GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

Bachelor's Thesis

Berechnung der Hydrationshüllenentropie eines Anti-Frost-Proteins

Calculation of the Hydration Layer Entropy of an Antifreeze Protein

prepared by

Laura Müller

from Fritzlar

at the Max Planck Institute for Biophysical Chemistry

Thesis period: 31st May 2021 until 6th September 2021

First referee: Hon.-Prof. Dr. Karl Helmut Grubmüller

Second referee: Prof. Dr. Stefan Klumpp

Abstract

Antifreeze proteins (AFPs) are found in many animal and plant species. They protect the organisms from freezing damage by preventing the freezing of the bodily fluids. Their working mechanism is not yet fully understood, but has been subject of many studies in the past years. More information about how AFPs work would give insight into the life of organisms in cold temperatures, and could be used for food and tissue preservation.
In this thesis, the ordering of hydration water at the ice-binding site (IBS) of the Tenebrio molitor AFP (TmAFP) was studied and compared to a mutant using molecular dynamics simulations. For the mutant, threonine (Thr) residues at the IBS, which have been shown to be vital for AFP-function, were replaced with valine (Val). To study the temperature-dependence of the protein's function, all simulations were carried out at different temperatures, above and below the freezing point. The simulations were then used as a basis for the calculation of the spatially resolved entropy of the water surrounding the protein with the new method Per|Mut.
I observed ordered water molecules at the IBS of the TmAFP, which is in agreement with results of other studies, as well as a lowered entropy around the Thr. The ordered water molecules were not found for the mutant, and the entropy around the Val was higher compared to the Thr, supporting that replacing the Thr leads to a loss in antifreeze activity. I also found that the IBS of the TmAFP influences the entropy of water up to a distance of about 6 Å, which is slightly farther than the mutant. In contrast to other sides of the protein, the entropy at the IBS grows very uniformly with distance. The results show that the protein's influence on the water is temperature-dependent and the protein works better for colder temperatures, which is in agreement with its biological function.

Contents

Abbreviations

AFP antifreeze protein
IBS ice-binding site
kNN k-nearest neighbor
mdp molecular dynamics parameters
MDS molecular dynamics simulation
MID maximum influence distance
MIE mutual information expansion
Per|Mut Permutation Reduction and Mutual Information Expansion
PDB Protein Data Bank [1]
Ser serine
Thr threonine
TmAFP Tenebrio molitor antifreeze protein
Val valine
VMD Visual Molecular Dynamics [2]

1. Introduction

Antifreeze proteins (AFP) are found in insects [3], fish [4], plants [5], and other ectoterm organisms [6] that have to tolerate freezing temperatures. By lowering the freezing point of the bodily fluids, the protein protects the organism from freezing damage [7].
Even though AFPs have such an important role, their working mechanism is not yet fully understood. More knowledge about them would not only give more insight into the life of animals in cold temperatures, but it could also be applied in commercial use, such as food and tissue preservation.

There are different types of AFPs. For this study, the hyperactive AFP of the species Tenebrio molitor is used. Hyperactive AFPs are seen in insects, as they sometimes have to tolerate very low temperatures of -30°C and lower [7]. Because of this, the freezing point has to be lowered by several degrees, which is why these AFPs are called hyperactive [8]. Because of the substantial lowering of the freezing point, hyperactive AFPs are often used for studies.

Tenebrio molitor, also known as the mealworm beetle, is found worldwide. It often lives in and feeds on wheat products, and is considered a pest [9]. The Tenebrio molitor antifreeze protein (TmAFP) was chosen, because it is relatively small (2.5 nm [3]) compared to other insect AFPs [10], which means calculations can be carried out more efficiently. The protein is also used in many other studies [11] [12] [13]. This allows the comparison of the results, like the arrangement of the water molecules around the protein, to those of other studies.

Studies have shown that the water molecules at the so-called ice-binding site (IBS) are ordered. Regularly ordered threonine (Thr) residues are suggested to be responsible for the ordering [14] [11]. The water molecules bind to them and display ice-like structures, which are believed to be needed for the AFP to be able to bind

to ice. One study suggested that the ordering of the water only takes place when the protein is next to ice [13]. Other studies found the opposite [14] [11]. By simulating the protein in liquid water, I am able to test this hypothesis and see if the ordering takes place without ice. Studies also found that replacing the Thr results in a loss of antifreeze activity [8].

In this simulation study, the hydration layer entropy is calculated for the TmAFP and a mutant, which has the Thr replaced with valine. Entropy is often interpreted as a "measure of disorder". It gives more information about the ordering of the water molecules than just looking at their positions. This enables me to get further insight into the influence of the AFP on its surroundings. The focus will be on the IBS because it is the most interesting part of the protein. By calculating the spatially resolved entropy, I analyze the entropy distribution right at the IBS to find out more about the function of the Thr residues. It also allows me to asses how far the protein influences the water, which might give new information about the binding process together with further studies. The simulation with the mutant is used for comparison and allows me to learn about how vulnerable the TmAFP is to mutation.
The entropy is calculated for different temperatures, above and below the freezing point, to get information about how temperature influences the protein's function and at which temperatures it works best.

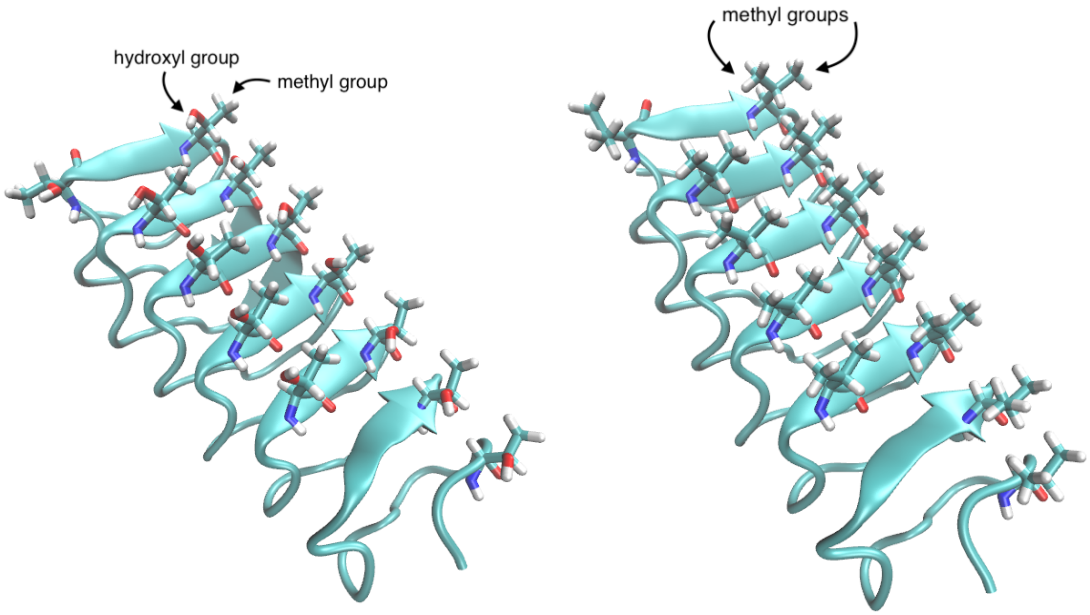
2. Background

2.1. The Tenebrio molitor antifreeze protein

A ribbon plot of the TmAFP is shown in figure 2.1 (a). It is a right-handed beta-helical protein, which consists of 84 residues. When folded, it is made up of seven 12-amino-acid loops that show extreme regularity. Both the backbone structure, and the orientations of the shown threonine (Thr) residues (sticks in the ribbon plot), are almost identical. According to Liou et al. [3], it may be the most regular protein structure observed so far (published in 2000). With its beta-helix structure, the protein is similar to other insect AFPs, for example that of the spruce budworm [10].

The protein has the shape of a flattened cylinder with a length of about 2.5 nm. This structure is an optimal foundation to achieve a regular arrangement, like the Thr residues have. The two sides of the protein are held together by disulphide bonds, which part the core into two channels. The core is further stabilized by hydrogen bonds and has nearly no hydrophobicity [3].

The Thr residues are part of a flat beta-sheet with regular ordered threonine-cystenine-threonine motifs. It is highly constrained by hydrogen bonds. The Thr residues are aligned in two parallel arrays and are pointing outwards, fully exposed to the solvent (see figure 2.1 (a)). Their precise position makes the spacing between their hydroxyl groups a near-perfect match to the ice lattice. The adjacent water molecules are ordered and have the same distance as the residues [15]. Their arrangement is especially close to the spacing between the oxygen atoms on the primary prism plane of the ice lattice and also a good match to the basal plane of ice [3]. This two-dimensional match is believed to facilitate the ice recognition process and the adsorption of the protein to the ice lattice [16], which is why this part of the protein is called the ice-binding site (IBS).



(a) TmAFP (b) mutant

Figure 2.1.: The TmAFP (left) and the mutant (right) in comparison. The proteins are drawn in cartoon style. On the IBS (arrows), the Thr residues for the wildtype and the Val residues for the mutant are drawn additionally as sticks. Their hydroxyl (OH) and methyl (CH_3) groups are labelled. White represents hydrogen, red oxygen and blue nitrogen.
The images were made using VMD.

2.2. The working mechanism of antifreeze proteins

There have been a number of recent studies on antifreeze proteins. This section is a summary about the current knowledge about AFPs, as well as possible working mechanisms, which are not yet fully understood. Because the different types of AFPs have different structures and therefore slightly different working mechanisms, the focus will be on insect AFPs, like the TmAFP.

It is currently assumed that AFPs work by an adsorption-inhibition mechanism, according to which the protein first adsorps onto the ice surface, and then inhibits ice growth [17]. The inhibition of ice growth is due to the Kelvin effect: After the proteins bind to ice nuclei, a curved ice-water interface begins to form between the proteins. This initial growth forms a convex surface, which grows until it reaches an equilibrium point, where the rate at which the water molecules bind to the ice surface is the same as their unbinding rate [7]. This is because the more the surface is curved, the fewer neighbors a molecule on the ice surface has, that hold it into place. This leads to growth inhibition up to a certain temperature and thus lowers the freezing point [17].

Another assumption is that AFPs also inhibit melting to a certain degree when a concave ice-water interface forms [18]. The difference between the melting and freezing point is called thermal hysteresis. Fish only have to tolerate temperatures that are not much lower than 1.9°C, the freezing point of sea water [7]. Insects, in contrast, sometimes have to tolerate temperatures of -30°C and lower [7]. Therefore, the thermal hysteresis gap for hyperactive insect AFPs is particularly big in contrast to other AFPs, like those of fish [19].

One side of the protein is the so called ice-binding surface or ice-binding site (IBS), with which the protein binds to ice. For the protein to be able to bind to the ice surface, the ice-binding site of the protein has to meet very complex requirements, which are not completely known to date. It is known that the IBS is rich in threonine (Thr), which occurs in regular arrays [20]. Thr consists of a non-polar, hydrophobic methyl group (CH_3) and a polar hydroxyl group (OH), which are shown in figure 2.1 (a). This mix of ordered polar and non-polar groups is needed for the hydration water to form an ordered structure, as two studies [14] [11] on the TmAFP show:

Through hydrophobic interactions, the non-polar groups are cause the formation of locally ordered waters. These waters form clathrate-like shells around the methyl groups by forming bonded rings around it. The polar groups then integrate these waters into a quasi-ice-like layered strucure through hydrogen-bonding interactions by being a member of these rings. It is to note that the structure of the ordered waters at the IBS is not identical to ice, but rather quasi-ice-like [14]. Even though the ordered water right at the IBS has been subject of multiple studies [14] [11] [16], it is not known to which extent the protein influences further layers of water.

If the Thr residues are replaced by other amino acids, the resulting mutant interacts differently with the surrounding waters, and the water structuring pattern changes. As a result, a change, or even loss, in antifreeze activity is seen [8].
The size of the IBS is also important. Ice nucleating proteins, which have the opposite effect on water than AFPs, have a similar surface, but are larger than AFPs. This allows the ordered waters on the surface to grow into ice nuclei [21], which is further discussed in section 5.1.

A third proposal is that the ordering of water at the IBS only takes place after the AFP is right next to the ice surface and in parallel alignment [13]. Before that, the protein slowly diffuses until it reaches its optimal position. For the binding, the ice-like waters at the IBS merge with the ice-like waters at the ice surface. According to this hypothesis, this occurs through a collective reorganization of the water molecules between the ice and the protein [13].
The driving force of the binding is not known, though a reduction of interfacial free energy is very likely [16].

3. Theory

3.1. Molecular dynamics simulations

Molecular systems are complex and consist of a very high number of particles. To get information about the functional motions of these systems, the trajectories of the individual particles have to be calculated. But the more particles a system has, the harder it becomes to calculate these trajectories because each particle is influenced by the other particles. For systems with more than two particles, it is impossible to solve the equations of motion analytically. Molecular dynamics simulations (MDS) are a method used to solve the movements of the particles numerically. To calculate the trajectories, the newtonian equations of motion

$$\dot{\mathbf{x}}_i(t) = \mathbf{v}_i(t),$$

$$m_i \dot{\mathbf{v}}_i(t) = \mathbf{F}_i(\{\mathbf{x}_i(t)\}) = -\vec{\nabla} V_i(\{\mathbf{x}_i(t)\})$$

are solved.
$\mathbf{x}_i(t)$, $\mathbf{v}_i(t)$ and m_i are the position, velocity and mass of an atom i at a given time t, respectively. The force $\mathbf{F}_i$ on a single atom is derived from the potential V_i, which is composed of the potentials of all interatomic interactions. $\{\mathbf{x}_i\}$ is the set of all positions.
For the integration, a leap-frog integrator [22] is used, because it has about the same computational effort as the Euler-method, but is a second-order method, whereas the Euler-method is a first-order method. The higher order is achieved by updating the positions and velocities at alternating time points:

$$\mathbf{v}(t + \Delta t/2) = \mathbf{v}(t - \Delta t/2) + \mathbf{F}(t)/m \cdot \Delta t + \mathcal{O}(\Delta t^3),$$

$$\mathbf{x}(t + \Delta t) = \mathbf{x}(t) + \mathbf{v}(t + \Delta t/2) \cdot \Delta t + \mathcal{O}(\Delta t^3).$$

Δt is the size of the integration time step.
Here, molecular dynamics simulations were carried out to calculate the trajectories of the protein and the water molecules, which were the basis of the entropy calculation.

3.2. Entropy calculation

To be able to analyze how the entropy changes with distance from the protein, and up to which distance the water molecules are influenced, the spatially resolved entropies are calculated. They also make it possible to see the impact of the Thr and Val on the first hydration layer. A possible entropy difference between the TmAFP and mutant could be observed. Together with the comparison of the IBS and non-IBS, this gives information about the nature of the TmAFP IBS and its influence on the hydration water as presented in chapter 5.

The spatially resolved entropies of the N hydration water molecules were calculated using the method Per|Mut [23]. The method uses a Boltzmann ensemble of structures, as generated by MD, and can be split into two main parts. The first part is a permutation reduction which increases sampling by $N!$ (see section 3.2.1). This drastically reduces computational efforts and makes the entropy calculation possible. In the next step, the entropy is calculated via a mutual information expansion (see section 3.2.2).

