sup>15,47</sup> We investigated the determinants of the binding enthalpies from simulations, giving accurate results. Table 2 includes subcomponents of binding enthalpies for 11

Table 2. Subcomponents of the Binding Enthalpies<sup>a</sup>

PDB	Val	Coul	LJ
1DPU	$-1.80 \pm 0.01$	$3.83 \pm 0.14$	$-17.27 \pm 0.09$
1RST	$0.01 \pm 0.01$	$-1.45 \pm 0.13$	$-8.62 \pm 0.08$
2LQC	$2.92 \pm 0.01$	$-0.66 \pm 0.14$	$-9.89 \pm 0.09$
2MNU	$-2.13 \pm 0.01$	$7.33 \pm 0.16$	$-9.00 \pm 0.10$
2MWY	$3.95 \pm 0.01$	$-2.39 \pm 0.11$	$-16.89 \pm 0.07$
4F14	$-0.49 \pm 0.01$	$3.69 \pm 0.14$	$-11.18 \pm 0.09$
4Q6F	$-5.86 \pm 0.01$	$12.24 \pm 0.12$	$-14.05 \pm 0.07$
5E0M	$-3.18 \pm 0.01$	$10.75 \pm 0.14$	$-14.20 \pm 0.09$
SOVC	$-1.16 \pm 0.01$	$8.90 \pm 0.14$	$-16.46 \pm 0.08$
6EVO	$2.34 \pm 0.01$	$3.59 \pm 0.12$	$-13.90 \pm 0.07$
6H8C	$3.09 \pm 0.01$	$8.83 \pm 0.14$	$-19.15 \pm 0.09$

<sup>a</sup>Coul: Coulombic electrostatic contribution. Val: contribution from changes in bond-stretch, angle-bend, and dihedral terms. LJ: Lennard-Jones contribution. All values are in kcal/mol.

protein–peptide systems into three physical determinants: changes in the van der Waals forces (LJ); changes in electrostatic interactions (Coul); changes in valence terms, which include bond-stretch, bond-angle, and dihedral terms (Val). The Lennard-Jones term was consistently negative for all complexes and makes a major contribution to the binding enthalpy. Unlike the Lennard-Jones contribution, the electrostatic contribution was mostly unfavorable, while valence terms varied in sign and magnitude. Interestingly, despite unfavorable contributions, there were moderate correlations between electrostatic contributions and experimental enthalpy values with 0.52 Pearson's *r*. Hydrogen bonds (in the force field model employed here) are primarily electrostatic attractions between molecules. Thus, we checked for any correlations between enthalpy values and different aspects of the hydrogen bonds. We found modest correlations for the average total number of hydrogen bonds between the receptor and peptide with 0.48 Pearson's *r* for experimental enthalpy values (Table S4). Perhaps surprisingly, there was no relationship between experimental enthalpy values and absolute hydrogen bond difference (i.e., all H-bonds in the system) between bound and unbound states ( $\Delta H$  bond), but  $\Delta H$  bond was highly correlated with Val (*r* = 0.73).

## DISCUSSION

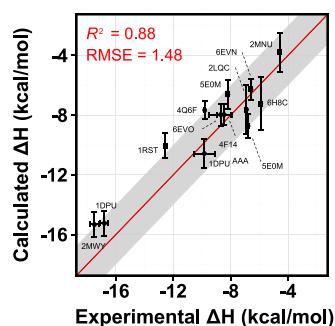
The combination of MD simulation, 3D structures, and experimental calorimetric data can provide a greater understanding of the driving forces for protein–protein binding.<sup>7,48,49</sup> Ideally, one would like to be in a position to predict the absolute thermodynamic properties ( $\Delta G$ ,  $\Delta H$ , and  $T\Delta S$ ) for any protein with its interaction, be that a small molecule or another protein, RNA, or DNA. Although much progress has been recently made on absolute  $\Delta G$  predictions,<sup>49</sup> it remains extremely challenging to do as well for  $\Delta H$  calculations (though see ref 43 for some recent progress in this area) and progress has been modest. It is very challenging to do this in a prospective manner.

One possible reason for progress has been modest is the assumption that the amount of sampling required to ensure a converged result is simply too large and not currently obtainable. That is to some extent confirmed by this study whereby the initial results showed very poor agreement with the experimentally derived enthalpy values and indeed in some instances initiating calculations from one state does not move to a second conformational state, suggesting a significant energy barrier between the two states. One might well ask, what is the point of these calculations then?

There are several aspects to the answer. First, the lack of agreement for certain complexes suggests that their dynamical behavior is not correctly being captured or rather the simulation may be trapped in a conformational state that is not compatible with the enthalpy of that complex. Thus, if one is interested in the dynamics of a complex, the enthalpy prediction, if not in good agreement with the experiment, can be used as an indicator that there are likely to be other conformations that are not being sampled in the simulation. In this scenario, the use of any number of enhanced sampling techniques (such as metadynamics) could be usefully employed to improve the degree of conformational sampling. Thus, one can imagine a scenario where the ITC calculation is first performed, and if there is poor agreement then one would switch to metadynamics to improve the sampling and discover previously hidden but important states, but these are not necessarily a global panacea to the sampling problem. Many enhanced sampling methods require several order parameters, or collective variables (CVs).<sup>50</sup> However, if important CVs are missing, such methods may suffer from the “hidden energy barrier” problem and converge slowly.<sup>51</sup> If no experimental or theoretical information is available, then one can use enhanced sampling methods without predefined reaction coordinates or CVs for binding enthalpy calculations. Such methods must accurately preserve both the thermodynamic and the kinetic properties of the system. However, their ability to reproduce accurate thermodynamic information has been questioned,<sup>52</sup> and thus, more systematic investigation is still needed in this regard.

Second, the work shows that if the correct conformations are being sampled well, then in fact the estimated enthalpies can be accurate. Although an artificial exercise, taking the best predictions from the simulations gives a very good agreement with the experiment (Figure 7). This shows that a commonly used force field (AMBER ff14SB) at least has the potential to predict enthalpies to a reasonable level. Of course, one should remember that ultimately the force field should predict  $\Delta G$  and  $\Delta H$  (and by inference  $T\Delta S$ ) correctly, and good





**Figure 7.** Comparison of calculated binding enthalpies from experimental values for 13 different protein–peptide systems (the 11 benchmarks plus 2 mutations). The line of equivalence is shown in red, and the gray shadow indicates the 2 kcal/mol error limit. Error bars were drawn by using the highest SEM values from the reblocking analysis.

performance in one aspect does not necessarily mean so for the other.

