

Reviewed Preprint

v1 • May 12, 2026

Not revised

✉ For correspondence:

stephanie@wankowiczlab.com

Competing interests: No

competing interests declared

Funding: See [page 16](#)

Reviewing editor: Shozeb Haider,
University College London, United
Kingdom

© 2026, Miller & Wankowicz. This article is distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use and redistribution provided that the original author and source are credited.

Binding Entropy Can Be Predicted by Crystallographic Ensembles

Charlotte A Miller¹, Stephanie A Wankowicz^{1,2,3,4} ✉

¹Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, United States • ²Departments of Biochemistry & Computer Science, Vanderbilt University, Nashville, United States • ³Center for AI in Protein Dynamics, Vanderbilt University, Nashville, United States • ⁴Center for Structural Biology, Vanderbilt University, Nashville, United States

eLife Assessment

This study provides a **useful** demonstration that, at least for the systems examined, aspects of the entropic contribution to protein-ligand binding can be inferred directly from crystallographic data. In doing so, it strengthens a view of crystal structures as heterogeneous ensembles that are amenable to statistical-mechanical analysis rather than purely static models. The analytical approaches are carefully developed and transparently discussed, with thoughtful consideration of both successful and less effective methods, lending **solid** support to the central conclusions. However, because the analysis is based on a relatively small and narrowly sampled set of protein-ligand complexes, the generality of these findings remains speculative and will require broader validation.

<https://doi.org/10.7554/eLife.111298.1.sa2>

Abstract

Protein-ligand binding is governed by free energy, comprising both enthalpic and entropic contributions. Yet structural interpretations of binding thermodynamics have predominantly focused on enthalpic interactions, largely neglecting entropy because it is difficult to quantify from static structural models. Here, we developed multiconformer ensemble models to analyze high-resolution X-ray crystallography structures and estimate both protein and solvent conformational entropies. These ensemble models successfully predicted experimental binding entropies measured by isothermal titration calorimetry for over 70 protein-ligand pairs across 12 proteins, revealing a strong linear correlation. Protein entropy, estimated using crystallographic order parameters that capture both harmonic and anharmonic motion, correlates linearly with experimental binding entropy. Incorporating resolution-corrected differences in water-molecule counts substantially improves predictions, demonstrating that protein and solvent contributions must be considered jointly. Analysis of water-protein hydrogen bonding networks partially explains entropic differences across complexes. These results establish that crystallographic ensembles can quantify binding entropy, enabling explicit entropic considerations in structure-based studies of molecular recognition for both functional analysis and drug design.

Introduction

Protein–ligand interactions orchestrate enzyme function and signaling throughout biology. Much of modern medicine harnesses these interactions to manipulate proteins with synthetic ligands, thereby altering disease pathways. We have long sought to understand the physical principles underlying protein-ligand interactions. However, our static structural models, and thus our physical interpretations, have historically emphasized enthalpic interactions[1]. Yet binding affinity is determined by free energy, which comprises both enthalpic and entropic contributions. The entropic component of free energy has received far less attention, primarily because entropy

cannot be directly visualized or quantified from static structural models, instead requiring ensemble representations [2,3]. Quantifying binding entropy directly from structural data would enable systematic entropic optimization and help reveal the structural determinants of binding thermodynamics.

Explanations of the entropic component of ligand binding often focus on solvent molecule release [4][5]. Yet accumulating evidence indicates that protein entropy is also a major determinant of binding entropy, with contributions differing substantially between proteins and between different ligands binding to the same protein [2,6–11]. Protein conformational entropy arises from the many degrees of freedom proteins possess, much of which is derived from side-chain sampling within and between rotamer states [12–15]. This conformational multiplicity, along with solvent entropy, creates an entropic reservoir that is perturbed upon ligand binding, contributing to both protein stability and binding thermodynamics [6,12,16–21].

Developing quantitative metrics to estimate protein entropy has been challenging. Protein entropy encompasses both conformational entropy, which reflects anharmonic transitions between discrete rotamer states, and harmonic vibrational entropy, which reflects fluctuations within individual rotamer states. Accurate entropy calculations require the complete enumeration of the partition function over all conformational states, weighted by their Boltzmann factors [22]. However, this is computationally intractable due to the high dimensionality of conformational space, leading to various quantitative approximations. In molecular dynamics (MD), quasi-harmonic analyses of trajectories approximate entropy from covariance matrices but fail to capture the anharmonic entropic contributions [23,24]. Further, many MD simulations do not span biologically relevant timescales [21]. Nuclear Magnetic Resonance (NMR) relaxation techniques provide site-specific measurements of disorder in methyl or amide groups on picosecond timescales [25,26]. These measurements capture probability distributions of spatial coordinates and have been correlated with entropy [7,14,27]; but face challenges of throughput, timescale, and structural degeneracy.

X-ray crystallography and cryo-electron microscopy (cryo-EM) collect data from millions of protein copies, capturing many conformational states, therefore potentially containing significant information on the multiplicity of states and, thus, entropy. However, structural models typically represent only a single state, discarding information about the multiplicity of states that drive entropy [28]. We previously demonstrated algorithmic advances that enable the modeling of high-resolution X-ray crystallography data as multiconformer ensembles [28,29]. These ensembles explicitly capture anharmonic heterogeneity, including differences in rotamer states, through alternative conformations, as well as harmonic heterogeneity through B-factors, thereby demonstrating an improved fit to experimental data [29]. Critically, these models enable quantitative estimates of protein conformational entropy via crystallographic order parameters [30], yielding values that agree with NMR relaxation order parameters and thus capture the spatial distribution of side chains.

Previous studies identified a quantifiable relationship between binding entropy and protein conformational entropy as measured by NMR order parameters [30,31], leading to the development of the ‘entropy meter’ [7]. Recently, we demonstrated that within a single system, protein and solvent entropy derived from crystallographic ensembles estimates correlated with binding entropy [8]. We also showed that protein-solvent hydrogen-bonding networks also correlate with binding entropy, linking protein conformational entropy to solvent entropy through coordinated, system-level redistribution. However, whether these findings generalize beyond individual systems remains an open question.

Here, we establish that crystallographic ensembles provide quantitative entropy estimates that correlate with binding entropy across multiple systems. We show that accounting for both protein and solvent entropy is key to predicting binding entropy. These structure-based entropy metrics enable integration of this critical free energy component into both the intuition and quantification of binding thermodynamics.

Results

Dataset

We collected thermodynamic binding measurements from isothermal titration calorimetry (ITC) from our previous study and from 14 datasets representing 12 different proteins (SARS-CoV-2 macrodomain [Mac1], Cyclin-dependent kinase 2 [CDK2], Thermolysin, Major Urinary Protein 1 [MUP-1], Bromodomain-containing protein 4 [BRD4], tRNA Guanine Transglycosylase [TGT], human Carbonic Anhydrase II [HCAII], Farnesyl pyrophosphate synthase [FPS], Galectin, and Transthyretin [TTR], Bovine Trypsin [Trypsin]), totaling 75 unique protein-ligand interactions (Supplementary Table 1 [\[8,9,32–41\]](#)). For protein-ligand pairs with multiple ITC measurements, we averaged each measurement for subsequent analysis. For each ITC measurement, we required a corresponding structure that contained the same protein-ligand pair measured by ITC, along with an apo structure with the same sequence, space group, and resolution (within 0.3 Angstroms), and unit cell dimensions/angles (within 10%) (Supplementary Table 2 [\[42–45\]](#)). While we looked at additional protein-ligand pairs, most notably the extensive work on HIV protease and Thrombin [\[42–45\]](#), these did not meet the stringent crystallographic criteria (see all datasets examined but omitted in Supplementary Table 3 [\[46–47\]](#)). All structures had resolutions better than 2.1 Angstroms and had their structure factors deposited (Supplementary Figure 1A/B [\[48\]](#)). We remodeled all structures using a pipeline of multiconformer models from qFit and Phenix refinement (Methods [\[29,46,47\]](#)). Multiconformer models were required to have an R_{free} of better than 0.25 (Supplementary Figure 1C/D [\[49\]](#)). Three pairs were removed due to refinement errors.

Our dataset comprised proteins with bound chains ranging from 115 to 316 residues (median = 258; Supplementary Figure 2A [\[50\]](#)). All bound and unbound structure pairs exhibited minimal conformational changes, with alpha carbon root mean squared deviations (RMSD) below 1.75 Å (Supplementary Figure 2B [\[51\]](#)). This confirmed the absence of large conformational changes, as previously investigated by NMR [\[31\]](#). Ligands ranged in molecular weight from 102.46 to 461.50 Da (median = 284.68), with hydrogen bond acceptor counts of 1–7 (median = 4) and hydrogen bond donor counts of 0–5 (median = 2; Supplementary Figure 3A [\[52\]](#)).

Binding Entropy Linearly Correlates with Protein Conformational Entropy

We first aimed to determine if protein entropy, estimated by crystallographic ensembles, correlated with binding entropy (Figure 1A [\[53\]](#)). To estimate protein entropy from multiconformer structures, we calculated crystallographic order parameters that measure the first torsion angle of each alternative conformation, weighted by its B-factor. We measured them for all residues except glycine and proline (Methods [\[48\]](#)). We calculated the difference in protein entropy as the average difference between residues multiplied by the number of residues in the structure, as previously done [\[7\]](#). When we compared these structural estimates to binding entropy measurements from ITC, we observed a linear relationship but found that our structural estimates systematically overestimated entropy changes by approximately fourfold. After applying a correction factor of 0.25, the structural estimates aligned much more closely with experimental values along a 1-to-1 line. Proteins showing increased conformational entropy in the bound state exhibited more entropically driven binding (Slope=0.85, $R^2=0.59$; Figure 1B [\[54\]](#); Supplementary Table 4 [\[55\]](#)). We also explored the correlation between protein conformational entropy and the individual measured ITC values, observing that a significant correlation persisted, though with greater scatter around the median trend line, as expected given experimental variation in replicate measurements (Supplementary Figure 4 [\[56\]](#)). This correlation indicates that binding entropy increases by approximately 0.71 kcal/mol-K per unit of protein conformational entropy. The magnitude of these changes provides insight into their structural distribution. Given that the proteins examined

average 258 residues and that the maximum conformational entropy change per residue is 0.25 kcal/mol·K (with a median closer to 0.05 kcal/mol·K), the observed entropy changes likely arise from tens to hundreds of residues shifting their conformational sampling.