3.2.1. Permutation reduction

The size of the configuration space for a system of N particles is $\propto N^N \times SO(3)^N$, where the first part is the translational contribution and the second part the rotational contribution. $SO(3)$ is the 3-dimensional rotation group. This size of the configuration space comes from the fact that normally, in a naive integration of the configuration space, the permutation symmetry of the identical particles is not taken into account and all $N!$ physically identical microstates, that only differ by a permutaion of the labels, are counted. This makes the configuration space volume unneccessarily large and numerical calculations impossible. In analytical treatment, the problem is solved by dividing by the Gibbs factor $N!$. But this cannot be

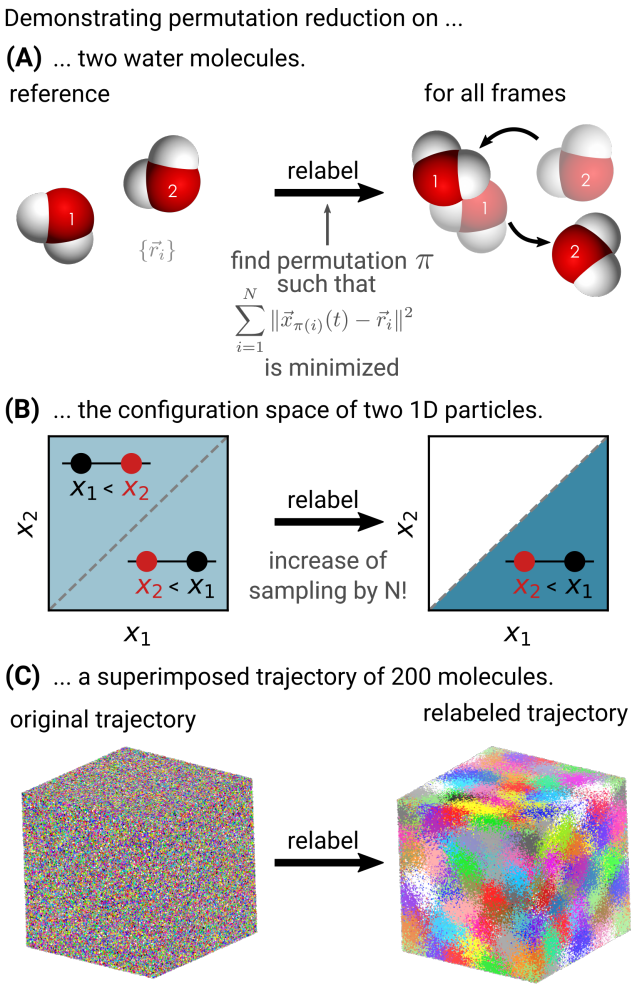


Figure 3.1.: Demonstration of permutation reduction. (A) Relabeling of two water molecules with a set reference frame (left). The molecules are relabelled so that distances to the reference molecules with the same labels are minimized. (B) The two microstates only differ by a permutation of the labels. This makes it possible to map the trajectory into one of the two identical subvolumes of the configuration space. (C) The original and relabeled trajectory for an ensemble of water molecules demonstrated with colors.
The picture was taken from [24] with permission.

done in numerical treatment. As an alternative, permutation reduction takes advantage of the fact that each microstate is physically indistinguishable from the $N! - 1$ microstates that result from swapping the labels of particles. The permutation symmetry implies that the configuration space can be subdivided into $N!$ identically subvolumes, where the configuration space density functions are identical. By permuting the labels of the identical particles, the trajectory is mapped into

one of these $N!$ identical subvolumes of the configurational space (demonstrated for two particles in figure 3.1 B), producing a permutationally reduced trajectory. This reduces the required amount of sampling by $N!$ [23]. The transformed trajectory then makes density-based entropy estimation methods possible.
Permutation reduction is especially made for diffusive systems of identicial particles, like the hydration water around the protein [25].

To define the permutation for each frame, first, a reference frame is set, for example the positions of the molecules at the beginning of the simulation. Then, for each time frame a permutation π is chosen, so that the distances to the reference molecules with the same labels are minimized, and the molecules are relabelled accordingly, like visualized in figure 3.1 A. The result is the total trajectory mapped into one of the identical subvolumes of the configuration space that is smaller by a factor of $N!$. A consequence is that the positions of each molecule over time are located in a small three-dimensional space around the reference positions. This reduces statistical errors and makes it possible to calculate spatially resolved entropies [23].

3.2.2. Mutual information expansion

To make the calculation of the entropies of the permutated trajectories possible, the high-dimensional integral for the entropy

$$S = -k_{\mathrm{B}} \langle \log \rho \rangle$$

is expanded into multiple integrals in lower dimensions over marginal distributions. This is called mutual information expansion (MIE) [24].

The total entropy S_{tot} is first split into a kinetic term S_{kin} and a configurational term S_{conf}. The kinetic part of the entropy can be calculated analytically. The configurational term then is further split into translational contributions S_{trans}, rotational contributions S_{rot} and a mutual information term $-I_{\mathrm{trans\text{-}rot}}$. The translational and

rotational terms are both approximated by a third-order MIE

$$S_{\text{conf}} \approx \sum_{i=1}^{N} S_1(i) - \sum_{(j,k)\in\text{pairs}} (S_1(j) + S_1(k) - S_2(j,k)) +$$

$$\sum_{(l,m,n)\in\text{triples}} (S_1(l) + S_1(m) + S_1(n) - S_2(l,m) - S_2(m,n) - S_2(l,n) + S_3(l,m,n)) \text{[23]}, \tag{3.1}$$

which consists of three terms. The first term is the sum of the entropies of the single molecules. The second term is the loss of entropy because of correlations between molecule pairs and the third term is the triple correlation entropy contribution [26]. Higher terms are neglected because they would make the calculation impossible and make negligible contributions compared to the first three terms. i, j, k, l, m, n are the labels of the molecules and have values from 1 to N. S_p are the entropies of the marginal configuration space probability densities of p molecules, i.e., the density, projected down onto the configurational subspace of p molecules. The distinction between the different terms gives information about the physical origin of entropy changes.

The marginal entropy of a trajectory is approximated using a k-nearest neighbor (kNN) density estimator [24]:

$$\begin{aligned} \frac{S_p}{k_\text{B}} &= -\langle \log \rho_p \rangle \\ &\approx -\left\langle \log\left(\frac{k}{(n_f-1)V(\boldsymbol{r}_{i,k})}\right)\right\rangle \\ &\approx -\frac{1}{n_f}\sum_{i=1}^{n_f} \log\left(\frac{k}{(n_f-1)V(\boldsymbol{r}_{i,k})}\right) - \Psi(k) + \log k \quad \text{[23]}, \end{aligned} \tag{3.2}$$

where n_f is the number of configurations in a $3p$-dimensional configuration space. k_B is the Boltzmann constant. For the translational part, $\boldsymbol{r}_{i,k}$ is the distance from the configuration $\boldsymbol{x}_i \in \mathbb{R}^{3p}$ to its kth neighbor in the $3p$-dimensional configuration space using the Euclidean metric. k is a fixed positive integer. V is the volume of the $(3p-1)$-dimensional sphere with radius $\boldsymbol{r}_{i,k}$. For the rotational part, quaternions $\boldsymbol{q}_i \in SO(3)$ are used as a parametrization of orientations in 3-dimensional space together with the quaternion metric. The volumes V are calculated corresponding to the quaternion metric. $-\Psi(k) + \log k$ is a correction term introduced by Kraskov et al. [27], which takes into account that the kth neighbor is on the edge of the spheres. Ψ is the digamma function, the logarithmic derivative of the gamma function.

The kNN density estimator works by finding the k-nearest neighbor in configuration space for each frame, then dividing k by $V(\boldsymbol{r}_{i,k})$. This is based on the assumption that the density in V is approximately constant. Therefore, it is important that the frames of the molecules' positions do not leave a detectable trace in the configuration space. This uneven distribution throughout the frames would lead to an inaccurate density estimation and consequently an inaccurate entropy. The detection of a possible trace and avoidance of the problem are further explained in section 4.3.2, where the transformed coordinates are reviewed.

An MIE is also used for the translation-rotation correlation term

$$I_{\text{trans-rot}} \approx \sum_{(j,\tilde{k})\in\text{pairs}} (S_1(j) + S_1(\tilde{k}) - S_2(j,\tilde{k})) \quad [26],$$

which takes the correlation between translational and rotational motions into account. j and $\tilde{k}$ are the translational and rotational degrees of freedom, respectively. For the calculation of S_p, a composite metric and the corresponding volume are used in equation (3.2).

In the last step, all the separately calculated terms are added up to get the total entropy:

$$S_{\text{tot}} = S_{\text{trans, kin}} + S_{\text{trans, conf}} + S_{\text{rot, kin}} + S_{\text{rot, conf}} - I_{\text{trans-rot}}. \tag{3.3}$$

The use of the quaternion and composite metric and the corresponding volumes are explained in [26] in detail.

4. Methods

4.1. Mutant

The regularly arranged Thr residues have been shown to be responsible for the ordering of the water molecules at the IBS [28] [14] [11]. The amino acid valine (Val) is structurally similar to Thr, but contains two methyl groups instead of one methyl and one hydroxyl group (see figure 2.1), and is consequently fully hydrophobic. It is known that changing the Thr to Val therefore results in a loss of antifreeze activity [8].
To quantify the role of the Thr residues on the IBS and how they influence the water molecules around the protein, all simulations were carried out with the wild type, as well as a mutant, which had the Thr replaced with Val as shown in figure 2.1. The residues were mutated using the residue mutation option in VMD [2]. Four Thr residues, that are not on the IBS, were kept.

Another interesting mutation would be the substitution of Thr by serine (Ser), which is also structurally similar to Thr, but has no methyl group. However, within the scope of this thesis, I limited myself to the simulation and analysis of only one mutant.

4.2. Molecular dynamics simulation

All MD simulations were carried out using the software package GROMACS [29], which is especially designed to simulate proteins and other biochemical molecules. The structure for the TmAFP was downloaded from the RCSB protein data bank (PDB code: 1EZG [1]).

4.2.1. Preparation

First, the protein has to be fully hydrated. The solvation was carried out with GROMACS. TIP3P [30] was used for the water. It is a 3-site model, which means that each of the three atoms of the water molecule has a point charge and acts as an interaction point. The molecule has an angle of 104.52 degrees, which corresponds to the observed angle. For the force-field, the CHARMM36m model [31] was used. The proteins were put in a box of size 6.4 nm x 4.8 nm x 4.3 nm. To the wildtype, 3982 water molecules were added; to the mutant, 3990 water molecules. To neutralize the net charge of the system, two water molecules were substituted by two Na^+ ions for each system. The added Na^+ ions correspond to concentrations of approximately 28 mmol/l in both the TmAFP and mutant system.

In the next step, an energy minimization was carried out to make sure that the net force on each atom is at minimum and as close to zero as possible. Because the atoms at the mutated residues were very far apart from each other right after the mutation, this was especially important for the mutant.
Next, the potential and kinetic energies were equilibrated and the systems were brought to the right temperatures for the main simulations. For that, a 10 ns simulation at 273 K with velocity generation on was carried out. In order to study the protein at different realistic temperatures (above, at and below the freezing point), as well as to study a possible temperature-dependence of the protein function, four simulation setups at 253, 263, 273 and 283 K, respectively, were prepared. Therefor four 100 ns simulations with restrained C-alpha-atoms and activated pressure coupling were carried out.
For the pressure coupling, the Berendsen algorithm [32] was used. For both simulations, V-Rescale [33], a modified Berendsen thermostat, was used for temperature coupling. The time step for the simulations was 2 fs. Sample mdp-files with all simulation parameters are included in appendix A.

4.2.2. Main simulation

After the equilibration runs, the main simulations were carried out. For each of the four temperatures, there were two different simulations, each lasting 2 μs. To simulate a proper NPT-ensemble, the Parinello-Rahman barostat [34] was used. All

further parameters remained identical to the equilibration simulations. A sample mdp-file is included in appendix A.

4.3. Entropy calculation

4.3.1. Permutation reduction and mutual information expansion

Permutation reduction was carried out using g_permute [35]. The software choses the permutations of the water molecules for each time frame that is closest to a reference structure, as described in section 3.2.1. This mapping provides a maximally compact trajectory in configuration space.
The permutated trajectories of the protein then were separated from the trajectories of the water molecules using GROMACS. For the separation of the translational and rotational entropies, a transformation of coordinates into a translational part and quaternions for the rotational coordinates followed. As a last preparation for the MIE, the transformed coordinated were reviewed, as described in section 4.3.2. After a list of the pairs and triples was created from the permuted and transformed trajectories, the terms from eq. (3.1), the kinetic entropy terms, and the translation-rotation correlation term were calculated separately and added up like shown in eq. (3.3). The k-nearest-neighbor search was performed using the Non-Metric Space Library [36], which was developed for non-metric space searching.

4.3.2. Review of the transformed coordinates

To get an accurate estimate of the configuration space density, the molecules must not leave a detectable trace, as explained in section 3.2.2. To verify that this is not the case, the positions $\boldsymbol{x}_i(t)$ were examined for each molecule i. For each point in time t, the distance d_n between $\boldsymbol{x}_i(t)$ and $\boldsymbol{x}_i(t+\Delta t)$ and the distance d_nn between $\boldsymbol{x}_i(t)$ and $\boldsymbol{x}_i(t+2\Delta t)$ were determined. For both distances, the time-based average was calculated. The quotient of these two average distances has to be one for the trajectory not to leave a detectable trace. A quotient of one means that the distance between two positions is independent of the time, a quotient bigger or smaller than one means that d_n is bigger or smaller than d_nn, respectively, and the molecules leave a detectable trace.
For the translational coordinates, the euclidean metric $\|d\| = \sqrt{\sum d_j^2}$ was used. For the quaternions $\boldsymbol{q}$, the metric $\min\left(\|\boldsymbol{q}_j - \boldsymbol{q}_k\|, \|\boldsymbol{q}_j + \boldsymbol{q}_k\|\right)$ [37] was used.

To visualize the results, the quotients of d_{n} and d_{nn} for each molecule were put in a plot together with a constant horizontal line at one, which are shown in figures 4.1 and 4.2 for the TmAFP and mutant, respectively.

For both the TmAFP and the mutant, it can be seen that for the translational coordinates, most quotients are scattered around one, which corresponds to what is expected through statistical considerations. A few points are farther below one, which suggest a slight time-dependence. The higher the temperature, the fewer points are farther away from one. For the rotational coordinates, the mean quotient is below one, which means that the distance between two points is not time-independent. One can see that, the higher the temperature is, the higher the mean quotient is, which is because of more movement of the molecules at higher temperatures.

If the trajectory has a detectable trace, the nearest-neighbor estimator gets distorted. Therefore, to see if the time-dependence has any influence on the first order entropies, the first order entropies are calculated using only every second value, which is 50000 values. This is done to use a different, bigger time step. To get a fair comparison, as well as exclude a possible convergence phase, the first order entropies are calculated again, using the first and second half of the values, respectively. This gives 50000 values again, but with the initial time step. Because all entropies are calculated with the same amount of data, they can be compared. The entropies are shown in table 4.1, which shows the first order entropies for all water molecules, as well as the mean single molecule first order entropies.

To exclude the possibility of a convergence phase in the beginning, the entropies calculated with the first and second 50000 values are compared. It can be seen that there is no significant difference between the entropy values.
When comparing the entropies that were calculated with a bigger time-step, by only using every second value, to the entropies that were calculated with the initial time-step (first and second 50000 values), no significant differences stand out.
As a last step, the entropies calculated with 50000 values are compared to those calculated with 100000 values. The differences here are bigger than the differences between the entropy values that were calculated with different, but the same amount of values, for most rows. This suggests a sampling problem but confirms that the

differing values of the quotients do not have a significant impact on the entropy calculation. Therefore, the accuracy of the entropies is still sufficient for my purposes and the entropies calculated with 100000 values will be used.

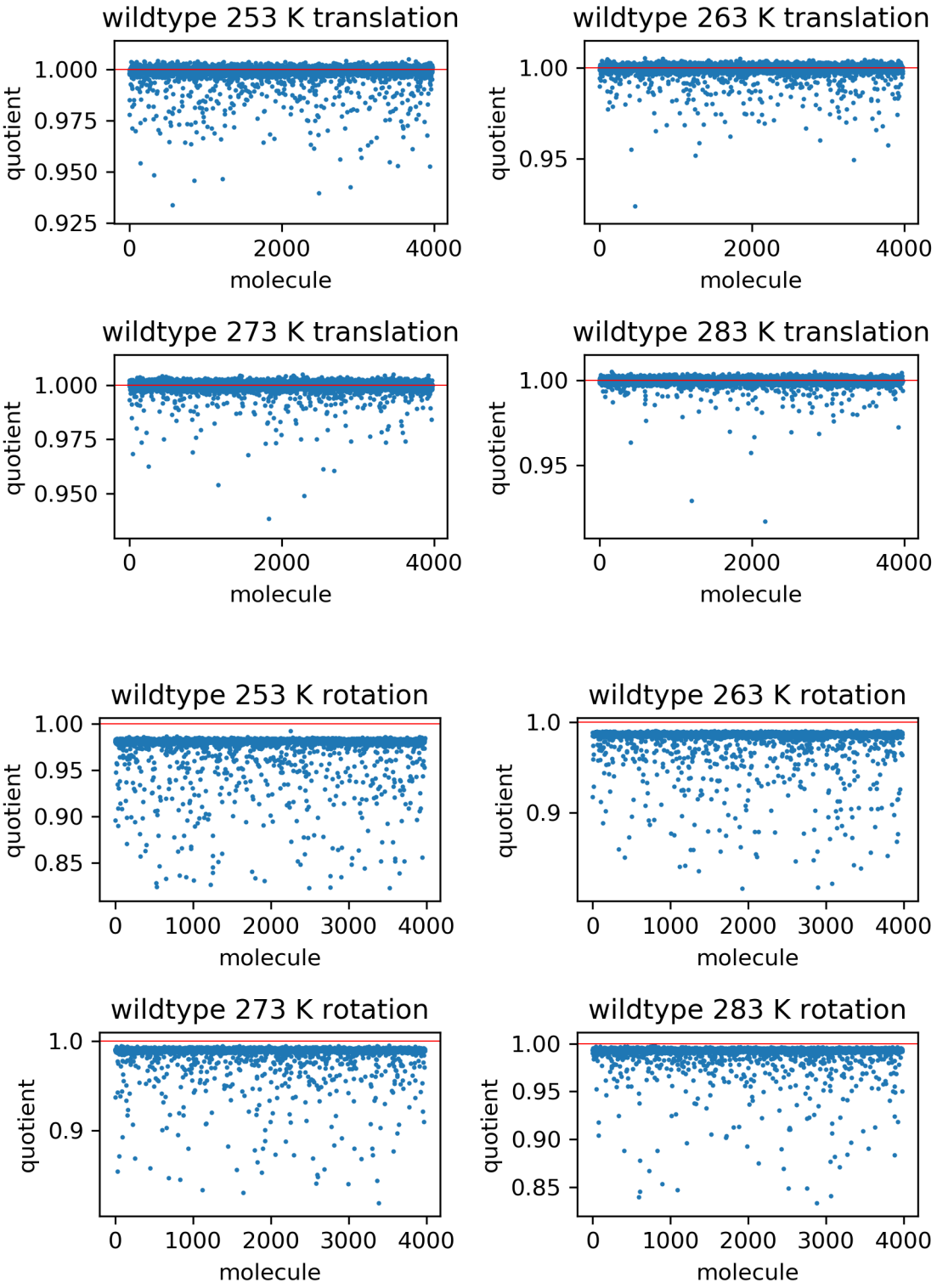


Figure 4.1.: The quotients of the time-based averages of d_{n} and d_{nn} for each molecule for the TmAFP, as explained in section 4.3.2. The molecules are labeled from 1 to 3990. A red horizontal line is added at one for better comparison.