Accurate predictions (even posthoc corrected ones as summarized in Figure 7) may also be useful in terms of interpreting calorimetric calculations with atomistic details because calorimetric data often apparently contradict accepted scientific dogmas.<sup>11,53,54</sup> For example, it was often thought that the enthalpic contribution of the binding is dominated by polar interactions in protein–ligand binding;<sup>54,55</sup> however, the calculations here emphasize the importance of the van der Waals interactions for protein–protein systems (Table 2). Indeed the role of van der Waals interactions in stabilizing the protein–protein complexes has been the focus of recent studies.<sup>56</sup> Moreover, a gain in the number of hydrogen bonds upon the binding was usually thought to be related to more favorable enthalpies, but we found no such correlation between overall change in  $\Delta H$  bond and the binding enthalpy at least for these systems. These observations are intriguing and require further investigation, but a cursory inspection suggests that the peptide binding increases the number of vdW interactions.

Any calculations related to the free energy landscape of the binding between proteins will most likely be computationally very expensive since proteins can adopt many alternative conformations.<sup>57</sup> Thus, insufficient conformational sampling is expected to hamper the accurate calculation of calorimetric components. Moreover, the free energy is a combination of many terms, which also increases the risks of errors and thus decreases precision and accuracy. Yet, a current meta-analysis of 853 studies from 34 different research groups showed that free energy methods were close to chemical accuracy (error = 1.58 kcal/mol).<sup>58</sup> On the other hand, binding enthalpy–entropy calculations have generally not been given much consideration with MD simulations, and there are limited studies, especially for complex systems.<sup>10,11,14,15,59</sup> It has been shown that an accurate binding enthalpy calculation is possible with a simplistic model, host–guest systems.<sup>47,60,61</sup> Notable efforts have also been made for complex systems to obtain relative binding enthalpies.<sup>11,14</sup> All of these studies have demonstrated that binding enthalpy calculations are sensitive to the choice of force field and sufficient sampling.

Previously, we reported a comprehensive assessment of force fields and water models using a set of 25 cucurbit[7]uril–guest (CB7) pairs and showed that the TIP3P water model combination with general AMBER force field provided good

agreement with the experimental enthalpy values.<sup>61</sup> From the sampling perspective, it is unclear how much simulation data is needed to obtain accurate binding enthalpies of complex systems. Roy et al. ran 40 independent 10 ns simulations to get sufficient sampling with 1 ps writing frequency for energy data.<sup>14</sup> Furthermore, Li and Gilson reached over 250  $\mu$ s of simulation time by seeding every 200 ns and recording energies every 2 ps.<sup>11</sup> So far, these were the only examples using the direct method for complex systems from the literature. Here, we performed 20 completely independent repeats of 100 ns simulations and recorded energy data every 100 fs for each system, and as discussed above, inaccurate enthalpies came mostly from incorrect global sampling exemplified by conformational states not seen experimentally.

It is well-known that receptors can adopt alternative conformations upon peptide binding,<sup>62</sup> as reported here. Here, we observed the C-terminal loop region having two major conformations in RPA32 (Figure 3b) and the PSB domain of human C-P4H-II proteins (Figure 4a). In the NMR structure (1DPU), RPA32 exhibits conformational heterogeneity in the C-terminal, but in the simulations, two distinct conformations were populated (Figure S2c), and RPA32 provided accurate enthalpies for UNG2<sup>WT</sup> and UNG2<sup>RNK/AAA</sup> peptides only when it had two different conformations in bound and unbound states. Likewise, the PSB domain of human C-P4H-II also provided accurate enthalpies for two different peptides (6EVO and 6EVN) in pure water and experimental buffer conditions (20 mM Tris, 50 mM NaCl, and 50 mM glycine) when having two different conformations in bound and unbound states. Conversely, the PSB domain has these conformations in X-ray crystal structures (Figure S3a): **tail1** conformation in 6EVN, 6EVO, and 6EVP, **tail2** conformation in 6EVL (apo) and 6EVM, which contains a weak binder (P)<sub>9</sub> peptide. These suggest that the C-terminal of the PSB domain tends to have the **tail1** conformation with good binders but the **tail2** conformation with weak binders or when in the apo form. This observation of conformational differences is mirrored in our simulations and enthalpy calculations. Interestingly, there are charged residues around the C-terminal in both cases of the RPA32 (Figure S2a,b) and PSB domains of human C-P4H-II (Figure S3b,c), and it is possible that they drive the conformational change via electrostatic interactions. We also note that, paradoxically, the choice of small systems to aid the sampling problem could also have meant that termini near the interface could be more prone to setup dependencies.

In contrast to large folded domains, peptides are usually in the form of unstructured/disordered molecules in the unbound state.<sup>63</sup> They sometimes undergo folding upon binding, which results in better structural adaptation with a much lower level of conformational freedom. For example, the Bak peptide gets an  $\alpha$ -helix when complexed to Bcl-x<sub>L</sub>, although it is an unstructured random coil in solution.<sup>64</sup> Importantly, free p53 TAD, which we used in our benchmark as a peptide (2MWY), is intrinsically disordered. Furthermore, NMR studies showed that the p53 TAD did not tend to create a helical structure in the free form.<sup>65</sup> However, the p53 TAD gets more helical upon binding to MDM2, MDMX,<sup>66,67</sup> and the nuclear receptor coactivator binding domain of CREB binding protein.<sup>68</sup> These structural stabilizations are concomitant with a large enthalpic gain.<sup>69</sup> In the same way, we observed large enthalpic gains in two different protein–peptide systems upon helix formation in MDM2–p53 (2MWY) and GABARAPL2–UBA5 (6H8C;

