Title: Molecular dynamics simulations on the Ice-Crystal inhibition of the Type III Antifreeze Protein RD1
Author: Jimmy J. Li

Introduction

The Antarctic eel pout's (Lycodichthys dearborni) Type III Antifreeze Protein (AFP), which corresponds to PDB entry 1UCS, is an intriguing illustration of how nature has adapted to cold climates.[2] The compact globular structure of this protein is primarily made up of β-sheets (about 60–70%) with shorter α-helices and turns scattered throughout. Asn14, Thr18, and Gln44 are among the polar residues that line the flat topology of 1UCS's ice-binding surface (IBS), which forms important hydrogen bonds with ice crystals. [1] Through an adsorption-inhibition mechanism, the domains and motifs are important to its function, which prevents additional crystal growth by binding to the nascent ice.[3] What it does is that the antifreeze binds irreversibly to the ice, the AFP blocks water molecules from adding to the crystal lattice, and the crystal growth is restricted to the unblocked surface between the impurities. Eventually, it leads to a localized freezing point depression, which, from a microscopic scale, is a halt in crystal growth. The primary function of RD1 is that it moves through cold-water fish's bloodstream. In there, it binds with ice and prevents the process of crystallization, which effectively lowers the freezing point of water. This mechanism prevents the formation of ice that will damage fish's blood and tissues, which is a vital survival tactic in polar environments.

Molecular dynamics (MD) simulations are a very powerful tool that allows researchers to link between static structural snapshots and the dynamic of the protein. They provide atomic-level observation of the atom interactions and conformational changes under various conditions.[4] Those characteristics make the MD simulation a good fit for the investigation of critical phenomena such as the antifreeze activity of RD1s.

In order to understand more about this type of antifreeze activity, two complementary molecular dynamics (MD) simulations were carried out. To isolate the contribution of ice nucleation and crystal growth, two different systems were frozen: The first one was the water box with the protein, and the second one was a bulk water box. This method addresses an important question: How does the antifreeze protein's conformational affect its ability to bind ice and the solvent environment? Furthermore, we can compare RD 1 with other antifreeze proteins to see how their unique structure determines the antifreeze ability.

Below are some high-resolution images of the 1UCS protein that help with this report which are rendered in orthographic mode with a contrasting white background. They highlighted some important features, such as the important residues and the distributions of hydrophilic and hydrophobic regions.