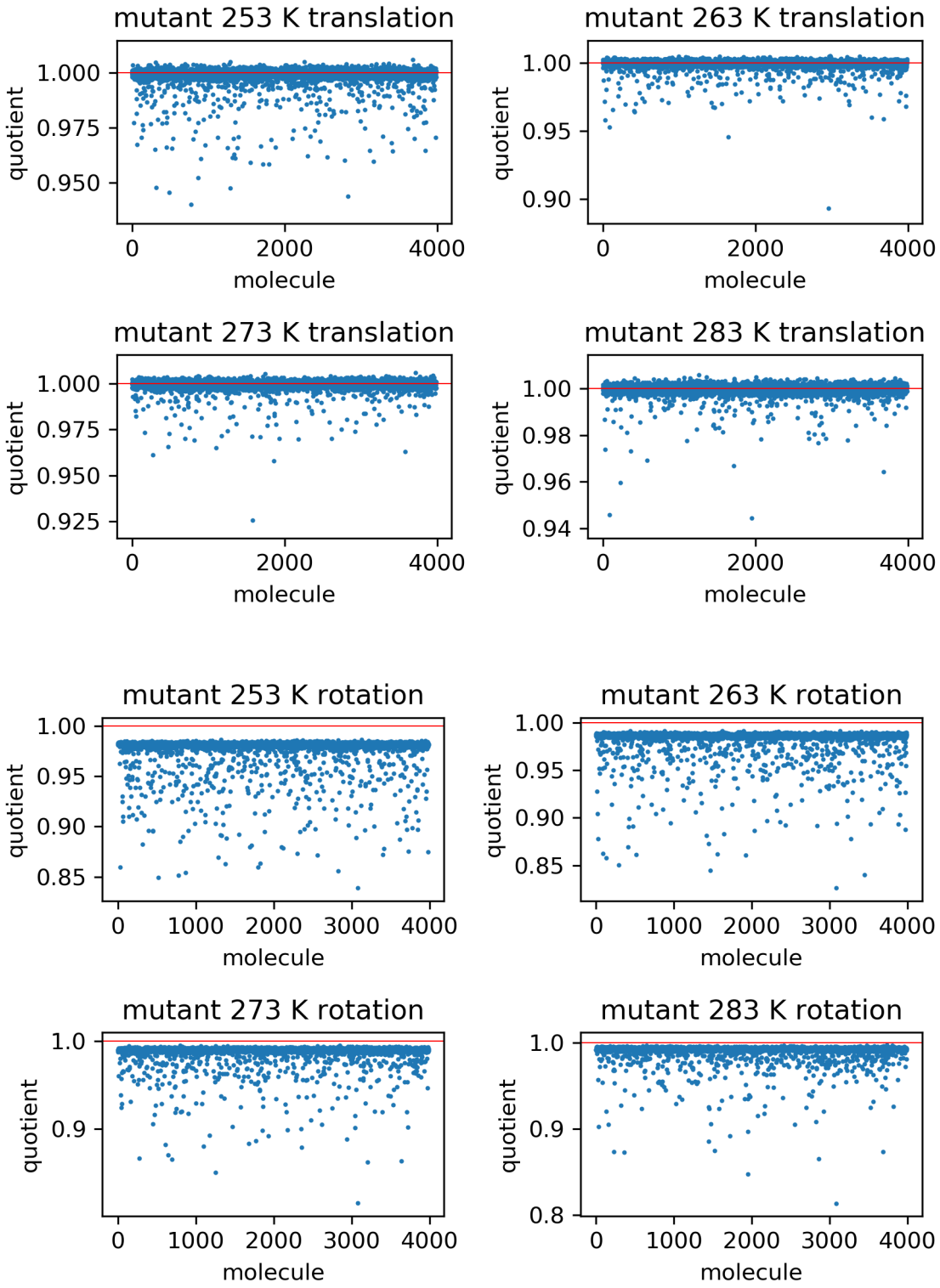


Figure 4.2.: The quotients of the time-based averages of d_{n} and d_{nn} for each molecule for the mutant, as explained in section 4.3.2. The molecules are labeled from 1 to 3982. A red horizontal line is added at one for better comparison.

First order rotational entropies

	Every second value	First half	Second half	100000
W 253 K	165038	165037	165039	165043
	41.467	41.467	41.467	41.468
W 263 K	167242	167239	167235	167243
	42.021	42.020	42.019	42.021
W 273 K	169175	169175	169171	169176
	42.506	42.506	42.505	42.506
W 283 K	171113	171117	171112	171116
	42.993	42.994	42.993	42.994
M 253 K	166320	166335	166281	166323
	41.705	41.709	41.695	41.706
M 263 K	168321	168324	168302	168324
	42.207	42.208	42.202	42.208
M 273 K	170297	170303	170281	170300
	42.702	42.704	42.698	42.703
M 283 K	172158	172162	172133	172156
	43.169	43.170	43.163	43.169

First order translational entropies

	Every second value	First half	Second half	100000
W 253 K	316417	316417	316417	316470
	79.502	79.502	79.502	79.515
W 263 K	318972	318966	318967	319020
	80.144	80.142	80.142	80.156
W 273 K	321293	321296	321291	321345
	80.727	80.728	80.726	80.740
W 283 K	323703	323700	323700	323754
	81.332	81.332	81.332	81.345
M 253 K	318968	318978	318910	319017
	79.982	79.985	79.967	79.994
M 263 K	321320	321316	321289	321371
	80.572	80.571	80.564	80.585
M 273 K	323705	323712	323672	323755
	81.170	81.172	81.161	81.182
M 283 K	325971	325968	32594	326022
	81.738	81.737	81.731	81.751

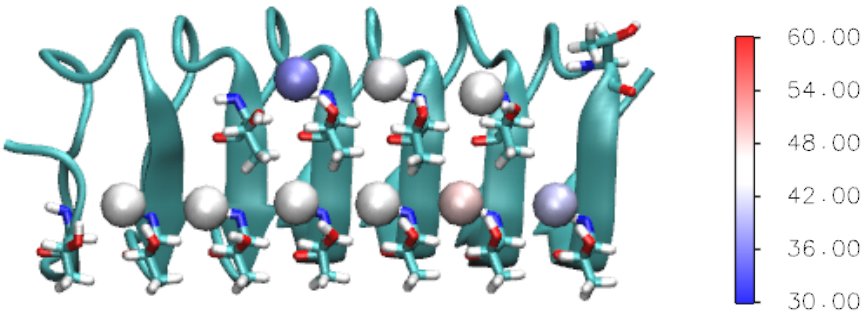
Table 4.1.: The first order entropies and the mean first order entropies per molecule. The entropies are calculated with different sets of 50000 values and 100000 values for the wildtype (W) and the mutant (M). The entropies are given in $\mathrm{Jmol^{-1}K^{-1}}$.

5. Results and discussion

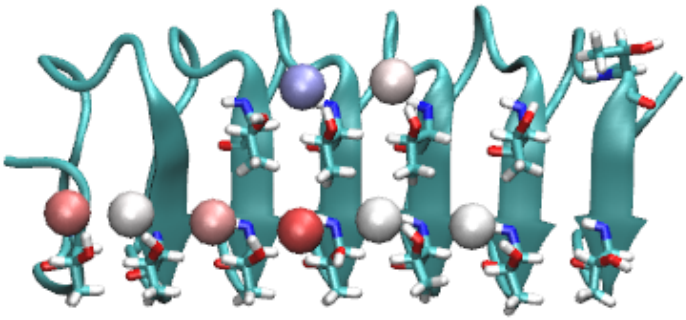
5.1. Ordered water molecules at the IBS of the TmAFP

To quantify whether pre-ordering of water molecules at the IBS takes place, I performed MD simulations, as described in section 4.2. Spatially resolved water entropies were then calculated using permutation reduction and a mutual information expansion, as described in sections 3.2.1 and 3.2.2, respectively. First, the average positions of the water molecules at the IBS are analyzed. For the TmAFP, the water molecules are arranged regularly between the Thr residues, as portrayed in picture 5.1, which shows the TmAFP and the water molecules at the IBS together with their entropies for different temperatures. This observation is consistent with findings from other studies [12] [11] [14], but disagrees with one study that proposed that ordering only takes place when the protein is next to ice [13].

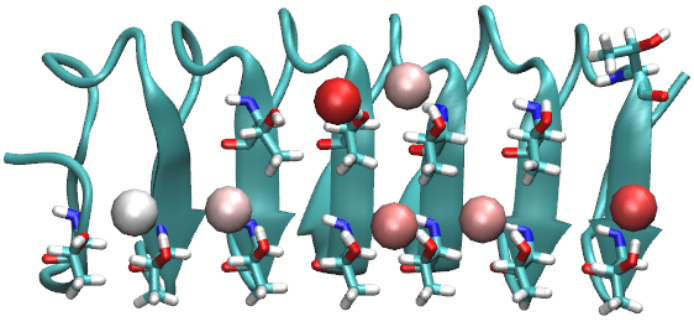
To inspect whether the ordering of the water molecules also takes place if the Thr is replaced, the simulations were also performed with a mutated version, which had the Thr replaced with Val. Previous studies have shown that a replacement of the Thr leads to a loss in antifreeze activity [8]. Figure 5.2 shows the TmAFP and the mutant from the side view, together with the average positions of the water molecules at the IBS and their entropies. Here, the ordered water at the IBS of the TmAFP can be seen from another perspective. Around the Val residues of the mutant, however, no such order is seen. There are no water molecules between the Val rows because of the hydrophobic properties of Val due to the methyl groups. Because these results coincide with previous studies, they support that my simulations are sufficiently accurate to learn something new about the antifreeze mechanism. It can also be observed that the hydration layer entropy of the TmAFP reaches much lower values than the mutant, which will be further discussed in section 5.2.



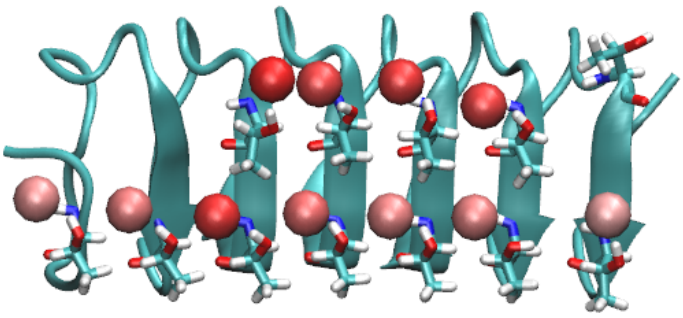
253 K



263 K



273 K



283 K

Figure 5.1.: The water molecules at the IBS of the TmAFP are arranged regularly between the Thr residues. Only molecules that fit into this regular arrangement are portrayed in this image. The positions are their average positions and their color corresponds to their entropy, which is given in $\mathrm{Jmol}^{-1}\mathrm{K}^{-1}$.
The pictures were made using VMD.

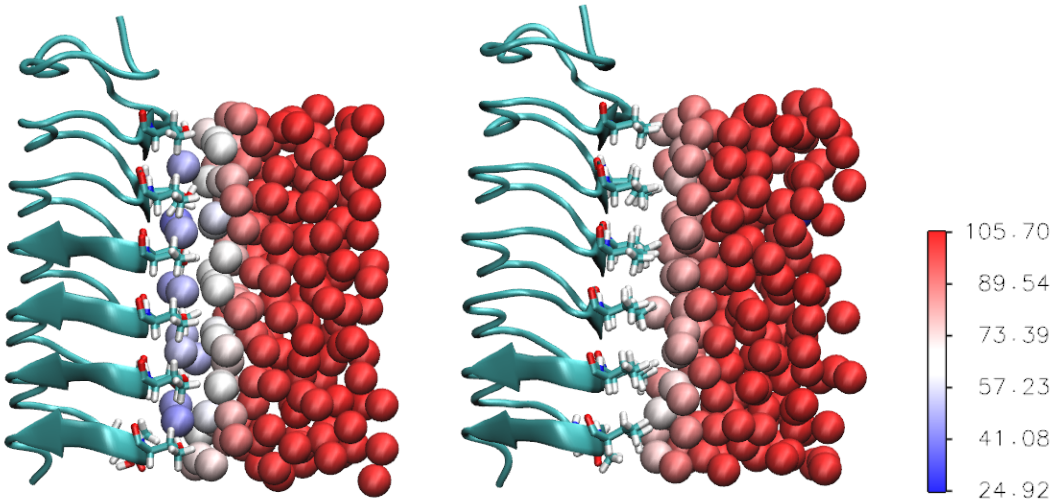


Figure 5.2.: The TmAFP and the mutant with the IBS pointing to the right. Also, the average positions of the water molecules are shown. It can be seen that the ordered water molecules at the IBS are only found for the TmAFP. Here, 253 K is used for both proteins. The colors of the water molecules correspond to their entropies, which are given in $\mathrm{Jmol^{-1}K^{-1}}$. The pictures were made using VMD.

In figure 5.2, it can also be seen that there is only one layer of ordered water molecules at the IBS of the TmAFP, suggesting that the IBS does not act as an ice nucleation site, meaning that the ice-like structures at the IBS do not continue to grow. Qiu et al. [15] found in their study that this is because of the size of the TmAFP, which has a length of 2.5 nm. For an ice nucleus to be able to grow larger, the nucleus has to be a certain size, which is called the critical ice nucleus. This size is temperature dependent. With its small size, the TmAFP is able to nucleate ice at about 240 K, which is much lower than the temperatures used in these simulations and will not be reached in nature where the mealworm beetle lives. This makes sense, because ice nucleation would be against the protein's biological function. In this study, the TmAFP therefore is not able to stabilize the critical ice nucleus and does not promote ice growth.

To find out more about the arrangement of the water molecules and its similarity to ice, the distances, which are measured in VMD, are compared to the lattice constants of hexagonal ice. The positions of the water molecules are the average

positions after the permutation reduction, which are shown in figure 5.1. For 253 K, the distances for the upper row are 4.96 and 5.70 Å and for the lower row 4.64, 5.07, 4.56, 4.64 and 5.03 Å from left to right. For 263 K, the distance for the upper row is 4.37 Å and 3.52, 4.46, 4.93, 4.41 and 5.09 Å are the distances for the lower row. For 273 K, the distance for the upper row is 4.31 Å and 5.04, 4.78 and 8.13 Å for the lower row (the distance between the second and third molecule from the right was not measured because one molecule between them is missing). For 283 K, the distances for the upper row are 2.59, 4.64 and 4.77 Å and 5.13, 4.83, 4.87, 4.61, 4.64 and 7.28 Å for the lower row.
It is noticeable that, for all temperatures, the measured distances in the lower row are similar to the lattice constant $a = 4.4969(2)$ Å of hexagonal ice crystals [38], although they mostly are slightly larger. The distances between the atoms from the upper and lower row are 6.92 and 7.08 Å for 253 K, 6.75 and 7.79 Å for 263 K, 7.22 and 7.59 Å for 273 K, and 8.44, 7.51, 7.54 and 6.42 Å for 283 K, which is similar to the lattice constant $c = 7.3211(3)$ Å [38]. The measurements again have small deviations. These distance mismatches between the ordered water and hexagonal ice have been reported before and were found to be another factor that influences the temperature at which the protein is able to nucleate ice [15]. Because of the mismatches, the molecules do not show perfect ice structures, but suggest quasi-ice-like structures, as mentioned in section 2.2.