Figure 5). The p53-MDM2/MDMX complexes have been well-characterized as a target for cancer therapy.<sup>44,67,70</sup> The UBA5 peptide (<sup>337</sup>EDNEWGIELVSEVSE<sup>351</sup>) is an unstructured region in crystal structures and the AlphaFold model,<sup>71</sup> but it formed a small helix at the N-terminal in our simulations. Consequently, complex simulations having more helix formation in the peptides provided more accurate results for the binding enthalpies of the MDM2-p53 and GABARAPL2-UBA5 complexes. Interestingly, there is an artificial cloning artifact (GAM) in the ITC construct of the UBA5 peptide (Figure S4b). Thus, this artificial extension in the N-terminal possibly drives helix formation in the UBA5 peptide upon binding. On the other hand, the p53 TAD was already a helix in the PDB structure, but it was shorter than we observed in the simulations and the AlphaFold model (Figure S5). Similarly, more helix formation ensured more accurate binding enthalpy for the MDM2-p53 complex as well. At this point, it is also worth remembering that the underlying tendency of the force field to adopt helical conformations is an aspect that has received attention in recent years<sup>72</sup> and that force field development is an ongoing exercise.<sup>73</sup>

Considering complete experimental conditions in simulations might increase the complications during setup and computational time. The choice of the buffer can affect the thermodynamic signatures if there are proton movements between solute and buffer during binding.<sup>74</sup> Moreover, it has been suggested that buffer salts might increase the hydration of the ligands on the binding of agonists and antagonists at the histamine H3 receptor.<sup>75</sup> Nonetheless, the buffer conditions did not appear to dramatically affect the enthalpy calculations in these calculations since we got highly accurate calculations with simulations in pure water (Figure 7). In addition, we reported accurate binding enthalpies for bromodomain–ligand complexes in pure water after considering ZA-loop dynamics.<sup>43</sup> On the contrary, Gao et al. reported the strong sensitivity of the computed binding enthalpies to the concentration of NaCl (0–500 mM) for the CB7-B2 system.<sup>47</sup> As was discussed,<sup>74</sup> this may be attributed to cation binding at the portals CB7 that are displaced upon binding of the guest. Although strong sensitivities to salt concentrations were observed in calculations with different water models, there was no significant effect of the experimental buffer condition according to our observations. Here, we reported similar results for binding enthalpies of the PSB domain of human C-P4H-II with two different peptides (6EVO and 6EVN) in pure water and also experimental buffer conditions (20 mM TRIS, 50 mM NaCl, and 50 mM glycine). Although we observe no real effect of buffer in this study, this is an area that requires a more systematic and complete approach that is beyond the scope of this paper. It should not be forgotten that differences between the simulations and ITC conditions can also contribute significantly to the overall error.

More than 50% of proteins include metal ions, making them prevalent in biological systems.<sup>76</sup> Metalloproteins can perform vital tasks like scavenging free radicals and catalyzing respiration reactions in the cell.<sup>77</sup> Consequently, two metalloproteins were included in our benchmark; calmodulin having two Ca<sup>2+</sup> (PDB: 2LQC) and the PHD zinc finger domain of BAZ2A having two Zn<sup>2+</sup> (PDB: 4Q6F). 2LQC gave us an accurate  $\Delta H$ , while 4Q6F provided an incorrect  $\Delta H$  when we used the default parameters in AMBER ff14SB. Here, the nonbonded model, the simplest model among nonpolarizable models, was used for metals.<sup>78</sup> It consists only of the

electrostatic and van der Waals terms. However, it was not reproducible for the hydration-free energy and ion–oxygen distance of the first solvation shell for the zinc ion when using the nonbonded model.<sup>22,79</sup> Later, Li and Merz proposed bonded models for different zinc-coordinated systems (ZAFF), which reproduced the experimental hydration-free energy, coordination number, and ion–oxygen distance simultaneously.<sup>22,23,80</sup> Obtaining better results after using ZAFF pointed out that further improvement in force field accuracy for Zn<sup>2+</sup> was needed, as might be expected.

## CONCLUSIONS

In conclusion, we have shown that the predicted enthalpy is a straightforward way to assess whether a protein–peptide system is likely to sample the correct ensemble of states. Furthermore, posthoc explicit consideration of this in further simulations suggests that at least the AMBER force field used here is capable of producing enthalpy values in agreement with the experiment. However, we are still some way off from being able to perform these calculations prospectively, and to that end our assembled data set of diverse protein–peptide complexes complete with binding enthalpy–entropy data could serve as a valuable benchmark for the community. Such benchmarks are extremely useful for drug discovery studies by providing opportunities for comparing algorithms.<sup>16,81</sup> The data set is available at doi: <http://10.0.20.161/zenodo.7635945>, including the PDB IDs and ITC details. It should be an invaluable resource for computational structural biologists trying to predict thermodynamic components from the structure and to test new methods for protein–protein interactions.

## ASSOCIATED CONTENT

### Data Availability Statement

Complete input and parameter files, along with ITC summary details are available at doi: <http://10.5281/zenodo.7635945>.

### Supporting Information

The Supporting Information is available free of charge at <http://pubs.acs.org/doi/10.1021/acs.jcim.3c01041>.

Breakdown of potential energies, experimental buffer conditions for each system, enthalpies for additional calculations and detailed hydrogen bond analysis; individual convergence patterns, details of specific tail interactions, detailed conformational analysis of specific systems, and pictorial representations of hydrogen-bond analysis (PDF)

## AUTHOR INFORMATION

### Corresponding Author

Philip C. Biggin – Department of Biochemistry, University of Oxford, Oxford OX1 3QU, U.K.; [orcid.org/0000-0001-5100-8836](https://orcid.org/0000-0001-5100-8836); Email: [philip.biggin@bioch.ox.ac.uk](mailto:philip.biggin@bioch.ox.ac.uk)

### Author

Süleyman Selim Çınaroğlu – Department of Biochemistry, University of Oxford, Oxford OX1 3QU, U.K.; [orcid.org/0000-0001-7120-3540](https://orcid.org/0000-0001-7120-3540)

Complete contact information is available at:

<https://pubs.acs.org/doi/10.1021/acs.jcim.3c01041>

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

Computing was supported via Advanced Research Computing (Oxford), the ARCHER UK National Supercomputing Service, and JADE (EP/P020275/1) granted via the High-End Computing Consortium for Biomolecular Simulation, (HECBioSim—<http://www.hecbiosim.ac.uk>), supported by EPSRC (EP/L000253/1). S.S.C. thanks The Ministry of National Education of The Republic of Turkey for student sponsorship.

