

RESEARCH ARTICLE

Total Free Energy Analysis of Fully Hydrated Proteins

Jas Kalayan*¹ | Arghya Chakravorty² | Jim Warwicker³ | Richard H. Henchman⁴

¹Division of Pharmacy and Optometry,
Faculty of Biology, Medicine and Health,
University of Manchester, Manchester, UK

²Department of Chemistry and Biophysics,
University of Michigan, Ann Arbor, Ann
Arbor, Michigan, USA

³Manchester Institute of Biotechnology and
School of Biological Sciences, Faculty of
Biology, Medicine and Health, University
of Manchester, Manchester, UK

⁴Sydney Medical School, Faculty of
Medicine and Health, University of Sydney,
Sydney, Australia

Correspondence

*Jas Kalayan, Division of Pharmacy and
Optometry, University of Manchester,
Manchester, UK. Email:
jas.kalayan@manchester.ac.uk

Abstract

The total free energy of a hydrated biomolecule and its corresponding decomposition of energy and entropy provides detailed information about regions of thermodynamic stability or instability. The free **energies of four** hydrated globular **proteins with different net charges are** calculated from a molecular dynamics simulation, with the energy coming from the system Hamiltonian and entropy using multiscale cell correlation. Water is found to be most stable around anionic residues, intermediate around cationic and polar residues, and least stable near hydrophobic residues, especially when more buried, with stability displaying moderate entropy-enthalpy compensation. Conversely, anionic residues in the proteins are energetically destabilised relative to singly solvated amino acids, while trends for other residues are less clear-cut. Almost all residues lose intra-residue entropy when in the protein, **enthalpy changes are negative on average but may be positive or negative, and the resulting overall stability is moderate for some proteins and negligible for others.** The free energy of water around single amino acids is found to closely match existing hydrophobicity scales. **Regarding the effect of secondary structure, water is slightly more stable around loops, of intermediate stability around β strands and turns, and least stable around helices.** An interesting asymmetry observed is that cationic residues stabilise a residue when bonded to its N-terminal side but destabilise it when on the C-terminal side, with a weaker reversed trend for anionic residues.

KEYWORDS:

free energy, entropy, statistical mechanics, protein hydration

1 | INTRODUCTION

The structure and stability of proteins are controlled not only by their sequence but also by their solution environment^{1,2,3,4,5,6} **comprising** salts, solutes, other biomolecules and numerous water molecules. The influence is mutual, with the properties of these molecules in turn affected by the solvated protein. To understand biomolecular stability and related properties such as folding, binding and function, it is helpful to be able to quantify and characterise stability in terms of all the constituent molecules

~~This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/prot.26411~~

of the system, solvent and solute molecules alike. Stability of a protein with respect to another reference state can be measured experimentally in terms of an equilibrium constant, but this only works if a convenient reference state exists. The absolute stability of a system has no such requirement, depends directly and exclusively on the system itself, and is quantified by the free energy, with lower free energy meaning greater stability. While free energy cannot be measured experimentally, it can be computed from a molecular dynamics (MD) simulation, either from the free energy itself or from energy minus entropy times temperature^{7,8,9,10,11,12}, which can be referred to as energy-entropy (EE) methods. Energy relates to the strength of molecular interactions and entropy relates to molecular flexibility or the probability distribution of configurations.

To calculate total system entropy and energy for biomolecular systems, all-atom, force-field-based simulations are the most useful because they extensively sample the ensemble of configurations over all atoms in the system, unlike coarser-grain methods, which miss important atomic interactions and degrees of freedom, or electronic-structure methods, which are much more expensive and still too slow to achieve sufficient sampling. Theoretical approaches to calculate free energy are available for the solvent, such as Poisson-Boltzmann (PB) or Generalised-Born (GB) continuum solvent methods^{13,14} or the three-dimensional reference interaction site model (3D-RISM), which uses an approximation to the Ornstein-Zernike equation solved on a 3D grid around the protein.^{15,16,17} While these methods avoid the need for ensemble averages over many solvent configurations, they do not yield explicit solvent energy and entropy and nor do they provide the contribution from the protein. The free-energy method of Meirovitch and coworkers can be applied to both solvent and protein alike,^{18,19,20} but requires a specific Monte Carlo simulation to grow in the molecules as opposed to a standard MD simulation.

Regarding EE methods, energy can be calculated in a straightforward manner from the average system Hamiltonian in a molecular dynamics simulation. Calculating the entropy is more difficult because it requires determining the probability distribution of quantum states of the whole system, and it is a particular challenge for heterogeneous systems comprising both large proteins and large ensembles of small molecules. Of the methods to calculate the entropy of a protein,^{21,22,23,24,25,26} two of the most popular are the multidimensional harmonic approximation, namely normal mode analysis (NMA)^{27,28,29} based on the curvature at the energy minimum and quasi-harmonic analysis (QHA)^{30,31,32,33} based on coordinate covariances in MD simulations. Methods using non-harmonic probability distributions may be more accurate but their correlations are limited to low dimensionality and fail to account for the quantum nature of high-frequency vibrations in covalently bonded systems. Individual dihedrals^{34,35,36} omit non-negligible correlations, which can be included either by using distributions along eigenvectors of dihedral covariance,^{37,38,39,40} mutual information expansions (MIE)^{41,42} and the minimum spanning tree (MIST) variant.^{43,44} However, none of these protein-entropy methods are applicable to water because they do not account for translational and rotational entropy which is small for a protein but the main component for collections of small molecules. The solvent contribution can be included using hybrid methods such as MM-PBSA and MM-GBSA^{13,14} which combine molecular mechanics (MM) energy with either a PB or GB solvent free energy, a non-polar surface-area (SA) term, and either NMA or QHA entropy.

Of the entropy methods that are applicable to water around proteins,^{45,46,47,48,49,50,51,52,53} a widely used method is inhomogeneous solvation theory (IST)^{54,55,56,45,57,58,59,60,50,61} based on the solute-solvent density distribution, as well as the closely related PerlMut method^{62,63}. The two-phase thermodynamics (2PT) method, which is based on the velocity autocorrelation function, can be used to calculate water entropy in a wide range of hydration environments.^{64,65,66,67,68,69,70,71,48,49,52} The effect of protein flexibility on hydration can be addressed by considering multiple protein conformations^{72,73,74,75,52,63} possibly supported by conformational clustering,⁷⁶ but this becomes expensive for IST-type calculations.^{77,78} None of these hydration methods are viable for proteins because having distributions over such a large number of internal degrees of freedom would make them prohibitively expensive.

It would be advantageous to have a single, general method that can account for the free energy of all molecules in hydrated protein systems from an MD simulation. A recently developed EE method that is able to do this is EE-MCC. This uses the multiscale cell correlation (MCC) method to evaluate entropy of all molecules in the system using data from an equilibrium MD simulation^{79,80,81}, together with energy provided in the usual way from the system Hamiltonian. MCC arose from a multiscale synthesis of the entropy of single flexible molecules using covariance matrices to capture correlations within a molecule⁸² and the entropy of aqueous solutions using cell theory that accounts for multimolecular entropy in a mean-field manner.^{83,84,85,86} The system is represented as a collection of units of atoms at multiple length scales, with non-bonded units treated in a mean-field fashion and bonded units as correlated units. The multiscale treatment enables scalability, fast-convergence and interpretability, [which are all desirable features](#) for large systems. The size of the energy well for each unit or collection of units are parametrised in the harmonic approximation from forces, which have been found to have a strongly Gaussian distribution,^{87,82,80} and the quantum harmonic oscillator provides an accurate way to account for entropy in high-frequency bonded systems. Forces moreover are readily available in a MD simulation and provide an efficient and accurate representation of the average environment of a unit without having to explicitly refer to [positions of the surrounding atoms](#). This is in contrast to coordinate-based methods which have to define coordinates with respect to external units, which is problematic in continually changing liquid-phase systems. Entropy is also included for the distribution of different energy wells for each unit, relating to conformations, hydrogen-bond arrangements etc., which are defined using unit contacts. EE-MCC has been used to calculate free energy changes in chemical reactions,⁸⁸ octanol-water partition coefficients,⁸⁹ host-guest binding,⁹⁰ and protein-excipient stabilisation.⁸¹ Here we use EE-MCC to provide a detailed analysis of the free energy of proteins and their first-shell water molecules. We examine protein stability, water stability, how they compare with hydrophobicity scales, how they correlate with each other, and whether there is any dependence on various structural features of the proteins.

2 | METHODS

2.1 | EE-MCC Method

The Gibbs free energy G is calculated using the EE-MCC method via the equation $G = H - TS$ where H is enthalpy, S is entropy and T is temperature. We explain first the MCC method to calculate entropy of a protein and of water, followed by the calculation of the enthalpy.

2.2 | MCC Protein Entropy

In the MCC method all coordinates are discretised into one or more energy wells, leading to separate terms for the entropy over the different energy wells, termed topographical entropy, and the average entropy within the energy wells, termed vibrational entropy. Moreover, this discretisation is done for collections of atoms, termed units, at multiple length scales. For hydrated proteins, the entropy is a sum over the vibrational entropy of units at the polymer (P), monomer (M) and united-atom levels (UA), and over the topographical entropy at the united-atom level of theory in terms of sets of conformational states in each residue⁸⁰ and summarised here. Protein entropy is calculated using

$$S_{\text{prot}}^{\text{total}} = S_{\text{P}}^{\text{transvib}} + S_{\text{P}}^{\text{rovib}} + S_{\text{M}}^{\text{transvib}} + S_{\text{M}}^{\text{rovib}} + S_{\text{UA}}^{\text{transvib}} + S_{\text{UA}}^{\text{rovib}} + S_{\text{UA}}^{\text{topo}} \quad (1)$$

where the vibrational entropy relates to the number of states within each energy well and the topographical entropy to the distribution over different energy wells. The polymer corresponds to the whole protein, the monomer to each residue in the protein, and the united atom to each heavy atom and all its bonded hydrogens, such as hydroxyl groups, methyl groups, carbonyl carbons and oxygens, and water molecules. The vibrational entropy comprises the translations and rotations of a single or collection of correlated units, labelled as *transvib* and *rovib* respectively. The vibrational entropy component for each unit or collection of correlated units is calculated from its vibrational frequencies using the equation for a quantum harmonic oscillator

$$S^{\text{vib}} = k_{\text{B}} \sum_{i=1}^{N_{\text{vib}}} \left(\frac{h\nu_i/k_{\text{B}}T}{e^{h\nu_i/k_{\text{B}}T} - 1} - \ln(1 - e^{-h\nu_i/k_{\text{B}}T}) \right) \quad (2)$$

where k_{B} is Boltzmann's constant, h is Planck's constant, T is temperature, ν_i are vibrational frequencies and N_{vib} is the number of vibrational modes for each level. A level is defined as a set of smaller, covalently bonded units: for the polymer level $N = 1$ corresponding to the single protein in the simulation; at the monomer level N is the number of residues in the protein; at the united-atom level N is the number of united atoms in a residue. Vibrational frequencies are calculated for the set of N units from the eigenvalues (λ_i) of separate force and torque covariance matrices using

$$\nu_i = \frac{1}{2\pi} \sqrt{\frac{\lambda_i}{k_{\text{B}}T}} \quad (3)$$

These covariance matrices are constructed from the net forces and torques on each unit derived from the atomic forces outputted from the molecular dynamics simulation. At each structural level, the matrix elements $\langle F_i F_j / \sqrt{m_i m_j} \rangle$ are mass-weighted forces for translations and inertia-weighted torques $\langle \tau_i \tau_j / \sqrt{I_i I_j} \rangle$ for rotations, which are both rotated into the appropriate coordinate system, described below. The force covariance matrices at each length scale have $3N$ eigenvalues, where N is the number of units. At all but the highest polymer level, the six smallest eigenvalues, which correspond to translation and rotation of the collection of units, are removed to avoid double counting these same degrees of freedom of the single unit they comprise at the higher length scale. For torques, depending on the linearity of each constituent unit, the number of contributed eigenvalues is 3, 2 and 0 for non-linear, linear and point constituents, respectively. Concerning coordinate systems, for the polymer, the x, y, z axes are taken as the principal axes with the origin at the center of mass. For residue translations, the same polymer principal axes are used, and for residue rotations a local frame is used: the origin is the average position of the three backbone atoms, and the x, y, z directions of the residue-level principal axes are defined as the N-C vector, the vector orthogonal to the $NC_\alpha C$ plane, and the vector orthogonal to both these vectors, respectively. The united-atom translations use the same local residue axes, and for united-atom rotation a more localised frame is used with the origin at the heavy atom and x, y, z axes defined by the average vector of covalent bonds to hydrogens and two axes orthogonal to this.⁸⁰

The last component of protein entropy in Eq. 1 is the topographical entropy at the united-atom level S_{UA}^{topo} , also called the conformational entropy. As done earlier⁸⁰, for each dihedral angle comprising four heavy atoms in a residue, the probability distribution is discretised into conformers. Conformers are assigned for each dihedral angle according to the nearest peak in the distribution, where the distribution is constructed with a 30° bin width, which was found to have sufficient resolution to resolve different conformers. The entropy is calculated from the probability p_i of each set of conformers over all dihedrals in the residue using

$$S_{UA}^{\text{topo}} = -k_B \sum_{i=1}^{N_{\text{conf}}} p_i \ln p_i \quad (4)$$

where N_{conf} is the number of unique sets.