The strong observed correlation between protein conformational entropy and binding entropy, without explicitly accounting for solvent, likely arises for two reasons. First, we and others have observed that protein and solvent entropic contributions tend to covary [4,8] with structural rearrangements that alter protein conformational entropy, often simultaneously reorganizing interfacial and proximal water networks. Second, it is possible that for the protein-ligand pairs studied here, protein conformational entropy may simply have a larger effect than solvent entropy [4].

We also noted that protein-ligand pairs from the same protein tended to cluster within a relatively narrow range of both structural and binding entropy values, with TGT being an exception [35,49]. We suspect that this limited dynamic range is driven by the ligand series used in each system, as we obtained ITC data from papers that often explore closely related ligands bound to the same protein, leading to similar binding thermodynamics. It is also possible that this pattern reflects intrinsic thermodynamic preferences of individual proteins, with some systems favoring more entropically driven binding mechanisms and others relying more on enthalpic stabilization. Further binding-entropy analysis across a broader range of ligands is needed to determine whether this hypothesis is true.

Solvent Entropy Estimates are not Strongly Correlated with Binding Entropy

Having observed a strong linear correlation between protein entropy and binding entropy, we next investigated whether incorporating estimates of solvent entropy from crystallographic structures could further refine these predictions, as binding entropic effects arise from the collective behavior of protein, solvent, and ligand [50–52]. We first estimated changes in solvent-accessible surface area (SASA) between bound and unbound structures (Methods). Across all protein-ligand pairs, bound structures exhibited, on average, slightly greater solvent-accessible surface area (Supplementary Figure 5). This increase was primarily driven by apolar SASA, which has been hypothesized to enhance solvent entropy by displacing ordered water molecules from apolar surfaces into the bulk solvent [53]. However, we observed no strong correlation between any SASA measure and binding entropy (total Δ SASA: $R^2 = 0.0058$; apolar Δ SASA: $R^2 = 0.017$; polar Δ SASA: $R^2 = 0.0027$; Supplementary Figure 6). Previous work has linked solvent-derived heat capacity estimates to entropy estimates using the relationship (Δ Heat Capacity = $(-0.26)\Delta$ SASA_{apolar} + $(0.45)\Delta$ SASA_{polar}) [7,54], prompting us to explore whether we could identify a similar relationship. However, we found no strong correlation between changes in heat capacity and binding entropy ($R^2 = 0.066$; Figure 2A).

We reasoned that SASA measures alone may be insufficient, as they only indicate whether a residue is solvent-exposed, without capturing the complex thermodynamic behavior of water, particularly its hydrogen-bonding capacity or ‘iceberg’ like interactions at non-polar areas [50,51,55,56]. Therefore, we analyzed changes in the number of water molecules bound per residue, defined as water molecules within 3.2 Å of each residue (Methods). Across all protein-ligand pairs, most residues showed no change in bound water molecules (60.7%), while 22.8% gained and 16.5% lost water molecules (Supplementary Figure 7A). Notably, the mean difference in water molecule counts correlated with resolution differences between apo and bound structures, with the higher-resolution structure exhibiting more bound water molecules ($R^2 = 0.38$; Supplementary Figure 7B). To address this confounding factor, we performed residual analysis to regress out resolution effects (Methods; Supplementary Figure 8). After correction, the distribution reversed, with 56.7% of residues losing water molecules while 43.3% gained them. We then determined if the differences in the number of water molecules, controlling for resolution differences, would correlate with binding entropy. While this slightly improved the correlation with binding entropy relative to estimates of heat capacity, it still had little power to explain binding entropy (Figure 2B; $R^2 = 0.11$).

Figure 1.

A. We obtained binding thermodynamic parameters by ITC from 14 proteins, comprising 72 protein-ligand pairs. Using multiconformer models, we calculated estimates of protein and solvent conformational entropies to identify correlations between ITC binding entropy and structural estimates of entropy. **B.** Distribution of the thermodynamic binding contributions across all protein-ligand pairs ordered by entropy ($n = 72$). **C.** Correlation between protein conformational entropy as measured by crystallographic multiconformer structures and ITC-measured binding entropy. Each dot represents an individual protein-ligand complex, with each color representing a different protein (slope = 0.85, $R^2 = 0.59$). Higher binding entropy/protein conformational entropy indicates a more unfavorable binding entropy.

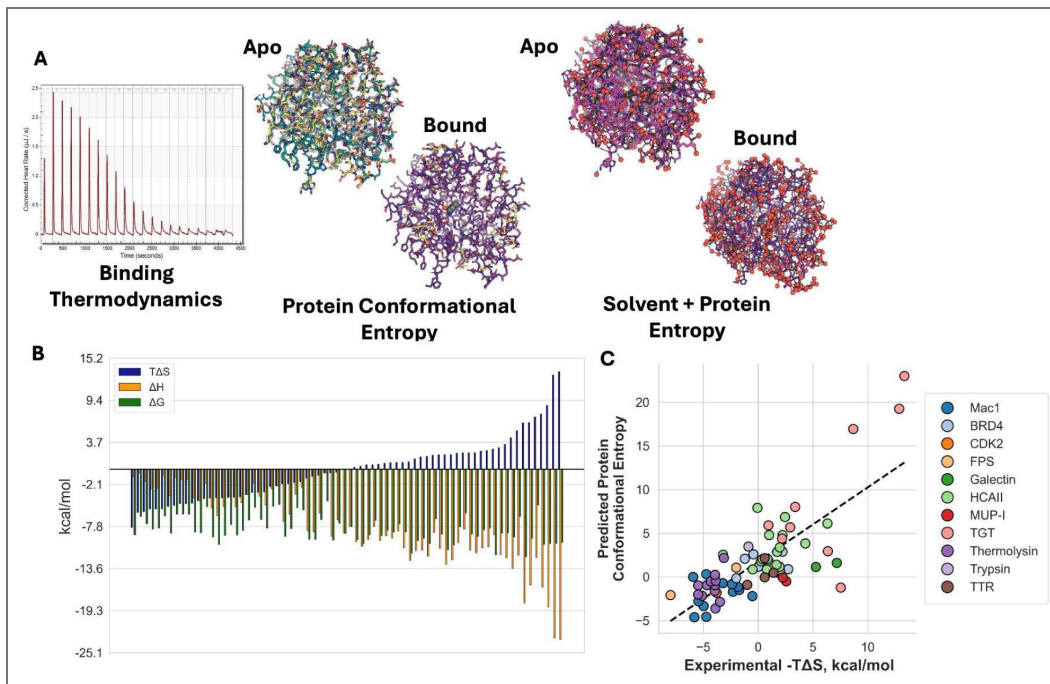
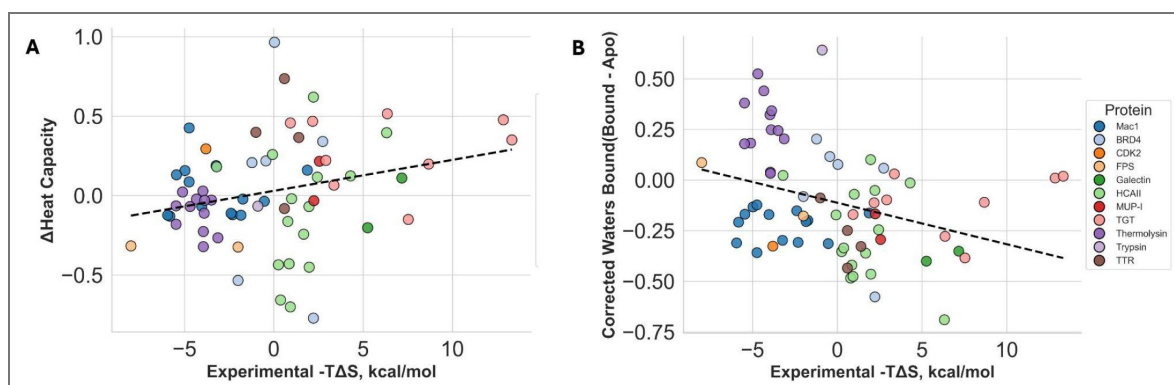


Figure 2.

A. Estimates of differences in heat capacity with binding entropy (Slope = 0.020, $R^2 = 0.066$); **B.** Difference in the number of water molecules bound corrected for resolution and correlation with binding entropy (Slope = -0.021, $R^2 = 0.11$)



Binding Entropy can be Predicted from Protein and Solvent Entropy Estimates from Crystallographic Ensembles

We then aimed to test whether estimates in solvent entropy, combined with protein entropy, improved the prediction of binding entropy. We first used SASA values, including heat capacity; however, we did not observe any improvement over protein conformational entropy estimates alone ($R^2 = 0.58$; [Figure 3A](#); [Supplementary Figure 9](#)). Finally, we evaluated whether changes in water molecule counts improved binding entropy predictions from crystallographic ensembles. Using a linear model, we found that resolution-corrected water counts improved prediction accuracy ($R^2 = 0.69$; $-1.44 + 0.64 \times \text{Protein Conformational Entropy} - 3.3 \times \text{Difference in Number of Waters (Corrected)}$; [Figure 3B](#)). Slightly better improvements were observed with uncorrected differences in water counts ($R^2 = 0.704$, [Supplementary Figure 10](#)). The consistent trends observed across proteins and ligands demonstrate that protein conformational entropy estimated from high-resolution crystallographic ensembles can provide semi-quantitative estimates of protein conformational entropy upon ligand binding.