## REFERENCES

- (1) (a) Vidal, M.; Cusick, M. E.; Barabási, A. L. Interactome networks and human disease. *Cell* **2011**, *144* (6), 986–998. (b) Fukao, Y. Protein–protein interactions in plants. *Plant Cell Physiol.* **2012**, *53* (4), 617–625. (c) Hayes, S.; Malacrida, B.; Kiely, M.; Kiely, P. A. Studying protein–protein interactions: progress, pitfalls and solutions. *Biochem. Soc. Trans.* **2016**, *44* (4), 994–1004.
- (2) (a) Kuzmanov, U.; Emili, A. Protein–protein interaction networks: probing disease mechanisms using model systems. *Genome Med.* **2013**, *5* (4), 37. (b) Rosell, M.; Fernández-Recio, J. Hot-spot analysis for drug discovery targeting protein–protein interactions. *Expert Opin. Drug Discovery* **2018**, *13* (4), 327–338.
- (3) (a) Watkins, P. Drug safety sciences and the bottleneck in drug development. *Clin. Pharmacol. Ther.* **2011**, *89* (6), 788–790. (b) Gatti, M.; De Ponti, F. Drug repurposing in the COVID-19 era: Insights from case studies showing pharmaceutical peculiarities. *Pharmaceutics* **2021**, *13* (3), 302.
- (4) (a) Chang, C.-k.; Lin, S.-M.; Satange, R.; Lin, S.-C.; Sun, S.-C.; Wu, H.-Y.; Kehn-Hall, K.; Hou, M.-H. Targeting protein–protein interaction interfaces in COVID-19 drug discovery. *Comput. Struct. Biotechnol. J.* **2021**, *19*, 2246–2255. (b) Ciccone, L.; Tonalì, N.; Nencetti, S.; Orlandini, E. Application of PROTAC strategy to TTR-A $\beta$  protein–protein interaction for the development of Alzheimer's disease drugs. *Neural Regen. Res.* **2021**, *16* (8), 1554. (c) Karunakaran, K. B.; Yanamala, N.; Boyce, G.; Becich, M. J.; Ganapathiraju, M. K. Malignant pleural mesothelioma interactome with 364 novel protein–protein interactions. *Cancers* **2021**, *13* (7), 1660. (d) Patel, S.; Das, A.; Meshram, P.; Sharma, A.; Chowdhury, A.; Jariyal, H.; Datta, A.; Sarmah, D.; Nalla, L. V.; Sahu, B.; et al. Pyruvate kinase M2 in chronic inflammations: a potpourri of crucial protein–protein interactions. *Cell Biol. Toxicol.* **2021**, *37* (5), 653–678. (e) Gu, Y.; Zhu, D. nNOS-mediated protein–protein interactions: promising targets for treating neurological and neuropsychiatric disorders. *J. Biomed. Res.* **2021**, *35* (1), 1. (f) Lasala, F.; García-Rubia, A.; Requena, C.; Galindo, I.; Cuesta-Geijo, M. A.; García-Dorival, I.; Bueno, P.; Labiod, N.; Luczkowiak, J.; Martínez, A.; et al. Identification of potential inhibitors of protein–protein interaction useful to fight against Ebola and other highly pathogenic viruses. *Antiviral Res.* **2021**, *186*, 105011.
- (5) (a) Plattner, N.; Doerr, S.; De Fabritiis, G.; Noé, F. Complete protein–protein association kinetics in atomic detail revealed by molecular dynamics simulations and Markov modelling. *Nat. Chem.* **2017**, *9* (10), 1005–1011. (b) Tse, C.; Wickstrom, L.; Kvaratskhelia, M.; Gallicchio, E.; Levy, R.; Deng, N. Exploring the free-energy landscape and thermodynamics of protein–protein association. *Biophys. J.* **2020**, *119* (6), 1226–1238.
- (6) (a) Brandsdal, B. O.; Österberg, F.; Almlöf, M.; Feierberg, I.; Luzhkov, V. B.; Åqvist, J. Free energy calculations and ligand binding. *Adv. Protein Chem.* **2003**, *66*, 123–158. (b) Pohorille, A.; Jarzynski, C.; Chipot, C. Good practices in free-energy calculations. *J. Phys. Chem. B* **2010**, *114* (32), 10235–10253. (c) Lapelosa, M.; Gallicchio, E.; Levy, R. M. Conformational transitions and convergence of absolute binding free energy calculations. *J. Chem. Theory Comput.* **2012**, *8* (1), 47–60. (d) Bhati, A. P.; Coveney, P. V. Large scale study of ligand–protein relative binding free energy calculations: Actionable predictions from statistically robust protocols. *J. Chem. Theory Comput.* **2022**, *18* (4), 2687–2702.
- (7) 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.* **2017**, *139* (2), 946–957.
- (8) Russell, R. B.; Alber, F.; Aloy, P.; Davis, F. P.; Korkin, D.; Pichaud, M.; Topf, M.; Sali, A. A structural perspective on protein–protein interactions. *Curr. Opin. Struct. Biol.* **2004**, *14* (3), 313–324.
- (9) (a) Pierce, M. M.; Raman, C.; Nall, B. T. Isothermal titration calorimetry of protein–protein interactions. *Methods* **1999**, *19* (2), 213–221. (b) Frederick, K. K.; Marlow, M. S.; Valentine, K. G.; Wand, A. J. Conformational entropy in molecular recognition by proteins. *Nature* **2007**, *448* (7151), 325–329.
- (10) Sun, Z.; Yan, Y. N.; Yang, M.; Zhang, J. Z. Interaction entropy for protein–protein binding. *J. Chem. Phys.* **2017**, *146* (12), 124124.
- (11) Li, A.; Gilson, M. K. Protein–ligand binding enthalpies from near-millisecond simulations: Analysis of a preorganization paradox. *J. Chem. Phys.* **2018**, *149* (7), 072311.
- (12) (a) Stone, J. E.; Hardy, D. J.; Ufimtsev, I. S.; Schulten, K. GPU-accelerated molecular modeling coming of age. *J. Mol. Graphics Modell.* **2010**, *29* (2), 116–125. (b) Liu, W.; Schmidt, B.; Voss, G.; Müller-Wittig, W. Accelerating molecular dynamics simulations using Graphics Processing Units with CUDA. *Comput. Phys. Commun.* **2008**, *179* (9), 634–641. (c) Bowers, K. J.; Chow, E.; Xu, H.; Dror, R. O.; Eastwood, M. P.; Gregersen, B. A.; Klepeis, J. L.; Kolossvary, I.; Moraes, M. A.; Sacerdoti, F. D. Scalable algorithms for molecular dynamics simulations on commodity clusters. In *Proceedings of the 2006 ACM/IEEE Conference on Supercomputing*, 2006; p 84.
- (13) (a) Raman, E. P.; Paul, T. J.; Hayes, R. L.; Brooks, C. L., III Automated, accurate, and scalable relative protein–ligand binding free-energy calculations using lambda dynamics. *J. Chem. Theory Comput.* **2020**, *16* (12), 7895–7914. (b) Zavitsanos, S.; Tsengenes, A.; Papadourakis, M.; Amendola, G.; Chatzigoulas, A.; Dellis, D.; Cosconati, S.; Cournia, Z. FEPprepare: A Web-Based Tool for Automating the Setup of Relative Binding Free Energy Calculations. *J. Chem. Inf. Model.* **2021**, *61* (9), 4131–4138. (c) Wang, L.; Chambers, J.; Abel, R. Protein–ligand binding free energy calculations with FEP+. *Biomolecular Simulations*; Springer, 2019; pp 201–232. (d) Chen, H.; Maia, J. D.; Radak, B. K.; Hardy, D. J.; Cai, W.; Chipot, C.; Tajkhorshid, E. Boosting free-energy perturbation calculations with GPU-Accelerated NAMD. *J. Chem. Inf. Model.* **2020**, *60* (11), 5301–5307.
- (14) Roy, A.; Hua, D. P.; Ward, J. M.; Post, C. B. Relative binding enthalpies from molecular dynamics simulations using a direct method. *J. Chem. Theory Comput.* **2014**, *10* (7), 2759–2768.
- (15) Fenley, A. T.; Henriksen, N. M.; Muddana, H. S.; Gilson, M. K. Bridging calorimetry and simulation through precise calculations of cucurbituril–guest binding enthalpies. *J. Chem. Theory Comput.* **2014**, *10* (9), 4069–4078.
- (16) (a) Liu, Z.; Li, Y.; Han, L.; Li, J.; Liu, J.; Zhao, Z.; Nie, W.; Liu, Y.; Wang, R. PDB-wide collection of binding data: current status of the PDBbind database. *Bioinformatics* **2015**, *31* (3), 405–412. (b) Wang, R.; Fang, X.; Lu, Y.; Yang, C.-Y.; Wang, S. The PDBbind database: methodologies and updates. *J. Med. Chem.* **2005**, *48* (12), 4111–4119.
- (17) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The protein data bank. *Nucleic Acids Res.* **2000**, *28* (1), 235–242.
- (18) Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T. J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Söding, J.; et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **2011**, *7* (1), 539.
- (19) (a) Ortiz, A. R.; Strauss, C. E.; Olmea, O. MAMMOTH (matching molecular models obtained from theory): an automated method for model comparison. *Protein Sci.* **2009**, *11* (11), 2606–2621. Herbert, A.; Sternberg, M. *MaxCluster: A tool for Protein Structure Comparison and Clustering*, 2008.
- (20) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25* (13), 1605–1612.