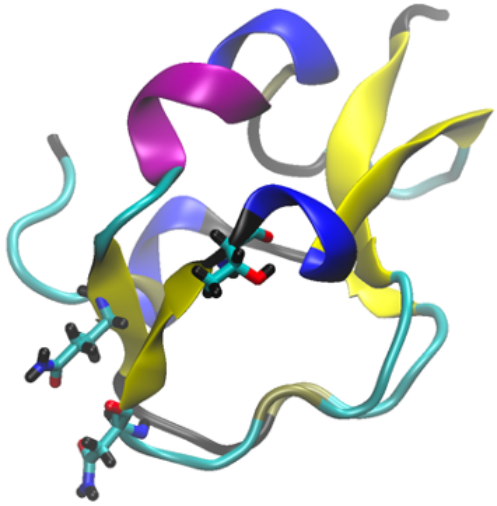

Figure 1. A picture shows the important residue of the protein

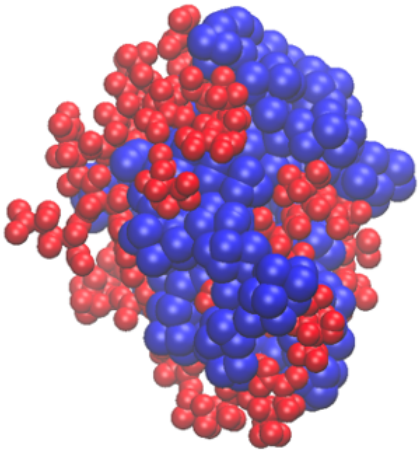

Figure 2. A picture shows the hydrophobicity of 1UCS,

Blue is the hydrophobic residues

Method

All the files used in this report are generated through VMD, besides the initial PDB and the water topology file. Two systems were prepared for this simulation: the protein with a water box and the pure water box. The initial PDB file was downloaded from The Research Collaboratory for Structural Bioinformatics Protein Data Bank (***RCSB*** PDB) website. First, we check if there aren't any missing residues in the PDB file and remove all the ions and water molecules to generate a cleaner version of the PDB for the research. Load this file to the *Auto PSF Generation Plugin* inside the VMD, creating two PDB and PSF files. Having those two files allows us to use the *Solvate Plugin*. In there, made a 100 x 100 x 100 Å cubic water box. Due to the limited calculating power, the box chosen is smaller than it should be because the nucleation process requires a large system to form a clear crystalized pattern (Hexagon). For the accuracy of this simulation, a nonstandard solvent, *TIP4P/2005*, is used. The topology file *toppar_water_ions_tip4p_2005_Modified_Shriya.str* is used in this simulation along with the choice of the solvent box side length to be *60*, and *name OH2* is used for the solvent box key selection. Next, add ion to the Auto PSF Generation's output PSF and PDB files. The addition of salt such as NaCl is not needed, because it could potentially lower the freezing point of water due to colligative properties [5]. Lastly, the protein is centered and restrained to prevent the protein from leaving the box and prevent structural distortions during the Initial Stages. Now the Protein system is prepared. In contrast, the pure water system was built following a similar procedure except for the centering and restraint steps. After generating the solvated water box, the system was left unrestrained to mimic bulk water conditions. This approach directly compared the protein-containing and pure water simulations under nearly identical environmental parameters. For the molecular dynamics simulations, the prepared system was run on the Midway2 supercomputer using NAMD 2.11*. The simulation consists of several stages: Energy minimization, heat-up equilibrium, cooldown, and 10ns stabilize. The minimization is conducted

using a one fs timestep for 2000 steps. Reducing the special clashes and optimizing the local geometry allows the later simulation to be performed from a stable starting structure and the input file. Long-range electrostatic interactions were managed using the Particle Mesh Ewald (PME) method with a grid spacing of 0.6 Å, (*PMEGridSpacing 0.6*), which is critical for accurately adapting the TIP4P/2005 water model. [5] In addition, the simulation employed the *rigidBonds all* setting to constrain all bonds involving hydrogen atoms, thereby allowing for stable integration with a one fs timestep [5]. The temperature controlled in the later simulations (NPT ensemble) was maintained through Langevin dynamics. After the minimization stage, the system was kept under NPT conditions with a target pressure of 1.01326 bar. The pressure was regulated using the *Langevin piston* method [5]. The following parameters are used: *LangevinPistonTarget 1.01325, LangevinPistonPeriod 100, LangevinPistonDecay 50*. Using these parameters, the pressure was set where the piston target pressure was set to 1.01325 bar, the oscillation period was set to 100 fs, and the damping time (decay) was set to 50 fs. In addition, the Langevin piston noise temperature was set equal to the target simulation temperature to ensure consistent thermal control. The heating and equilibration stage began with a temperature ramp from 240 K to 300 K; The temperature was incremented in 12 steps, each with a 10,000-step run, accounting for roughly 0.12 ns of heating. Followed by 0.1ns of stabilization. After that, several equilibration runs were performed in sequence to gradually relax the positional restraints on the protein over time, totaling approximately 0.9 ns. Overall, the heating and equilibration stage lasted about 1.12 ns, all under continuous NPT control.
For the cooldown phase, although the freezing point is 249K because the water does not freeze instantaneously. Instead, it undergoes supercooling, remaining in a metastable liquid state before nucleation occurs. [5] That was the reason why this phrase goes from 300 to 186K. In this phase, the temperature was decreased in 1 K increments, with each iteration running for 200,000 steps. Cooling from 300 K down to 186 K (114 iterations) thus required approximately 22.8 ns for the cooling stage—again, under the same NPT conditions. Also, the restrained and equilibration parameters were getting rid of in this stage due to the decreasing temperature acting as a restraint, which allows the protein to relax naturally. Finally, the system was stabilized at the target conditions during the cooldown stabilization stage with a 10,000,000-step run at a one fs timestep, corresponding to 10 ns. During this stage, both pressure and temperature control remained active, which led to roughly 33ns of the simulation. In the end, the *pbc unwrap -first 0* and *catdcd* file were used to unwrap the system and concatenate the different trajectoryies.

Results

Since this is primarily a cooldown simulation, the protein structure didn't experience much of a change in the conformation. For example, the salt bridge, GLU36 – LYS39, that holds the 3-10 helix or GLU 25 – LYS23, that is connecting the beta strand, has merely been broken throughout the simulation.