The most dramatic changes between binding entropies of the same protein were observed in TGT, with binding entropies ranging from 0.95 to 13.34 kcal/mol [[35,49](#)]. All entropies impose a penalty on free energy, though to varying degrees. We aimed to visually investigate what was driving these changes. In the apo state, nearly every residue exhibited alternative conformations, predominantly involving the backbone. Upon ligand binding, both complexes (4Q8T and 4PUK) showed substantial reductions in conformational heterogeneity, contributing to the entropic penalty of binding. Notably, 4PUK exhibited a larger estimated protein entropy change (11.31 vs. 5.91 kcal/mol·K) and a correspondingly larger experimental entropic penalty (13.34 kcal/mol). This complex also showed increased water occupancy (0.018 additional resolution-corrected waters per residue), whereas 4Q8T exhibited net water release (-0.17 fewer resolution-corrected waters per residue). These results demonstrate that both protein conformational restriction and differential solvent reorganization contribute to the distinct binding thermodynamics observed for these two ligands binding the same protein ([Figure 3C](#)).

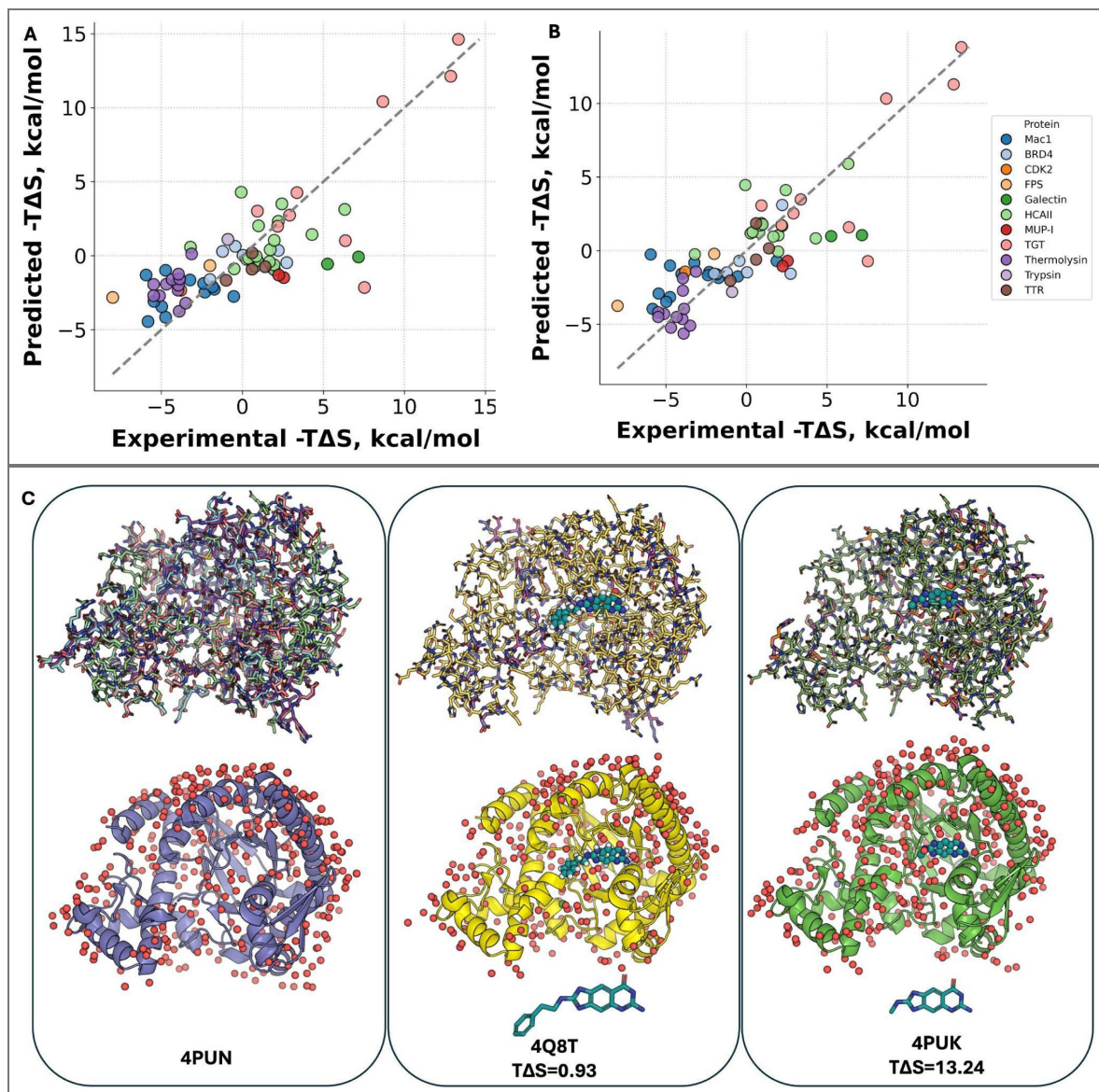
Hydrogen Bond Networks Show Moderate Correlation with Binding Entropy

We previously found that decreased residue packing and fewer protein–solvent hydrogen bonds in the bound state are associated with decreased binding entropy [[8](#)]. We aimed to assess whether these patterns generalize across different proteins. We excluded the Mac1 protein–ligand pairs from this analysis, as they were the source of our initial observations. To investigate the structural basis for entropic differences, we built a hydrogen bond network and computed graph-theoretic metrics that capture network connectivity, with nodes corresponding to hydrogen-bond-forming residues (**Methods**). Graph parameters, including the number of nodes and edges were computed. When examining protein only features, we observed weak correlations between structural metrics and binding entropy, with the number of nodes ($R^2 = 0.073$) and edges ($R^2 = 0.050$) showing minimal predictive power ([Supplementary Figure 11](#)). Given that local packing density restricts conformational sampling and correlates with residue-level entropy in folding studies [[13](#)], we tested whether packing metrics would correlate with binding entropy. We quantified packing by counting neighboring residues within 5.0 Å of each residue's heavy atoms. Packing density showed only weak correlation with experimental binding entropy ($R^2 = 0.13$; [Supplementary Figure 12](#)), suggesting that the entropic penalties of binding are likely not very driven by static structural crowding.

However, incorporating water molecules into the hydrogen bond network analysis (considering protein or water molecules as potential hydrogen bonding nodes) revealed stronger correlations with binding entropy (nodes [$R^2 = 0.38$] and edges [$R^2 = 0.24$]; [Supplementary Figure 13](#)). This indicates that binding events that generate more extensive hydrogen-bond networks, encompassing both protein and water, tend to be less entropically driven. The notable improvement observed when water is included in hydrogen-bonding networks further supports

Figure 3.

A. Correlation between predicted binding entropy values using estimates of heat capacity ($R^2 = 0.59$; $\Delta S = -1.044 + 0.891 * [\text{Protein Entropy}] + 1.097 * [\text{change in heat capacity}]$) **B.** Correlation between predicted binding entropy values using change in number of bound water molecules corrected for resolution ($R^2 = 0.69$; $\Delta S = -1.45 + 0.64 * [\text{Protein Entropy}] - 3.8 * [\text{Number of Water}]$) **C.** Protein conformational entropy and solvent differences between apo TGT (PDB: 4PUN) and two ligand-bound structures with dramatically different changes in binding entropy. Both ligands pay an entropic penalty, with 4Q8T paying a lower entropic penalty, derived from a smaller reduction in side chain alternative conformers (observed based on color differences of stick representations in the first row) and a smaller difference in the number of water molecules in the bound structure. 4PUK has more water molecules bound per residue with corrected resolution, further explaining the larger entropic penalty.



the idea that protein and solvent jointly mediate the entropic cost of binding. These features may also help differentiate water patterns driven by distinct thermodynamic processes. For example, a more extensive hydrogen-bonding network involving solvent and protein would more strongly restrict solvent translation than a weaker network, thereby reducing translational entropy [57]. While the correlations remain moderate, the substantial increase in predictive power over protein-only metrics underscores the need to consider the entire protein-water system when analyzing binding thermodynamics.

Discussion

We demonstrate that binding entropy can be predicted from crystallographic ensembles across diverse protein-ligand systems, enabling its quantification in structure-based drug design. We observe that accounting for both protein and solvent entropy is key to predicting this highly variable role in binding. This variability demonstrates that explicit entropy quantification is essential for effective design. Our approach addresses this gap by simultaneously optimizing both contributions using transferable metrics across globular proteins.

While entropy calculations require enumerating the partition function over all conformational states, they are computationally intractable due to high dimensionality and rugged energy landscapes. The crystallographic order parameter approach provides a practical alternative by measuring side-chain angular disorder as a proxy for conformational entropy [25,48]. Drawing on the Lipari-Szabo formalism [25], which relates order parameters to restricted angular motion, we capture local structural heterogeneity that reflects the system's protein entropy. Further, our solvent estimations showed that using metrics based on explicitly placed solvent yielded a large improvement over implicit heat-capacity measurements [58]. Water molecules resolved in crystal structures represent those with sufficient occupancy and order to be detected, likely corresponding to waters that form stable, enthalpically favorable hydrogen bonds with the protein. The loss of water molecules upon ligand binding increases their rotational and translational entropy, contributing favorably to binding entropy. Conversely, the gain in ordered waters might stabilize specific conformational states, coupling solvent organization to protein entropy [51,59,60]. The improved binding-entropy predictions achieved by combining protein conformational entropy with water counts, compared with using either metric alone, indicate that these contributions are at least partially independent. However, we also found that changes in solvent-protein hydrogen-bonding networks contribute to binding entropy variability beyond what protein conformational changes alone explain. This finding suggests that protein and solvent entropy changes may reinforce each other [51,59].

The strong linear correlation with binding entropy validates our approach as an “entropy meter” for crystallography, extending NMR-based methods to the more accessible crystallographic domain [7]. However, several limitations warrant improvement in the future. Current metrics focus on single-residue angular motion rather than whole-side-chain descriptors and do not account for collective motions between residues or between protein and solvent. Improved protein conformational entropy metrics could be more directly grounded in statistical mechanics and could help illuminate correlations between protein and solvent entropy. Further, crystallographic ensemble models currently do not fully capture the conformational ensembles in the diffraction data, even before accounting for crystallographic artifacts. This limitation is particularly acute for water molecules, where resolution-dependent effects obscure the accurate quantification of solvent entropy [25,48,61]. Although our residual regression approach corrects for systematic resolution biases, it assumes uniform impact across all waters, an assumption that may fail if resolution differences correlate with genuine changes in structural heterogeneity rather than purely technical limitations. These challenges underscore the broader need for improvements in ensemble modeling and solvent modeling methods. Finally, our dataset is dominated by closely related ligand series targeting the same proteins, potentially overweighting certain protein families and limiting the chemical diversity explored. Expanding this analysis to include more structurally diverse proteins and ligands would further test the generalizability of our findings.