- (21) (a) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11* (8), 3696–3713. (b) Mark, P.; Nilsson, L. Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. *J. Phys. Chem. A* **2001**, *105* (43), 9954–9960.
- (22) Yu, Z.; Li, P.; Merz, K. M. Extended zinc AMBER force field (EZAFF). *J. Chem. Theory Comput.* **2018**, *14* (1), 242–254.
- (23) Peters, M. B.; Yang, Y.; Wang, B.; Fusti-Molnar, L.; Weaver, M. N.; Merz, K. M. Structural survey of zinc-containing proteins and development of the zinc AMBER force field (ZAFF). *J. Chem. Theory Comput.* **2010**, *6* (9), 2935–2947.
- (24) (a) Páll, S.; Zhmurov, A.; Bauer, P.; Abraham, M.; Lundborg, M.; Gray, A.; Hess, B.; Lindahl, E. Heterogeneous parallelization and acceleration of molecular dynamics simulations in GROMACS. *J. Chem. Phys.* **2020**, *153* (13), 134110. (b) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1–2*, 19–25. (c) Berendsen, H. J.; van der Spoel, D.; van Drunen, R. GROMACS: a message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* **1995**, *91* (1–3), 43–56. (d) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. GROMACS: fast, flexible, and free. *J. Comput. Chem.* **2005**, *26* (16), 1701–1718.
- (25) Ponder, J. W.; Richards, F. M. An efficient newton-like method for molecular mechanics energy minimization of large molecules. *J. Comput. Chem.* **1987**, *8* (7), 1016–1024.
- (26) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126* (1), 014101.
- (27) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52* (12), 7182–7190.
- (28) (a) Hess, B.; Bekker, H.; Berendsen, H. J.; Fraaije, J. G. LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997**, *18* (12), 1463–1472. (b) Hess, B. P-LINCS: A parallel linear constraint solver for molecular simulation. *J. Chem. Theory Comput.* **2008**, *4* (1), 116–122.
- (29) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103* (19), 8577–8593.
- (30) Flyvbjerg, H.; Petersen, H. G. Error estimates on averages of correlated data. *J. Chem. Phys.* **1989**, *91* (1), 461–466.
- (31) (a) Lee, R.; Conduit, G.; Nemec, N.; López Ríos, P.; Drummond, N. Strategies for improving the efficiency of quantum Monte Carlo calculations. *Phys. Rev.* **2011**, *83* (6), 066706. (b) Wolff, U.; Collaboration, A. Monte Carlo errors with less errors. *Comput. Phys. Commun.* **2004**, *156* (2), 143–153.
- (32) Xie, S.; Lu, Y.; Jakoncic, J.; Sun, H.; Xia, J.; Qian, C. Structure of RPA 32 bound to the N-terminus of SMARCA1 redefines the binding interface between RPA 32 and its interacting proteins. *FEBS J.* **2014**, *281* (15), 3382–3396.
- (33) Schmidt, T. G.; Koepke, J.; Frank, R.; Skerra, A. Molecular interaction between the Strep-tag affinity peptide and its cognate target, streptavidin. *J. Mol. Biol.* **1996**, *255* (5), 753–766.
- (34) Liu, Z.; Vogel, H. J. Structural basis for the regulation of L-type voltage-gated calcium channels: interactions between the N-terminal cytoplasmic domain and Ca<sup>2+</sup>-calmodulin. *Front. Mol. Neurosci.* **2012**, *5*, 38.
- (35) Yu, T. K.; Shin, S. A.; Kim, E. H.; Kim, S.; Ryu, K. S.; Cheong, H.; Ahn, H. C.; Jon, S.; Suh, J. Y. An unusual protein–protein interaction through coupled unfolding and binding. *Angew. Chem., Int. Ed.* **2014**, *53* (37), 9784.
- (36) Grace, C. R.; Ban, D.; Min, J.; Mayasundari, A.; Min, L.; Finch, K. E.; Griffiths, L.; Bharatham, N.; Bashford, D.; Kiplin, R.; et al. Monitoring ligand-induced protein ordering in drug discovery. *J. Mol. Biol.* **2016**, *428* (6), 1290–1303.
- (37) Eulitz, S.; Sauer, F.; Pelissier, M.-C.; Boisguerin, P.; Molt, S.; Schulz, J.; Orfanos, Z.; Kley, R. A.; Volkmer, R.; Wilmanns, M.; et al. Identification of Xin-repeat proteins as novel ligands of the SH3 domains of nebulin and nebulin and analysis of their interaction during myofibril formation and remodeling. *Mol. Biol. Cell* **2013**, *24* (20), 3215–3226.