When looking at the positions of the water molecules relative to the Thr in figure 5.1, it becomes apparent that the Thr residues all have the same orientation with the methyl group away from the water molecule. This makes sense because the methyl group is hydrophobic. The water molecules seem to bind to the hydroxyl group through hydrogen bonding, as Miya and Bandyopadhyay also suggested in their papers [14] [11].
When looking at the images and distances between the water molecules, it is noticeable that some water molecules are slightly out of line. Consequently, the distances to these water molecules are larger. To find out more about where these irregularities come from, I take a look at the corresponding Thr residues. The hydroxyl groups of the corresponding Thr residues are rotated. Some of the hydroxyl groups, that are rotated and at the edge, do not permanently bind to a water molecule. For 283 K, the second water molecule from the left in the lower row is not bound to the hydroxyl group, but still in line with the other water molecules. This is also seen for

the third water molecule from the left in the lower row for 263 K. These observations indicate that the orientation of the Thr, especially the hydroxyl groups, is more important at the edge of the protein than in the middle. They also suggest that the water molecules are not only influenced by one Thr, but also the neighbouring water molecules and/or Thr residues.

It is also noticeable that, the lower the temperature, the more ordered the water molecules appear. For 253 K, the distances between the water molecules appear to be the most consistent. This is also seen in the entropy of the water molecules, which are mostly blue and white in the image. For 263 K, the molecules are not in a near perfect line anymore and some molecules are red, showing a higher entropy. For 273 K, there are no more molecules with an entropy below 42 $\mathrm{Jmol^{-1}K^{-1}}$ (shown blue in figure 5.1) and the distances between them are very irregular. For 283 K, most molecules have an entropy of 50 $\mathrm{Jmol^{-1}K^{-1}}$ and higher (shown dark red in figure 5.1). These results are expected as higher temperatures coincide with more particle movement and a higher entropy.

Summarizing, the results show that Thr residues play an important role in the antifreeze activity of the protein, because they are responsible for the ordering of water molecules, which is needed for the protein to be able to bind to ice according to Graether et al. [8]. If the Thr residues are replaced, the water molecules are not ordered, showing that not only the regular arrangement of the amino acids, but also the amino acid itself, is vital.
The arrangement and orientation of the Thr are important, because they determine the arrangement of the water molecules. However, it depends on the position of the Thr on the IBS if a water molecule still gets in line with the other water molecules even though the corresponding hydroxyl group is rotated. For the TmAFP, the arrangement of the Thr is so that the resulting arrangement of the water molecules resembles hexagonal ice, although mismatches to the ice lattice constants are found. A different arrangement of the Thr may influence the binding to ice.
The ordering of the water also depends on the temperature. The lower the temperature is, the more consistent the water is ordered and the lower is its entropy.

5.2. Hydration layer entropy at the IBS

In order to look into the function of the Thr residues, I compare the entropy at the IBS of the TmAFP and the mutant. Therefor, the entropy of the hydration water molecules that are directly next to the protein was mapped onto the surface of the protein. The results can be seen in figure 5.3, where 253 K is used as an example. The images for the other temperatures are included in appendix B. The entropy is separated into translational and rotational contributions to get a better understanding of where the entropy difference between the TmAFP and the mutant comes from.
It is to note that the entropy values calculated with Per|Mut are not accurate. The Per|Mut entropy for bulk water is about 105 $Jmol^{-1}K^{-1}$ whereas the experimental value is about 70 $Jmol^{-1}K^{-1}$. The entropy differences, however, are far more accurate [23], which is why not the entropy values but the differences will be discussed.

It can be seen that the entropy around the rows of Thr is lower compared to the entropy right where the Thr is, for both the translational and rotational entropy. This is consistent with the observation that the water molecules around the Thr are ordered, as shown shown in section 5.1. The water molecules around the Val rows are not ordered, which is also consistent with the fact that the entropy around the Val is higher than the entropy around the Thr.
When comparing the translational and rotational entropy, it is noticeable that the translational entropy is higher than the rotational entropy. This is expected because it is also the case for the entropy of the bulk water, which is why another presentation of the results is needed to get more insight into whether the entropy difference between the TmAFP and the mutant comes from the rotational or translational contribution. For a closer inspection, the entropy difference between TmAFP and mutant is calculated. For that, the positions of the water molecules around the TmAFP are compared to those around the mutant and for each water molecule around the TmAFP a closest neighbor around the mutant is found. The entropy of the mutant then is substracted from the entropy of the TmAFP for each molecule pair and again mapped onto the protein structure. Here, the TmAFP structure is used. The resulting images are presented in figure 5.4. The entropy values go from about -24 $Jmol^{-1}K^{-1}$ (blue) over white to about 24 $Jmol^{-1}K^{-1}$ (red).

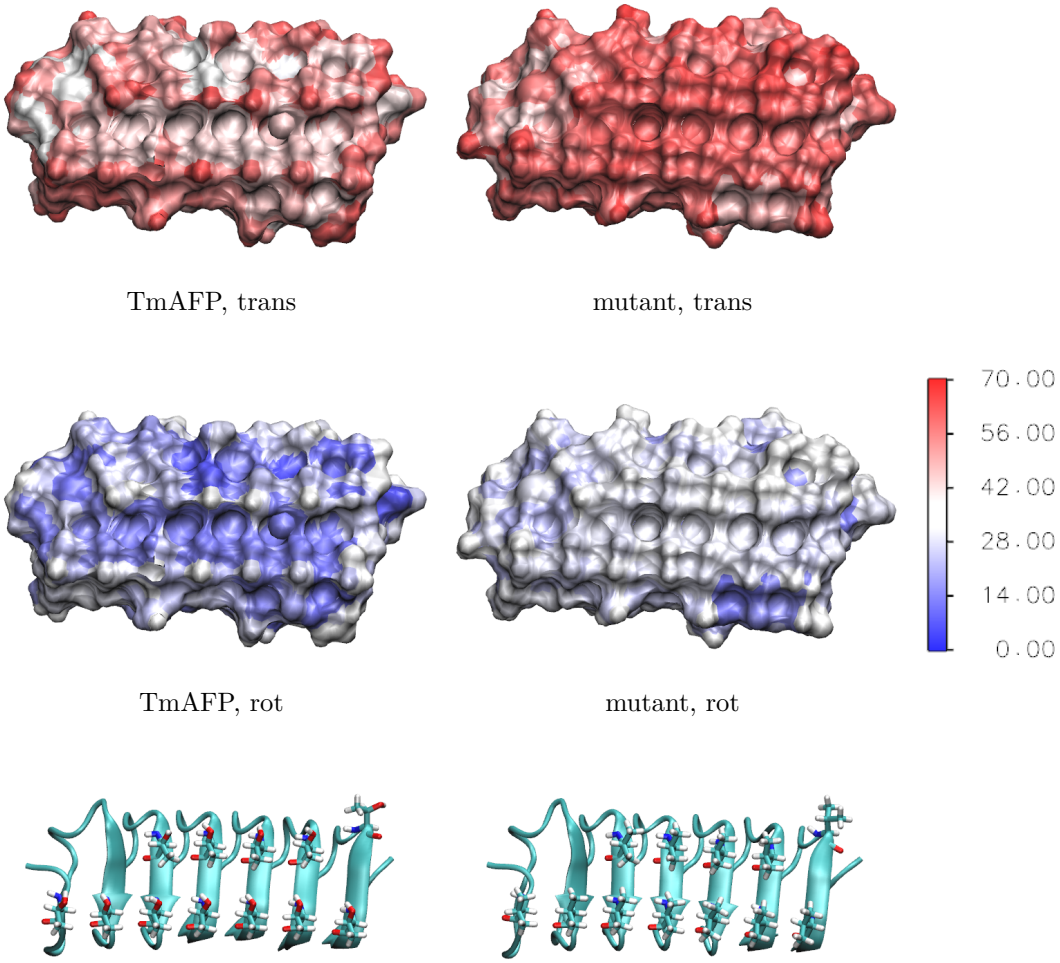


Figure 5.3.: The left and right side show the IBS of the TmAFP and mutant, respectively. The top four images show the entropy of the water molecules surrounding the protein mapped onto the surface of the protein. The first row shows the translational (trans) entropy, the second row the rotational (rot) entropy. The colors go from red ($0\,\mathrm{Jmol^{-1}K^{-1}}$) to blue ($70\,\mathrm{Jmol^{-1}K^{-1}}$). In the last row, images of the proteins are added to demonstrate the orientation.
The images correspond to a temperature of 253 K. The figures for the other temperatures are included in appendix B.
The images were made using VMD.

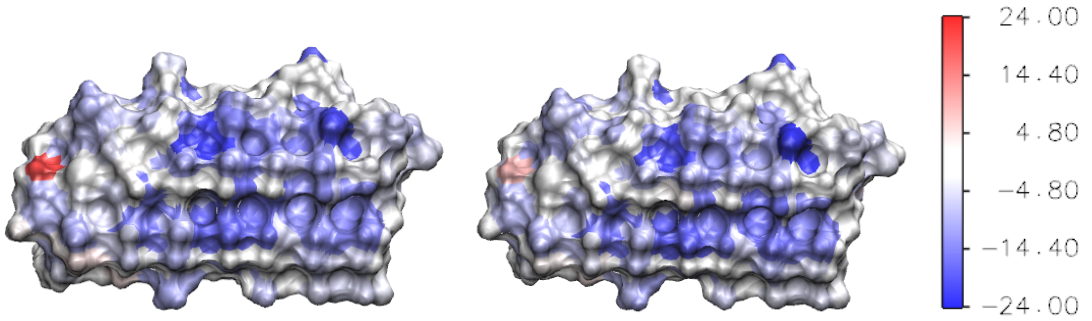


Figure 5.4.: The entropy differences at the IBS between the TmAFP and the mutant for 253 K. The translational (trans) and rotational (rot) entropy are on the left and right side, respectively, and are given in $Jmol^{-1}K^{-1}$. The images for the other temperatures are included in appendix B. The images were made using VMD.

The images again show that there is an entropy difference around the Thr/Val, where the entropy of the TmAFP is lowered. Hydration layer entropy differences in other areas are very local and are due to amino residues that are differently orientated on both proteins, which was checked in VMD.
When comparing the values of the translational and rotational entropy differences, it is noticeable that the colors are about the same. This suggests that neither the translational, nor the rotational entropy, makes a larger contribution to the entropy difference between the TmAFP and the mutant.
Because the entropy of the hydration water around the IBS is only lowered around the Thr and not the Val, the results suggest that it is lowered because of the Thr. This coincides with the results from section 5.1 and further supports the assumption that the Thr is needed for the antifreeze activity of the protein.

5.3. Hydration layer entropy at the non-IBS

To compare the entropy distribution and values at the IBS to those of an area which does not have regularly arranged amino residues, the non-IBS is used.
The entropy of the water around the non-IBS is looked at in two parts. In the first part, the entropies around the non-IBS of the TmAFP and the mutant are compared (section 5.3.1). In the second part, the entropy around the non-IBS is compared to the entropy around the IBS of the TmAFP (section 5.3.2).

5.3.1. Comparison of the entropy at the non-IBS of the TmAFP and the mutant

First, the non-IBS of the TmAFP and the mutant are compared as a control. This is used to rule out possible errors during the simulation or entropy calculation. There are no mutations on the non-IBS of the mutant, predicting no significant differences between the hydration layer entropies at the non-IBS of the TmAFP and mutant. The results, which can be seen in figure 5.6, which uses 253 K as an example, are in accordance with these expectations, showing very similar images for the TmAFP and the mutant.

For a closer inspection, the differences between the hydration layer entropies around both proteins are shown for 253 K in figure 5.5. There are no significant entropy differences, therefore the colors are mostly white. There are a few red and blue spots, indicating entropy differences. Closer inspection shows that these differences are due to rotated amino residues.

Similar results were obtained for the other temperatures, which are included in appendix C.

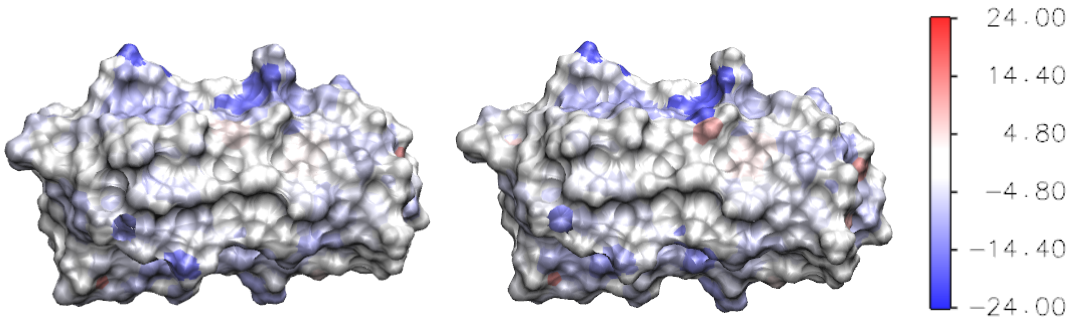


Figure 5.5.: The entropy differences at the non-IBS between the TmAFP and the mutant for 253 K. The translational (trans) and rotational (rot) entropy are on the left and right side, respectively, and are given in $Jmol^{-1}K^{-1}$. The images for the other temperatures are included in appendix C. The images were made using VMD.

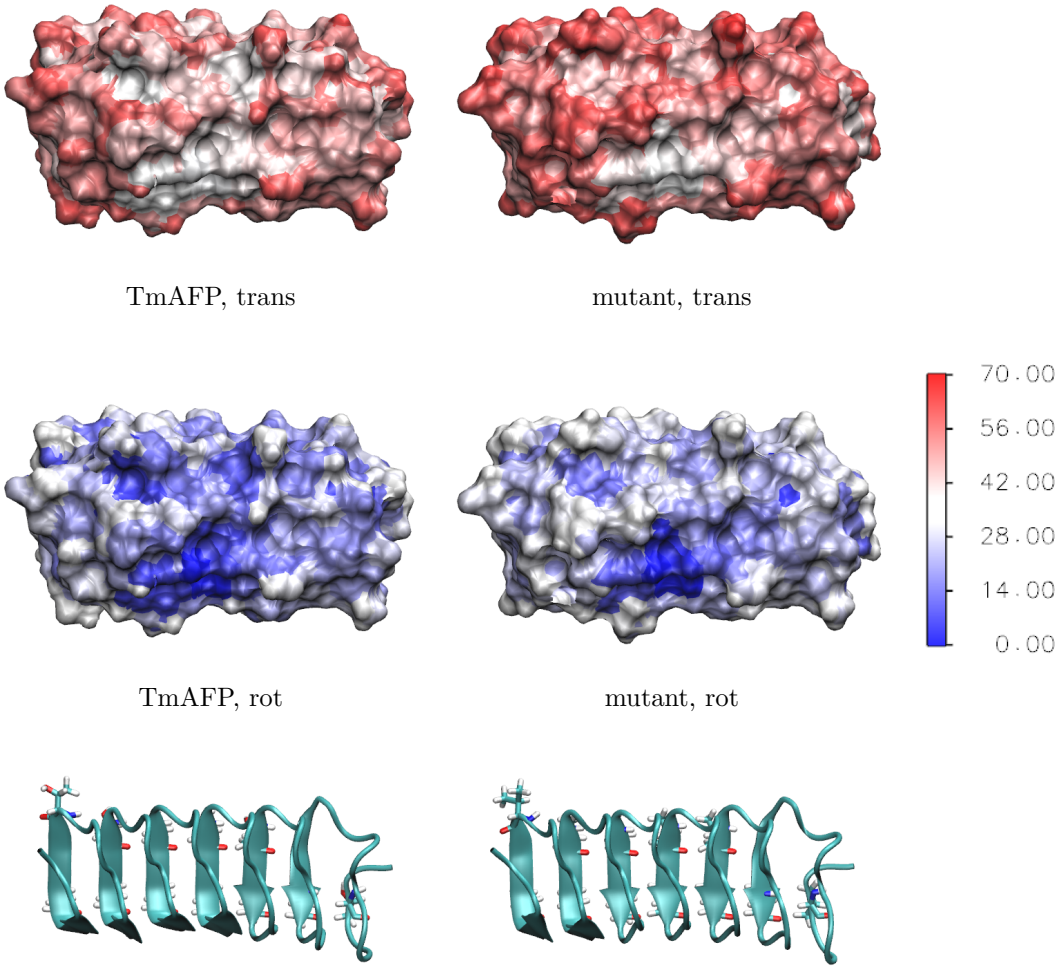


Figure 5.6.: The left and right side show the non-IBS of the TmAFP and mutant, respectively. The top four images show the entropy of the water molecules surrounding the protein mapped onto the surface of the protein. The first row shows the translational (trans) entropy, the second row the rotational (rot) entropy. The colors go from blue ($0\,\mathrm{Jmol^{-1}K^{-1}}$) to red ($70\,\mathrm{Jmol^{-1}K^{-1}}$). In the last row, images of the proteins are added to demonstrate the orientation.
The images correspond to a temperature of 253 K. The images for the other temperatures are included in appendix C.
The images were made using VMD.