Making entropy quantifiable from high-resolution crystallographic data enables thermodynamically balanced ligand optimization rather than enthalpy-focused design. We reveal that the dynamic ensemble, captured through multiconformer modeling, provides a practical means to quantify entropic contributions in structure-activity relationship campaigns, integrating this critical yet often overlooked free-energy component into rational drug design workflows. Our findings establish a foundation for entropically informed drug design, bringing the vision of designing ligands with optimized free energy landscapes, rather than enthalpic interactions alone, increasingly achievable.

Methods

Creation of multiconformer ensemble models

All structures were initially downloaded from PDBRedo[62] and were re-modeled using the qFit (version 2025.2)[29,47,63]. qFit structures were then placed through Phenix.refine (version 1.20) [47].

Crystallographic Order Parameters

Crystallographic order parameter values were computed for each residue (excluding glycine and proline) as a proxy for conformational entropy [48]. These values integrate both side-chain and backbone flexibility using two components: (1) the angle of alternative conformers (s_{2angle}), calculated from χ_1 dihedral distributions, and (2) atomic displacement data derived from normalized B-factors of alpha or beta carbons bonded to a hydrogen (s_{2ortho}) [48]. The s_{2ortho} term was corrected for resolution-dependent variation in B-factors using the following normalization:

$$s_{2ortho_{normalized}} = s_{2ortho} * B_{factor_{\alpha\text{phacarbon}}} / 10(\text{resolution})$$

The final s_{2calc} value was then obtained by combining both components multiplicatively:

$$s_{2calc} = s_{2ortho_{normalized}} * s_{2ang}$$

This combined metric captures both thermal motion and conformational variability, offering a robust estimate of per-residue rigidity. Values range from 0 (highly flexible) to 1 (fully rigid). Protein conformational entropy was estimated as $0.25 * \text{number of residues} * \text{average crystallographic order parameter}$.

Solvent Accessible Surface Area

For each atom, we computed the SASA using FreeSasa [64]. We then divided polar versus apolar surface area based on atom type. When calculating differences, we used continuous information

Water Molecule Analysis

We quantified local solvation by counting the number of water molecules within 3.2 Å of each residue in the structure. For each PDB file, all non-hydrogen atoms of protein residues were compared against the oxygen atoms of water molecules, and any water whose oxygen atom lay within 3.2 Å of any heavy atom in a given residue was counted once. This analysis generated residue-level water contact counts, providing a quantitative measure of local hydration and solvent accessibility across the protein surface.

Water Molecule Resolution Correction

To remove systematic bias in water counts arising from differences in crystallographic resolution, we applied a resolution correction to all water count differences using linear residualization. The change in mean water count upon binding (bound minus apo) was modeled as a linear function of

the resolution difference between paired structures. The expected contribution due solely to resolution was estimated by linear regression and subtracted from the observed water count change, yielding a resolution-corrected water count difference for each structure pair.

Hydrogen Bonds

Hydrogen bonds were identified if the donor–acceptor distance ≤ 3.5 Å or hydrogen–acceptor distance ≤ 2.6 Å, and donor–hydrogen–acceptor angle $\geq 120^\circ$. Networks were constructed as undirected graphs with residues as nodes and hydrogen bonds as edges, analyzed both with and without water molecules.

Network metrics included: basic properties (node/edge counts, density, degree distribution), connectivity (component analysis, clustering coefficients, transitivity), path metrics (average shortest path, diameter, radius), and centrality measures (betweenness, closeness, degree, eigenvector). Water-specific metrics quantified water-mediated versus direct protein–protein interactions. Per-residue metrics captured local connectivity and centrality within the hydrogen bonding network. All analyses were performed separately for networks that included and excluded water molecules to distinguish direct and water-mediated interactions.

Residue Packing

Local packing density was quantified for each residue by counting neighboring residues within 5.0 Å using heavy-atom distances (excluding hydrogens and water). For each residue, we computed the number of neighboring residues, contact density (neighbors normalized by the number of heavy atoms in the residue), and average B-factor and occupancy values.

Ligand Properties

Ligand properties were obtained using RDKit via SDF files obtained from PDB structures.

Data

All multiconformer modeled structures are in the Zenodo depository:
<https://zenodo.org/records/18209769>

Code

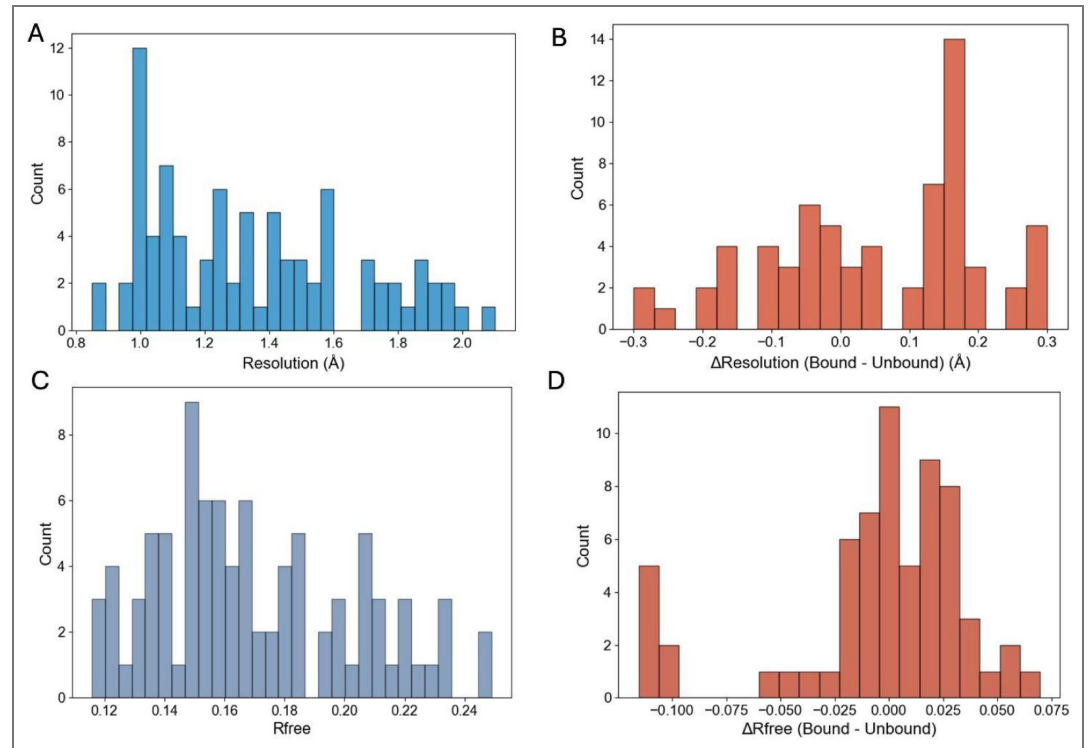
<https://github.com/ExcitedStates/qfit-3.0>

https://github.com/Wankowicz-Lab/ensemble_bioinformatic_toolkit

Data availability

Data availability is listed in [Supplementary Table 1](#). PDBs and thermodynamic data were all previously published.

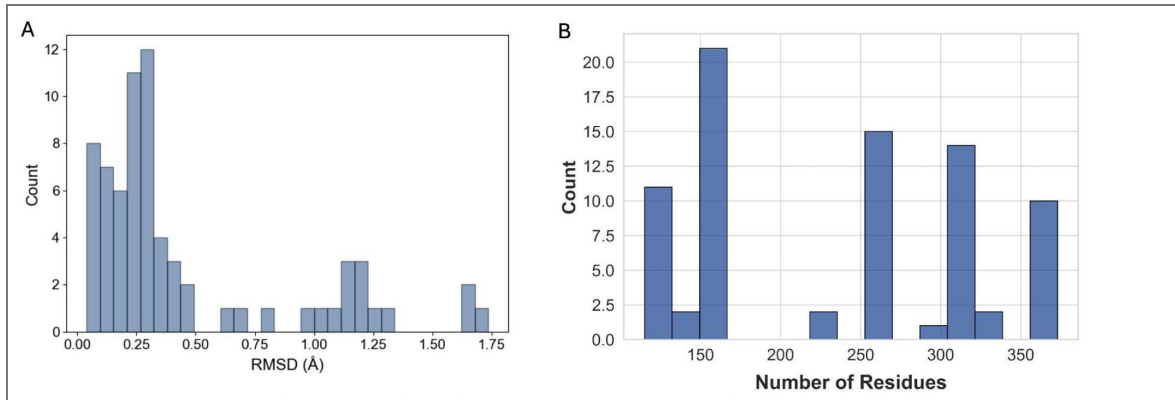
Supplementary Figures



Supplementary Figure 1. **A.** Resolution distribution across all PDBs (Median: 1.30 [IQR: 1.08-1.60]). **B.** Resolution differences between paired PDBs (Median: 0.06 [IQR: -0.06 - 0.15]). **C.** Rfree distribution across all PDBs (Median: 0.16 [IQR: 0.15-0.19]). **D.** Rfree differences between paired PDBs (Median: 0.010 [IQR: -0.010-0.027]).