- (38) Tallant, C.; Valentini, E.; Fedorov, O.; Overvoorde, L.; Ferguson, F. M.; Filippakopoulos, P.; Svergun, D. I.; Knapp, S.; Ciulli, A. Molecular basis of histone tail recognition by human TIP5 PHD finger and bromodomain of the chromatin remodeling complex NoRC. *Structure* **2015**, *23* (1), 80–92.
- (39) Clark, S.; Nyarko, A.; Löhr, F.; Karplus, P. A.; Barbar, E. The anchored flexibility model in LC8 motif recognition: insights from the chica complex. *Biochemistry* **2016**, *55* (1), 199–209.
- (40) Ponna, S. K.; Ruskamo, S.; Myllykoski, M.; Keller, C.; Boeckers, T. M.; Kursula, P. Structural basis for PDZ domain interactions in the post-synaptic density scaffolding protein Shank3. *J. Neurochem.* **2018**, *145* (6), 449–463.
- (41) Murthy, A. V.; Sulu, R.; Koski, M. K.; Tu, H.; Anantharajan, J.; Sah-Teli, S. K.; Myllyharju, J.; Wierenga, R. K. Structural enzymology binding studies of the peptide-substrate-binding domain of human collagen prolyl 4-hydroxylase (type-II): High affinity peptides have a PxGP sequence motif. *Protein Sci.* **2018**, *27* (9), 1692–1703.
- (42) Huber, J.; Obata, M.; Gruber, J.; Akutsu, M.; Löhr, F.; Rogova, N.; Güntert, P.; Dikic, I.; Kirkin, V.; Komatsu, M.; et al. An atypical LIR motif within UBA5 (ubiquitin like modifier activating enzyme 5) interacts with GABARAP proteins and mediates membrane localization of UBA5. *Autophagy* **2020**, *16* (2), 256–270.
- (43) Çınaroğlu, S. S.; Biggin, P. C. The role of loop dynamics in the prediction of ligand–protein binding enthalpy. *Chem. Sci.* **2023**, *14* (24), 6792–6805.
- (44) Wade, M.; Li, Y.-C.; Wahl, G. M. MDM2, MDMX and p53 in oncogenesis and cancer therapy. *Nat. Rev. Cancer* **2013**, *13* (2), 83–96.
- (45) Zhang, A.; Yu, H.; Liu, C.; Song, C. The Ca<sup>2+</sup> permeation mechanism of the ryanodine receptor revealed by a multi-site ion model. *Nat. Commun.* **2020**, *11* (1), 922.
- (46) Peters, M. B.; Yang, Y.; Wang, B.; Fusti-Molnar, L.; Weaver, M. N.; Merz, K. M., Jr. Structural Survey of Zinc-Containing Proteins and Development of the Zinc AMBER Force Field (ZAFF). *J. Chem. Theory Comput.* **2010**, *6* (9), 2935–2947.
- (47) Gao, K.; Yin, J.; Henriksen, N. M.; Fenley, A. T.; Gilson, M. K. Binding enthalpy calculations for a neutral host–guest pair yield widely divergent salt effects across water models. *J. Chem. Theory Comput.* **2015**, *11* (10), 4555–4564.
- (48) (a) Takemura, K.; Matubayasi, N.; Kitao, A. Binding free energy analysis of protein–protein docking model structures by evERdock. *J. Chem. Phys.* **2018**, *148* (10), 105101. (b) Perthold, J. W.; Oostenbrink, C. Simulation of reversible protein–protein binding and calculation of binding free energies using perturbed distance restraints. *J. Chem. Theory Comput.* **2017**, *13* (11), 5697–5708.
- (49) Aldeghi, M.; Heifetz, A.; Bodkin, M. J.; Knapp, S.; Biggin, P. C. Accurate calculation of the absolute free energy of binding for drug molecules. *Chem. Sci.* **2016**, *7* (1), 207–218.
- (50) Valsson, O.; Tiwary, P.; Parrinello, M. Enhancing Important Fluctuations: Rare Events and Metadynamics from a Conceptual Viewpoint. *Annu. Rev. Phys. Chem.* **2016**, *67* (1), 159–184.
- (51) Abrams, C.; Bussi, G. Enhanced sampling in molecular dynamics using metadynamics, replica-exchange, and temperature-acceleration. *Entropy* **2013**, *16*, 163–199.
- (52) (a) Yang, L.; Liu, C.-W.; Shao, Q.; Zhang, J.; Gao, Y. Q. From thermodynamics to kinetics: Enhanced sampling of rare events. *Acc. Chem. Res.* **2015**, *48* (4), 947–955. (b) Miao, Y.; McCammon, J. A. Unconstrained enhanced sampling for free energy calculations of biomolecules: a review. *Mol. Simul.* **2016**, *42* (13), 1046–1055.
- (53) (a) Krishnamurthy, V. M.; Bohall, B. R.; Semetey, V.; Whitesides, G. M. The paradoxical thermodynamic basis for the interaction of ethylene glycol, glycine, and sarcosine chains with bovine carbonic anhydrase II: an unexpected manifestation of enthalpy/entropy compensation. *J. Am. Chem. Soc.* **2006**, *128* (17), 5802–5812. (b) DeLorbe, J. E.; Clements, J. H.; Teresk, M. G.;