2.3 | Water Entropy

The entropy of water molecules is divided into vibrational and topographical components at just one structural level

$$S_W^{\text{total}} = S_W^{\text{transvib}} + S_W^{\text{rovib}} + S_W^{\text{or}} \quad (5)$$

The vibrational components, S_W^{transvib} and S_W^{rovib} , are calculated in the same way as described earlier for protein vibrational entropy at the polymer level because they both correspond to the translation and rotation of a single molecule. The topographical entropy for water here manifests as orientational entropy, S_W^{or} , described next.

S_{W}^{or} accounts for the probability of water molecules accepting and donating hydrogen bonds (HBs) with neighbors and the directional bias of these interactions.⁸¹ Neighbors are defined based on what solvation shell they are in with respect to the protein surface. A schematic example of neighbor definitions is shown in Supporting Figure S1. The theory builds on previous work for the orientational entropy of flexible liquids⁹¹ which assumed an isotropic distribution of orientations and previous work on dilute solutions of hydrated ions.⁹² S_{W}^{or} is calculated as a weighted sum of the logarithm of the effective number of orientations over all observed coordination shells c

$$S_{\text{W}}^{\text{or}} = k_{\text{B}} \sum_c p(c) \ln \left[(N_{\text{eff}} \pi)^{\frac{3}{2}} p(\text{HB}_{\text{av}}) / \sigma \right] \quad (6)$$

where $\sigma = 2$ is the symmetry number of water and N_{eff} is the effective number of available neighbors that can be hydrogen-bonded to without bias, and $p(\text{HB}_{\text{av}})$ expresses the probability that neighboring molecules are in the correct orientation for the molecule of interest to form its HBs. In the homogeneous water case,⁹¹ $p(\text{HB}_{\text{av}}) = 0.25$. More generally, it is calculated as a weighted average of the probability of forming HBs to each type of neighbor n

$$p(\text{HB}_{\text{av}}) = \frac{\sum_n p(\text{HB}_n) N_n}{N_c} \quad (7)$$

where N_n is the number of neighbors of type n , N_c is the total number of neighbors in the coordination shell such that $N_c = \sum_n N_n$, and $p(\text{HB}_n)$ is the probability for a water molecule to donate or to accept from a neighbor of type n , calculated as

$$p(\text{HB}_n) = \frac{p(\text{D}_n)}{p(\text{D}_n) + p(\text{A}_n)} \times \frac{p(\text{A}_n)}{p(\text{D}_n) + p(\text{A}_n)} \quad (8)$$

where $p(\text{D}_n)$ and $p(\text{A}_n)$ are the probabilities of donating to or accepting from a neighbor of type n , respectively. Thus $p(\text{D}_n) = N_{\text{D}_n} / \sum_n N_{\text{D}_n}$, where N_{D_n} is the number of donations to neighbour n and similarly $p(\text{A}_n) = N_{\text{A}_n} / \sum_n N_{\text{A}_n}$. The effective number of neighbors available to form HBs to in Eq. 6 takes into account how often the neighbor is involved in a HB. It is calculated using

$$N_{\text{eff}} = \sum_n \frac{p(\text{HB}_n) N_n}{0.25} \quad (9)$$

where 0.25 is the ideal number for two HBs with 0.5 probability to accept and donate. Hydrogen bonds are defined topologically^{93,84,92} between a hydrogen and the most favorably interacting acceptor, namely the acceptor for which $q_{\text{D}} q_{\text{A}} / r^2$ is most negative, where q_{D} and q_{A} are the charges of the donor and acceptor atoms, respectively, and r is the distance between them.

2.4 | Protein Hydration Shell

To calculate the free energy of hydrated proteins, we only consider water molecules in the coordination shell of the protein as depicted in Figure 1 A. Although longer-range effects may be present, causing water molecules to be perturbed beyond the protein surface,^{68,94,95} they are weak and difficult to account for because of statistical noise involving a large number of

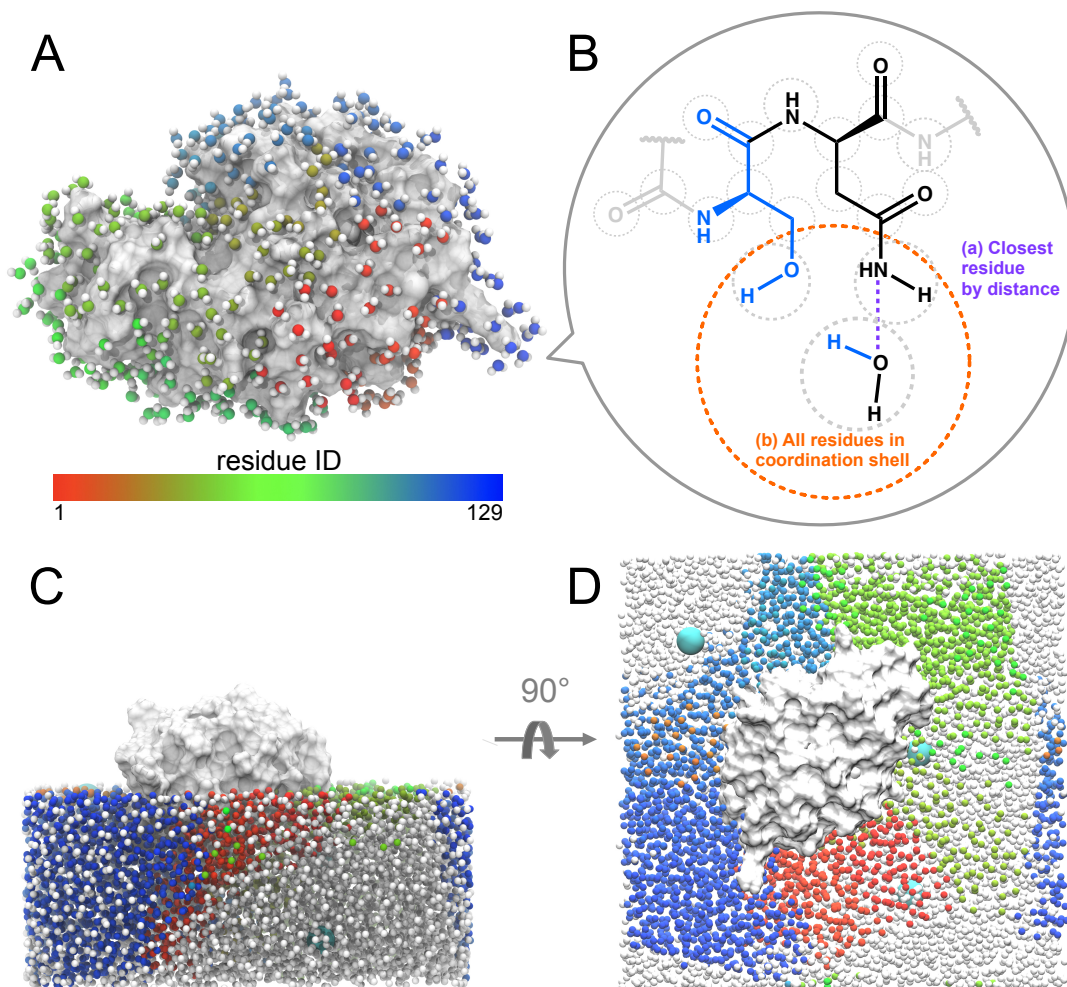


FIGURE 1 **A:** waters in the first coordination shell of lysozyme, colored by their closest residue united atom (continuously shaded from red to green to blue to represent residue IDs in the protein sequence). **B:** water molecules grouped by (a) the closest residue (dashed purple line) in its coordination shell (dashed orange circle) and (b) all residues pairs in the coordination shell. United atoms are marked by dashed grey circles. **C** and **D:** two orientations of the lysozyme protein (white) and half the water, residue IDs colored as in part **A**. Waters nearest to ions are colored white.

molecules. To analyse the different types of water molecules, we group them according to the nearest non-water united atom (Figure 1 B), whether in the protein or an ion, and the nearest residue, defined as the residue that contains the nearest united atom. This is similar to the so-called proximity criterion,^{96,97} which groups waters based on the distance to their closest solute atom, as depicted in Figure 1 C and D. As well as considering a hydration water's closest residue, we also consider its nearest two residues when a second such residue exist. United-atoms in a molecule are defined to lie in the coordination shell of another united atom if they are not blocked by a closer united atom using the [Relative Angular Distance \(RAD\)](#) algorithm

$$\frac{1}{r_{ij}^2} > \frac{1}{r_{ik}^2} \cos \theta_{jik} \quad (10)$$

TABLE 1 PDB ID, Net Charge and Numbers of Residues and United Atoms for the Proteins

Protein	PDB ID	Q	N_R	N_{UA}
lysozyme	2vb1	+8	129	1001
α -chymotrypsin	1yph	+3	241	1751
α -lactalbumin	1f6r	-6	123	993
ribonuclease Sa	1rgg	-7	96	746

where r_{ik} and r_{ij} are distances of particle i from particles k and j respectively, θ_{jik} is the angle between k and j formed at i .

The RAD method is applied at the united-atom level, which is defined as a heavy atom and its bonded hydrogen atoms, and the centre point of a united atom is at the heavy atom. We also consider the stability of water according to protein secondary structure, grouping water molecules according to their nearest secondary structure elements as defined by STRIDE.⁹⁸

2.5 | Calculation and Partition of the Enthalpy

The enthalpy H of the whole system is directly accessible in an equilibrium MD simulation as the ensemble average of the system Hamiltonian as specified by the force field plus a pressure-volume term PV . The total PV term at ambient pressure is on the order of tens of kJ mol^{-1} but effectively the same on a per-water or per-residue basis, on the order of 2 J mol^{-1} for water and 10 J mol^{-1} for residues, and so is ignored in this analysis. Consequently in this work we use the terms enthalpy and energy interchangeably. H can be partitioned into atomic terms U_i which equals the sum of all energy terms in which that atom participates, be it bonded or non-bonded, divided up equally between contributing atoms. The enthalpy of any unit j , whether polymer, monomer or united-atom, is calculated as a sum over all constituent atoms using

$$H_j = \sum_i^{N_j} (K_i + U_i) \quad (11)$$

where K_i is the atom's kinetic energy and N_j is the number of atoms in the unit.

2.6 | Simulation Protocol

All systems are generated and minimised with AMBER 18 (Assisted Model Building and Energy Refinement)⁹⁹ and then subject to MD simulations in LAMMPS (Large-Scale Atomic/Molecular Massive Parallel Simulator),¹⁰⁰ which we make use of because it outputs atom-specific energies. Simulations are performed of each of the four proteins lysozyme, α -chymotrypsin, α -lactalbumin, and ribonuclease Sa. The protein structures are taken from the Protein Databank using the IDs given in Table 1 which also lists the number of residues N_R and united atoms N_{UA} in each protein.¹⁰¹ Protonation states of titratable groups in the protein are set appropriate to a pH of 7 using the PDB2PQR online server,¹⁰² with the only non-standard protonation of Asp88 in α -lactalbumin. For histidine, the neutral HIE tautomers are used. The resulting net protein charges Q are listed in Table 1. Each

system is constructed in a cubic box by solvating with water and either sufficient Na^+ or Cl^- ions using Packmol to neutralise the system,¹⁰³ with 20 Å of water around each protein. MD simulations of each of the twenty amino acids in water are also run to examine their solvation and provide reference data to compare with residues in the protein. Each simulation contains one amino acid, which is acetyl-capped at the N-terminus and methyl-amide-capped at the C-terminus, and 900 water molecules. Each amino-acid system and box of 900 water molecules are run in triplicate to provide an estimate of the errors involved. Each system is constructed in a cubic box by solvating with water out to 10 Å using Packmol. The four proteins and the single capped amino acids are modeled using the Amber FF14SB force field,¹⁰⁴ TIP3P for water,¹⁰⁵ and the Joung and Cheatham TIP3P parameters¹⁰⁶ for the Na^+ and Cl^- ions. System topology files are generated using AMBER 18⁹⁹ and then minimised for 5000 steps using steepest descents. The files are converted into LAMMPS formatted input files using InterMol.¹⁰⁷ Using LAMMPS,¹⁰⁰ temperature is slowly increased to 298 K under constant NVT (number volume temperature) for 0.2 ns followed by NPT (number pressure temperature) equilibration for 5 ns under NPT conditions with a 1 fs time step. Production simulations are run for 50 ns under the same NPT conditions and a 2 fs time-step. Temperature and pressure are controlled using a Nose-Hoover thermostat and barostat respectively. Temperature is relaxed every 0.2 ps and pressure is relaxed every 0.5 ps with an isotropic stress tensor across box dimensions. Nonbonded interactions are cut off at 9 Å, and long-range interactions beyond this distance are calculated with particle-particle particle-mesh (PPPM).¹⁰⁸ The SHAKE algorithm is used to constrain all bonds and angles to hydrogen atoms.¹⁰⁹ The *pe/atom* and *ke/atom* flags in LAMMPS are set to output trajectories of potential and kinetic energies per atom for molecular energy analysis. Force and coordinate trajectories are outputted for the entropy analysis. Trajectories are saved every 10 ps to give 5000 frames for analysis for each simulation run. Previous work on proteins has shown that this sampling protocol is sufficient to obtain converged entropies.⁸⁰ Output files are read in using the MDAnalysis python library.¹¹⁰ The energy and entropy of water and protein energy are analysed using an in-house python program POSEIDON Beta V2.0 available at <https://github.com/jkalayan/PoseidonBeta>. Protein entropy is analysed using earlier [software CodeEntropy](#)⁸⁰ which is available at <https://github.com/arghya90/CodeEntropy>. This requires converting the LAMMPS topology and trajectory files to PSF and DCD formats using CPPTRAJ.¹¹¹

2.7 | Analysis of Structure-Thermodynamics Correlations

To assess the trends in energy and entropy in the protein simulations, several other quantities are calculated on a per-residue basis:

1. HR: Hydrophobicity of each type of amino acid ranked by the calculated free energy of water molecules around amino-acid side chains, where HR value increases with hydrophobicity .