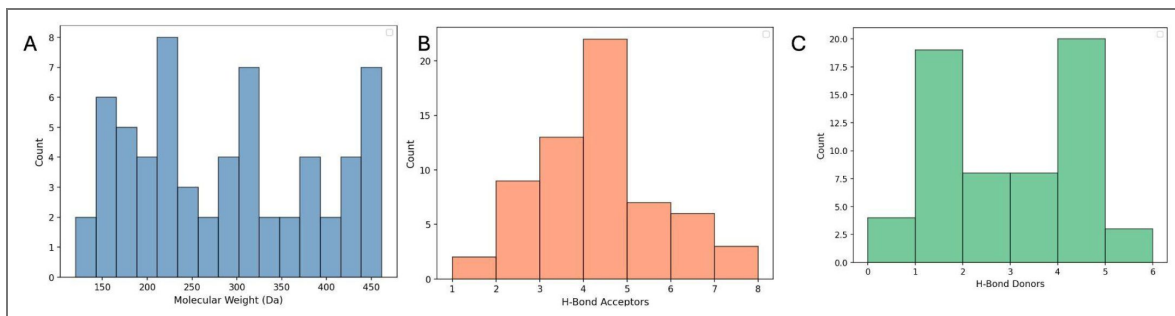
Supplementary Figure 2.

A. Distribution of protein size across the dataset by number of residue numbers. **B.** Alpha carbon RMSD between bound and unbound PDB pairs (Median: 0.27 [IQR: 0.19-0.46]).



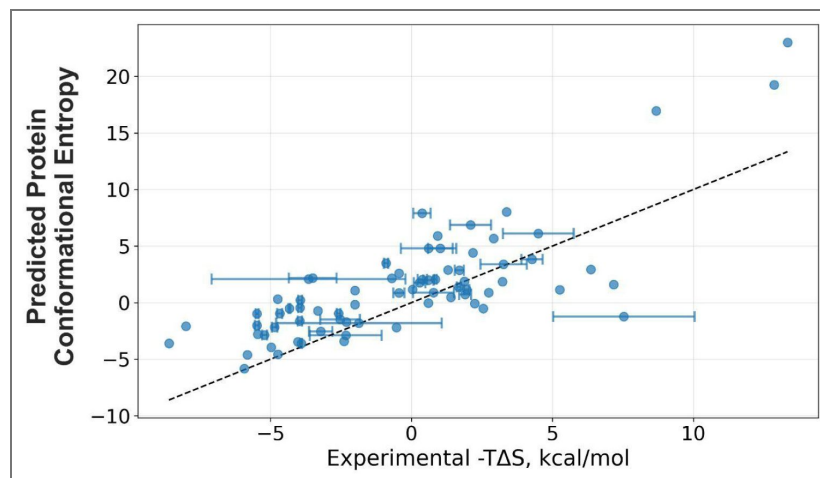
Supplementary Figure 3. Ligand information

A. Molecular weight distribution across the dataset (Median: 284.68[102.46-461.50]) **B.** Hydrogen Bond Acceptor (Median: 4 [1-7]) **C.** Hydrogen Bond Donor (Median: 2 [0-5])



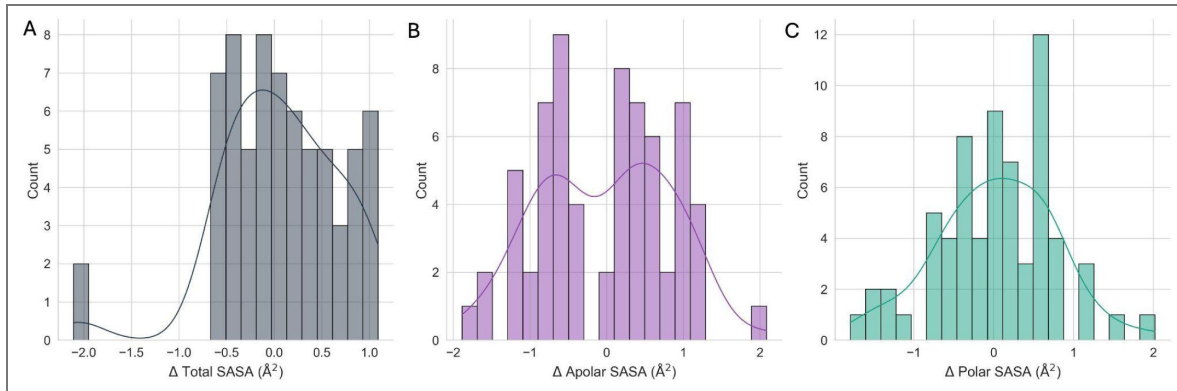
Supplementary Figure 4. Correlation between ITC-measured binding entropy and protein conformational entropy metrics.

Each point represents the mean binding entropy from ITC experiments, with error bars indicating the range of all measurements. Linear regression parameters: slope = 0.88, intercept = 1.63, $R^2 = 0.61$.



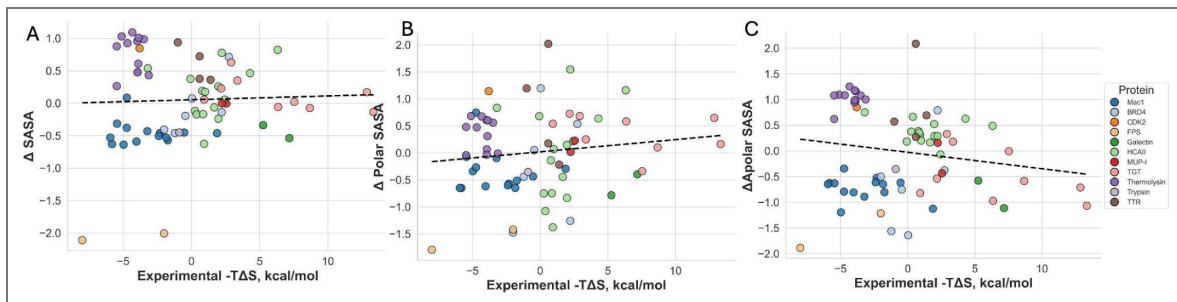
Supplementary Figure 5. Distribution of change in SASA values across pairs

A. Total SASA changes (Median: 0.06, Mean: 0.05, Std: 0.62); B. Apolar SASA changes (Median: 0.16, Mean:-0.03, Std: 0.84); C. Polar SASA changes (Median: 0.07, Mean: 0.03, Std: 0.74)



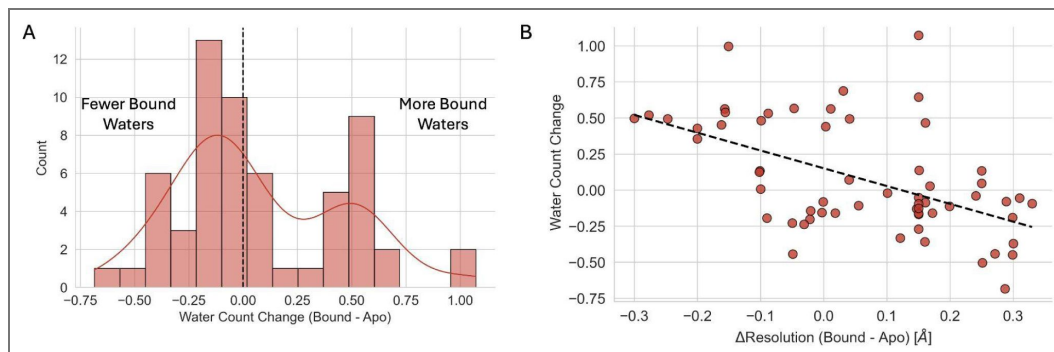
Supplementary Figure 6. Correlation of SASA changes with binding entropy

A. Total Δ SASA: $R^2=0.0058$; B. apolar Δ SASA: $R^2=0.0174$; C. polar Δ SASA: $R^2=0.0027$.



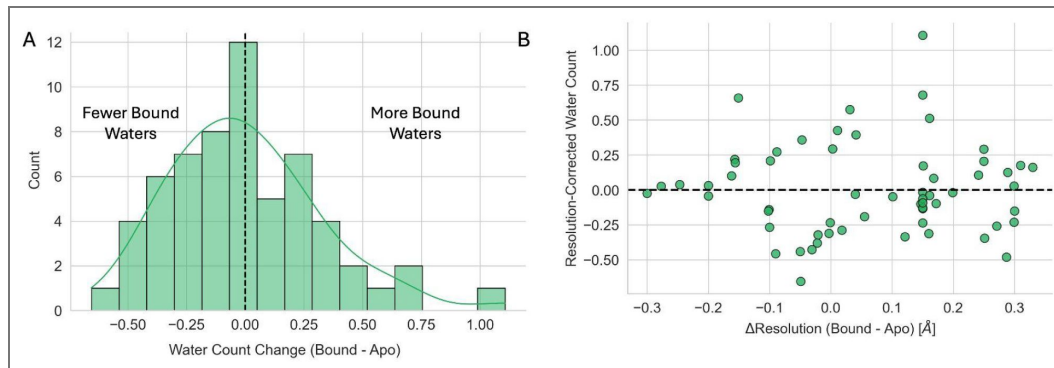
Supplementary Figure 7.

A. Distribution of average number of waters gained or lost between bound and unbound. B. Correlation between resolution differences and mean water count change per residue ($R^2=0.53$, Slope=-1.232).



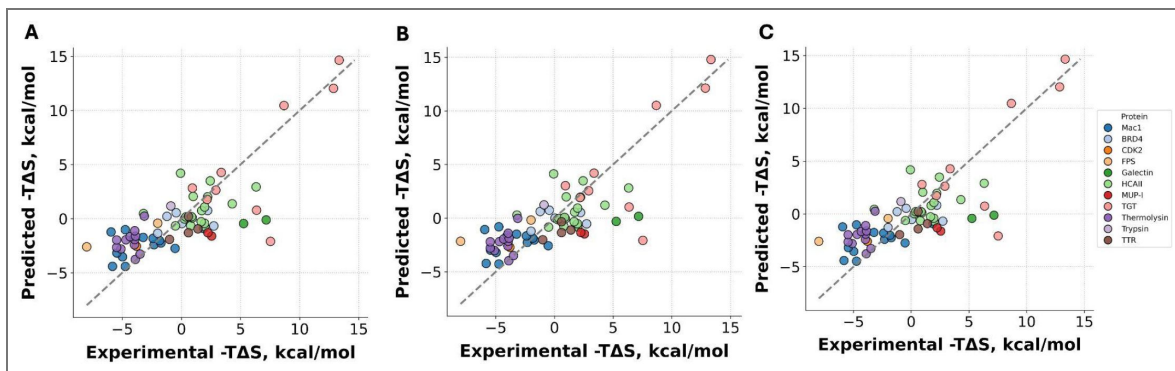
Supplementary Figure 8.