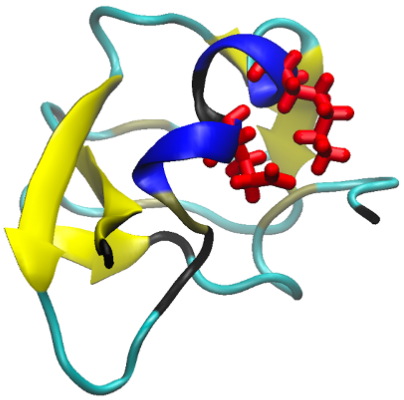

Figure 3. 1UCS protein with
the salt bridge, GLU36 – LYS39

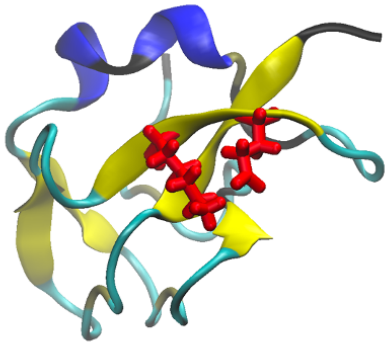

Figure 4. 1UCS protein with
the salt bridge, GLU 25 – LYS23

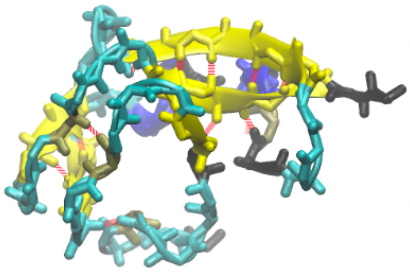

Figure 5. 1UCS protein at the beginning
of the simulation with the H-bonds displayed

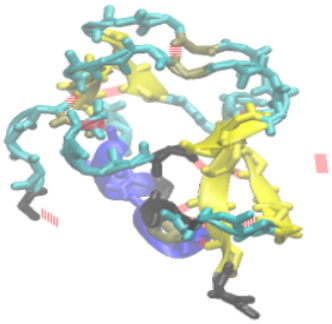

figure 6. 1UCS protein at the end
of the simulation with the H-bonds displayed

Also, as shown in the Radius of Gyration (RG) graph, which describes the overall size and compactness of a molecule, such as a protein, the graph shows the trend that matches the simulation. The radius should expand in the beginning because the protein is being heated up, which loses some hydrogen bonds, but as the cooldown system is being crystalized, it is reasonable that the protein is starting to decrease in radius.

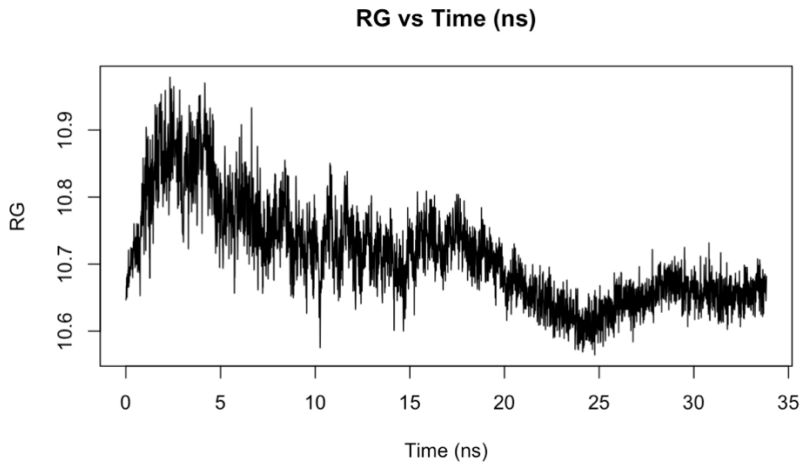

figure7. The RG plot shows the compactness of the protein

Below are the results of a quantitative measure of the effect of the 1UCS ability to inhibit ice crystal formation, which is the way of how RD1 works.

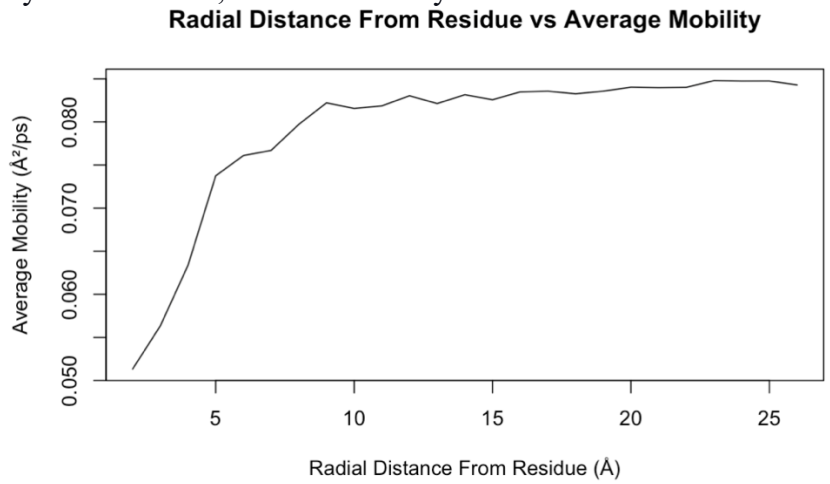

Figure 8. Average mobility as a function of Radial Distance from Protein (concatenate run)