TABLE 2 Water Entropy Components, Enthalpy, Free Energy and [Count in the Protein Coordination Shell](#)

System	S_W^{vib}	S_W^{or}	S_W^{total}	H_W	G_W	N_{wcp}
	(J K ⁻¹ mol ⁻¹)			(kJ mol ⁻¹)		
lysozyme	67.2	6.8	74.0	-33.2	-55.3	464
α -chymotrypsin	67.1	6.5	73.7	-33.3	-55.2	739
α -lactalbumin	66.9	6.6	73.4	-33.5	-55.4	477
ribonuclease Sa	67.2	6.4	73.5	-33.4	-55.4	396
Bulk water	68.1	10.7	78.8	-32.3	-55.8	-

2. RMSD: Average root-mean-square deviation of side-chain heavy-atoms on each residue aligned and compared to the simulation starting structure using CPPTRAJ.¹¹¹
3. N_{wc} : [Average number of water molecules in the coordination shell of the closest residue.](#)
4. N_{rc} : [Average number of residue united atoms in the coordination shell of a residue.](#)

Correlations between protein structural features and thermodynamic properties of residues and water molecules are determined using a covariance matrix with elements

$$\text{cov}(X, Y) = \frac{1}{N} \sum_{i=1}^N (X_i - \bar{X})(Y_i - \bar{Y}) \quad (12)$$

where for each value X_i in feature X containing a total of N values, the deviation from the mean \bar{X} is assessed simultaneously with the deviation from the mean of values in feature Y . Each value is normalised so that the mean is zero and standard deviation σ is one. This gives matrix elements between 1 (fully correlated covariance) and -1 (fully anti-correlated covariance), where 0 is no correlation in the covariance of two features.

3 | RESULTS

3.1 | Free Energy of Water around Proteins

The free energy, energy and entropy per water molecule for water in the coordination shell of each protein and in bulk are shown in Table 2, with superscript "vib" encapsulating both "transvib" and "rovib" in Eq. 1. [Per-water errors in enthalpy were calculated from their standard deviations and found to be in the range 0.01-0.02 kJ mol⁻¹. With errors in entropy having been found elsewhere to be slightly smaller⁸⁸, this means that the numbers shown are accurate to the precision used. The total number of water molecules in the protein coordination shell \$N_{\text{wcp}}\$ in Table 2 is all water molecules that have a residue UA in their coordination shell. Consistent with what has been seen elsewhere for hydration,^{112,113,114} there is mild energy-entropy](#)

compensation across all proteins, the highest entropy and enthalpy being for lysozyme and the lowest for α -lactalbumin. This trend could be related to the charge of the proteins, given that the entropy and energy per water molecule are lower near the two negatively charged proteins, α -lactalbumin and ribonuclease Sa (net charges in Table 1). Moreover, both these proteins have overall free energies that are slightly more stable, that is to say, more negative, because negatively charged residues have fewer but stronger HBs because of the local deficiency of donors. This is consistent with the preferred solvation of anions over cations in water due to water's asymmetric structure^{115,116,117,92}, making it is easier for a water molecule to donate to two strong hydrogen bonds with its well-spaced hydrogens than it is to accept from two hydrogens via its single oxygen. The trend in N_{wc} reflects the size of the protein: the larger the protein, the more first-shell water molecules there are around the protein. Note that the values of H and S for bulk TIP3P are higher than the experimental values of $-34.1 \text{ kJ mol}^{-1}$ and $69.9 \text{ J K}^{-1} \text{ mol}^{-1}$ as noted elsewhere^{118,119}, with the orientational entropy in this work being even higher and in better agreement with other methods¹¹⁹ compared to the earlier value using the tetrahedral model¹¹⁸.

Water molecules around each of the four proteins are further assessed in Figure 2 based on their closest residue. Again, the energy-entropy compensation of water molecules is clearly present. Hypothetical energy and entropy values required to give the bulk water free energy of $-55.8 \text{ kJ mol}^{-1}$ are shown as a dashed line in Figure 2 A. Molecules to the left of this line are stabilised and those to the right are destabilised. We first note that there is much more variation in energy than in entropy. Generally, water around negative residues are more stable than bulk water, while water molecules around positive and polar residues are similarly stable or slightly less stable. These kinds of water molecules show little dependence on the number of water molecules in their shell, which relates to the degree of burial. Water around non-polar residues are the most diverse and scattered in entropy and energy because of the contribution of orientational entropy which depends in a more multi-body fashion on the anisotropy and size of the coordination shell, rather than solely on the strength of interactions. Water molecules that have lower N_{wc} are most likely more buried in the protein. They are typically near hydrophobic residues and have a higher free energy. The burial effect may be examined in more detail from the free energy, enthalpy, entropy and entropy components versus the number of surrounding protein united atoms N_{UAc} (Supporting Figure S2). The two main trends are that orientational entropy decreases strongly with the degree of burial, likely due to asymmetry and confinement, and that the spreads of the other terms increase for more buried waters, going from strongly bulk-like to either higher or lower, again with the majority of waters less stable than in bulk. The variation in non-polar water free energy in Figure 2 A is different according to the proteins, with ribonuclease Sa having less scatter than the other proteins. This appears to correlate with the size of the protein, which also relates to the number of non-polar and buried residues, with α -chymotrypsin having the most and ribonuclease-Sa the least. The location of each water around the flexible protein surface (Figure 2 C) does not display any notable trends, except that water molecules are seen to have higher free energy in more buried regions, as noted earlier.

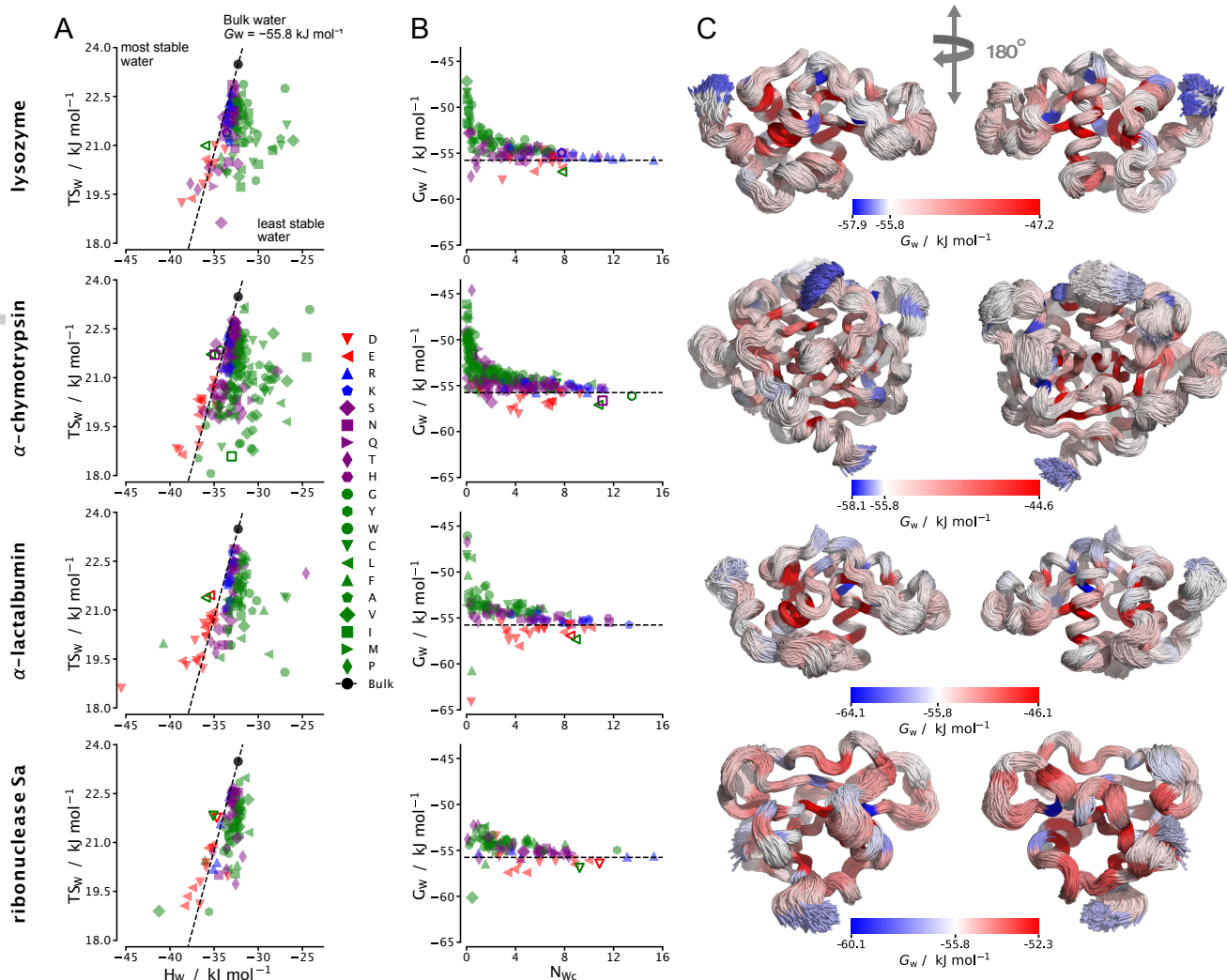


FIGURE 2 A: energy versus entropy for water around each residue in each protein. Dashed lines correspond to hypothetical values of enthalpy and entropy that give the free energy of bulk water ($-55.8 \text{ kJ mol}^{-1}$) and the black circle is the actual computed bulk-water value. B: water free energy versus number of water molecules closest to a particular protein residue. Negative residues are red, positive residues are blue, polar residues are purple and non-polar residues are green. Terminal residues are represented as open markers. C: superimposed protein structures of 2500 frames from 50 ns trajectories shown from the front and back (left and right). Each residue is colored by the free energy of water molecules closest to that residue relative to bulk water (blue more stable and red less stable).