A. Distribution of average number of waters gained or lost between bound and unbound corrected for resolution differences between bound and apo. B. Correlation between resolution differences and mean water count change per residue ($R^2=0$, Slope=0).

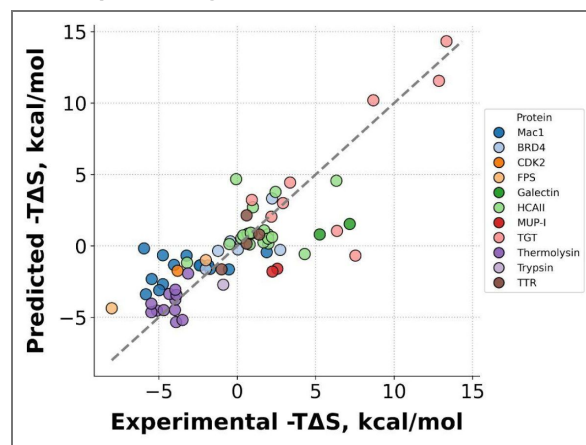


Supplementary Figure 9. Contribution of solvent-accessible surface area components along with protein conformational entropy to experimental binding entropy.

A. Model including total SASA ($R^2 = 0.588$). B. Model including apolar SASA ($R^2 = 0.589$). C. Model including polar SASA ($R^2 = 0.587$).

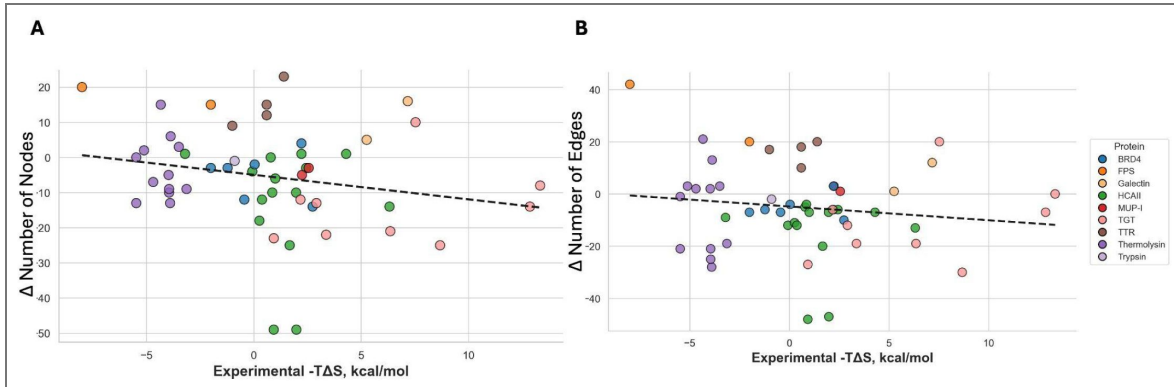


Supplementary Figure 10. Linear model relating non-corrected protein conformational entropy and non-corrected water count to the entropic binding ($R^2 = 0.681$)



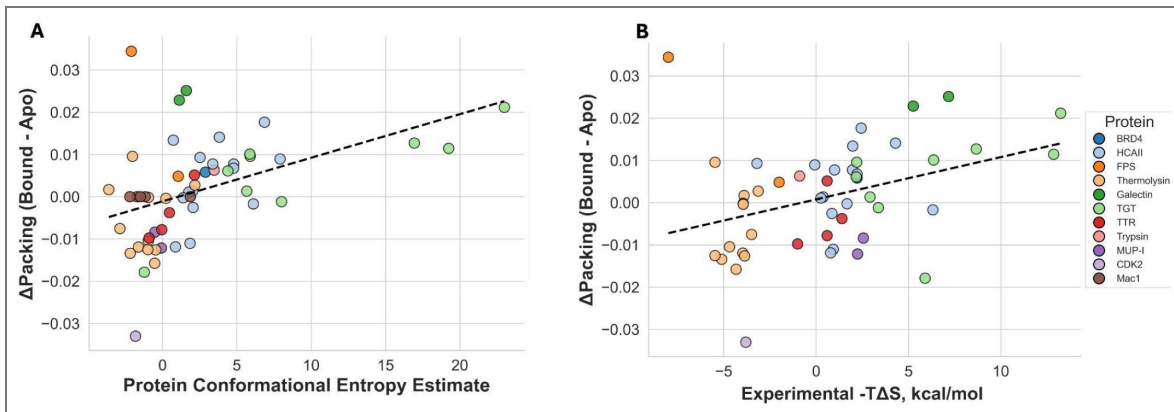
Supplementary Figure 11. Hydrogen bond protein only

A. Correlation between the number of nodes and experimental entropy measurement ($R^2=0.0734$) **B.** Correlation between the number of edges and experimental entropy measurement ($R^2=0.0489$)



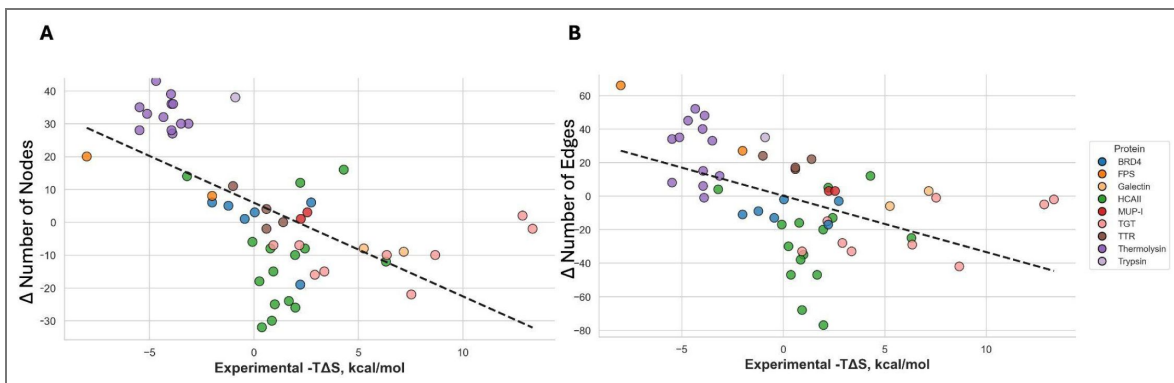
Supplementary Figure 12. Packing and entropy measurement

A. Correlation between the difference in packing and protein conformational entropy estimates ($R^2=0.1976$). **B.** Correlation between average number of waters gained or lost between bound and unbound with protein structural entropy estimates ($R^2=0.1318$).



Supplementary Figure 13. Hydrogen bond protein and solvent

A. Correlation between the number of nodes and experimental entropy measurement ($R^2=0.3765$) **B.** Correlation between the number of edges and experimental entropy measurement ($R^2=0.2423$)



Additional files

[Supplementary Table 1](#) 

[Supplementary Table 2](#) 

[Supplementary Table 3](#) 

[Supplementary Table 4](#) 

Additional information

Funding

| Funder | Grant reference number |
|-------------------------|------------------------|
| American Cancer Society | |

Author ORCID iDs

Stephanie A Wankowicz:  <https://orcid.org/0000-0002-4225-7459>

References

1. Bissantz C, Kuhn B, Stahl M (2010) A medicinal chemist's guide to molecular interactions. *J Med Chem* **53**:5061-5084 <https://doi.org/10.1021/jm100112j> | PubMed
2. Wankowicz SA, Fraser JS (2025) Advances in uncovering the mechanisms of macromolecular conformational entropy. *Nat Chem Biol* **21**:623-634 <https://doi.org/10.1038/s41589-025-01879-3> | PubMed
3. Ladbury JE, Klebe G, Freire E (2010) Adding calorimetric data to decision making in lead discovery: a hot tip. *Nat Rev Drug Discov* **9**:23-27 <https://doi.org/10.1038/nrd3054> | PubMed
4. Verteramo ML, Stenström O, Ignjatović MM, Caldararu O, Olsson MA, Manzoni F, et al. (2019) Interplay between Conformational Entropy and Solvation Entropy in Protein-Ligand Binding. *J Am Chem Soc* **141**:2012-2026 <https://doi.org/10.1021/jacs.8b11099> | PubMed
5. Olsson TSG, Williams MA, Pitt WR, Ladbury JE (2008) The thermodynamics of protein-ligand interaction and solvation: insights for ligand design. *J Mol Biol* **384**:1002-1017 <https://doi.org/10.1016/j.jmb.2008.09.073> | PubMed
6. Tzeng S-R, Kalodimos CG (2012) Protein activity regulation by conformational entropy. *Nature* **488**:236-240 <https://doi.org/10.1038/nature11271> | PubMed
7. Caro JA, Harpole KW, Kasinath V, Lim J, Granja J, Valentine KG, et al. (2017) Entropy in molecular recognition by proteins. *Proc Natl Acad Sci U S A* **114**:6563-6568 <https://doi.org/10.1073/pnas.1621154114> | PubMed
8. Seo L, Farran I, Aslam A, Li X, Jaishankar P, Fraser JS, et al. (2025) Crystallographic Ensembles Reveal the Structural Basis of Binding Entropy in SARS-CoV2 Macrodomein. *bioRxiv* <https://doi.org/10.1101/2025.11.25.690589> | PubMed
9. Diehl C, Engström O, Delaine T, Håkansson M, Genheden S, Modig K, et al. (2010) Protein flexibility and conformational entropy in ligand design targeting the carbohydrate recognition domain of galectin-3. *J Am Chem Soc* **132**:14577-14589 <https://doi.org/10.1021/ja105852y> | PubMed
10. Capdevila DA, Edmonds KA, Campanello GC, Wu H, Gonzalez-Gutierrez G, Giedroc DP (2018) Functional Role of Solvent Entropy and Conformational Entropy of Metal Binding in a Dynamically Driven Allosteric System. *J Am Chem Soc* **140**:9108-9119 <https://doi.org/10.1021/jacs.8b02129> | PubMed
11. Bonin JP, Sapienza PJ, Lee AL (2022) Dynamic allostery in substrate binding by human thymidylate synthase. *eLife* **11** <https://doi.org/10.7554/eLife.79915> | PubMed