- Benfield, A. P.; Plake, H. R.; Millspaugh, L. E.; Martin, S. F. Thermodynamic and structural effects of conformational constraints in protein–ligand interactions. Entropic paradox associated with ligand preorganization. *J. Am. Chem. Soc.* **2009**, *131* (46), 16758–16770.
- (54) Chodera, J. D.; Mobley, D. L. Entropy–enthalpy compensation: role and ramifications in biomolecular ligand recognition and design. *Annu. Rev. Biophys.* **2013**, *42*, 121–142.
- (55) Geschwindner, S.; Ulander, J.; Johansson, P. Ligand binding thermodynamics in drug discovery: still a hot tip? *J. Med. Chem.* **2015**, *58* (16), 6321–6335.
- (56) Nilofer, C.; Sukhwil, A.; Mohanapriya, A.; Sakthar, M. K.; Kanguane, P. Small protein–protein interfaces rich in electrostatic are often linked to regulatory function. *J. Biomol. Struct. Dyn.* **2020**, *38* (11), 3260–3279.
- (57) (a) Tyka, M. D.; Keedy, D. A.; André, I.; DiMaio, F.; Song, Y.; Richardson, D. C.; Richardson, J. S.; Baker, D. Alternate states of proteins revealed by detailed energy landscape mapping. *J. Mol. Biol.* **2011**, *405* (2), 607–618. (b) Gershenson, A.; Gierasch, L. M.; Pastore, A.; Radford, S. E. Energy landscapes of functional proteins are inherently risky. *Nat. Chem. Biol.* **2014**, *10* (11), 884–891.
- (58) Fu, H.; Zhou, Y.; Jing, X.; Shao, X.; Cai, W. Meta-analysis reveals that absolute binding free-energy calculations approach chemical accuracy. *J. Med. Chem.* **2022**, *65* (19), 12970–12978.
- (59) Suárez, D.; Díaz, N. Conformational and entropy analyses of extended molecular dynamics simulations of  $\alpha$ - $\beta$ - and  $\gamma$ -cyclodextrins and of the  $\beta$ -cyclodextrin/nabumetone complex. *Phys. Chem. Chem. Phys.* **2017**, *19* (2), 1431–1440.
- (60) (a) Yin, J.; Henriksen, N. M.; Muddana, H. S.; Gilson, M. K. Bind3p: optimization of a water model based on host–guest binding data. *J. Chem. Theory Comput.* **2018**, *14* (7), 3621–3632. (b) Henriksen, N. M.; Fenley, A. T.; Gilson, M. K. Computational calorimetry: high-precision calculation of host–guest binding thermodynamics. *J. Chem. Theory Comput.* **2015**, *11* (9), 4377–4394.
- (61) Çınaroğlu, S. S.; Biggin, P. C. Evaluating the performance of water models with host–guest force fields in binding enthalpy calculations for cucurbit [7] uril–guest systems. *J. Phys. Chem. B* **2021**, *125* (6), 1558–1567.
- (62) (a) Qasba, P. K.; Ramakrishnan, B.; Boeggeman, E. Substrate-induced conformational changes in glycosyltransferases. *Trends Biochem. Sci.* **2005**, *30* (1), 53–62. (b) Zhang, Z.; Liu, F.; Chen, J. Conformational changes of CFTR upon phosphorylation and ATP binding. *Cell* **2017**, *170* (3), 483–491.e8. e488. (c) Henchman, R. H.; Wang, H.-L.; Sine, S. M.; Taylor, P.; Andrew McCammon, J. Ligand-induced conformational change in the  $\alpha 7$  nicotinic receptor ligand binding domain. *Biophys. J.* **2005**, *88* (4), 2564–2576. (d) Léger, C.; Di Meo, T.; Aumont-Nicaise, M.; Velours, C.; Durand, D.; Li de la Sierra-Gallay, I.; van Tilbeurgh, H.; Hildebrandt, N.; Desmadril, M.; Urvoas, A. Ligand-induced conformational switch in an artificial bidomain protein scaffold. *Sci. Rep.* **2019**, *9* (1), 1178. (e) Latz, E.; Verma, A.; Visintin, A.; Gong, M.; Sirois, C. M.; Klein, D. C.; Monks, B. G.; McKnight, C. J.; Lamphier, M. S.; Duprex, W. P.; et al. Ligand-induced conformational changes allosterically activate Toll-like receptor 9. *Nat. Immunol.* **2007**, *8* (7), 772–779. (f) Zarutskie, J. A.; Sato, A. K.; Rushe, M. M.; Chan, I. C.; Lomakin, A.; Benedek, G. B.; Stern, L. J. A conformational change in the human major histocompatibility complex protein HLA-DR1 induced by peptide binding. *Biochemistry* **1999**, *38* (18), 5878–5887.
- (63) Ho, B. K.; Dill, K. A. Folding very short peptides using molecular dynamics. *PLoS Comput. Biol.* **2006**, *2* (4), No. e27.
- (64) (a) Sattler, M.; Liang, H.; Nettessheim, D.; Meadows, R. P.; Harlan, J. E.; Eberstadt, M.; Yoon, H. S.; Shuker, S. B.; Chang, B. S.; Minn, A. J.; et al. Structure of Bcl-xL-Bak peptide complex: recognition between regulators of apoptosis. *Science* **1997**, *275* (5302), 983–986. (b) Petros, A. M.; Nettessheim, D. G.; Wang, Y.; Olejniczak, E. T.; Meadows, R. P.; Mack, J.; Swift, K.; Matayoshi, E. D.; Zhang, H.; Thompson, C. B.; et al. Rationale for Bcl-xL/Bad peptide complex formation from structure, mutagenesis, and biophysical studies. *Protein Sci.* **2000**, *9* (12), 2528–2534.