Average mobility in this context means how much of a collection of atoms moves (vibrates) over a given time frame. In this case figure, it represents the average movement of water molecules within a shell of thickness 1Å at the radial distance indicated on the x-axis. From the graph, we observe that outside the immediate vicinity of the protein (beyond 5 Å), the mobility of water molecules remains largely unaffected by the protein's presence. This is consistent with the behavior of bulk water, where molecules are free to move without significant constraints. This can be supported in Figure 9. However, within 5 Å of the protein, the mobility of water molecules decreases significantly, indicating that the RD1 (AFP) influences the dynamics of nearby water. This behavior aligns with the known function ofRD1s. These proteins inhibit ice formation by binding to water molecules and disrupting their ability to form a crystalline lattice. The decrease in mobility of water molecules near the protein, as shown in Figure 8, supports the hypothesis that the RD1 restricts the movement of water in its vicinity, thereby preventing ice nucleation and growth.

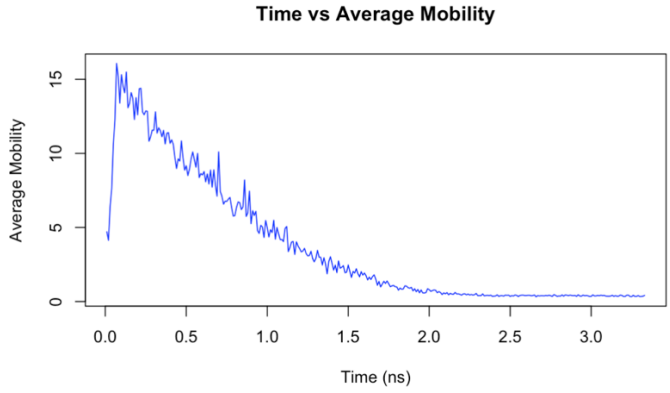

Figure 9. Average Mobility though time with the bulk water box (concatenate run)

The Average number of Hydrogen bonds can be an indication of how icy a system is. In ice, water can have a maximum of 4 hydrogen bonds with the neighboring water molecule. (In liquid, it is 1-2). So, if theRD1 is effective in the following data, it should clearly show that the bulk water box will always have more hydrogen bonds at all times.

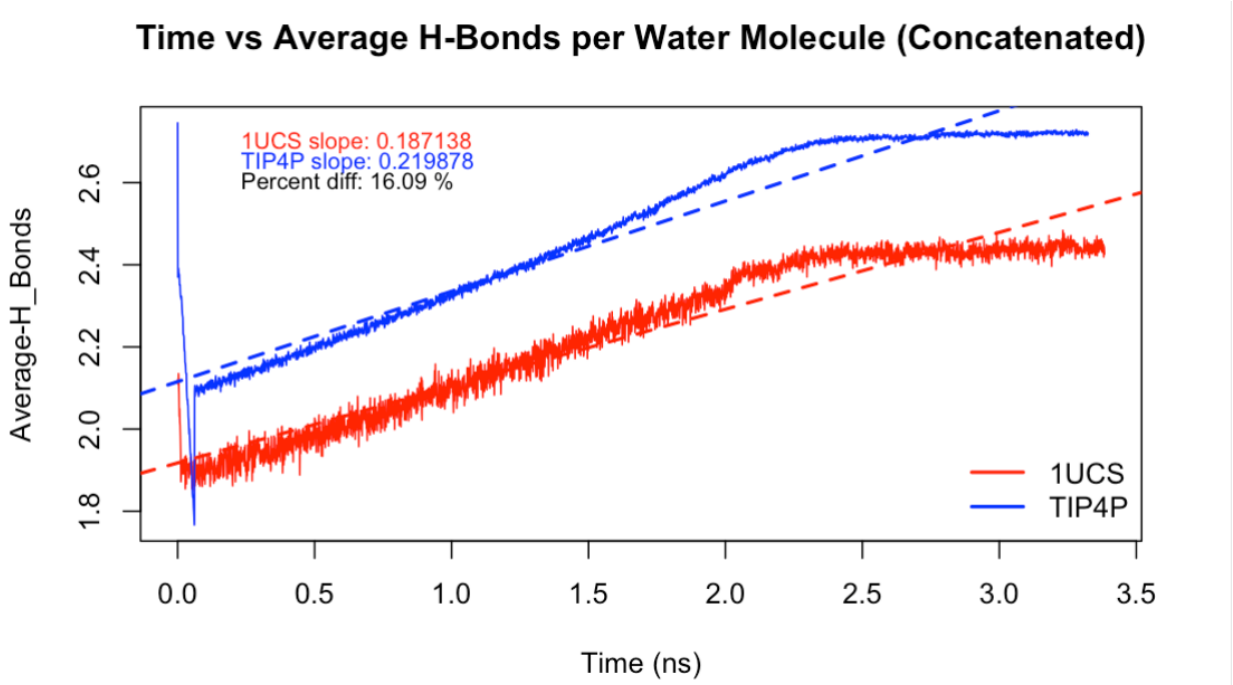

figure10. Average Number of Hydrogen Bonds Over Time with and without 1UCS (Concatenated)