3.2 | Free Energy of Water Contacting Two Protein Residues

To understand how the free energy of water is affected by being near multiple residues, plotted in Figure 3 A is the free energy of water molecules that are in contact with the united atoms of two or more residues for the four proteins. This plot also reveals which residues are close enough to interact with the same water molecule and whether the amino acids are adjacent or distant in sequence, the latter being more likely in a binding site. Equivalent plots in terms of enthalpy and entropy are provided in Supporting Figure S3. The trends are found to be fairly similar to those for the nearest single residues. For example, the active site of lysozyme¹²⁰ (Glu35 and Asp52) is surrounded by more stable water molecules than bulk, which is partly because of

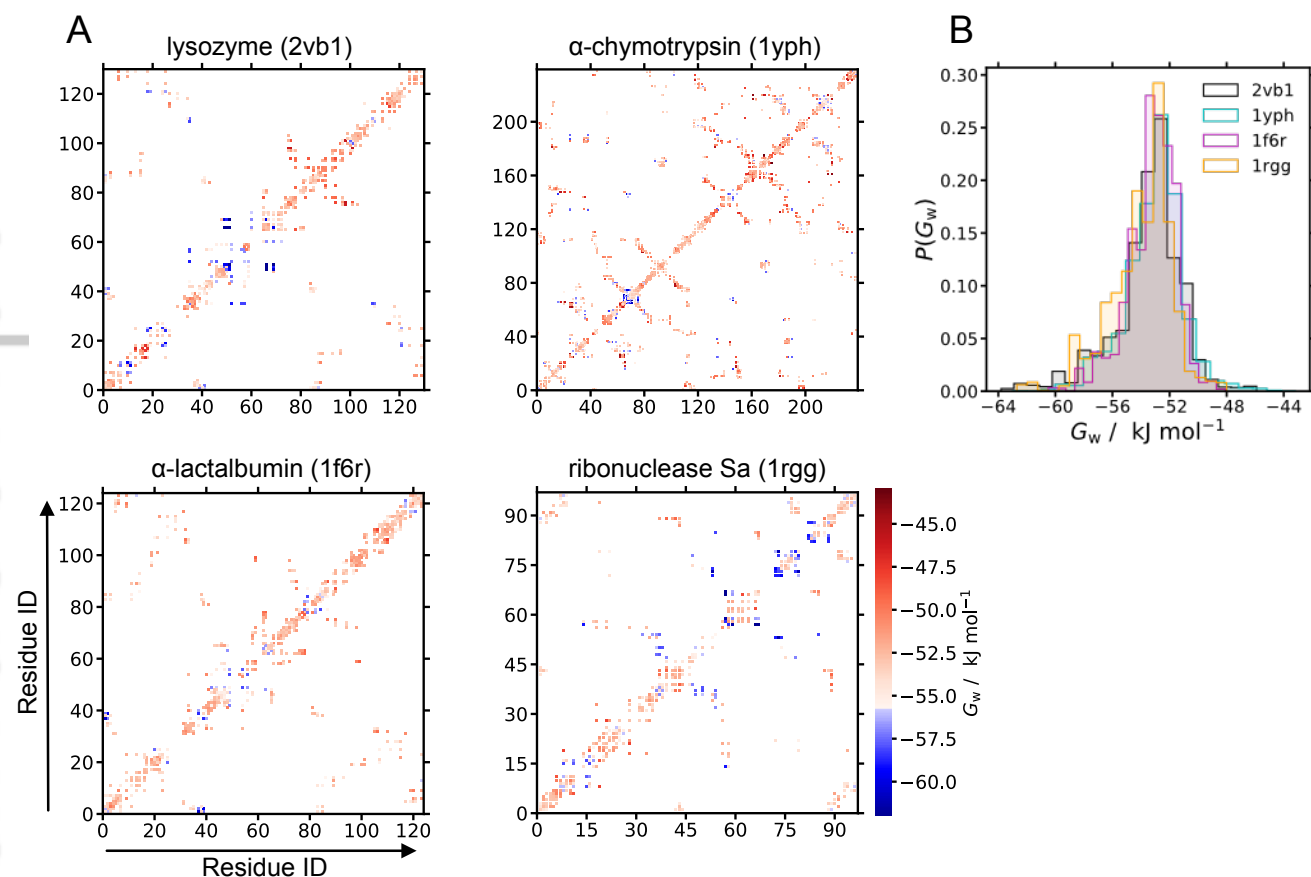


FIGURE 3 **A:** free energy G_W of water in contact with pairs of protein residues colored in red for less stable and blue for more stable than bulk water. **B:** normalised probability histograms of G_W around each protein.

the negatively charged residues. Cation-binding sites also generally display stabilised water. For α -lactalbumin, some stabilised water molecules are found in calcium-binding sites, one site being at residues Asp82, Asp87 and Asp88, and the backbone carbonyls of residues Lys79, Asp4, and Asp8, and the other site at Thr38, Gln39, Asp83 and the backbone carbonyl of Leu81,¹²¹ with the water molecules in the more exposed second binding site being more stabilised. Some zinc-binding sites are also surrounded by stabilised water molecules and again generally appear at more solvent-exposed regions. On the other hand, water molecules in some non-polar regions are destabilised and would therefore be easier to displace by non-polar molecules. The histograms of G_W values (Figure 3 B) are similar for each protein, but ribonuclease-Sa has a greater number of stabilised waters, consistent with it having the most negative charge, as noted earlier.

3.3 | Free Energy of Water according to Protein Secondary Structure

Table 3 contains the values of TS_W , H_W and G_W around each type of secondary structure averaged over all four proteins, together with the number percentage of residues and water molecules associated with each type. It can be seen that G_W has

TABLE 3 Water Entropy, Enthalpy, Free Energy, and Residue and Water Percentages for Protein Secondary Structure

Secondary structure	TS_W (J K ⁻¹ mol ⁻¹)	H_W (kJ mol ⁻¹)	G_W	% residues	% water
bridge	20.1	-34.0	-54.1	3	3
coil	20.2	-33.9	-54.1	17	20
extended	20.3	-33.5	-53.8	19	12
turn	20.4	-33.4	-53.8	18	17
3_{10} helix	20.3	-32.9	-53.2	4	4
α helix	20.4	-32.9	-53.3	18	17

a ranking from most to least stable of bridge, coil < extended, turn < 3_{10} helix, α helix. Most of this trend is governed by a less negative enthalpy that is slightly offset by a larger entropy. Evidently, water appears to be less stable around the more ordered helices and sheets and is more stable around the less ordered coils and isolated bridges. This is consistent with the finding elsewhere using GIST⁵⁸ whose authors rationalised the effect in terms of hydrogen-bonding groups in the protein being less favourably placed in more constrained environments such as helices and strands. It could also reflect different amino-acid propensities in different types of secondary structure.

3.4 | Free Energy of Water around Capped Amino Acids

It is insightful to examine the free energy, energy and entropy of water molecules around single solvated amino acids as a point of comparison with residues in the protein and to compare with a range of widely used hydrophobicity scales for amino acids.^{122,123,124,125,126,127} These values are illustrated in Figure 4 A, ordered from most stable to least stable free energy, together with the water orientational, rovibrational and transvibrational entropy components. Water molecules closest to capping groups or ions are not included in the analysis. The water energy is most stable around the negative amino acids Asp and Glu due to strong, polar interactions between negatively charged side-chain oxygens and water-molecule hydrogen atoms. At the same time, the orientational, rovibrational and transvibrational entropy components of water are smaller because of the associated stronger forces and torques, partly offsetting the more negative energy. Such energy-entropy compensation for water around solutes has been observed in previous studies.^{112,113,114} These effects are present but not as strong for water near positively charged Lys and Arg side chains. This trend of progressively less negative free energy continues as amino acids become more hydrophobic, driven more by energy, less by vibrational entropy and with little dependence on orientational entropy. Only for the most hydrophobic side-chains does orientational entropy decrease again because of the anisotropic bias of hydrogen donation towards water molecules and away from the hydrophobic solutes. Interestingly, orientational entropy around Pro is most heavily reduced, indicating that water molecules near Pro must have smaller solvation shells, as shown by a low N_{eff} Eq. 9 for water around Pro in Supporting Table S1.

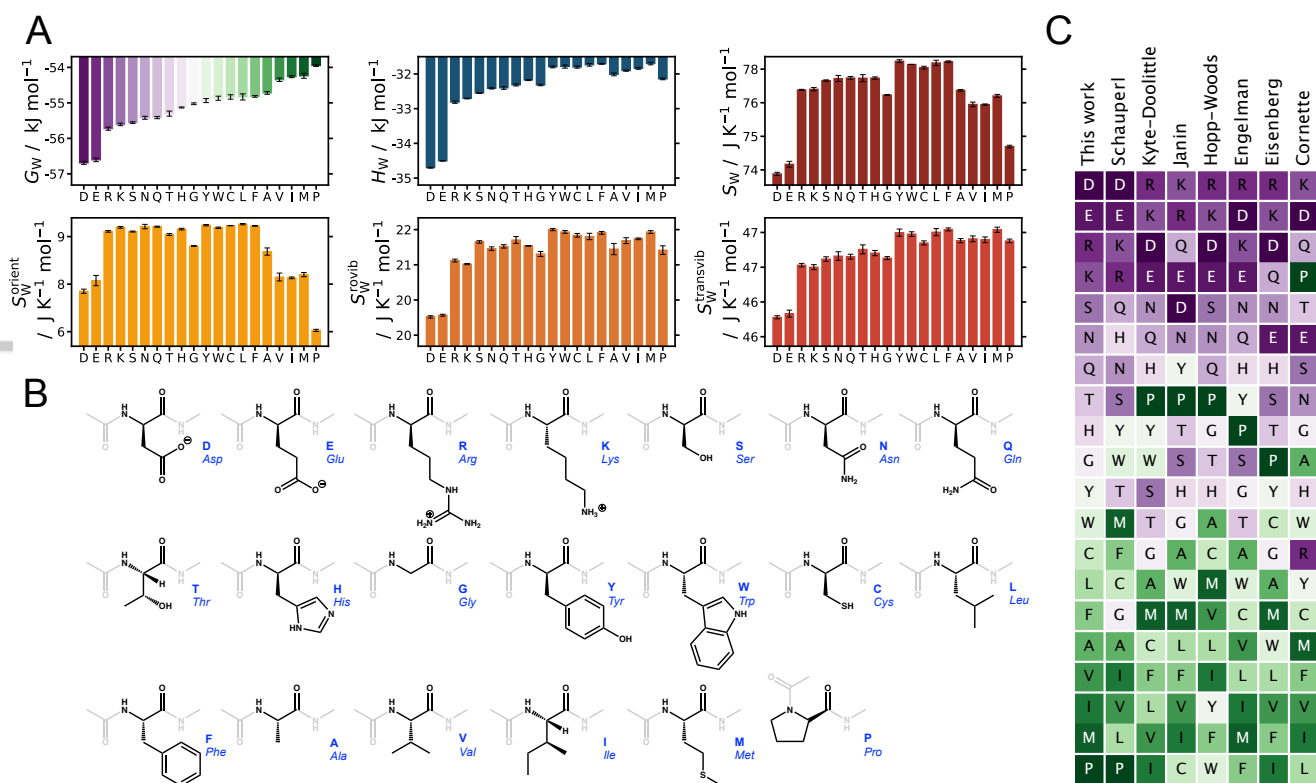


FIGURE 4 **A:** average free energy G_w , enthalpy H_w and entropy S_w per water molecule over all water molecules in the first shell of each capped amino acid, ordered from most stable to least stable free energy, and entropy components of orientational S_w^{orient} , rotational S_w^{rot} and translational S_w^{trans} (bottom left to right). Error bars show the standard deviation for each data point. **B:** structure of each amino acid with capping groups greyed out, ordered from most to least stable hydration shell. **C:** water molecule free energy ordered from most stable/hydrophilic (purple) to least stable/hydrophobic (green) around each amino acid side-chain calculated here (first column) versus hydrophobicity scales from various other works.^{122,123,124,125,126,127}

Figure 4 C makes clear that water free energies in the hydration shell of amino acids show similar trends to those of other hydrophobicity scales, with similar groupings of hydrophobic amino acids at one end and hydrophilic amino acids at the other. One notable point of disagreement between scales is the ranking of Pro, which is most hydrophobic here in agreement with Schauperl's scale¹²² whereas most other scales rank it as more hydrophilic. As noted earlier, this greater hydrophobicity of Pro arises from its low orientational entropy because water molecules near Pro have fewer hydrogen-bonded neighbors. Interestingly, Leu is ranked as more hydrophilic than Ile, Val and Ala, even though it is larger than Val and Ala and the same composition as Ile. The dominant contribution to the difference is the orientational entropy, which is markedly higher for Leu and comparable to that of the more polar amino acids, suggesting there are more hydrogen-bond groups near Leu than for the other three aliphatic amino acids. Also, Met is ranked more hydrophobic here than in most scales, again because of its lower orientational entropy and water coordination number. These effects arising from orientational entropy are subtle and non-obvious but may be related to the number of hydrocarbon methyl termini of the branched-chain amino acids which can fit better into water's structure in a clathrate-like arrangement than bulkier hydrocarbon chains or sulfur atoms of methionine which are more disruptive.

