Quantifying solvation entropy in molecular recognition

Recognition between macromolecules and ligands is at the foundation of every biological system. While we can catalogue the enthalpic interactions between ligands and receptors, the entropic consequences of solvent rearrangements are difficult to study in atomic detail. This need is highlighted in Calmodulin (CaM), a flexible protein that interacts with a diverse set of ligands. While Nuclear Magnetic Resonance (NMR) experiments have quantified the side chain, backbone, and ligand entropies[1], but the residual solvent entropy has resisted description. Solvent accessible surface area (ASA) has served as a proxy for solvent entropy, but this method falls short of accounting for the gap in entropy between the binding of smooth muscle myosin light chain kinase (smMLCK), a 26-peptide long natural ligand of CaM, to wildtype and E84K CaM[1]. The total entropies of smMLCK binding to wildtype and E84K are very different (-25 kJ/mol, -13 kJ/mol, respectively), but the experimentally measured protein and ligand entropies are similar. Furthermore, solvent ASA alone predicts similar solvation entropy for these two systems (0.54 kJ/mol, 0.57kJ/mol, respectively). If there was an improved way to model water molecules, then we could better predict solvent entropy, potentially explaining the experimentally measured difference in total entropy. High resolution X-ray crystallography is a rich source for information on water molecule positions around macromolecules. In particular, high resolution electron density maps allow us to model spatially overlapping alternative protein conformations, but there are no automated methods to model the corresponding partially occupied water molecules. **I hypothesize that displacement or addition of such water molecules could represent an important untapped reservoir of entropy of binding, thus providing a better estimation of the driving forces of molecular recognition.** I will design a method to simultaneously model protein alternative conformations and partially occupied water molecules. I will then use this method to investigate the disparity in solvation entropy between wildtype and E84K CaM binding affinities to smMLCK that I hypothesize is due to the release of partially occupied water molecules.

Aim 1. Design, implement, and validate a method to simultaneously place alternative side chain and water configurations. My goal is to take in an electron density map and a conventional X-ray model and to return an improved model that has alternative conformations for selected side chains and water molecules focusing on the first shell of water molecules directly hydrogen bonded or arranged on the protein's surface. One limitation of my method is that it requires electron densities with resolutions better than 2Å, as maps at lower resolution do not contain accurate information about water molecule occupancies. Building on my lab's previous work on side chain and ligand sampling[2], I will use linear algebra-based methods with strong constraints on minimum and total occupancy to assign water molecule positions and occupancies. For each solvent exposed residue, I will use a combination of real space filters and mixed-integer quadratic programming to determine an optimal model of 1-5 conformations chosen from ~10^{12} configurations based on an existing protein rotamer library[3] expanded with water molecule positions and geometries relative to protein functional groups[4]. To resolve clashes between neighboring residues or water molecules, I will use monte carlo sampling to assign alternative conformation labels and, finally, use existing X-ray refinement protocols[5] to optimize positions in hydrogen bonding networks. I will use

synthetic datasets to tune the behavior of my method under increasing noise. I will then test my method on multiple high-resolution X-ray structures to determine if it can place known alternative water molecule positions. Updated models will be assessed by cross validation metrics such as R_{free} and local density correlation metrics.

Aim 2. Quantify the solvation entropy difference across Calmodulin (CaM) structures.

Applying my method built in Aim 1, I will examine the water molecule placement and occupancies in CaM without any ligand (PDB:1XO2) and CaM bound to various peptides, including smMLCK. Using updated X-ray models that account for partially occupied solvent, I will calculate the entropic contributions of peptide binding to each complex. To assess water molecule positional entropy, I will use previously calculated values of displacing each water molecule, weighted based on occupancy[6]. As an alternative, I would use an ensemble refinement method[5], but that would come with caveats about correlated movements and the inability to iteratively improve models. Additionally, I will crystalize E84K CaM bound to smMLCK to assess water molecule placement and occupancies, and thus the solvation entropy. For the complexes previously measured with experimental NMR and isothermal titration calorimetry (ITC) data[7], I expect my updated solvation entropy estimation will help explain the gap between the experimental and predicted total binding entropy.

Aim 3. Experimentally measure the total binding entropy of smMLCK to wildtype and mutant Calmodulin (CaM).

Based on the analysis above, I will predict residues that when mutated will displace partially occupied water molecules in the bound complex with smMLCK. I will measure the binding entropies of smMLCK to these CaM variants using ITC. Alternatively, I will use surface plasmon resonance (SPR) to measure binding affinities. However, this technique would not be ideal as it requires inferring the binding entropies in a less direct manner. In structures with partially occupied water molecules displaced due to mutation, I expect the solvation entropy to be favorable for binding, resulting in greater binding entropy (ITC) and tighter binding affinity (ITC or SPR). If the solvation entropy is not favorable, I will crystalize the mutant CaM structure bound to smMLCK to determine if other aspects of the complex changed, including side chain, backbone, or ligand entropy, due to the mutation.

Broader impacts: Solvation entropy is one of the major missing pieces to fully understand molecular recognition. My open source method to place water molecule positions and occupancies in X-ray structures will help fill this void. This method will be available on Github and as part of the widely distributed PHENIX[5] software package.

References:

1. Marlow, M. S., et al. Nat. Chem. Biol. 6, 352–358 (2010).
2. Keedy, D. A., et al.PLoS Comput. Biol. 11, e1004507 (2015).
3. Lovell, S. C., et al. Proteins 40, 389–408 (2000).
4.Inhester,T., et al. J. Chem. Inf. Model. 57, 2132–2142 (2017).
5.Adams, P. D., et al. International Tables for Crystallography 539–547 (2012).
6.Luque, I., et al. Proteins 49, 181–190 (2002).
7.Lee, A. L Nat. Struct. Biol. 7, 72–77 (2000).