5.3.2. Comparison of the entropy at the IBS and the non-IBS of the TmAFP

Next, the entropy of the hydration water around the non-IBS is compared to the entropy of the water around the IBS. Here, only the TmAFP is used to examine if the extensive uniform entropy distribution and the low entropy values between the Thr are also found in other areas.

The images of the entropies and the proteins are shown in figure 5.7. It appears that the entropy around the non-IBS is lowered as well. However, in contrast to the IBS, no regularities can be seen at the non-IBS and the colors are more patchy. In the areas where the entropy is lowered, lysine, aspartic acid, serine, threonine and asparagine are found, which are all polar or charged amino acids, making them hydrophilic. This suggests that the entropy is lowered, because water molecules are bound to amino acids. In the areas with higher entropy, a lot of methyl groups are pointing outwards, which are hydrophobic and do not bind to water molecules.

The difference in the hydration layer entropy between the IBS and non-IBS implies that both sides influence the water differently, which was expected because only the IBS has rows of regularly arranged Thr. Although low entropy values are found on both sides, extensive and evenly distributed low entropy values are only found on the IBS. Together with the observation that ordered water molecules are only found at the IBS, this means that the orientation of the protein is important and only one side of it is able to bind to ice, which is why it is called the ice-binding site.

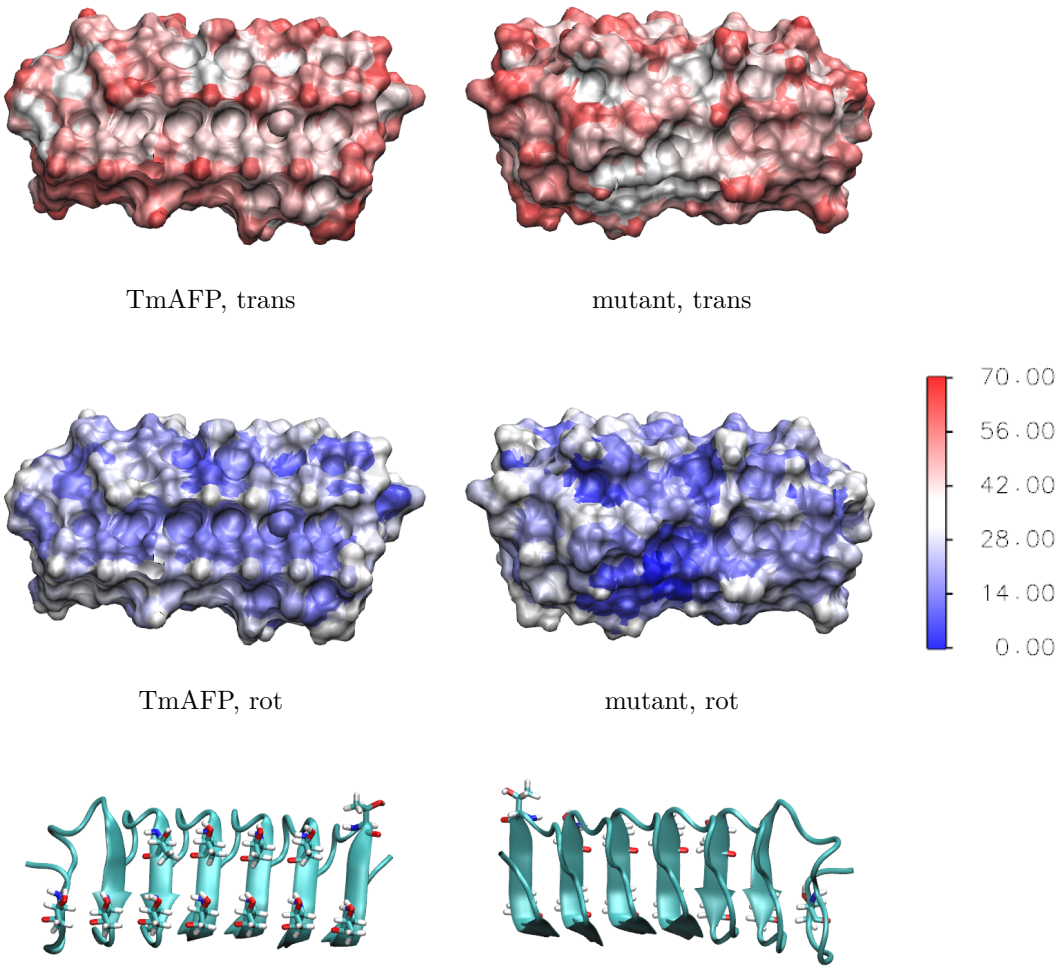


Figure 5.7.: The left and right side show the IBS and non-IBS, respectively. The top four images show the entropy of the water molecules surrounding the protein mapped onto the surface of the protein. The first row shows the translational (trans) entropy, the second row the rotational (rot) entropy. The colors go from red ($0\,\mathrm{Jmol^{-1}K^{-1}}$) to blue ($70\,\mathrm{Jmol^{-1}K^{-1}}$). In the last row, images of the proteins are added to demonstrate the orientation.
The images correspond to a temperature of 253 K. The figures for the other temperatures are included in appendix C.
The images were made using VMD.

5.4. Temperature dependence of the hydration layer entropy at the IBS of the TmAFP

To get a better understanding of the temperature-dependence of the TmAFP's function, the images of the entropy mapped onto the IBS at different temperatures are put next to each other. They are shown in figures 5.8 and 5.9 for the translational and rotational entropy, respectively. For both, it can be seen that the entropy gets higher with the temperature, as expected. This is in agreement with the results from section 5.1, where it was observed that the water molecules at the IBS are more uniformly ordered the lower the temperature is.

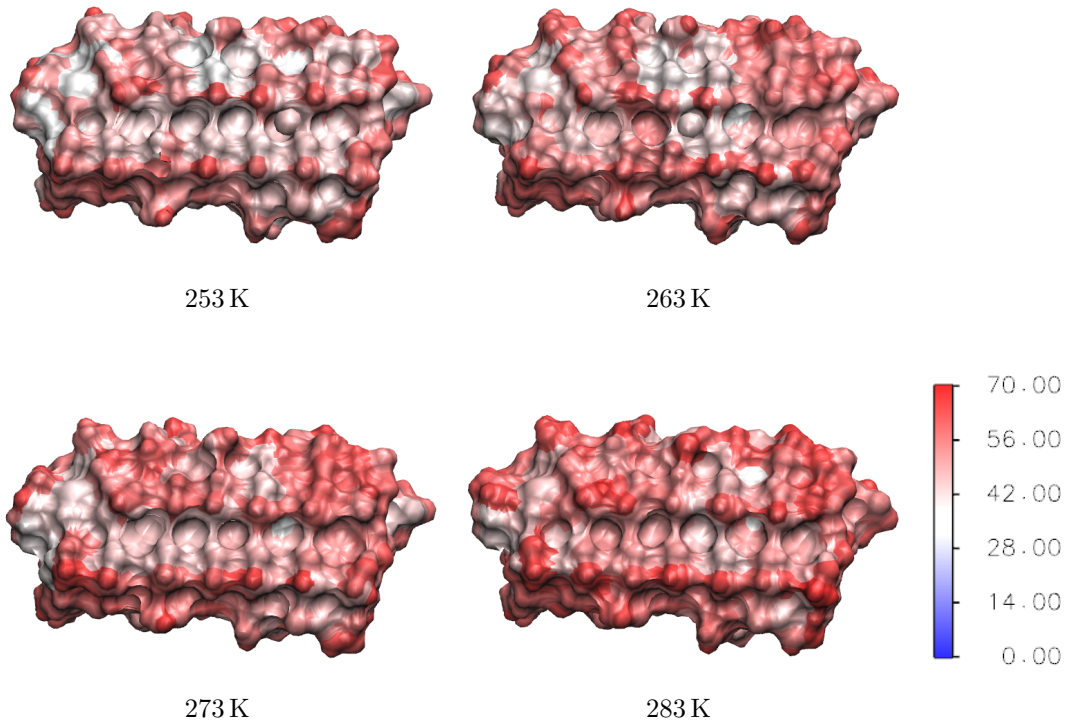


Figure 5.8.: Comparison of the translational entropy at the IBS of the TmAFP. The color bar shows entropies between 0 (blue) and 70 $\mathrm{Jmol^{-1}K^{-1}}$ (red). The images were made using VMD.

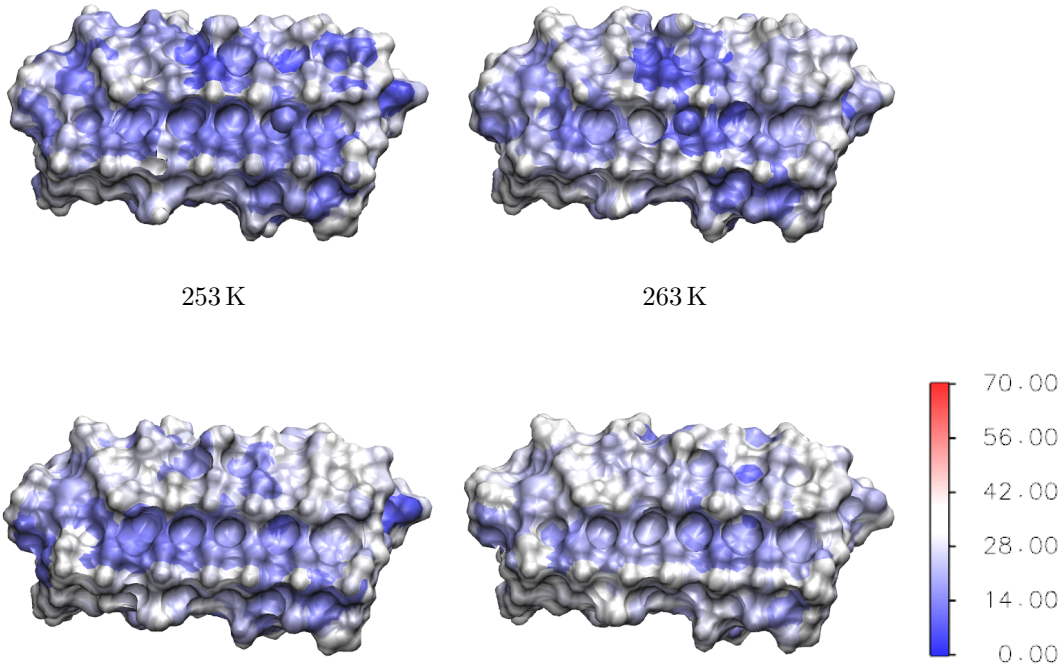


Figure 5.9.: Comparison of the rotational entropy at the IBS of the TmAFP. The color bar shows entropies between 0 (blue) and 70 $\mathrm{Jmol^{-1}K^{-1}}$ (red). The images were made using VMD.

5.5. Entropy as a function of distance

To get a better understanding of how far the protein influences water order, the entropies were calculated as a function of protein-distance, as shown in figures 5.11 and 5.12 for the TmAFP and mutant, respectively. The figures are for 253 K; the other temperatures can be found in appendix D. Each figure consists of four plots, which show the distances from the IBS, non-IBS and top and bottom of the protein. The water molecules that were used for each plot are demonstrated in figure 5.10. Each dot in the plots corresponds to one of these water molecules.

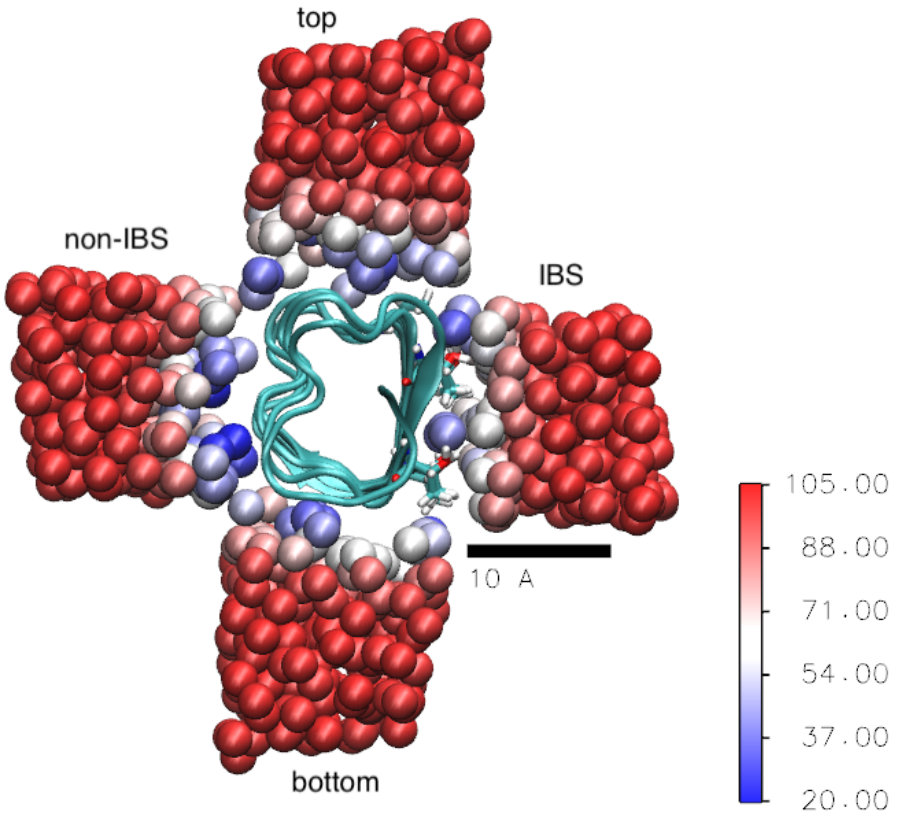


Figure 5.10.: Demonstration of the water molecules used for each side of the protein for the plots in this section. A ruler is added to give a better idea of the distances. The image shows the TmAFP at 253 K, which is used as an example. The entropy is given in $\mathrm{Jmol^{-1}K^{-1}}$.
The image was made using VMD.

5.5.1. Comparison of the distance dependence of the entropy at the IBS between the TmAFP and the mutant

To answer the question if the mutation changes the course of the entropy as a function of distance, the IBS (upper left in all figures) is discussed for the TmAFP and mutant. For the wildtype in figure 5.11, a few dots can be seen right at the IBS, which correspond to the ordered water molecules. This coincides with the low entropy of these molecules. For the mutant, the water molecules are slightly farther away (about 1 Å), as shown in figure 5.12, because no water molecules are directly at the IBS as discussed in section 5.1. The entropy values start at 70 to 80 $\mathrm{Jmol^{-1}K^{-1}}$. The entropy values at the same distance for the TmAFP are between 50 and 60 $\mathrm{Jmol^{-1}K^{-1}}$. For both proteins, the entropy values are higher for higher temperatures.
The entropy then grows with the distance until it reaches about 105 $\mathrm{Jmol^{-1}K^{-1}}$, where it stays with only small statistical fluctuations. The exact value depends on the temperature. As mentioned before, 105 $\mathrm{Jmol^{-1}K^{-1}}$ does not coincide with the experimental entropy value of water, but with the Per|Mut water entropy. The entropy of bulk water is reached at a distance of about 5 to 6 Å for the TmAFP, where the smaller distances correspond to the higher temperatures. I will call this distance the maximum influence distance (MID). The MID for the mutant seems to be slightly smaller than the MID of the TmAFP for each temperature, suggesting that the mutation influences the MID. This fits with the observation that the entropy for the mutant is higher than for the TmAFP for all distances up to the MID.

To get a better look at the entropy differences between the mutant and the TmAFP, figure 5.13 shows the difference between the entropies of the mutant and TmAFP for 253 K. The figures for the other temperatures are included in appendix D. Positive and negative entropy values correspond to a higher entropy of the TmAFP and mutant, respectively. Nearly all entropy values are below zero up to the MID, supporting the observations that the mutant has higher entropy values for all distances and a shorter MID than the TmAFP. Because the differences decrease with distance, it shows that the TmAFP has a greater slope. The density of the dots in figures 5.11 and 5.12 up to the MID is higher for the mutant than for the TmAFP. This is due to the fact that the entropy interval for the TmAFP is larger for the same distance interval because of the larger slope, but about the same amount of

water molecules is viewed.