TABLE 4 Protein and Residue Entropy Components, and Enthalpy and Free Energy per Residue for Each Protein

Protein	$S_{\text{P}}^{\text{vib}}$	$S_{\text{R}}^{\text{vib}}/N_{\text{R}}$	$S_{\text{UA}}^{\text{vib}}/N_{\text{R}}$	$S_{\text{UA}}^{\text{topo}}/N_{\text{R}}$	$S_{\text{prot}}^{\text{total}}/N_{\text{R}}$	$H_{\text{prot}}^{\text{total}}/N_{\text{R}}$	$G_{\text{prot}}^{\text{total}}/N_{\text{R}}$
	(J K ⁻¹ mol ⁻¹)					(kJ mol ⁻¹)	
lysozyme	91.2	37.9	19.6	3.9	62.1	-71.4	-89.9
α -chymotrypsin	87.0	40.9	20.0	4.0	65.3	-12.4	-31.9
α -lactalbumin	86.2	37.7	20.7	5.6	64.7	-35.2	-54.5
ribonuclease Sa	87.2	32.9	17.7	4.0	55.5	-52.6	-69.2

3.5 | Protein Free Energy

A breakdown of the entropy, energy and free energy for each protein is presented in Table 4, including the entropy decomposition according to polymer, residue and united atom specified by Eq. 1, with the residue and united-atom terms divided by N_{R} to give per-residue values. Similar to before, per-residue errors in enthalpy were calculated from their standard deviations and found to be in the range 0.01-0.03 kJ mol⁻¹, again indicating that the numbers given are accurate to the precision used. Some strikingly large differences are seen for the different proteins but it should be noted that these values depend on the amino-acid composition of the protein. Lysozyme has the lowest free energy per residue and α -chymotrypsin the highest, driven mostly by energy but also by entropy. Ribonuclease Sa also has a lower entropy than the other proteins matched by a lower enthalpy. Before interpreting these values, it should be noted that there is a large range of free energies of protein residues (265 to -638 kJ mol⁻¹), as seen by the entropy and enthalpy of single capped solvated amino acids listed in Supporting Table S2. It can be seen that these values depend on a combination of the polarity of the atoms and the size of the amino acids. Evidently, the different kinds of amino acid in each protein mean that absolute values are difficult to compare on the same scale, a problem that does not occur for water molecules earlier, which are all the same. Therefore, to better understand the components of the protein free energies, the relative free energy of each protein residue is analysed by taking the free energy of the residue and subtracting off the free energy of the corresponding single solvated capped amino acid (Supporting Table S2), excluding the contributions from the capping groups. Terminal residues are not considered in this analysis due to their different numbers of atoms and charges. This referencing is not done for entropy at the protein or residue levels because we do not have reference values for these. Average referenced values ΔS_{UA} , ΔH_{UA} and ΔG_{UA} per residue obtained are listed in Table 5. It can be seen that these referenced values in per-residue form are now much more consistent across all proteins than the absolute values. Moreover, there is a strong enthalpy-entropy compensation in the stability, with negative changes in both enthalpy and entropy bringing about a marginal change in free energy, being stabilising for lysozyme and α -chymotrypsin, and neutral for α -lactalbumin and ribonuclease Sa. The specific referenced values of $T\Delta S_{\text{R}}$ and enthalpy ΔH_{R} of each residue are illustrated in Figure 5 A. When considering relative values, the dependence of relative free energy per residue with local environment is not as clear-cut as it was for water.

TABLE 5 Referenced Per-Residue Entropy, Enthalpy and Free Energy for Each Protein

Protein	$T\Delta S_{\text{UA}}/N_{\text{R}}$	$\Delta H_{\text{prot}}^{\text{total}}/N_{\text{R}}$	$\Delta G_{\text{prot}}^{\text{total}}/N_{\text{R}}$
	(J K ⁻¹ mol ⁻¹)	(kJ mol ⁻¹)	
lysozyme	-6.4	-8.0	-1.6
α -chymotrypsin	-5.3	-6.5	-1.2
α -lactalbumin	-6.3	-6.3	0.0
ribonuclease Sa	-6.2	-6.1	0.1

The entropy of almost all residues is smaller in proteins than for single capped amino acids, with $T\Delta S_{\text{R}}$ values being lower in proteins by up to -15 kJ mol^{-1} , (Figure 5 A). This is expected due to the loss of conformational flexibility brought about by restrained backbone atoms in the polymer but there must also be other vibrational contributions. For energy, the stabilisation is much more variable between residues of the same type, varying by up to $\pm 60 \text{ kJ mol}^{-1}$. Anionic residues tend to be less stable in proteins, cationic residues are more variable, and, polar and non-polar residues more often are stabilised but not always. This trend for negatively charged residues is somewhat opposite to water molecules as found earlier, which are stabilised near negative residues. When ΔG_{R} is mapped onto the 3D protein structures in Figure 5 B, there are no obvious trends, with increases or decreases seemingly randomly spread over the protein surface.

3.6 | Correlations between Free Energy of Bonded Residues

To understand what might affect residue free energies, the average relative free energy for each residue type in all four proteins is assessed based on what residue types are bonded to it on its C- and N-terminal sides and plotted in Figure 6 . Residues that are more stable than their single solvated amino acid equivalent are shown in blue and less stable residues are shown in red. Further decomposition of energy and entropy is shown in Supporting Figure S4. The trends overall are fairly weak. Most residues are destabilised compared to their single capped reference states, as noted earlier. As might be expected, cystines are more stable regardless of what they are bonded to due to covalent disulfide bonds. Pro, on the other hand, is generally stabilised regardless of the identity of the bonded residues, but if a Pro is on the C-side of a residue, then it destabilises that residue. Surprisingly, most residues are stabilised when a positive residue, Arg or Lys, is on the N-terminal side ($n - 1$) but destabilised if on the C-terminal side ($n + 1$). For negatively charged Asp and Glu there is a weaker but reverse trend. The reason for this is not clear but it is likely related to a stabilising or destabilising interaction between these charged residues and the backbone amide hydrogen or carbonyl oxygen.

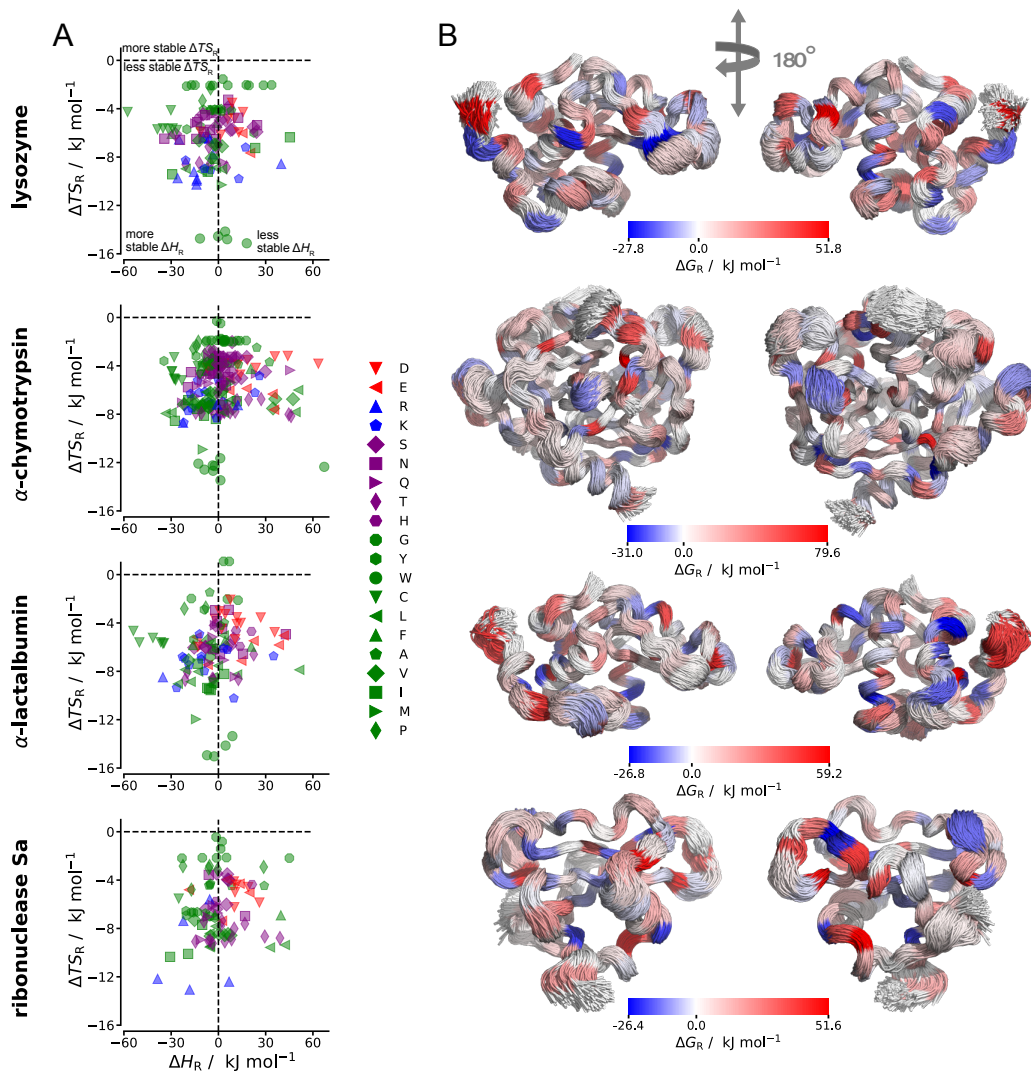


FIGURE 5 **A:** relative entropy $T\Delta S_R$ and energy ΔH_R of residues versus that of single solvated capped amino acids. **B:** per-residue change in free energy ΔG_R relative to single solvated capped amino acids mapped onto aligned structures from 2500 simulation frames.

3.7 | Correlations between Water and Protein Free Energy

Having examined the free energies of water and protein separately, we finally look for correlations between the free energy values of protein and water. In Figure 7, covariance matrices are presented between free energy, energy and entropy components together with RMSD, HR, N_{Wc} and N_{Rc} (defined in Section 2.7) for hydrophobic and hydrophilic residues separately. As might be expected, residue conformational entropy $\Delta TS_R^{\text{topo}}$ is anti-correlated with the number of residue-residue contacts N_{Rc} , more so for hydrophobic residues, because of greater restriction when buried. This effect is also evident in residues with a higher hydrophobic rank HR being more anti-correlated with ΔTS_R^{vib} because hydrophobic residues are more likely to be buried and so have a lower vibrational entropy due to greater confinement. To further investigate correlations between protein and water quantities, Supporting Figure S5 plots each of G_W , N_{Wc} , N_{Rc} or RMSD versus ΔG_R for the corresponding residue of each

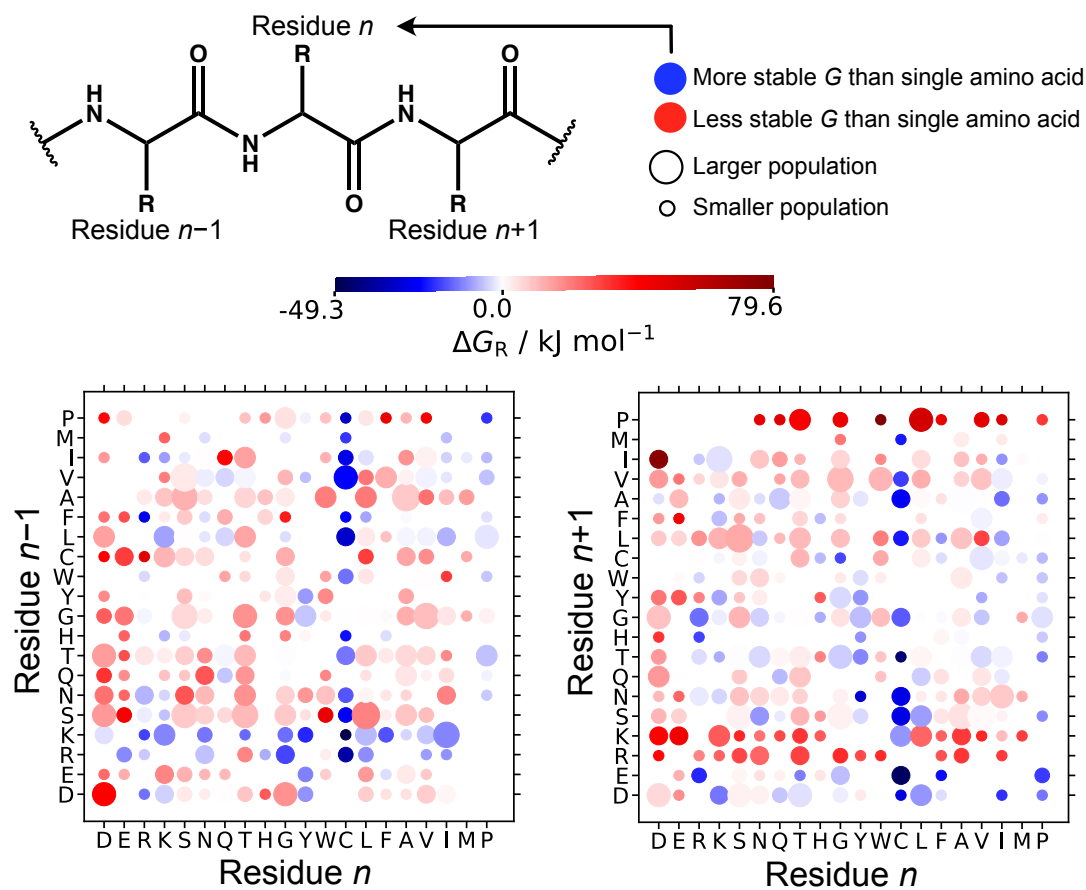


FIGURE 6 Average relative free energy ΔG_R per residue n over all four proteins when bonded to residue $n-1$ at the N-terminal side (left) and residue $n+1$ at the C-terminal side (right) of residue n . The number of times a residue pair occurs scales with the size of the dot.