- (65) (a) Botuyan, M. V. E.; Momand, J.; Chen, Y. Solution conformation of an essential region of the p53 transactivation domain. *Folding Des.* **1997**, *2* (6), 331–342. (b) Lee, H.; Mok, K. H.; Muhandiram, R.; Park, K.-H.; Suk, J.-E.; Kim, D.-H.; Chang, J.; Sung, Y. C.; Choi, K. Y.; Han, K.-H. Local structural elements in the mostly unstructured transcriptional activation domain of human p53. *J. Biol. Chem.* **2000**, *275* (38), 29426–29432. (c) Dawson, R.; Müller, L.; Dehner, A.; Klein, C.; Kessler, H.; Buchner, J. The N-terminal domain of p53 is natively unfolded. *J. Mol. Biol.* **2003**, *332* (5), 1131–1141.
- (66) Kussie, P. H.; Gorina, S.; Marechal, V.; Elenbaas, B.; Moreau, J.; Levine, A. J.; Pavletich, N. P. Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain. *Science* **1996**, *274* (5289), 948–953.
- (67) Böttger, V.; Böttger, A.; Garcia-Echeverria, C.; Ramos, Y. F.; van der Eb, A. J.; Jochemsen, A. G.; Lane, D. P. Comparative study of the p53-mdm2 and p53-MDMX interfaces. *Oncogene* **1999**, *18* (1), 189–199.
- (68) Lee, C. W.; Martinez-Yamout, M. A.; Dyson, H. J.; Wright, P. E. Structure of the p53 transactivation domain in complex with the nuclear receptor coactivator binding domain of CREB binding protein. *Biochemistry* **2010**, *49* (46), 9964–9971.
- (69) London, N.; Movshovitz-Attias, D.; Schueler-Furman, O. The structural basis of peptide-protein binding strategies. *Structure* **2010**, *18* (2), 188–199.
- (70) (a) Zhang, Q.; Zeng, S. X.; Lu, H. Targeting p53-MDM2-MDMX loop for cancer therapy. *Subcell. Biochem.* **2014**, *85*, 281–319. (b) Spiegelberg, D.; Mortensen, A. C.; Lundsten, S.; Brown, C. J.; Lane, D. P.; Nestor, M. The MDM2/MDMX-p53 antagonist PM2 radiosensitizes wild-type p53 tumors. *Cancer Res.* **2018**, *78* (17), 5084–5093. (c) Popowicz, G. M.; Dömling, A.; Holak, T. A. The structure-based design of MDM2/MDMX-p53 inhibitors gets serious. *Angew. Chem., Int. Ed.* **2011**, *50* (12), 2680–2688.
- (71) (a) Soudah, N.; Padala, P.; Hassouna, F.; Kumar, M.; Mashahreh, B.; Lebedev, A. A.; Isupov, M. N.; Cohen-Kfir, E.; Wiener, R. An N-terminal extension to UBA5 adenylation domain boosts UFM1 activation: Isoform-specific differences in ubiquitin-like protein activation. *J. Mol. Biol.* **2019**, *431* (3), 463–478. (b) Oweis, W.; Padala, P.; Hassouna, F.; Cohen-Kfir, E.; Gibbs, D. R.; Todd, E. A.; Berndsen, C. E.; Wiener, R. Trans-binding mechanism of ubiquitin-like protein activation revealed by a UBA5-UFM1 complex. *Cell reports* **2016**, *16* (12), 3113–3120.
- (72) Best, R. B.; Buchete, N.-V.; Hummer, G. Are current molecular dynamics force fields too helical? *Biophys. J.* **2008**, *95* (1), L07–L09.
- (73) Polêto, M. D.; Lemkul, J. A. Integration of experimental data and use of automated fitting methods in developing protein force fields. *Comms. Chem.* **2022**, *5* (1), 38.
- (74) Holdgate, G. A. Making cool drugs hot: isothermal titration calorimetry as a tool to study binding energetics. *Biotechniques* **2001**, *31* (1), 164.
- (75) Harper, E.; Black, J. Histamine H3-receptor agonists and imidazole-based H3-receptor antagonists can be thermodynamically discriminated. *Br. J. Pharmacol.* **2007**, *151* (4), 504–517.
- (76) (a) Zheng, X.; Cheng, W.; Ji, C.; Zhang, J.; Yin, M. Detection of metal ions in biological systems: A review. *Rev. Anal. Chem.* **2020**, *39* (1), 231–246. (b) Thomson, A. J.; Gray, H. B. Bio-inorganic chemistry. *Curr. Opin. Chem. Biol.* **1998**, *2* (2), 155–158.
- (77) (a) Waldron, K. J.; Rutherford, J. C.; Ford, D.; Robinson, N. J. Metalloproteins and metal sensing. *Nature* **2009**, *460* (7257), 823–830. (b) Yannone, S. M.; Hartung, S.; Menon, A. L.; Adams, M. W.; Tainer, J. A. Metals in biology: defining metalloproteomes. *Curr. Opin. Biotechnol.* **2012**, *23* (1), 89–95.
- (78) Stote, R. H.; Karplus, M. Zinc binding in proteins and solution: a simple but accurate nonbonded representation. *Proteins: Struct., Funct., Bioinf.* **1995**, *23* (1), 12–31.
- (79) Li, P.; Roberts, B. P.; Chakravorty, D. K.; Merz, K. M. Rational design of particle mesh Ewald compatible Lennard-Jones parameters for +2 metal cations in explicit solvent. *J. Chem. Theory Comput.* **2013**, *9* (6), 2733–2748.

(80) Li, P.; Merz, K. M. Taking into account the ion-induced dipole interaction in the nonbonded model of ions. *J. Chem. Theory Comput.* **2014**, *10* (1), 289–297.

(81) (a) Han, L.; Yang, Q.; Liu, Z.; Li, Y.; Wang, R. Development of a new benchmark for assessing the scoring functions applicable to protein–protein interactions. *Future Med. Chem.* **2018**, *10* (13), 1555–1574. (b) Mobley, D. L.; Gilson, M. K. Predicting binding free energies: frontiers and benchmarks. *Annu. Rev. Biophys.* **2017**, *46*, 531–558. (c) Kastiris, P. L.; Moal, I. H.; Hwang, H.; Weng, Z.; Bates, P. A.; Bonvin, A. M.; Janin, J. A structure-based benchmark for protein–protein binding affinity. *Protein Sci.* **2011**, *20* (3), 482–491.