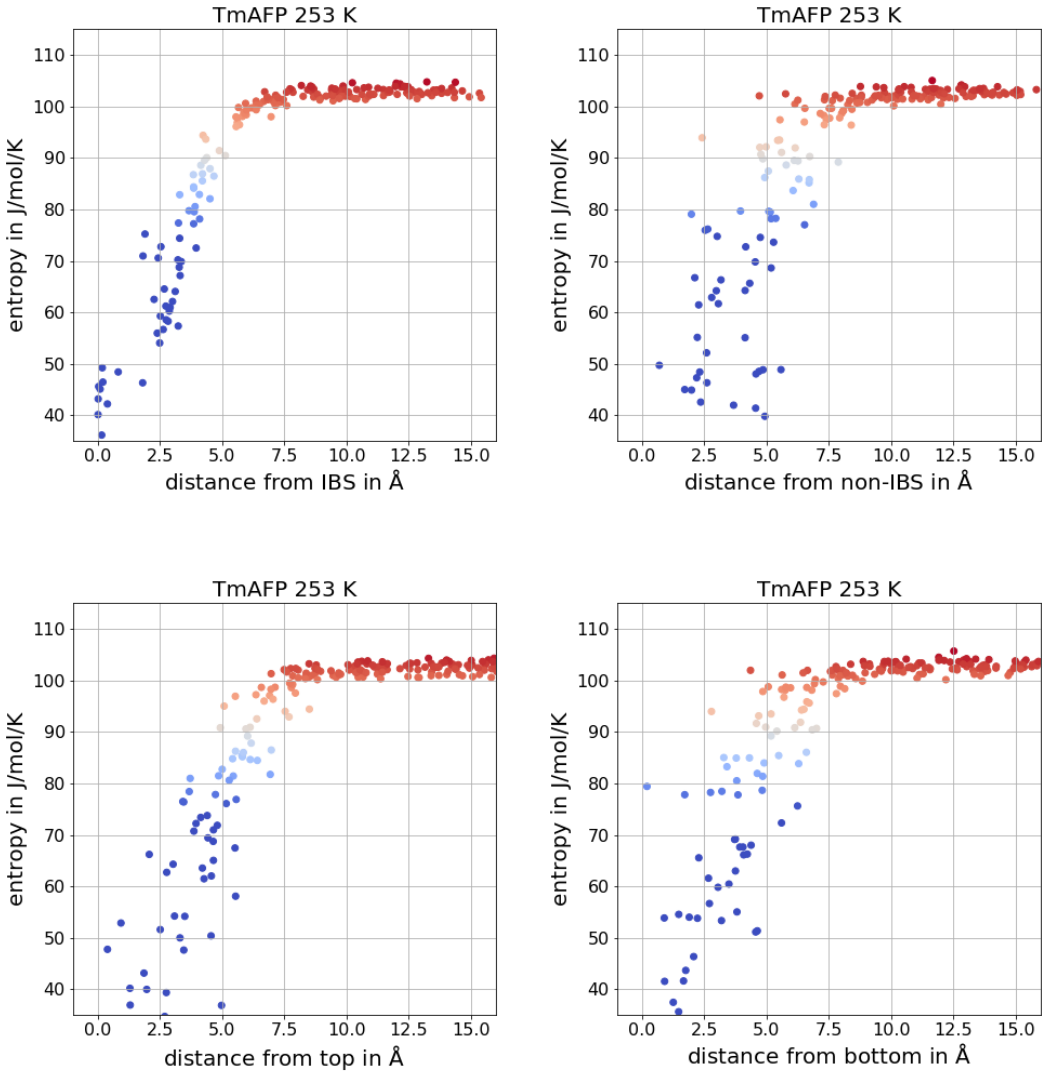


Figure 5.11.: The total entropy as a function of distance for the IBS, non-IBS, top and bottom of the TmAFP at 253 K. The figures for the other temperatures are included in appendix D. The entropy growth with distance until it reaches the entropy of water at the MID.

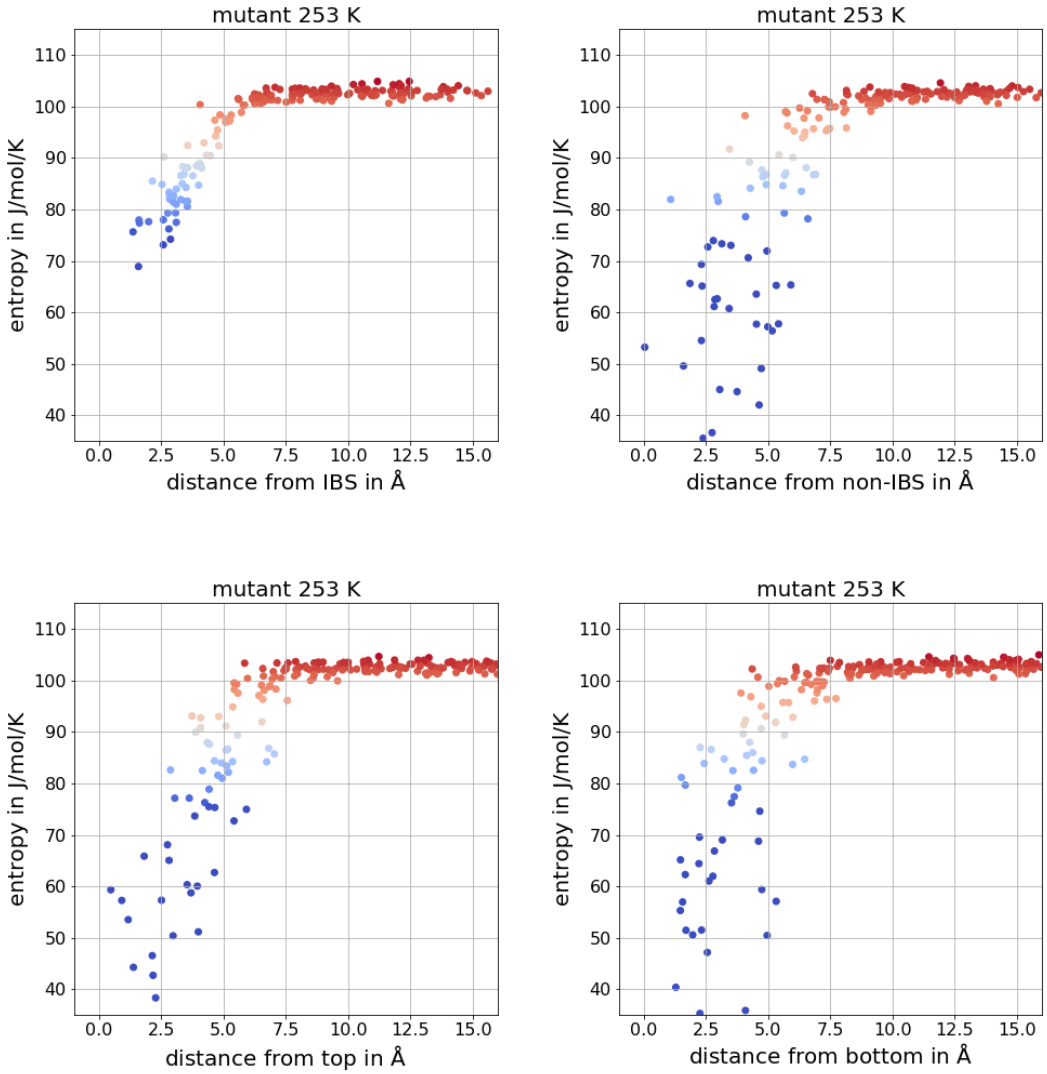


Figure 5.12.: The total entropy as a function of distance for the IBS, non-IBS, top and bottom of the mutant at 253 K. The figures for the other temperatures are included in appendix D. The entropy growth with distance until it reaches the entropy of water at the MID.

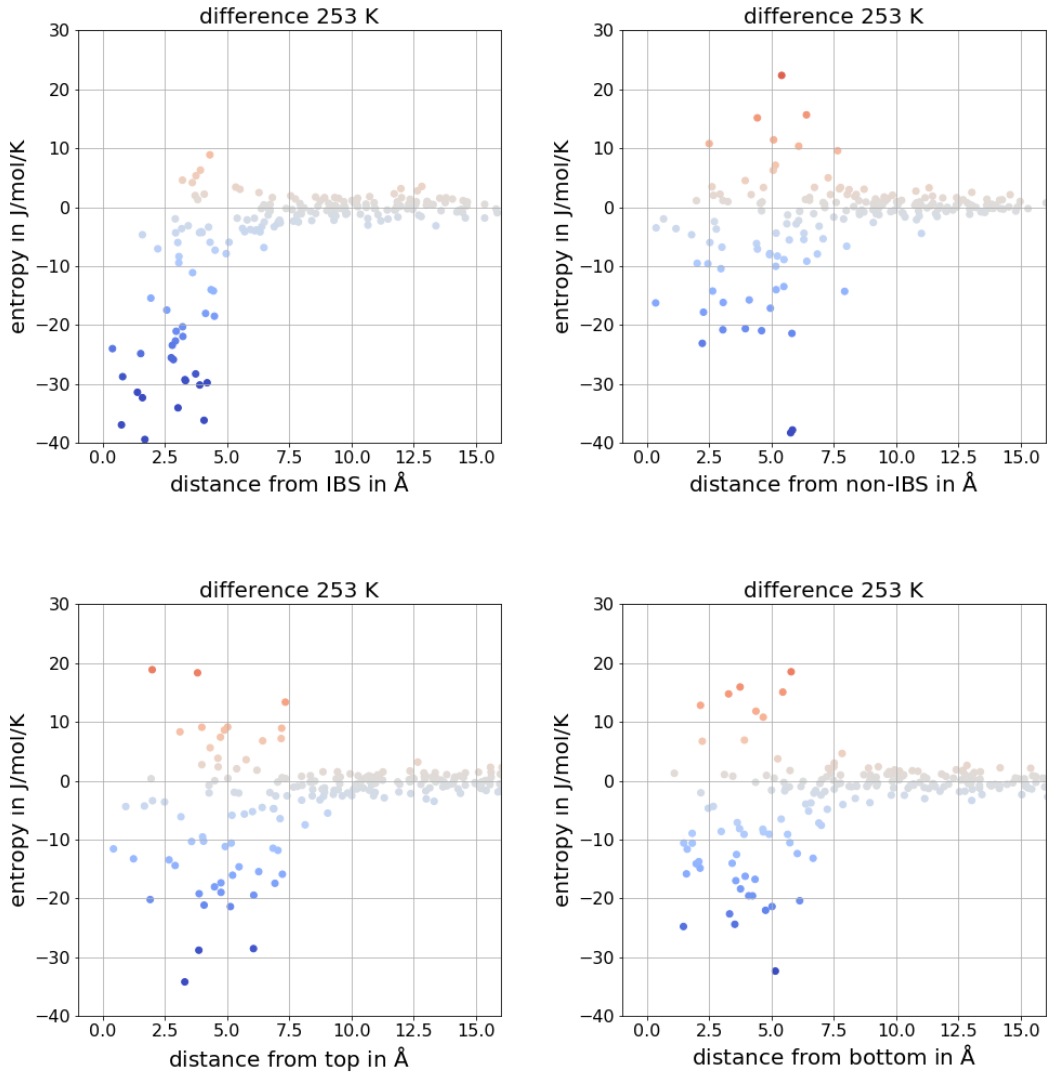


Figure 5.13.: The entropy difference between the TmAFP and the mutant as a function of distance for the IBS, non-IBS, top and bottom of the proteins at 253 K. The figures for the other temperatures are included in appendix D. Postitive and negative entropy values correspond to a higher entropy of the TmAFP and mutant, respectively. It can be seen that the entropy for the TmAFP is lower than for the mutant at all distances.

5.5.2. Comparison of the entropy course for different sides of the protein

To get an even better understanding of the course of the the entropy as a function of distance from the IBS, it will be compared to the other sides of the protein. The other sides are used because I do not have other proteins for comparison. The plots are shown in figures 5.11, 5.12 and 5.13. Because it is hard to find a comparable cutoff for all sides, only the course of the entropies will be discussed, but the starting points and MIDs cannot be compared qualitatively.

What first attracts attention, is the fact that the dots for the other sides show much more fluctuation than for the IBS, both for the TmAFP and the mutant. For the IBS, a steady increase of the entropy is seen. For the other sides, an increase is seen, too. However, it is less uniformly. A possible explanation is that the uniformly arranged identical amino acids on the IBS influence the water uniformly, whereas the different amino residues on the other sides influence the water differently in different areas of the side. When comparing the non-IBS and the top and bottom of the proteins, no statistically relevant differences between the TmAFP and mutant can be seen. This is expected because they are the same and in section 5.3.1 it was already shown, that right at the protein surface no differences can be seen between the non-IBS of both proteins. Between the different sides, there are also no observable differences, suggesting that as long as there are no ordered or identical amino acids, the overall entropy course is similar.
When comparing the different temperatures, it can be seen again, that the entropy values are higher for higher temperatures and the MID gets smaller for higher temperatures for all sides. This shows that the temperature dependence of the entropy at the IBS is the same as for other sides.

Summarizing the observations from section 5.5.1 and 5.5.2, it can be concluded that the course of the entropy as a function of distance from the IBS is special insofar that it is distributed very uniformly, probably due to the uniformly arranged Thr and Val. The difference between the proteins is that the entropy values for the mutant are higher and the MID smaller. The entropy also has a larger slope for the TmAFP. The overall behavior, that the entropy at the protein is low and grows with the distance, is the same as for all sides and both proteins. The temperature

influences the entropy by increasing it for higher temperatures. This also results in smaller MIDs for higher temperatures.

6. Conclusions

In this simulation study, I calculated and analyzed the entropy of the water surrounding the Tenebrio molitor antifreeze protein. The main focus was on the ice-binding site and its influence on water. I also assessed the role of the Thr residues on the IBS by using a mutant for comparison, which had the Thr residues on the IBS replaced with Val.

I have found that the water molecules at the IBS of the TmAFP are ordered and bound to the hydroxyl group of the Thr, which coincides with the results of other simulation studies, which have made the same observations [12] [11] [14]. However, it is in disagreement Hudait et al., who proposed that the water molecules only begin to order when the protein is right next to ice [13]. Because there was no ice next to the protein in my simulation, but the water molecules were still ordered, my results suggest that the ordering of the water molecules takes place without ice, too.

Ordered water was not observed for the mutant, supporting previous studies that found that the Thr plays a vital role in the ordering of the water and that exchanging it for Val, and possibly other amino residues, results in loss of the antifreeze activity [8] [14] [28] [39].
I also found that the orientation of the Thr influences how uniformly the water is ordered and that with higher temperature, the Thr residues were oriented less uniformly.
There was only one layer of ordered water, indicating that the TmAFP does not act as an ice nucleator. This is in accordance with another study that found that the TmAFP is too small to nucleate ice at the temperatures used in my simulations [15].

I analyzed the spatial distribution and variation of both the translational and rotational entropy contributions of the first hydration layer to get a better understanding of where the entropy difference compared to the bulk water entropy comes from. In-

deed, a marked difference between the entropy around the Thr and Val was seen: The entropy around the Thr residues was significantly lowered compared to the entropy around the Val. The translational and rotational entropy differences between both proteins were found to be of the same size. As a control, to rule out possible errors during the simulation or entropy calculation, I have also looked at the hydration layer entropy of the non-IBS, where no amino acids were exchanged for the mutant. As expected, no significant differences were seen between both proteins.
To compare the spatial distribution and values of the entropy to a protein surface without regularly ordered amino residues, the hydration layer entropy of the IBS and non-IBS of the TmAFP were compared. Even though areas with lowered entropy were observed on both sides, the extensive and evenly distributed lowered entropy around the Thr residues was only found on the IBS.

Besides the entropy of the hydration layer molecules right next to the protein, I also looked at how far the influence of the IBS reaches from the protein surface into the bulk water. I found that, for both the TmAFP and the mutant, the course of the entropy as a function of protein-distance looks very similar. However, the entropy for the TmAFP is lower than the mutant's entropy for all distances and the maximum influence distance is larger for the TmAFP, which is about 6Å.
By comparing the entropy at the IBS to the other sides of the protein, I showed that the IBS has a very uniform influence on the water entropy, whereas the entropy for the other sides is more unevenly distributed. But for all sides the entropy starts low and increases with distance, as expected. The uniform influence is assumed to be due to the uniformly arranged Thr and Val on the IBS.

Because all simulations were carried out at four different temperatures, which were above, at, as well as below, the freezing point, I also analysed the temperature-dependence of the above results. It was found that with rising temperature, the entropy difference to the bulk water entropy becomes smaller and the ordered water molecules are ordered less uniformly because the orientations of the Thr residues are more heterogeneously distributed. These results show that AFPs influence the water stronger the lower the temperature becomes, at least up to 253 K, which was the lowest temperature simulated in this study. This coincides with the fact that AFPs evolved to work at low temperatures and are not needed above freezing temperature.

The entropy of bulk water calculated with Per|Mut is about 105 $\mathrm{Jmol^{-1}K^{-1}}$, whereas the experimental value is about 70 $\mathrm{Jmol^{-1}K^{-1}}$. Because the method does not yield accurate entropy values, I was not able to analyze and discuss the entropy values of the hydration water. However, it yields more accurate entropy differences [23], which allowed me look into the differences between the TmAFP and mutant, as well as the differences between the hydration and bulk water entropy.
Because many observations I made coincide with those of previous studies, they support that my simulations are sufficiently accurate to obtain new insight, such as how far the protein influences water order.