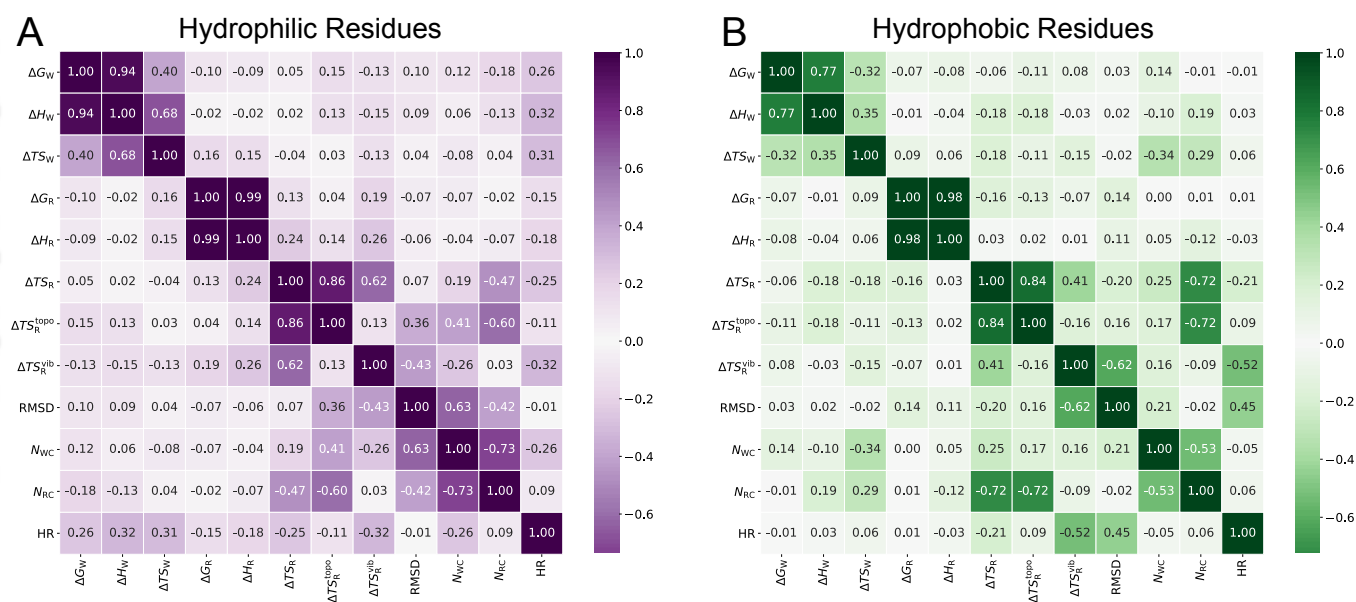


FIGURE 7 Covariance matrices for thermodynamic and structural properties of **A** hydrophilic residues, namely D, E, R, K, S, N, Q, T, H and **B** hydrophobic residues, namely G, Y, W, C, L, F, A, V, I, M, P. Darker shades imply larger correlations.

protein. No obvious trends are seen, apart from hydrophobic residues having surrounding water that is less stable, as has already been noted. Evidently, stability is likely a multi-body phenomenon that is not easily attributable to the stability of neighboring groups of atoms.

3.8 | Discussion

Computationally quantifying the thermodynamic properties of proteins in solution is a useful way to understand their behaviour in terms of all their constituent parts. Particularly for large, complex systems, it is important to both quantify and decompose free energy in order to go beyond qualitative descriptions such as hydrophobic and hydrophilic to comprehensively understand what atoms and interactions contribute to phenomena such as protein folding, allostery, binding or catalysis. Having a single consistent framework that addresses all system degrees of freedom that is free from parameters and operates at multiple length scales helps enable a general method for the thermodynamic analysis of such systems. Nevertheless, approximations are still necessary to make EE-MCC have these properties, such as the choice of a specific hierarchy of structures, the use of the harmonic approximation, which correlations within or between molecules to include, and a number of other issues. The orientational entropy of each protein is not considered here but could be calculated based on its size or coordination number as has been done elsewhere.^{128,91,79,88} This analysis only considers correlations between conformers within a residue and neglects those between residues, which are expected to be small, but should be considered in future work. Another omitted entropy term, the topographical entropy at the residue level, is assumed to be small for stable, relatively rigid proteins as those studied here, but would be larger for systems such as unfolded or intrinsically disordered proteins. The positional entropy arising from mixing with any counterions is not considered here, although in a first approximation one could use the ideal entropy of a mixture. The contributions of water molecules beyond the first shell of the protein have also been ignored because of their near-bulk-like properties and to reduce statistical noise. Other studies of hydrated proteins suggest that water contributions to entropy could be included out to 10 Å⁶⁸.

Examining the free energy of water molecules on a per-residue basis over an ensemble shows the expected trend for the stability of water molecules, being greater adjacent to hydrophilic residues than for hydrophobic residues, as has been reported elsewhere^{58,47,70,129,77,52,63}. Stability is further found to depend on the degree of burial of the water. Fewer water neighbours of a water molecule mostly means lower stability because of its diminished ability to form stabilising hydrogen bonds to the surrounding atoms in the more hydrophobic and asymmetric environment that biases orientations and lowers orientational entropy. A third factor affecting stability is having neighboring, negatively charged residues. A larger number of negative residues has previously been found to be suggestive of improved protein solubility.^{130,131} A fourth factor is secondary structure, with helices and strands being surrounded by less stable water.

Probing water molecules is typically the main way to find destabilised regions on a protein to which other molecules might bind. Protein free energy on a per-residue basis provides a further contribution to understand this process. The thermodynamics of proteins are more difficult to quantify due to their large size, flexible nature and correlated motions. The multiscale formulation employed here with three levels of hierarchy, polymer, monomer and united-atom length scales, enables protein and water entropy to be better understood in terms of multiple correlating units. We do not observe any significant correlations between protein free energy and surrounding water molecules. This may simply reflect the expectation that the free energy of protein residues is dominated by their surrounding residues, whether bonded or non-bonded, because of the strong covalent interactions for the bonded ones and close-packing of the others, whereas the free energy of water is governed by the atoms that their neighboring residues present to them. [However, using single-amino acid reference states, we are able to detect intriguing stabilisation for lysozyme and \$\alpha\$ -chymotrypsin not seen for \$\alpha\$ -lactalbumin and ribonuclease Sa.](#)

Despite this lack of correlation, we do make some intriguing findings. By assessing each residue based on what is bonded to its N- and C-terminal sides, greater residue stabilisation is found when a cationic residue is bonded to the N-side, compared to destabilisation when the cationic residue is on the C-side. The opposite trend is observed, albeit slightly weaker, for anionic residues. These asymmetric distributions of free energy changes may be due to how charged residues interact with atoms on the protein backbone. We hypothesise that cationic residues interact more favorably with the backbone carbonyl group on the next residue in the sequence. Conversely, anionic residues interact more favorably with the backbone amide group on the previous residue in the sequence. These asymmetric side-chain backbone interactions may cause a strain in the backbone on the opposite side of the direction in which the interaction takes place, resulting in instability. Further analysis of united-atom interactions between residues may highlight these side-chain backbone interactions. A similar asymmetric observation has been shown in other work of QM calculations that the $\alpha\text{C}-\text{C}'$ backbone bond length is reduced when an anionic residue is in the $n + 1$ position due to higher electron density, whereas a basic residue in the same position reduces the bond length.¹³²

3.9 | Conclusion

We present an EE method to calculate the free energy, energy and entropy of a hydrated protein from molecular dynamics simulations and apply it to individual hydrated amino acids and to four globular proteins, namely lysozyme, α -chymotrypsin, α -lactalbumin, and ribonuclease Sa. Entropy and energy are calculated for sets of atoms over a hierarchy of length scales from united atoms to residues to whole molecules. The decomposition of free energy of water molecules based on their neighboring residues gives detailed information about how protein interactions influence neighboring water molecules. Free energy decomposition also allows for the study of the interplay between the stability of water and their neighboring residues. Strong correlations were not observed between water and residue stability, although we did detect stabilisation of water next to anionic residues relative to cationic residues [and of water next to loops relative to that near helices and strands](#). Another intriguing finding was

the discovery of an asymmetry in the stability of residues depending on whether charged residues were on their N or C-terminal sides. EE-MCC with its insightful decomposition of free energy over groups of atoms, its hierarchy of length scales, and its single consistent formulation over all atoms in the system, should be readily scalable to larger and more flexible systems, such as protein-ligand complexes, protein assemblies and intrinsically disordered proteins, as well as many other kinds of molecular systems.

3.10 | Acknowledgments

The authors thank University of Manchester Research IT for access to the Computational Shared Facility. J.K. was supported by the EPSRC under grant codes EP/L015218/1 and EP/N025105/1.

References

1. Bellissent-Funel MC, Hassanali A, Havenith M, et al. Water Determines the Structure and Dynamics of Proteins. *Chemical Reviews* 2016; 116(13): 7673–7697. doi: 10.1021/acs.chemrev.5b00664
2. Spyraakis F, Ahmed MH, Bayden AS, Cozzini P, Mozzarelli A, Kellogg GE. The Roles of Water in the Protein Matrix: A Largely Untapped Resource for Drug Discovery. *Journal of Medicinal Chemistry* 2017; 60(16): 6781–6827. doi: 10.1021/acs.jmedchem.7b00057
3. Lee E, Choi JH, Cho M. The Effect of Hofmeister Anions on Water Structure at Protein Surfaces. *Physical Chemistry Chemical Physics* 2017; 19(30): 20008–20015. doi: 10.1039/C7CP02826A
4. Remsing RC, Xi E, Patel AJ. Protein Hydration Thermodynamics: The Influence of Flexibility and Salt on Hydrophobin II Hydration. *The Journal of Physical Chemistry B* 2018; 122(13): 3635–3646. doi: 10.1021/acs.jpcc.7b12060
5. Dahanayake JN, Mitchell-Koch KR. How Does Solvation Layer Mobility Affect Protein Structural Dynamics?. *Frontiers in Molecular Biosciences* 2018; 5. doi: 10.3389/fmolb.2018.00065
6. Friedman R. Simulations of Biomolecules in Electrolyte Solutions. *Advanced Theory and Simulations* 2019; 2(4): 1800163. doi: 10.1002/adts.201800163
7. Ahmad M, Gu W, Geyer T, Helms V. Adhesive Water Networks Facilitate Binding of Protein Interfaces. *Nature Communications* 2011; 2: 261. doi: 10.1038/ncomms1258

8. Hüfner-Wulsdorf T, Klebe G. Role of Water Molecules in Protein–Ligand Dissociation and Selectivity Discrimination: Analysis of the Mechanisms and Kinetics of Biomolecular Solvation Using Molecular Dynamics. *Journal of Chemical Information and Modeling* 2020; 60(3): 1818–1832. doi: 10.1021/acs.jcim.0c00156
9. Ribeiro SS, Samanta N, Ebbinghaus S, Marcos JC. The Synergic Effect of Water and Biomolecules in Intracellular Phase Separation. *Nature Reviews Chemistry* 2019; 3(9): 552–561. doi: 10.1038/s41570-019-0120-4
10. Darby JF, Hopkins AP, Shimizu S, et al. Water Networks Can Determine the Affinity of Ligand Binding to Proteins. *Journal of the American Chemical Society* 2019. doi: 10.1021/jacs.9b06275
11. Albrecht AV, Kim HM, Poon GMK. Mapping Interfacial Hydration in ETS-Family Transcription Factor Complexes with DNA: A Chimeric Approach. *Nucleic Acids Research* 2018; 46(20): 10577–10588. doi: 10.1093/nar/gky894
12. Privalov PL, Crane-Robinson C. Role of Water in the Formation of Macromolecular Structures. *European Biophysics Journal* 2017; 46(3): 203–224. doi: 10.1007/s00249-016-1161-y
13. Kollman PA, Massova I, Reyes C, et al. Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Accounts Chem. Res.* 2000; 33: 889-897.
14. Wang EC, Sun HY, Wang JM, et al. End-Point Binding Free Energy Calculation with MM/PBSA and MM/GBSA: Strategies and Applications in Drug Design. *Chem. Rev.* 2019; 119: 9478-9508.
15. Imai T, Harano Y, Kinoshita M, Kovalenko A, Hirata F. A theoretical analysis on hydration thermodynamics of proteins. *J. Chem. Phys.* 2006; 125: 024911.
16. Chong SH, Ham S. Dissecting Protein Configurational Entropy into Conformational and Vibrational Contributions. *J. Phys. Chem. B* 2015; 119: 12623-12631.
17. Nguyen C, Yamazaki T, Kovalenko A, et al. A Molecular Reconstruction Approach to Site-Based 3D-RISM and Comparison to GIST Hydration Thermodynamic Maps in an Enzyme Active Site. *PLOS ONE* 2019; 14(7): e0219473. doi: 10.1371/journal.pone.0219473
18. Meirovitch H, Vasquez M, Scheraga HA. Stability of Polypeptide Conformational States as Determined by Computer Simulation of the Free Energy. *Biopolymers* 1987; 26: 651-671.
19. Chelvaraja S, Meirovitch H. Simulation Method for Calculating the Entropy and Free Energy of Peptides and Proteins. *Proc. Natl. Acad. Sci. U.S.A.* 2004; 101: 9241-9246.