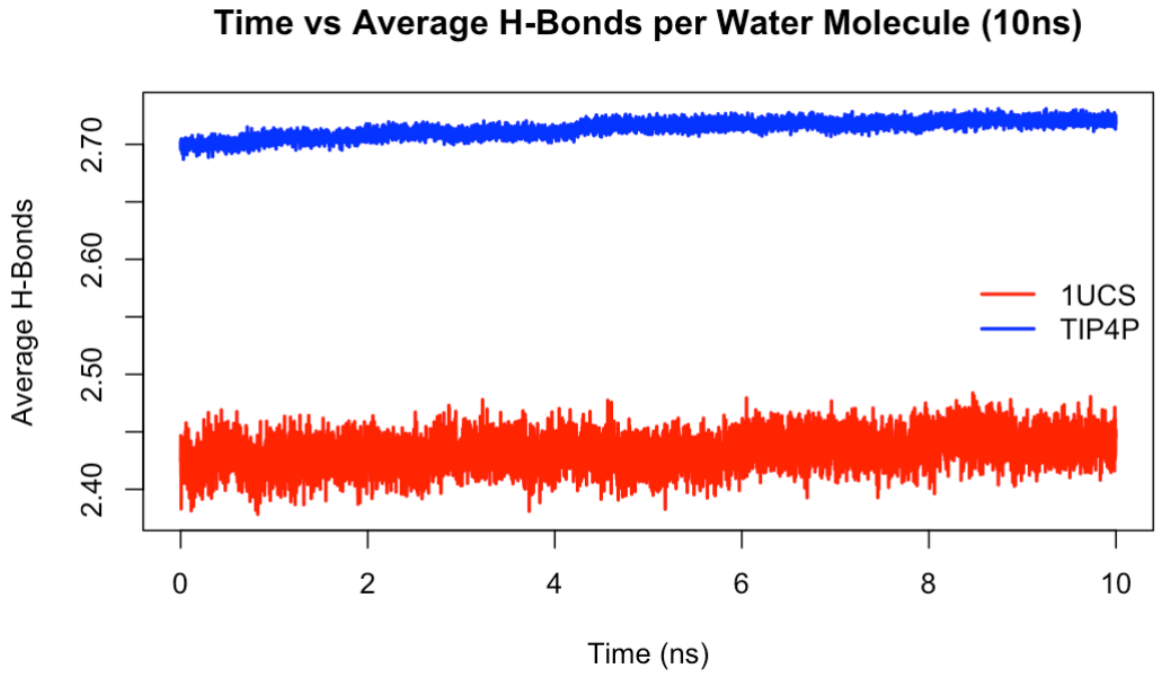

Figure 11. Average Number of Hydrogen Bonds Over Time with and without 1UCS (10ns)

From the figure above, the average hydrogen bonds in the bulk water system are far greater than those in the 1UCS system. In the concatenated graph, we can see that during the cooldown, on average, there are about 0.2 more Hydrogen bonds between the Bulk water system and the 1UCS system. There is a slope that is put in Figure 10; what makes it interesting is that the higher the slope is, the faster the system crystallizes. When comparing the two systems, we can clearly see that the bulk water box crystallizes way quicker than the system with the presence of the 1UCS protein, which is about 16%. When reaching the 10ns stabilization stage, the gap of the hydrogen bonds is even bigger. On average, the gap is roughly around 0.28.