Further studies on the entropy of other AFPs, as well as other proteins, could be used to get a better understanding of how the impact of the IBS on the water is compared to other proteins. Do all AFPs have a similar influence on water? Is the uniformly increasing entropy found for all AFPs? Is it also found for other proteins? Do AFPs have a stronger and farther influence on water than other proteins? Results like these could reveal more information about how the influence of AFPs on the hydration water might help the protein to enhance diffusion towards ice.
For an even better understanding of the threonine's role, further studies with other mutations than Val would be interesting, as suggested in section 4.1. If for all other mutations a loss in antifreeze activity is observed, the results would support that the ordering is only due to the methyl and hydroxyl group, which together are only found with Thr. A closer look at the binding of the water molecules to the Thr would also give a better idea about how the ordering of the water is achieved, which might eventually facilitate the design of artificial AFPs, which could be used in preservation and antifreeze agents.

A. Sample mdp-files for the simulations

mdp is short for molecular dynamics parameters. The file contains information for the simulation, like run parameters, boundary conditions, position restraints and more. The sample files here are the files used for the first and second equilibrium simulation, as well as the main simulation for 273 K.

A.1. mdp-file for the first equilibration simulation

```
define = -DPOSRES

; Run parameters
integrator = md ; leap-frog integrator
nsteps = 50000 ; 0.002ps * 1000000000 = 2000000 ps
dt     = 0.002 ; 2 fs

; Output control
nstxout =     0 ; save coordinates every 10.0 ps
nstvout =     0 ; save velocities every 10.0 ps
nstenergy =  5000 ; save energies every 10.0 ps
nstlog = 10000 ; update log file every 10.0 ps

compressed-x-grps      = System
nstxout-compressed     = 5000  ;
compressed-x-precision = 100000

; Bond parameters
continuation         = no ; first dynamics run
```

```
constraint_algorithm    = lincs     ; holonomic constraints
constraints             = h-bonds  ;
lincs_iter             = 1     ; accuracy of LINCS
lincs_order             = 4     ; also related to accuracy

; Neighborsearching
cutoff-scheme   = Verlet
ns_type     = grid ; search neighboring grid cells
nstlist     = 10 ; 20 fs, largely irrelevant with Verlet

; VdW
rvdw     = 1.2 ; short-range van der Waals cutoff (in nm)
vdwtype         = cut-off
vdw-modifier    = force-switch
rvdw-switch     = 1.0

; Electrostatics
coulombtype     = PME ; Particle Mesh Ewald for long-range electrostatics
pme_order     = 4 ; cubic interpolation
fourierspacing = 0.16 ; grid spacing for FFT
rcoulomb     = 1.2 ; short-range electrostatic cutoff (in nm)

; Temperature coupling is on
tcoupl = V-rescale             ; modified Berendsen thermostat
tc-grps = Protein Non-Protein  ; two coupling groups - more accurate
tau_t = 0.1   0.1           ; time constant, in ps
ref_t = 273    273           ; reference temperature, one for each group,
in K

; Pressure coupling is on
pcoupl         = no     ; Pressure coupling on in NPT
pcoupltype         = isotropic             ; uniform scaling of box vectors
tau_p         = 1.0                ; time constant, in ps
ref_p         = 1.0                ; reference pressure, in bar
compressibility     = 4.5e-5   ; isothermal compressibility of water
```

```
, bar^-1
refcoord_scaling    = com

; Periodic boundary conditions
pbc = xyz     ; 3-D PBC

; Dispersion correction
DispCorr = EnerPres ; account for cut-off vdW scheme

; Velocity generation
gen_vel = yes ; assign velocities from Maxwell distribution
gen_temp = 273 ; temperature for Maxwell distribution
gen_seed = -1 ; generate a random seed

; COM removal
comm-grps   = Protein Non-Protein
comm-mode   = Linear     ; Linear or Angular or None
nstcomm     = 100        ; frequency for center of mass motion removal
```

A.2. mdp-file for the second equilibration simulation at 273 K

```
define = -DPOSRESCA

; Run parameters
integrator = md ; leap-frog integrator
nsteps = 5000000 ; 0.002ps * 1000000000 = 2000000 ps
dt     = 0.002 ; 2 fs

; Output control
nstxout =     0 ; save coordinates every 10.0 ps
nstvout =     0 ; save velocities every 10.0 ps
nstenergy =  5000 ; save energies every 10.0 ps
nstlog = 10000 ; update log file every 10.0 ps

compressed-x-grps      = System
nstxout-compressed     = 5000  ;
compressed-x-precision = 100000

; Bond parameters
continuation         = yes ; first dynamics run
constraint_algorithm    = lincs     ; holonomic constraints
constraints             = h-bonds  ;
lincs_iter              = 1     ; accuracy of LINCS
lincs_order             = 4     ; also related to accuracy

; Neighborsearching
cutoff-scheme   = Verlet
ns_type     = grid ; search neighboring grid cells
nstlist     = 10 ; 20 fs, largely irrelevant with Verlet

; VdW
rvdw     = 1.2 ; short-range van der Waals cutoff (in nm)
vdwtype         = cut-off
```

```
vdw-modifier    = force-switch
rvdw-switch     = 1.0

; Electrostatics
coulombtype     = PME ; Particle Mesh Ewald for long-range electrostatics
pme_order     = 4 ; cubic interpolation
fourierspacing = 0.16 ; grid spacing for FFT
rcoulomb     = 1.2 ; short-range electrostatic cutoff (in nm)

; Temperature coupling is on
tcoupl = V-rescale             ; modified Berendsen thermostat
tc-grps = Protein Non-Protein  ; two coupling groups - more accurate
tau_t = 0.1   0.1           ; time constant, in ps
ref_t = 273    273           ; reference temperature, one for each group,
in K

; Pressure coupling is on
pcoupl         = Berendsen     ; Pressure coupling on in NPT
pcoupltype         = isotropic             ; uniform scaling of box vectors
tau_p         = 1.0               ; time constant, in ps
ref_p         = 1.0               ; reference pressure, in bar
compressibility     = 4.5e-5  ; isothermal compressibility of water
, bar^-1
refcoord_scaling    = com

; Periodic boundary conditions
pbc = xyz     ; 3-D PBC

; Dispersion correction
DispCorr = EnerPres ; account for cut-off vdW scheme

; Velocity generation
gen_vel = no ; assign velocities from Maxwell distribution
gen_temp = 273 ; temperature for Maxwell distribution
gen_seed = -1 ; generate a random seed
```

```
; COM removal
comm-grps    = Protein Non-Protein
comm-mode    = Linear      ; Linear or Angular or None
nstcomm      = 100         ; frequency for center of mass motion removal
```

A.3. mdp-file for the main simulation at 273 K

```
define = -DPOSRES

; Run parameters
integrator = md ; leap-frog integrator
nsteps = 1000000000 ; 0.002ps * 1000000000 = 2000000 ps
dt     = 0.002 ; 2 fs

; Output control
nstxout =     0 ; save coordinates every 10.0 ps
nstvout =     0 ; save velocities every 10.0 ps
nstenergy =  5000 ; save energies every 10.0 ps
nstlog = 10000 ; update log file every 10.0 ps

compressed-x-grps      = System
nstxout-compressed     = 5000  ;
compressed-x-precision = 100000

; Bond parameters
continuation         = no ; first dynamics run
constraint_algorithm    = lincs     ; holonomic constraints
constraints              = h-bonds  ;
lincs_iter              = 1     ; accuracy of LINCS
lincs_order              = 4     ; also related to accuracy

; Neighborsearching
cutoff-scheme   = Verlet
ns_type     = grid ; search neighboring grid cells
nstlist     = 10 ; 20 fs, largely irrelevant with Verlet

; VdW
rvdw     = 1.2 ; short-range van der Waals cutoff (in nm)
vdwtype         = cut-off
vdw-modifier    = force-switch
```

```
rvdw-switch      = 1.0

; Electrostatics
coulombtype      = PME ; Particle Mesh Ewald for long-range electrostatics
pme_order      = 4 ; cubic interpolation
fourierspacing = 0.16 ; grid spacing for FFT
rcoulomb      = 1.2 ; short-range electrostatic cutoff (in nm)

; Temperature coupling is on
tcoupl = V-rescale             ; modified Berendsen thermostat
tc-grps = Protein Non-Protein  ; two coupling groups - more accurate
tau_t = 0.1   0.1           ; time constant, in ps
ref_t = 273    273           ; reference temperature, one for each group,
in K

; Pressure coupling is on
pcoupl          = Parrinello-Rahman     ; Pressure coupling on in NPT
pcoupltype          = isotropic             ; uniform scaling of box vectors
tau_p          = 1.0                ; time constant, in ps
ref_p          = 1.0                ; reference pressure, in bar
compressibility      = 4.5e-5  ; isothermal compressibility of water,
bar^-1
refcoord_scaling    = com

; Periodic boundary conditions
pbc = xyz      ; 3-D PBC

; Dispersion correction
DispCorr = EnerPres ; account for cut-off vdW scheme

; Velocity generation
gen_vel = yes ; assign velocities from Maxwell distribution
gen_temp = 273 ; temperature for Maxwell distribution
gen_seed = -1 ; generate a random seed
```

```
; COM removal
comm-grps   = Protein Non-Protein
comm-mode   = Linear     ; Linear or Angular or None
nstcomm     = 100        ; frequency for center of mass motion removal
```

B. Hydration layer entropy at the IBS

B.1. Rotational and translational entropies

The following images show the entropy of the water surrounding the protein mapped onto the surface of the protein, as described in section 5.2. The left and right side show the IBS of the TmAFP and mutant, respectively. The top four images show the entropy of the water molecules surrounding the protein mapped onto the surface of the protein. The first row shows the translational entropy, the second row the rotational entropy. The colors go from blue ($0\,\mathrm{Jmol^{-1}K^{-1}}$) to red ($70\,\mathrm{Jmol^{-1}K^{-1}}$). In the last row images of the proteins are added to demonstrate the orientation. The images are for temperatures 263, 273 and 283 K.