12. Doig AJ, Sternberg MJ (1995) Side-chain conformational entropy in protein folding. *Protein Sci* **4**:2247-2251 <https://doi.org/10.1002/pro.5560041101> | PubMed
13. Bromberg S, Dill KA (1994) Side-chain entropy and packing in proteins. *Protein Sci* **3**:997-1009 <https://doi.org/10.1002/pro.5560030702> | PubMed
14. Akke M, Brueschweiler R, Palmer AG III (1993) NMR order parameters and free energy: an analytical approach and its application to cooperative calcium(2+) binding by calbindin D9k. *J Am Chem Soc* **115**:9832-9833 <https://doi.org/10.1021/ja00074a073>
15. Lee AL, Wand AJ (2001) Microscopic origins of entropy, heat capacity and the glass transition in proteins. *Nature* **411**:501-504 <https://doi.org/10.1038/35078119> | PubMed
16. Gavrillov Y, Kümmerer F, Orioli S, Prestel A, Lindorff-Larsen K, Teilum K (2022) Double Mutant of Chymotrypsin Inhibitor 2 Stabilized through Increased Conformational Entropy. *Biochemistry* **61**:160-170 <https://doi.org/10.1021/acs.biochem.1c00749> | PubMed
17. Capdevila DA, Braymer JJ, Edmonds KA, Wu H, Giedroc DP (2017) Entropy redistribution controls allostery in a metalloregulatory protein. *Proc Natl Acad Sci U S A* **114**:4424-4429 <https://doi.org/10.1073/pnas.1620665114> | PubMed
18. Lee AL, Kinnear SA, Wand AJ (2000) Redistribution and loss of side chain entropy upon formation of a calmodulin-peptide complex. *Nat Struct Biol* **7**:72-77 <https://doi.org/10.1038/71280> | PubMed
19. Jin M, Seed RI, Cai G, Shing T, Wang L, Ito S, et al. (2024) Dynamic allostery drives autocrine and paracrine TGF- β signaling. *Cell* **187**:6200-6219. <https://doi.org/10.1016/j.cell.2024.08.036> | PubMed
20. Stone MJ (2001) NMR relaxation studies of the role of conformational entropy in protein stability and ligand binding. *Acc Chem Res* **34**:379-388 <https://doi.org/10.1021/ar000079c> | PubMed
21. Genheden S, Ryde U (2012) Will molecular dynamics simulations of proteins ever reach equilibrium?. *Phys Chem Chem Phys* **14**:8662-8677 <https://doi.org/10.1039/c2cp23961b> | PubMed
22. Zhou H-X, Gilson MK (2009) Theory of free energy and entropy in noncovalent binding. *Chem Rev* **109**:4092-4107 <https://doi.org/10.1021/cr800551w> | PubMed
23. Chang C-E, Chen W, Gilson MK (2005) Evaluating the Accuracy of the Quasiharmonic Approximation. *J Chem Theory Comput* **1**:1017-1028 <https://doi.org/10.1021/ct0500904> | PubMed
24. Hikiri S, Yoshidome T, Ikeguchi M (2016) Computational Methods for Configurational Entropy Using Internal and Cartesian Coordinates. *J Chem Theory Comput* **12**:5990-6000 <https://doi.org/10.1021/acs.jctc.6b00563> | PubMed
25. Lipari G, Szabo A (1981) Nuclear magnetic resonance relaxation in nucleic acid fragments: models for internal motion. *Biochemistry* **20**:6250-6256 <https://doi.org/10.1021/bi00524a053> | PubMed
26. Baxa MC, Haddadian EJ, Jumper JM, Freed KF, Sosnick TR (2014) Loss of conformational entropy in protein folding calculated using realistic ensembles and its implications for NMR-based calculations. *Proc Natl Acad Sci U S A* **111**:15396-15401 <https://doi.org/10.1073/pnas.1407768111> | PubMed
27. Yang D, Kay LE (1996) Contributions to conformational entropy arising from bond vector fluctuations measured from NMR-derived order parameters: application to protein folding. *J Mol Biol* **263**:369-382 <https://doi.org/10.1006/jmbi.1996.0581> | PubMed
28. Kuzmanic A, Pannu NS, Zagrovic B (2014) X-ray refinement significantly underestimates the level of microscopic heterogeneity in biomolecular crystals. *Nat Commun* **5**:3220 <https://doi.org/10.1038/ncomms4220> | PubMed
29. Wankowicz SA, Ravikumar A, Sharma S, Riley B, Raju A, Hogan DW, et al. (2024) Automated multiconformer model building for X-ray crystallography and cryo-EM. *eLife* **12** <https://doi.org/10.7554/eLife.90606> | PubMed
30. Frederick KK, Marlow MS, Valentine KG, Wand AJ (2007) Conformational entropy in molecular recognition by proteins. *Nature* **448**:325-329 <https://doi.org/10.1038/nature05959> | PubMed

31. Marlow MS, Dogan J, Frederick KK, Valentine KG, Wand AJ (2010) The role of conformational entropy in molecular recognition by calmodulin. *Nat Chem Biol* **6**:352-358 <https://doi.org/10.1038/nchembio.347> | PubMed
32. Krimmer SG, Betz M, Heine A, Klebe G (2014) Methyl, ethyl, propyl, butyl: futile but not for water, as the correlation of structure and thermodynamic signature shows in a congeneric series of thermolysin inhibitors. *ChemMedChem* **9**:833-846 <https://doi.org/10.1002/cmdc.201400013> | PubMed
33. Glöckner S, Ngo K, Sager CP, Hüfner-Wulsdorf T, Heine A, Klebe G (2020) Conformational Changes in Alkyl Chains Determine the Thermodynamic and Kinetic Binding Profiles of Carbonic Anhydrase Inhibitors. *ACS Chem Biol* **15**:675-685 <https://doi.org/10.1021/acscchembio.9b00895> | PubMed
34. Yokoyama T, Matsumoto K, Ostermann A, Schrader TE, Nabeshima Y, Mizuguchi M (2019) Structural and thermodynamic characterization of the binding of isoliquiritigenin to the first bromodomain of BRD4. *Febs J* **286**:1656-1667 <https://doi.org/10.1111/febs.14736> | PubMed
35. Neeb M, Hohn C, Ehrmann FR, Härtsch A, Heine A, Diederich F, et al. (2016) Occupying a flat subpocket in a tRNA-modifying enzyme with ordered or disordered side chains: Favorable or unfavorable for binding?. *Bioorg Med Chem* **24**:4900-4910 <https://doi.org/10.1016/j.bmc.2016.07.053> | PubMed
36. Wood DJ, Korolchuk S, Tatum NJ, Wang L-Z, Endicott JA, Noble MEM, et al. (2019) Differences in the Conformational Energy Landscape of CDK1 and CDK2 Suggest a Mechanism for Achieving Selective CDK Inhibition. *Cell Chem Biol* **26**:121-130. <https://doi.org/10.1016/j.chembiol.2018.10.015> | PubMed
37. Krimmer SG, Cramer J, Betz M, Fridh V, Karlsson R, Heine A, et al. (2016) Rational Design of Thermodynamic and Kinetic Binding Profiles by Optimizing Surface Water Networks Coating Protein-Bound Ligands. *J Med Chem* **59**:10530-10548 <https://doi.org/10.1021/acs.jmedchem.6b00998> | PubMed
38. Kawasaki Y, Sekiguchi M, Kawasaki M, Hirakura Y (2014) Thermodynamic evaluation of the binding of bisphosphonates to human farnesyl pyrophosphate synthase. *Chem Pharm Bull (Tokyo)* **62**:77-83 <https://doi.org/10.1248/cpb.c13-00710> | PubMed
39. Iakovleva I, Brännström K, Nilsson L, Gharibyan AL, Begum A, Anan I, et al. (2015) Enthalpic Forces Correlate with the Selectivity of Transthyretin-Stabilizing Ligands in Human Plasma. *J Med Chem* **58**:6507-6515 <https://doi.org/10.1021/acs.jmedchem.5b00544> | PubMed
40. Szalai TV, Bajusz D, Börzsei R, Zsidó BZ, Ilaš J, Ferenczy GG, et al. (2024) Effect of Water Networks On Ligand Binding: Computational Predictions vs Experiments. *J Chem Inf Model* **64**:8980-8998 <https://doi.org/10.1021/acs.jcim.4c01291> | PubMed
41. Snyder PW, Mecinovic J, Moustakas DT, Thomas SW, Harder M, Mack ET, Lockett MR, Héroux A, Sherman W, Whitesides GM (2011) Mechanism of the hydrophobic effect in the biomolecular recognition of arylsulfonamides by carbonic anhydrase. *Proc Natl Acad Sci U S A* **108**:17889-17894 <https://doi.org/10.1073/pnas.1114107108> | PubMed
42. Velazquez-Campoy A, Todd MJ, Freire E (2000) HIV-1 protease inhibitors: enthalpic versus entropic optimization of the binding affinity. *Biochemistry* **39**:2201-2207 <https://doi.org/10.1021/bi992399d> | PubMed
43. Todd MJ, Freire E (1999) The effect of inhibitor binding on the structural stability and cooperativity of the HIV-1 protease. *Proteins* **36**:147-156 [https://doi.org/10.1002/\(sici\)1097-0134\(19990801\)36:2<147::aid-prot2>3.0.co;2-3](https://doi.org/10.1002/(sici)1097-0134(19990801)36:2<147::aid-prot2>3.0.co;2-3) | PubMed
44. Dullweber F, Stubbs MT, Musil D, Stürzebecher J, Klebe G (2001) Factorising ligand affinity: a combined thermodynamic and crystallographic study of trypsin and thrombin inhibition. *J Mol Biol* **313**:593-614 <https://doi.org/10.1006/jmbi.2001.5062> | PubMed
45. Czodrowski P, Sottriffer CA, Klebe G (2007) Protonation changes upon ligand binding to trypsin and thrombin: structural interpretation based on pK(a) calculations and ITC experiments. *J Mol Biol* **367**:1347-1356 <https://doi.org/10.1016/j.jmb.2007.01.022> | PubMed