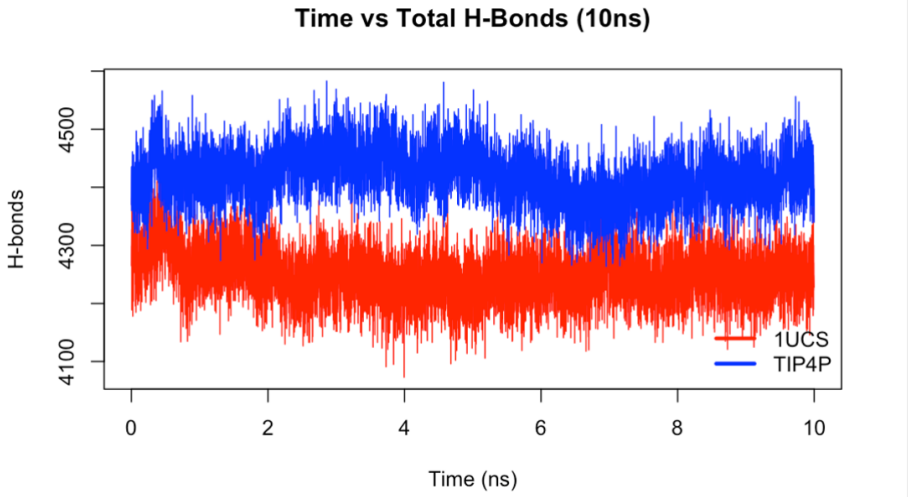

Figure 12. Total Number of Hydrogen Bonds over time with and without 1UCS (10 ns stabilization run) find the similar density of water molecules compared to the water within 10Å of protein

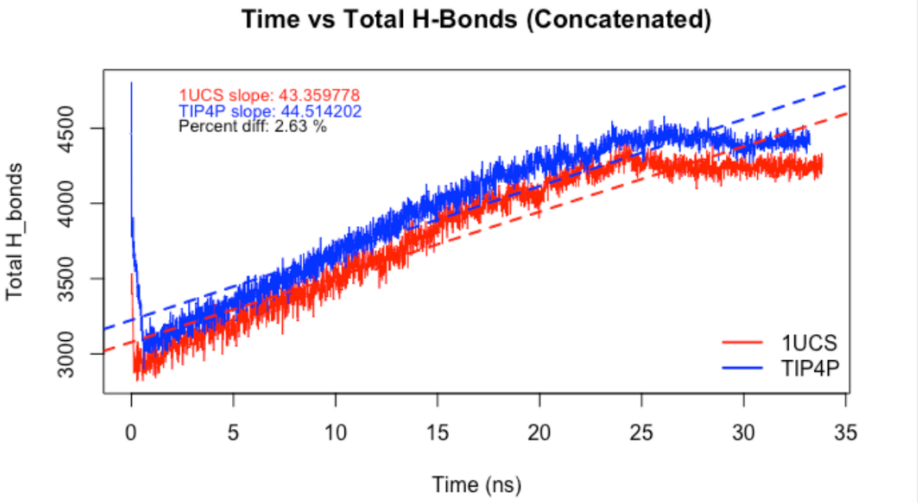

figure 13. Total Number of Hydrogen Bonds over time with and without 1UCS (concatenate); find the similar density of water molecule compared to the water within 10Å of protein

To further investigate how RD1 affects the dynamics during the crystallization, another analysis was introduced, which is to calculate the total hydrogen bonds in the same water density. What it means is that it is not the hydrogen bond over the entire water box, but rather a selection that has the same amount of water molecules that are selected with the command water within 10 of protein in the 1UCS system. The reason we do that is, as shown in Figure 8, only the water that is close to the RD1 will experience the inhabitation of crystallization, which, based on calculation, the box is about 35.6 Å^3. From the figure attached above, we can see that the difference in the hydrogen bonds between the two systems is about 200, which data suggests that RD1 does effectively lower the number of hydrogen bonds in a system and, therefore, has an inhibiting effect on ice crystallization.

Discussion

Throughout the simulation, multiple problems are presented; some are classic MD simulation problems, and some are not. Let's begin with the odd one. The first problem this simulation encountered is that the version of the NAMD. We first ran our simulation on the Midway3 supercomputer, but it turns out the NAMD 2.14 does not support the TIP4P model, which we chose for a more accurate representation of the water molecule. There is nothing we can do about it besides switch to the Midway2 supercomputer, where an older version of the NAMD supports the TIP4P model (NAMD 2.11). The rest of the challenges belong to the classic MD simulation pitfall: the Time Scale Problem and the Size Problem. Many biologically relevant processes, such as protein folding, occur on timescales longer than what is typically accessible with classical MD simulations, such as in nanoseconds or microseconds. Unfortunately, the MD simulation has to be performed in the femtosecond, which requires a gigantic amount of calculation power. The size problem has further increased this pressure because, typically, every biological system is very complex, consisting of millions or billions of atoms. As the number goes up, the calculation power needed will build up on top of the timescale problem.

Unfortunately, the cooldown system encounters those two problems more than anything else. In order for the water to be completely frozen, sufficient time and volume are required. Water freezing is a complex process that begins with nucleation, where small clusters of water molecules start to form ice-like, crystalized structures around impurities, dust, or other irregularities in the water. In very pure water, freezing can be delayed significantly, requiring much lower temperatures or external disturbances to initiate the process. So, to maximize the chance for our system to form the crystalized structure, we have to increase the size of the water box, or we need to let the simulation run longer.