20. Meirovitch H. Methods for Calculating the Absolute Entropy and Free Energy of Biological Systems Based on Ideas from Polymer Physics. *J. Mol. Recognit.* 2010; 23: 153-172.
21. Brady GP, Sharp KA. Entropy in Protein Folding and in Protein-Protein Interactions. *Curr. Biol.* 1997; 7: 215-221.
22. Meirovitch H, Cheluvarama S, White RP. Methods for Calculating the Entropy and Free Energy and their Application to Problems Involving Protein Flexibility and Ligand Binding. *Curr. Prot. Peptide Sci.* 2009; 10: 229-243.
23. Polyansky AA, Kuzmanic A, Hlevnjak M, Zagrovic B. On the Contribution of Linear Correlations to Quasi-harmonic Conformational Entropy in Proteins. *J. Chem. Theory Comput.* 2012; 8: 3820-3829.
24. Suárez D, Diaz N. Direct Methods for Computing Single-Molecule Entropies from Molecular Simulations. *Rev. Comput. Sci.* 2015; 5: 1-26.
25. Kassem S, Ahmed M, El-Sheikh S, Barakat KH. Entropy in Bimolecular Simulations: A Comprehensive Review of Atomic Fluctuations-based Methods. *J. Mol. Graph. Model.* 2015; 62: 105-117.
26. Chong SH, Chatterjee P, Ham S. Computer Simulations of Intrinsically Disordered Proteins. *Ann. Rev. Phys. Chem.* 2017; 68: 117-134.
27. Go N, Scheraga HA. Analysis of contribution of internal vibrations to statistical weights of equilibrium conformations of macromolecules. *J. Chem. Phys.* 1969; 51: 4751-4767.
28. Brooks B, Karplus M. Harmonic dynamics of proteins - normal-modes and fluctuations in bovine pancreatic trypsin-inhibitor. *Proc. Natl. Acad. Sci. U.S.A.* 1983; 80: 6571-6575.
29. Levitt M, Sander C, Stern PS. Protein normal-mode dynamics - trypsin-inhibitor, crambin, ribonuclease and lysozyme. *J. Mol. Biol.* 1985; 181: 423-447.
30. Karplus M, Kushick JN. Method for Estimating the Configurational Entropy of Macromolecules. *Macromolecules* 1981; 14(2): 325-332. doi: 10.1021/ma50003a019
31. Schlitter J. Estimation of Absolute and Relative Entropies of Macromolecules Using the Covariance-matrix. *Chem. Phys. Lett.* 1993; 215: 617-621.
32. Schäfer H, Daura X, Mark AE, Gunsteren vWF. Entropy Calculations on a Reversibly Folding Peptide: Changes in Solute Free Energy Cannot Explain Folding Behavior. *Proteins* 2001; 43: 45-56.
33. Andricioaei I, Karplus M. On the Calculation of Entropy from Covariance Matrices of the Atomic Fluctuations. *J. Chem. Phys.* 2001; 115: 6289-6292.

34. Edholm O, Berendsen HJC. Entropy Estimation from Simulations of Non-diffusive Systems. *Mol. Phys.* 1984; 51: 1011-1028.
35. Zhang JF, Liu JS. On Side-chain Conformational Entropy of Proteins. *PLoS Computational Biology* 2006; 2: e168.
36. Towse CL, Akke M, Daggett V. The Dynameomics Entropy Dictionary: A Large-Scale Assessment of Conformational Entropy across Protein Fold Space. *J. Phys. Chem. B* 2017; 121: 3933-3945.
37. Dinola A, Berendsen HJC, Edholm O. Free-energy Determination of Polypeptide Conformations Generated by Molecular Dynamics. *Macromolecules* 1984; 17: 2044-2050.
38. Harpole KW, Sharp KA. Calculation of Configurational Entropy with a Boltzmann-Quasiharmonic Model: The Origin of High-Affinity Protein-Ligand Binding. *J. Phys. Chem. B* 2011; 115: 9461-9472.
39. Wang J, Bruschweiler R. 2D Entropy of Discrete Molecular Ensembles. *J. Chem. Theory Comput.* 2006; 2: 18-24.
40. Nguyen PH. Estimating Configurational Entropy of Complex Molecules: A Novel Variable Transformation Approach. *Chem. Phys. Lett.* 2009; 468: 90-93.
41. Killian BJ, Kravitz JY, Somani S, Dasgupta P, Pang YP, Gilson MK. Configurational Entropy in Protein-Peptide Binding: Computational Study of Tsg101 Ubiquitin E2 Variant Domain with an HIV-Derived PTAP Nonapeptide. *J. Mol. Biol.* 2009; 389: 315-335.
42. Numata J, Knapp EW. Balanced and Bias-Corrected Computation of Conformational Entropy Differences for Molecular Trajectories. *J. Chem. Theory Comput.* 2012; 8: 1235-1245.
43. King BM, Silver NW, Tidor B. Efficient Calculation of Molecular Configurational Entropies Using an Information Theoretic Approximation. *J. Phys. Chem. B* 2012; 116: 2891-2904.
44. Goethe M, Gleixner J, Fita I, Rubi JM. Prediction of Protein Configurational Entropy (Popcoen). *J. Chem. Theory Comput.* 2018; 14: 1811-1819.
45. Abel R, Young T, Farid R, Berne BJ, Friesner RA. Role of the Active-Site Solvent in the Thermodynamics of Factor Xa Ligand Binding. *Journal of the American Chemical Society* 2008; 130(9): 2817–2831. doi: 10.1021/ja0771033
46. Chong SH, Ham S. Distinct Role of Hydration Water in Protein Misfolding and Aggregation Revealed by Fluctuating Thermodynamics Analysis. <http://pubs.acs.org/doi/abs/10.1021/acs.accounts.5b00032>; 2015
47. Gerogiokas G, Southey MWY, Mazanetz MP, et al. Assessment of Hydration Thermodynamics at Protein Interfaces with Grid Cell Theory. *The Journal of Physical Chemistry B* 2016; 120(40): 10442–10452. doi: 10.1021/acs.jpcc.6b07993

48. Heyden M. Heterogeneity of Water Structure and Dynamics at the Protein-Water Interface. *The Journal of Chemical Physics* 2019; 150(9): 094701. doi: 10.1063/1.5081739
49. Heyden M. Disassembling Solvation Free Energies into Local Contributions—Toward a Microscopic Understanding of Solvation Processes. *WIREs Computational Molecular Science* 2019; 9(2): e1390. doi: 10.1002/wcms.1390
50. Verteramo ML, Stenstrom O, Ignjatovic MM, et al. Interplay between Conformational Entropy and Solvation Entropy in Protein-Ligand Binding. *J. Am. Chem. Soc.* 2019; 141: 2012-2026.
51. Mitusinska K, Raczynska A, Bzowka M, Bagrowska W, Gora A. Applications of water molecules for analysis of macromolecule properties. *Comput. Struct. Biotechnol. J.* 2020; 18: 355-365.
52. Fajardo TN, Heyden M. Dissecting the Conformational Free Energy of a Small Peptide in Solution. *J. Phys. Chem. B* 2021; 125: 4634-4644.
53. Mukherjee S, Schafer LV. Spatially Resolved Hydration Thermodynamics in Biomolecular Systems. *J. Phys. Chem. B* 2022; 126: 3619-3631.
54. Lazaridis T. Inhomogeneous Fluid Approach to Solvation Thermodynamics. 1. Theory. *The Journal of Physical Chemistry B* 1998; 102(18): 3531–3541. doi: 10.1021/jp9723574
55. Li Z, Lazaridis T. Thermodynamics of buried water clusters at a protein-ligand binding interface. *J. Phys. Chem. B* 2006; 110: 1464-1475.
56. Young T, Abel R, Kim B, Berne BJ, Friesner RA. Motifs for Molecular Recognition Exploiting Hydrophobic Enclosure in Protein–Ligand Binding. *Proceedings of the National Academy of Sciences* 2007; 104(3): 808–813. doi: 10.1073/pnas.0610202104
57. Nguyen CN, Kurtzman Young T, Gilson MK. Grid Inhomogeneous Solvation Theory: Hydration Structure and Thermodynamics of the Miniature Receptor Cucurbit[7]Uril. *The Journal of Chemical Physics* 2012; 137(4). doi: 10.1063/1.4733951
58. Beuming Thijs , Che Ye , Abel Robert , Kim Byungchan , Shanmugasundaram Veerabahu , Sherman Woody . Thermodynamic Analysis of Water Molecules at the Surface of Proteins and Applications to Binding Site Prediction and Characterization. *Proteins: Structure, Function, and Bioinformatics* 2012; 80(3): 871–883. doi: 10.1002/prot.23244
59. Haider K, Cruz A, Ramsey S, Gilson MK, Kurtzman T. Solvation Structure and Thermodynamic Mapping (SSTMap): An Open-Source, Flexible Package for the Analysis of Water in Molecular Dynamics Trajectories. *Journal of Chemical Theory and Computation* 2018; 14(1): 418–425. doi: 10.1021/acs.jctc.7b00592

60. Irwin BWJ, Vukovic S, Payne MC, Huggins DJ. Large-Scale Study of Hydration Environments through Hydration Sites. *J. Phys. Chem. B* 2019; 123: 4220-4229.
61. Wallerstein J, Ekberg V, Kumar R, et al. Entropy-Entropy Compensation between the Protein, Ligand, and Solvent Degrees of Freedom Fine-Tunes Affinity in Ligand Binding to Galectin-3C. *JACS Au* 2021; 1: 484-500.
62. Heinz LP, Grubmueller H. PerlMut: Spatially Resolved Hydration Entropies from Atomistic Simulations. *J. Chem. Theory Comput.* 2021; 17: 2090-2098.
63. Heinz LP, Grubmueller H. Spatially resolved free-energy contributions of native fold and molten-globule-like Crambin. *Biophys. J.* 2021; 120: 3470-3482.
64. Lin ST, Maiti PK, Goddard WA. Dynamics and Thermodynamics of Water in PAMAM Dendrimers at Subnanosecond Time Scales. *The Journal of Physical Chemistry B* 2005; 109(18): 8663–8672. doi: 10.1021/jp0471958
65. Jana B, Pal S, Maiti PK, Lin ST, Hynes JT, Bagchi B. Entropy of Water in the Hydration Layer of Major and Minor Grooves of DNA. *The Journal of Physical Chemistry B* 2006; 110(39): 19611–19618. doi: 10.1021/jp061588k
66. Debnath A, Mukherjee B, Ayappa KG, Maiti PK, Lin ST. Entropy and Dynamics of Water in Hydration Layers of a Bilayer. *The Journal of Chemical Physics* 2010; 133(17): 174704. doi: 10.1063/1.3494115
67. Sinha SK, Jana M, Chakraborty K, Bandyopadhyay S. In Silico Studies of the Properties of Water Hydrating a Small Protein. *The Journal of Chemical Physics* 2014; 141(22): 22D502. doi: 10.1063/1.4895533
68. Fiset O, Paslack C, Barnes R, et al. Hydration Dynamics of a Peripheral Membrane Protein. *J. Am. Chem. Soc.* 2016; 138: 11526-11535.
69. Pascal TA, Goddard WA, Maiti PK, Vaidehi N. Role of Specific Cations and Water Entropy on the Stability of Branched DNA Motif Structures. *J. Phys. Chem. B* 2012; 116: 12159-12167.
70. Pattni V, Vasilevskaya T, Thiel W, Heyden M. Distinct Protein Hydration Water Species Defined by Spatially Resolved Spectra of Intermolecular Vibrations. *The Journal of Physical Chemistry B* 2017; 121(31): 7431–7442. doi: 10.1021/acs.jpcc.7b03966
71. Persson RAX, Pattni V, Singh A, Kast SM, Heyden M. Signatures of Solvation Thermodynamics in Spectra of Intermolecular Vibrations. *Journal of Chemical Theory and Computation* 2017; 13(9): 4467–4481. doi: 10.1021/acs.jctc.7b00184
72. Aggarwal L, Biswas P. Hydration Water Distribution around Intrinsically Disordered Proteins. *The Journal of Physical Chemistry B* 2018. doi: 10.1021/acs.jpcc.7b11091