46. Riley BT, Wankowicz SA, de Oliveira SHP, van Zundert GCP, Hogan DW, Fraser JS, et al. (2021) qFit 3: Protein and ligand multiconformer modeling for X-ray crystallographic and single-particle cryo-EM density maps. *Protein Sci* **30**:270-285 <https://doi.org/10.1002/pro.4001> | PubMed
47. Afonine PV, Grosse-Kunstleve RW, Echols N, Headd JJ, Moriarty NW, Mustyakimov M, et al. (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr* **68**:352-367 <https://doi.org/10.1107/S0907444912001308> | PubMed
48. Fenwick RB, van den Bedem H, Fraser JS, Wright PE (2014) Integrated description of protein dynamics from room-temperature X-ray crystallography and NMR. *Proc Natl Acad Sci U S A* **111**:E445-54 <https://doi.org/10.1073/pnas.1323440111> | PubMed
49. Neeb M, Betz M, Heine A, Barandun LJ, Hohn C, Diederich F, et al. (2014) Beyond affinity: enthalpy-entropy factorization unravels complexity of a flat structure-activity relationship for inhibition of a tRNA-modifying enzyme. *J Med Chem* **57**:5566-5578 <https://doi.org/10.1021/jm5006868> | PubMed
50. Fox JM, Zhao M, Fink MJ, Kang K, Whitesides GM (2018) The Molecular Origin of Enthalpy/Entropy Compensation in Biomolecular Recognition. *Annu Rev Biophys* **47**:223-250 <https://doi.org/10.1146/annurev-biophys-070816-033743> | PubMed
51. Breiten B, Lockett MR, Sherman W, Fujita S, Al-Sayah M, Lange H, et al. (2013) Water networks contribute to enthalpy/entropy compensation in protein-ligand binding. *J Am Chem Soc* **135**:15579-15584 <https://doi.org/10.1021/ja4075776> | PubMed
52. Wallerstein J, Ekberg V, Ignjatović MM, Kumar R, Caldararu O, Peterson K, et al. (2021) Entropy-Entropy Compensation between the Protein, Ligand, and Solvent Degrees of Freedom Fine-Tunes Affinity in Ligand Binding to Galectin-3C. *JACS Au* **1**:484-500 <https://doi.org/10.1021/jacsau.0c00094> | PubMed
53. Prabhu N, Sharp K (2006) Protein-solvent interactions. *Chem Rev* **106**:1616-1623 <https://doi.org/10.1021/cr040437f> | PubMed
54. Hilser VJ, Wrabl JO, Millard CEF, Schmitz A, Brantley SJ, Pearce M, et al. (2025) Statistical Thermodynamics of the Protein Ensemble: Mediating Function and Evolution. *Annu Rev Biophys* **54**:227-247 <https://doi.org/10.1146/annurev-biophys-061824-104900> | PubMed
55. Wang L, Berne BJ, Friesner RA (2011) Ligand binding to protein-binding pockets with wet and dry regions. *Proc Natl Acad Sci U S A* **108**:1326-1330 <https://doi.org/10.1073/pnas.1016793108> | PubMed
56. Ladbury JE (1996) Just add water! The effect of water on the specificity of protein-ligand binding sites and its potential application to drug design. *Chem Biol* **3**:973-980 [https://doi.org/10.1016/s1074-5521\(96\)90164-7](https://doi.org/10.1016/s1074-5521(96)90164-7) | PubMed
57. Harano Y, Kinoshita M (2005) Translational-entropy gain of solvent upon protein folding. *Biophys J* **89**:2701-2710 <https://doi.org/10.1529/biophysj.104.057604> | PubMed
58. Prabhu NV, Sharp KA (2005) Heat capacity in proteins. *Annu Rev Phys Chem* **56**:521-548 <https://doi.org/10.1146/annurev.physchem.56.092503.141202> | PubMed
59. Fenimore PW, Frauenfelder H, McMahon BH, Parak FG (2002) Slaving: solvent fluctuations dominate protein dynamics and functions. *Proc Natl Acad Sci U S A* **99**:16047-16051 <https://doi.org/10.1073/pnas.212637899> | PubMed
60. Dahanayake JN, Mitchell-Koch KR (2018) Entropy connects water structure and dynamics in protein hydration layer. *Phys Chem Chem Phys* **20**:14765-14777 <https://doi.org/10.1039/c8cp01674g> | PubMed
61. Carugo O, Bordo D (1999) How many water molecules can be detected by protein crystallography?. *Acta Crystallogr D Biol Crystallogr* **55**:479-483 <https://doi.org/10.1107/s0907444998012086> | PubMed
62. Joosten RP, Salzemann J, Bloch V, Stockinger H, Berglund A-C, Blanchet C, et al. (2009) PDB_REDO: automated re-refinement of X-ray structure models in the PDB. *J Appl Crystallogr* **42**:376-384 <https://doi.org/10.1107/s0021889809008784> | PubMed
63. Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**:2126-2132 <https://doi.org/10.1107/s0907444904019158> | PubMed

64. Mitternacht S (2016) FreeSASA: An open source C library for solvent accessible surface area calculations. *F1000Res* 5:189 <https://doi.org/10.12688/f1000research.7931.1> | PubMed

Peer reviews

Reviewer #1 (Public review):

Summary:

The authors show that if they generate a weighted multi-conformer ensemble of structural models to fit crystallographic electron density data, the application of statistical mechanical methodologies to that ensemble can provide reasonable estimates of configurational entropy terms related to protein-ligand binding.

Strengths:

A fair range of proteins (12) and ligands (70) is included in the study. The analytical methodologies are well described. Both successful and less successful analytical approaches are discussed, and the latter are frequently as insightful as the former.

Weaknesses:

Compared to the universe of protein-ligand complexes, this dataset is inevitably very limited, so the generality of the observations made here remains speculative. Though a fair range of proteins is studied, the dynamic range in the binding affinity data is limited. The practical utility of the approach is never really commented on.

<https://doi.org/10.7554/eLife.111298.1.sa1>

Reviewer #2 (Public review):

The manuscript by Miller and Wankowicz (M&W) develops a crystallographic approach to predict the contribution of protein conformational entropy to the total binding entropy using multi-conformer ensemble models. The approach loosely follows the path developed by Wand using NMR relaxation methods. Their approach is to generate local crystallographic order parameters (analogous to NMR order parameters) to estimate protein conformational entropy and then combine this with statements about water entropy. The static view of the ensemble is perhaps easier to grasp, with respect to entropy, than the NMR-based dynamical view. This approach is potentially ground-breaking and of great importance given the ease, relative to NMR, with which the source data can be obtained. However, the approach has several deficiencies, only some of which are noted by the authors.

Like the initial Wand approach (Frederick et al *Nature*, 2007), M&W develop a simple counting relationship between members of the ensemble and a statement about conformational entropy. For reasons that are not clear, M&W utilize "per residue" scaling, which was initially introduced by Wand but later discarded for the more physically meaningful "per torsion angle" scaling. As noted in the *Nature* 2007 paper, this assumes uncorrelated occupancy. The current Wand approach (Caro et al *PNAS*, 2017) subsumes correlated occupancy and potentially incomplete sampling of the ensemble into an empirically determined scaling parameter (sd). This is likely a major contributor to the mysterious 1/4 scaling factor that is introduced. It is not clear to me how discrete conformational states are counted from the qFit models. Using the B-factor, as opposed to a thermal factor, to account for motion in a rotamer well seems suspect. With some irony, M&W only look at chi-1 rotamers in distinct contrast to the NMR approach, which looks at the end of the side chain, which captures the entire disorder. On the other hand, the crystallographic approach "sees" all side chains, whereas the NMR approach, as currently

rendered, looks only at methyl-bearing side chains and requires coupling to neighbors to report on all side chains (see Kasinath JACS 2013 and Wand & Sharp ARB 2018).

Nevertheless, as noted by Nature 2007, the fact that a linear relationship is seen between the apparent conformational entropy and total binding entropy suggests that the former is a major component of the latter. It also reinforces the idea that dS_{rt} is constant for higher affinity complexes, i.e., residual rigid-body motion of protein relative to ligand is limited (a conclusion reached in PNAS 2017) but not mentioned. This is an important result.

The classic hydrophobic effect is potentially a significant component of total binding entropy. Here, the manuscript falls flat by focusing on crystallographically resolved waters. As shown in site-resolved detail (Nucci et al, NSMB 2011 and others), hydration water has a range of residual motion (entropy) that will modulate contributions to water entropy upon displacement from an interface. A very clear example of the potential for large contributions was demonstrated in the wet interface of a barnase-DNA complex (PNAS 2017). The fact that the classic dASA treatment failed, I think, points to problems elsewhere in the approach.

I note that the range of ligand types explored by M&W is quite limited as compared to PNAS 2017, making generalization somewhat difficult (see Wand Cur. Opin. Struct. Biol, 2013 for why this is important). Finally, it is disappointing that the authors chose not to examine systems common to PNAS 2017, making direct comparison to the NMR method impossible.

In summary, this manuscript sets the field in a new direction. It is a first serious look at conformational entropy using crystallographic approaches. If fully validated, this approach would permit an explosion of insight since the crystallography is now straightforward, very fast and capable of approaching larger systems, relative to the NMR approach. However, there are missing quantitative elements represented by a formal relationship that is fitted by the data. I do not think this is a fatal flaw for this manuscript, however. If the supplementary material is improved for clarity and completeness (e.g, include tables of thermodynamic data; conformer analysis; B-factors) such that all figures could be independently reproduced and therefore analyzed in different ways, and the comments made above are addressed, if not resolved, then I think this manuscript could become a keystone for this new direction.

<https://doi.org/10.7554/eLife.111298.1.sa0>