From the result we obtained from the experiment, the RD1 does have the effect of inhibiting the hydrogen bond formation, which led to the fact that it can slow down the crystallization process.

Comparison

Let's compare it with another Antifreeze protein, which my colleague is doing: The Snow flea Antifreeze Protein or, in the short term, sfAFP. (PDB ID: 2PNE). sfAFP is a type of AFP found in snow fleas. Although they don't live in the frozen environment, but since they are so small, that still live in a semi-frozen environment. All methods used to get the results for RD1 were mirrored with the research presented here on sfAFP. We established with RD1 that a good measure of its crystallization inhibiting function is the percentage difference of the lines of best

fit between the system with the protein and the bulk water box system on the graph of average hydrogen bonds.

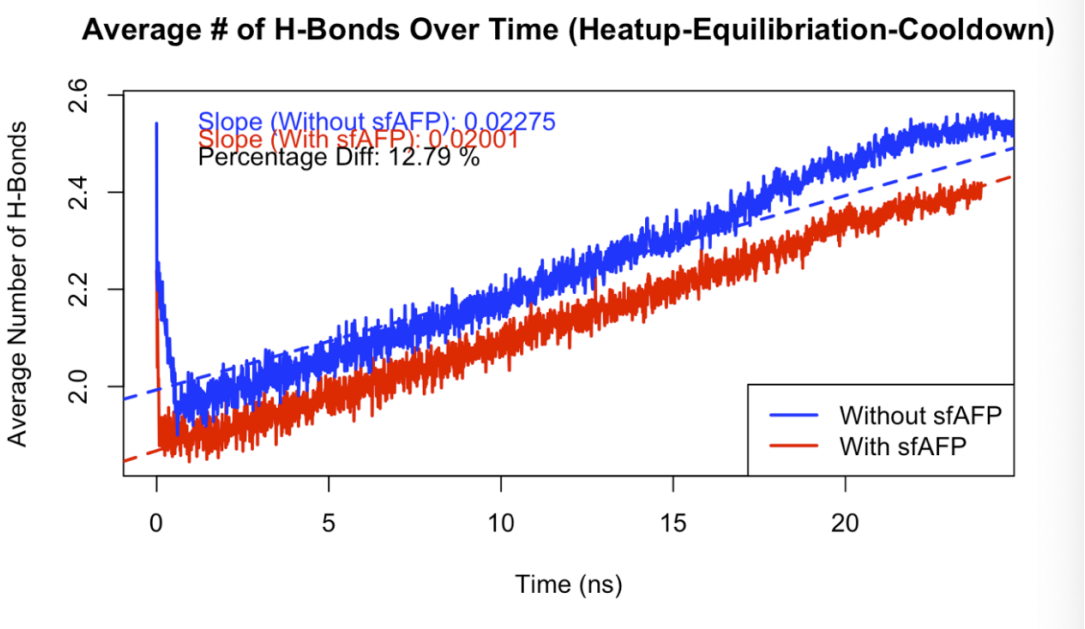

Figure 14. Average Number of Hydrogen bonds for sfAFP (heat up-equilibrium-cooldown)

The percentage difference that we found for sfAFP was 12.79%, and it is 16.09% forRD1. Since they are both being compared to the same bulk water trajectory this is a fair comparison between the two. The rate of crystallization in a waterbox with RD1 is lower than that of the sfAFP. This suggests that RD1 is more effective at reducing crystallization than sfAFP.