73. Chong SH, Ham S. Component Analysis of the Protein Hydration Entropy. *Chemical Physics Letters* 2012; 535: 152–156. doi: 10.1016/j.cplett.2012.03.033
74. Chong SH, Ham S. Interaction with the Surrounding Water Plays a Key Role in Determining the Aggregation Propensity of Proteins. *Angewandte Chemie International Edition* 2014; 53(15): 3961–3964. doi: 10.1002/anie.201309317
75. Chong SH, Park M, Ham S. Structural and Thermodynamic Characteristics That Seed Aggregation of Amyloid- β Protein in Water. *Journal of Chemical Theory and Computation* 2012; 8(2): 724–734. doi: 10.1021/ct200757a
76. Yang Y, Hu B, Lill MA. Analysis of Factors Influencing Hydration Site Prediction Based on Molecular Dynamics Simulations. *Journal of Chemical Information and Modeling* 2014; 54(10): 2987–2995. doi: 10.1021/ci500426q
77. Kraml J, Kamenik AS, Waibl F, Schauperl M, Liedl KR. Solvation Free Energy as a Measure of Hydrophobicity: Application to Serine Protease Binding Interfaces. *Journal of Chemical Theory and Computation* 2019; 15(11): 5872–5882. doi: 10.1021/acs.jctc.9b00742
78. Yoshidome T, Ikeguchi M, Ohta M. Comprehensive 3D-RISM Analysis of the Hydration of Small Molecule Binding Sites in Ligand-Free Protein Structures. *Journal of Computational Chemistry* 2020; 41(28): 2406–2419. doi: 10.1002/jcc.26406
79. Ali HS, Higham J, Henchman RH. Entropy of simulated liquids using multiscale cell correlation. *Entropy* 2019; 21: 750.
80. Chakravorty A, Higham J, Henchman RH. Entropy of Proteins Using Multiscale Cell Correlation. *Journal of Chemical Information and Modeling* 2020. doi: 10.1021/acs.jcim.0c00611
81. Kalayan J, A. CR, Warwicker J, Henchman RH. Thermodynamic origin of differential excipient-lysozyme interactions. *Front. Mol. Biosci.* 2021; 8: 536.
82. Hensen U, Gräter F, Henchman RH. Macromolecular entropy can be accurately computed from force. *J. Chem. Theory Comput.* 2014; 10: 4777-4781.
83. Irudayam SJ, Henchman RH. Solvation theory to provide a molecular interpretation of the hydrophobic entropy loss of noble gas hydration. *J. Phys.: Condens. Matter* 2010; 22: 284108.
84. Irudayam SJ, Plumb RD, Henchman RH. Entropic trends in aqueous solutions of common functional groups. *Faraday Discuss.* 2010; 145: 467-485.
85. Irudayam SJ, Henchman RH. Prediction and Interpretation of the Hydration Entropies of Monovalent Cations and Anions. *Molecular Physics* 2011; 109(1): 37–48. doi: 10.1080/00268976.2010.532162

86. Gerogiokas G, Calabro G, Henchman RH, Southey MWY, Law RJ, Michel J. Prediction of Small Molecule Hydration Thermodynamics with Grid Cell Theory. *Journal of Chemical Theory and Computation* 2014; 10(1): 35–48. doi: 10.1021/ct400783h
87. Klefas-Stennett M, Henchman RH. Classical and quantum Gibbs free energies and phase behavior of water using simulation and cell theory. *J. Phys. Chem. B* 2008; 112: 3769-3776.
88. Ali HS, Higham J, Visser dSP, Henchman RH. Comparison of Free-Energy Methods to Calculate the Barriers for the Nucleophilic Substitution of Alkyl Halides by Hydroxide. *J. Phys. Chem. B* 2020.
89. Falcioni F, Kalayan J, Henchman RH. Energy-entropy prediction of octanol-water logP of SAMPL7 N-acylsulfonamide bioisosters. *J. Comput. Aid. Mol. Des.* 2021; 35: 831-840.
90. Ali HS, Chakravorty A, Kalayan J, Visser dSJ, Henchman RH. Energy-entropy method using multiscale cell correlation to calculate binding free energies in the SAMPL8 Host-Guest Challenge. *J. Comput. Aid. Mol. Des.* 2021. doi: <https://doi.org/10.1007/s10822-021-00406-5>
91. Higham J, Chou SY, Gräter F, Henchman RH. Entropy of Flexible Liquids from Hierarchical Force–Torque Covariance and Coordination. *Molecular Physics* 2018; 0(0): 1–12. doi: 10.1080/00268976.2018.1459002
92. Irudayam SJ, Henchman RH. Prediction and interpretation of the hydration entropies of monovalent cations and anions. *Mol. Phys.* 2011; 109: 37-48.
93. Henchman RH, Irudayam SJ. Topological hydrogen-bond definition to characterize the structure and dynamics of liquid water. *J. Phys. Chem. B* 2010; 114: 16792-16810.
94. Persson F, Söderhjelm P, Halle B. The Spatial Range of Protein Hydration. *The Journal of Chemical Physics* 2018; 148(21): 215104. doi: 10.1063/1.5031005
95. Kalayan J, Henchman RH. Convergence behaviour of solvation shells in simulated liquids. *Phys. Chem. Chem. Phys.* 2021; 23: 4892-4900.
96. Mezei M, Beveridge DL. Structural Chemistry of Biomolecular Hydration via Computer Simulation: The Proximity Criterion. *Methods Enzymol.* 1986; 127: 21-47.
97. Ishii Y, Yamamoto N, Matubayasi N, Zhang BW, Cui D, Levy RM. Spatially-Decomposed Free Energy of Solvation Based on the Endpoint Density-Functional Method. *Journal of Chemical Theory and Computation* 2019; 15(5): 2896–2912. doi: 10.1021/acs.jctc.8b01309

98. Frishman D, Argos P. Knowledge-based protein secondary structure assignment. *Proteins* 1995; 23(4): 566–579.
99. Case DA, Ben-Shalom IY, Brozell SR, et al. *AMBER 2018*. University of California, San Francisco . 2018.
100. Plimpton S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *Journal of Computational Physics* 1995; 117(1): 1–19. doi: 10.1006/jcph.1995.1039
101. Berman HM, Westbrook J, Feng Z, et al. The Protein Data Bank. *Nucleic Acids Research* 2000; 28(1): 235–242.
102. Dolinsky TJ, Czodrowski P, Li H, et al. PDB2PQR: Expanding and Upgrading Automated Preparation of Biomolecular Structures for Molecular Simulations. *Nucleic Acids Research* 2007; 35(Web Server issue): W522–W525. doi: 10.1093/nar/gkm276
103. Martínez L, Andrade R, Birgin EG, Martínez JM. PACKMOL: A Package for Building Initial Configurations for Molecular Dynamics Simulations. *Journal of Computational Chemistry* 2009; 30(13): 2157–2164. doi: 10.1002/jcc.21224
104. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* 2015; 11: 3696–3713.
105. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 1983; 79: 926–935.
106. Joung IS, Cheatham TE. Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *The Journal of Physical Chemistry B* 2008; 112(30): 9020–9041. doi: 10.1021/jp8001614
107. Shirts MR, Klein C, Swails JM, et al. Lessons Learned from Comparing Molecular Dynamics Engines on the SAMPL5 Dataset. *Journal of Computer-Aided Molecular Design* 2017; 31(1): 147–161. doi: 10.1007/s10822-016-9977-1
108. Hockney RW, Eastwood JW. *Computer Simulation Using Particles*. CRC Press . 1988.
109. Ryckaert JP, Ciccotti G, Berendsen HJC. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *Journal of Computational Physics* 1977; 23(3): 327–341. doi: 10.1016/0021-9991(77)90098-5
110. Michaud-Agrawal N, Denning EJ, Woolf TB, Beckstein O. MDAAnalysis: A Toolkit for the Analysis of Molecular Dynamics Simulations. *Journal of Computational Chemistry* 2011; 32(10): 2319–2327. doi: 10.1002/jcc.21787
111. Roe DR, Cheatham TE. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *Journal of Chemical Theory and Computation* 2013; 9(7): 3084–3095. doi: 10.1021/ct400341p

112. Dragan AI, Read CM, Crane-Robinson C. Enthalpy–Entropy Compensation: The Role of Solvation. *European Biophysics Journal* 2017; 46(4): 301–308. doi: 10.1007/s00249-016-1182-6
113. Lumry R, Rajender S. Enthalpy–Entropy Compensation Phenomena in Water Solutions of Proteins and Small Molecules: A Ubiquitous Property of Water. *Biopolymers* 1970; 9(10): 1125–1227. doi: 10.1002/bip.1970.360091002
114. Breiten B, Lockett MR, Sherman W, et al. Water Networks Contribute to Enthalpy/Entropy Compensation in Protein–Ligand Binding. *Journal of the American Chemical Society* 2013; 135(41): 15579–15584. doi: 10.1021/ja4075776
115. Hummer G, Pratt LR, Garcia AE. Free energy of ionic hydration. *J. Phys. Chem.* 1996; 100: 1206-1215.
116. Grossfield A. Dependence of ion hydration on the sign of the ion’s charge. *J. Chem. Phys.* 2005; 122: 024506.
117. Mobley DL, Barber AE, Fennell CJ, Dill KA. Charge asymmetries in hydration of polar solutes. *J. Phys. Chem. B* 2008; 112: 2405-2414.
118. Henchman RH. Free Energy of Liquid Water from a Computer Simulation via Cell Theory. *The Journal of Chemical Physics* 2007; 126(6): 064504. doi: 10.1063/1.2434964
119. Wang L, Abel R, Friesner RA, Berne BJ. Thermodynamic Properties of Liquid Water: An Application of a Nonparametric Approach to Computing the Entropy of a Neat Fluid. *J. Chem. Theory Comput.* 2009; 5: 1462-1473.
120. Held J, van Smaalen S. The Active Site of Hen Egg-White Lysozyme: Flexibility and Chemical Bonding. *Acta Crystallographica Section D: Biological Crystallography* 2014; 70(Pt 4): 1136–1146. doi: 10.1107/S1399004714001928
121. Permyakov EA, Berliner LJ. α -Lactalbumin: Structure and Function. *FEBS Letters* 2000; 473(3): 269–274. doi: 10.1016/S0014-5793(00)01546-5
122. Schauerperl M, Podewitz M, Waldner BJ, Liedl KR. Enthalpic and Entropic Contributions to Hydrophobicity. *Journal of Chemical Theory and Computation* 2016; 12(9): 4600–4610. doi: 10.1021/acs.jctc.6b00422
123. Kyte J, Doolittle RF. A Simple Method for Displaying the Hydrophobic Character of a Protein. *Journal of Molecular Biology* 1982; 157(1): 105–132. doi: 10.1016/0022-2836(82)90515-0
124. Janin J. Surface and inside Volumes in Globular Proteins. *Nature* 1979; 277(5696): 491–492. doi: 10.1038/277491a0
125. Hopp TP, Woods KR. A Computer Program for Predicting Protein Antigenic Determinants. *Molecular Immunology* 1983; 20(4): 483–489. doi: 10.1016/0161-5890(83)90029-9

126. Engelman DM, Steitz TA, Goldman A. Identifying Nonpolar Transbilayer Helices in Amino Acid Sequences of Membrane Proteins. *Annual Review of Biophysics and Biophysical Chemistry* 1986; 15: 321–353. doi: 10.1146/annurev.bb.15.060186.001541
127. Cornette JL, Cease KB, Margalit H, Spouge JL, Berzofsky JA, DeLisi C. Hydrophobicity Scales and Computational Techniques for Detecting Amphipathic Structures in Proteins. *Journal of Molecular Biology* 1987; 195(3): 659–685. doi: 10.1016/0022-2836(87)90189-6
128. Irudayam SJ, Henchman RH. Entropic cost of protein-ligand binding and its dependence on the entropy in solution. *J. Phys. Chem. B* 2009; 113: 5871-5884.
129. Irwin BWJ, Vukovic S, Payne MC, Huggins DJ. Large-Scale Study of Hydration Environments through Hydration Sites. *The Journal of Physical Chemistry B* 2019. doi: 10.1021/acs.jpcc.9b02490
130. Trevino SR, Scholtz JM, Pace CN. Amino Acid Contribution to Protein Solubility: Asp, Glu, and Ser Contribute More Favorably than the Other Hydrophilic Amino Acids in RNase Sa. *Journal of Molecular Biology* 2007; 366(2): 449–460. doi: 10.1016/j.jmb.2006.10.026
131. DasSarma S, DasSarma P. Halophiles and Their Enzymes: Negativity Put to Good Use. *Current Opinion in Microbiology* 2015; 25: 120–126. doi: 10.1016/j.mib.2015.05.009
132. Dwyer DS. Nearest-neighbor effects and structural preferences in dipeptides are a function of the electronic properties of amino acid side-chains. *Proteins: Structure, Function, and Bioinformatics* 2006; 63(4): 939–948. doi: <https://doi.org/10.1002/prot.20906>

4 | SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