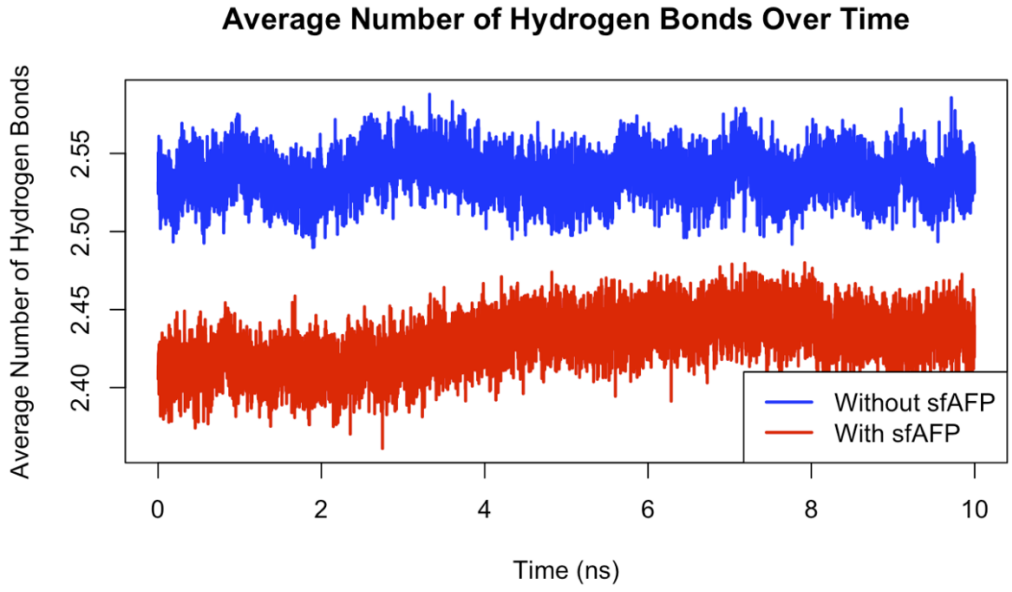

Figure 15 Average Number of Hydrogen Bonds for sfAFP (10 ns stabilization)

Indeed, this finding is reinforced by Fig.15. Where, in the stabilization run, sfAFP had 5% fewer hydrogen bonds than the bulk water box,RD1 has a 12.5% difference. This, combined with a lower rate of crystallization suggests that RD1 is a more effective antifreeze protein than sfAFP.

Conclusion

After all, RD1 does seem to have the ability to lower the rate of crystallization, which is supported by the fact that the average hydrogen bond per water molecule (within 10 of the protein) and the total hydrogen bonds in a given density of water. In comparison with the sfAFP (2PNE), theRD1 (1UCS) does show a greater potential in its antifreeze protein, likely due to the two different environments this organism lives in. However, it is worth mentioning that the amount of crystals shown is not sufficient. However, it does show potential in its function when observing the three different analyses performed in this report. Especially the various methods of measuring the total/average hydrogen bonds.

If the study can go further, without a computation limit, I would run the same analysis but under a different setting. For example, it is scarce to find a pure water environment. I would find a way to add some impurity into the water box, which seems like a way to form the ice crystal faster. Ideally, the exact environment where Lycodichthys dearborni lives, such as the PH level or the salinity in the Antarctic, because it simulates how the protein actually performs. Second, I would increase the size of the water box to 10,000 Å^3. The current water box is way too small, it's only 100 Å^3. It is not sufficient enough to simulate the real-life environment; the water that forms ice is always a big chunk. I would also like to experience it with the newest water model such as the TIP7P or the six-site water model, because it can better simulate the water in real life. Lastly, the ultimate test I want to run is to see how it would behave under absolute zero. However, it is unlikely for the classical MD to perform such a job because it relies on temperature-driven fluctuations, which would be nonexistent at absolute zero. Would the protein retain its native conformation, or would it collapse into a frozen structure?

Reference

[1] The Refined Crystal Structure of an Eel Pout Type III Antifreeze Protein RD1 at 0.62-Å Resolution Reveals Structural Microheterogeneity of Protein and Solvation Ko, Tzu-Ping et al. Biophysical Journal, Volume 84, Issue 2, 1228 – 1237

[2] https://doi.org/10.2210/pdb1UCS/pdb,

[3] Sönnichsen FD, DeLuca CI, Davies PL, Sykes BD. Refined solution structure of type III antifreeze protein: hydrophobic groups may be involved in the energetics of the protein-ice interaction. Structure. 1996 Nov 15;4(11):1325-37. doi: 10.1016/s0969-2126(96)00140-2. PMID: 8939756.

[4] Hollingsworth SA, Dror RO. Molecular Dynamics Simulation for All. Neuron. 2018 Sep 19;99(6):1129-1143. doi: 10.1016/j.neuron.2018.08.011. PMID: 30236283; PMCID: PMC6209097.

[5] James C. Phillips, David J. Hardy, Julio D. C. Maia, John E. Stone, Joao V. Ribeiro, Rafael C. Bernardi, Ronak Buch, Giacomo Fiorin, Jerome Henin, Wei Jiang, Ryan McGreevy, Marcelo C. R. Melo, Brian K. Radak, Robert D. Skeel, Abhishek Singharoy, Yi Wang, Benoit Roux, Aleksei Aksimentiev, Zaida Luthey-Schulten, Laxmikant V. Kale, Klaus Schulten, Christophe Chipot, and Emad Tajkhorshid.
Scalable molecular dynamics on CPU and GPU architectures with NAMD.
Journal of Chemical Physics, 153:044130, 2020. doi:10.1063/5.0014475
abstract, DOI