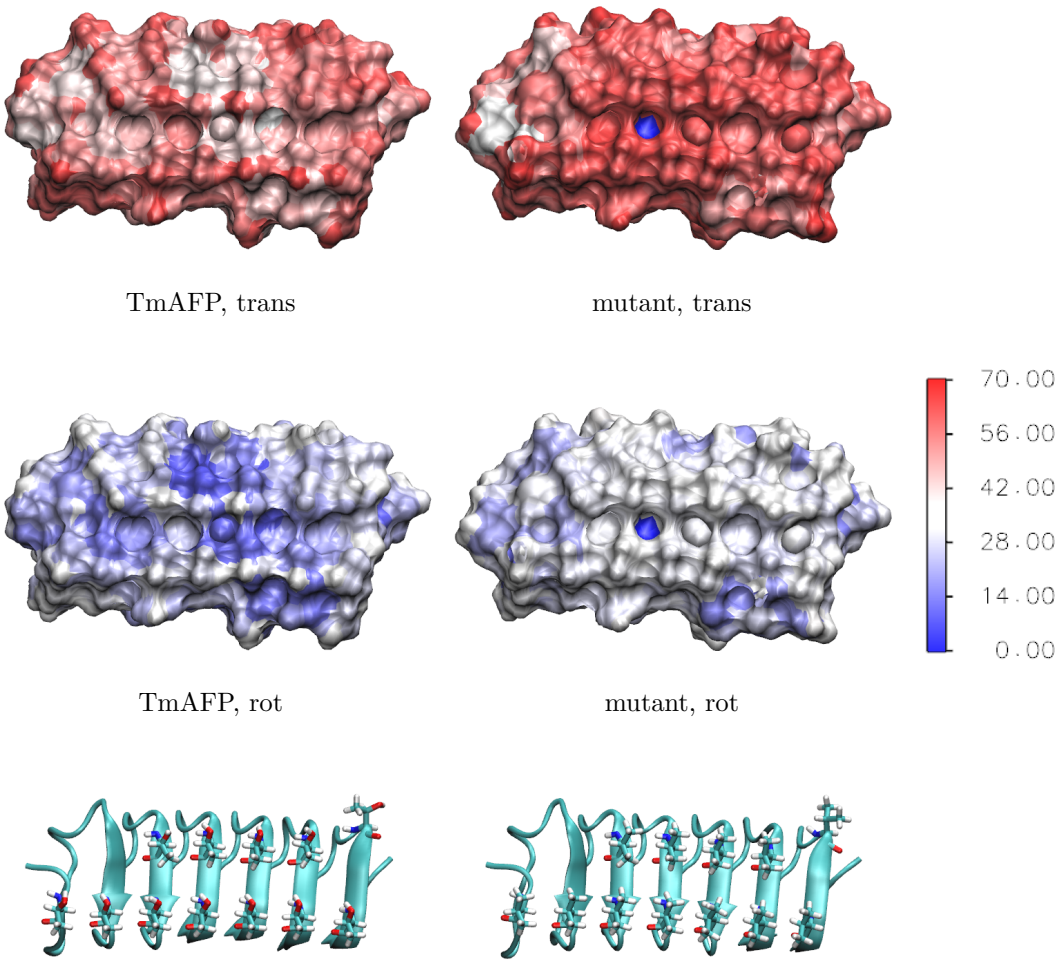


Figure B.1.: 263 K

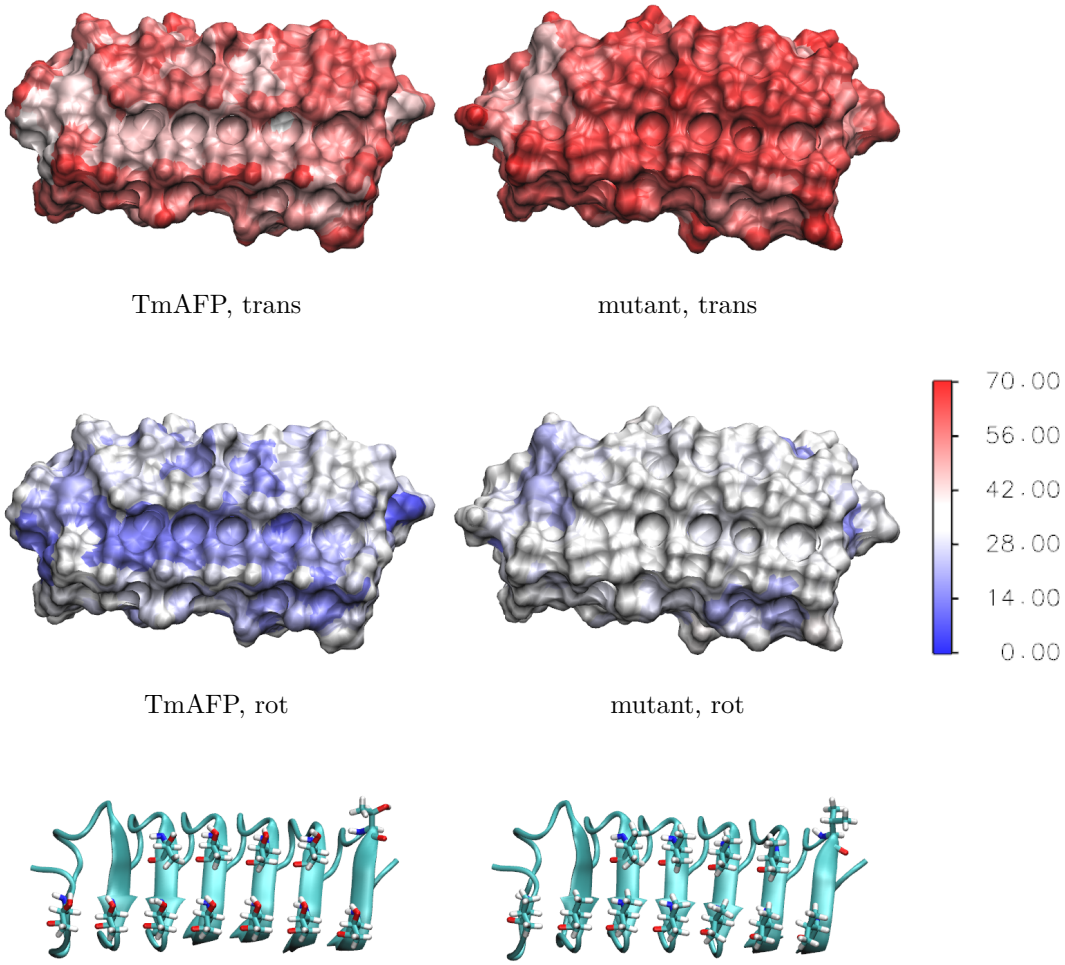


Figure B.2.: 273 K

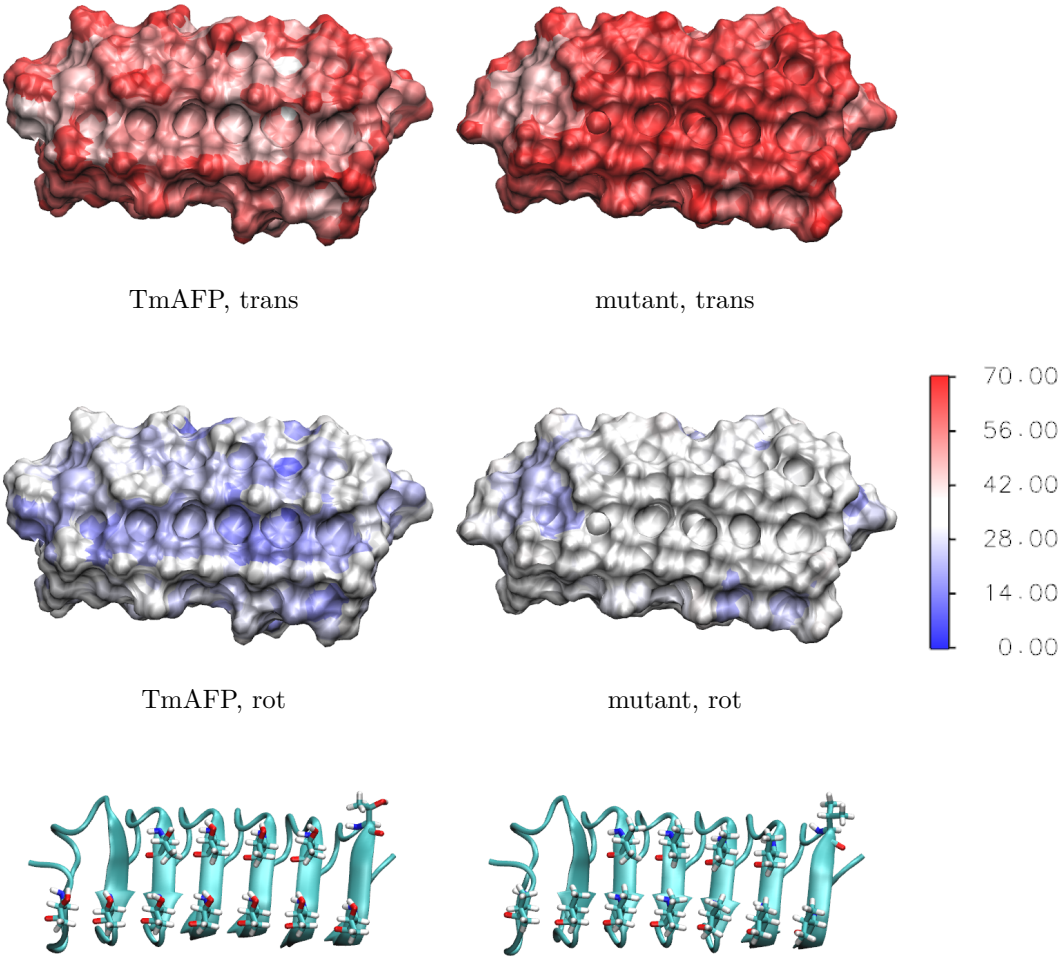
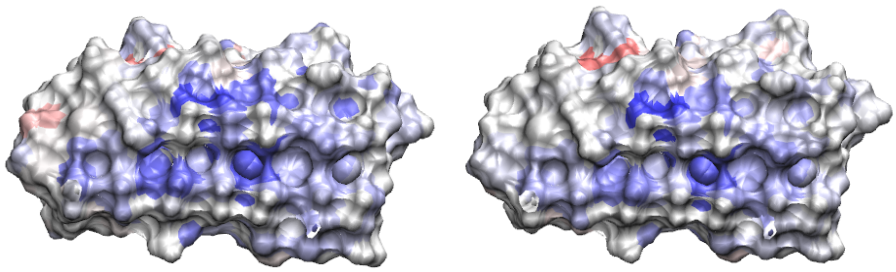


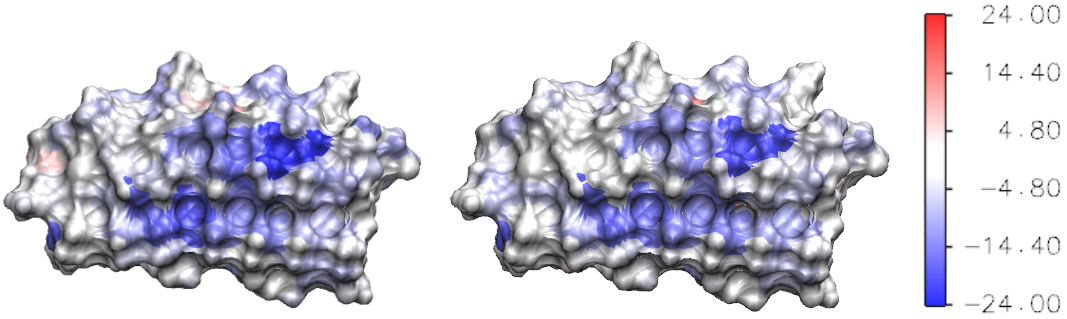
Figure B.3.: 283 K

B.2. Entropy differences

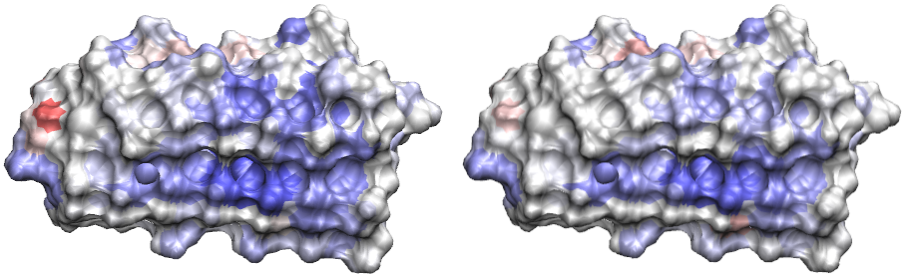
The images for the difference between the entropies around the TmAFP and mutant as described in section 5.2.



263 K, trans rot



273 K, trans rot



283 K, trans rot

C. Hydration layer entropy at the non-IBS

C.1. Rotational and translational entropies

The following images show the entropy of the water surrounding the protein mapped onto the surface of the protein, as described in section 5.3.1.The left and right side show the non-IBS of the TmAFP and mutant, respectively. The top four images show the entropy of the water molecules surrounding the protein mapped onto the surface of the protein. The first row shows the translational entropy, the second row the rotational entropy. The colors go from blue ($0\,\mathrm{Jmol^{-1}K^{-1}}$) to red ($70\,\mathrm{Jmol^{-1}K^{-1}}$). In the last row images of the proteins are added to demonstrate the orientation.

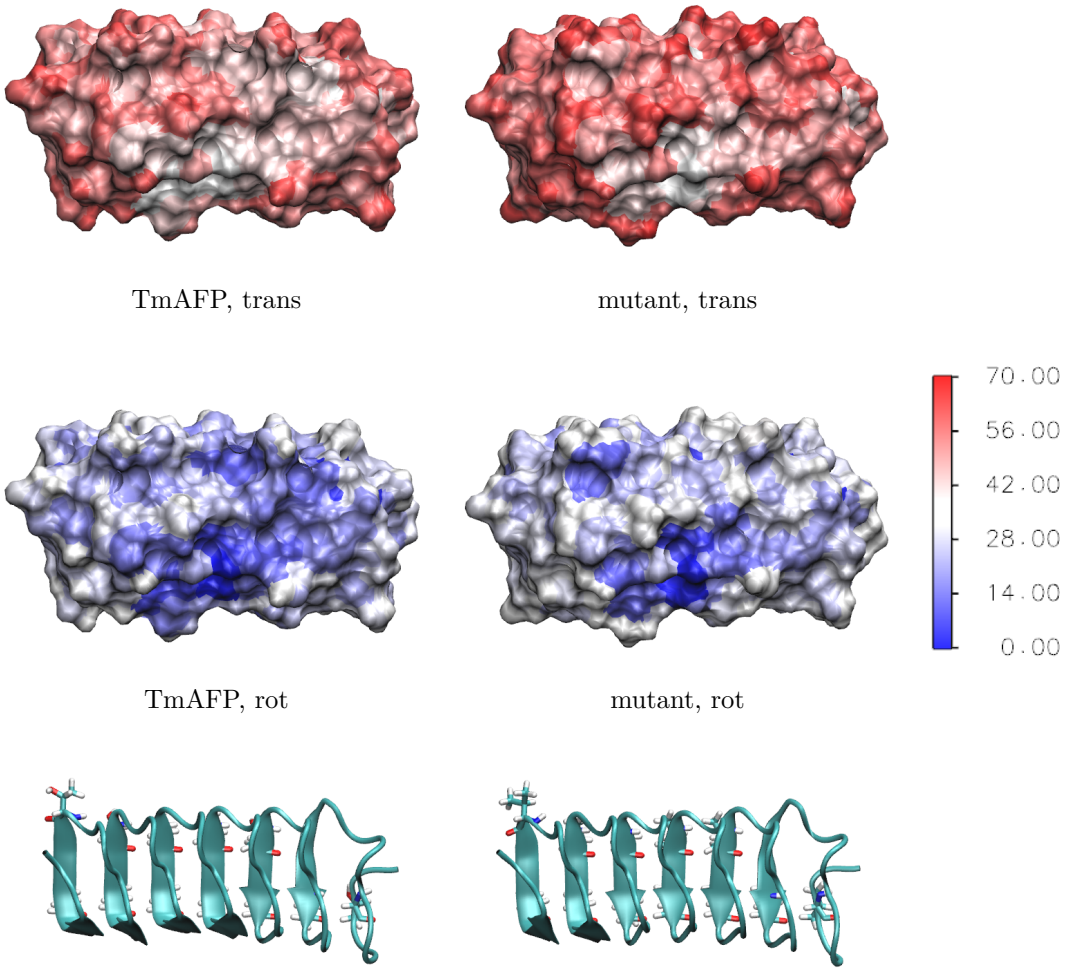


Figure C.1.: 263 K

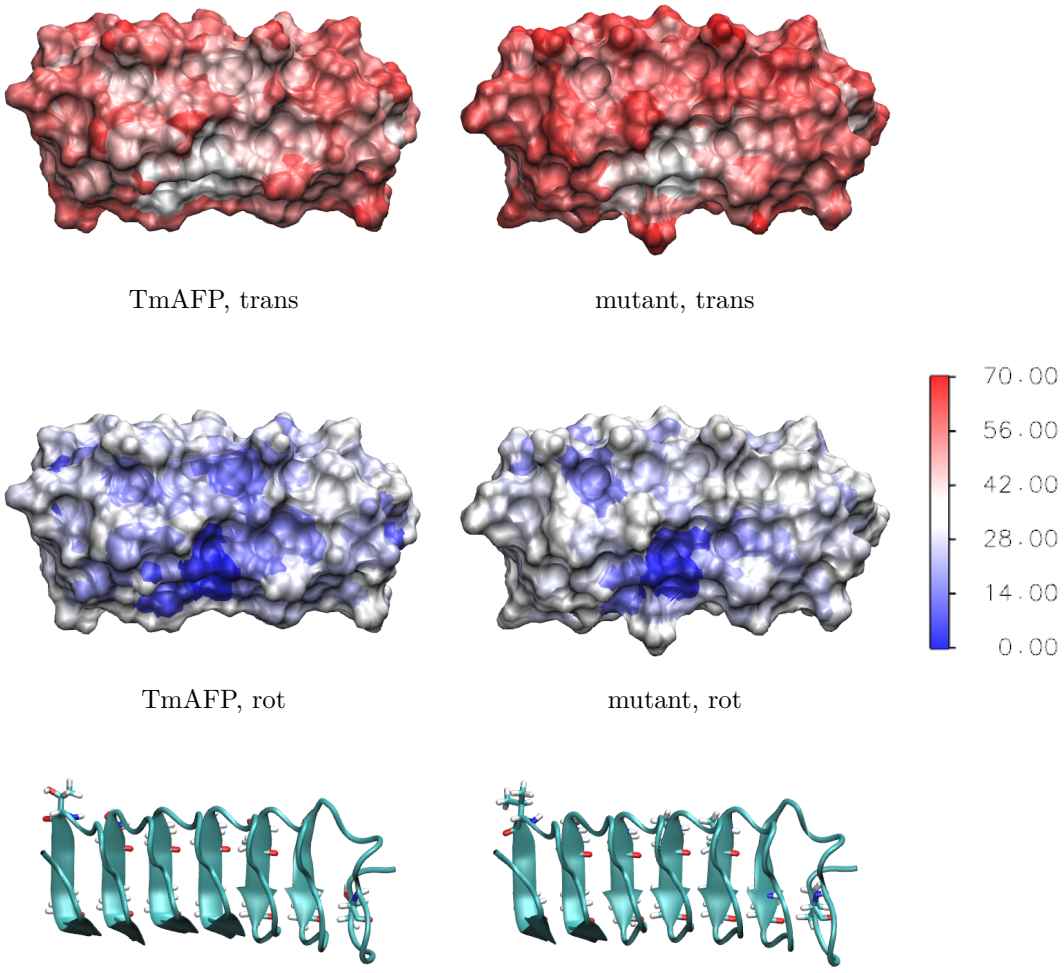


Figure C.2.: 273 K

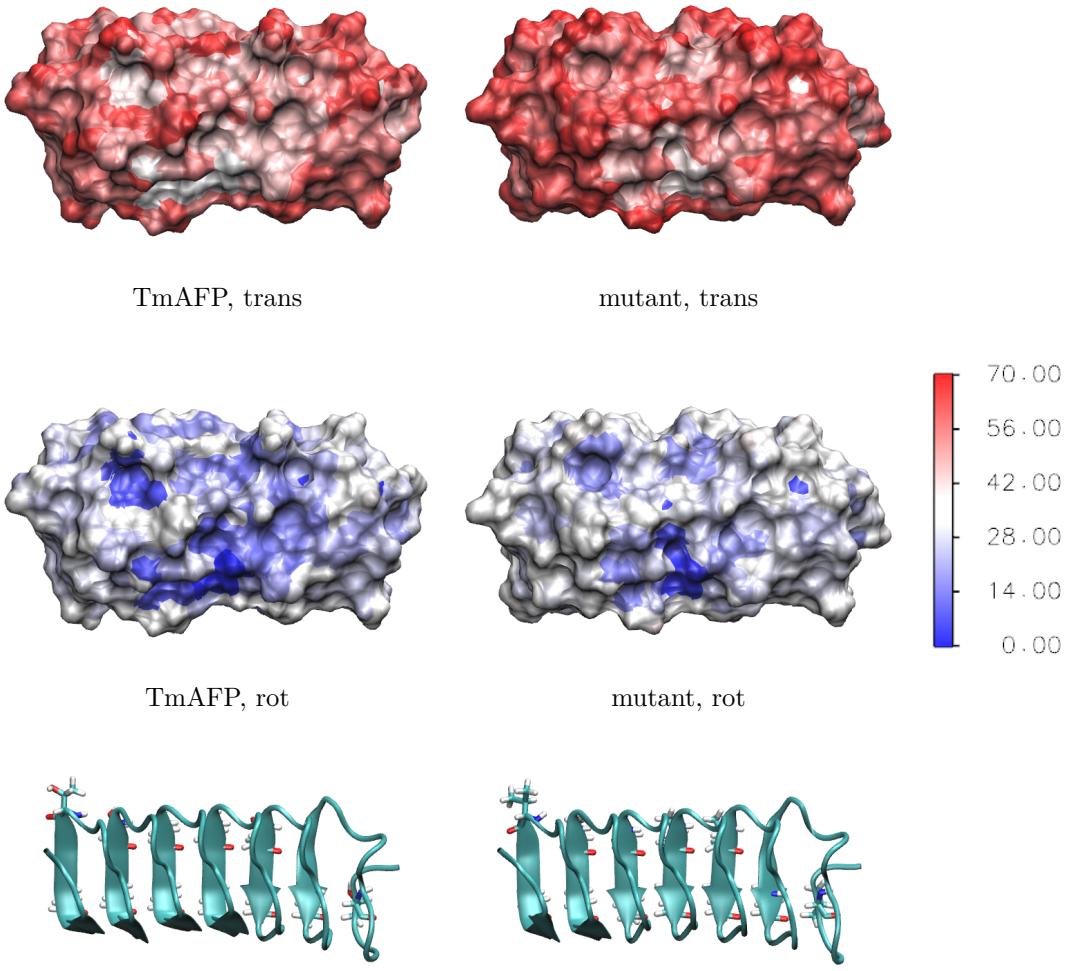
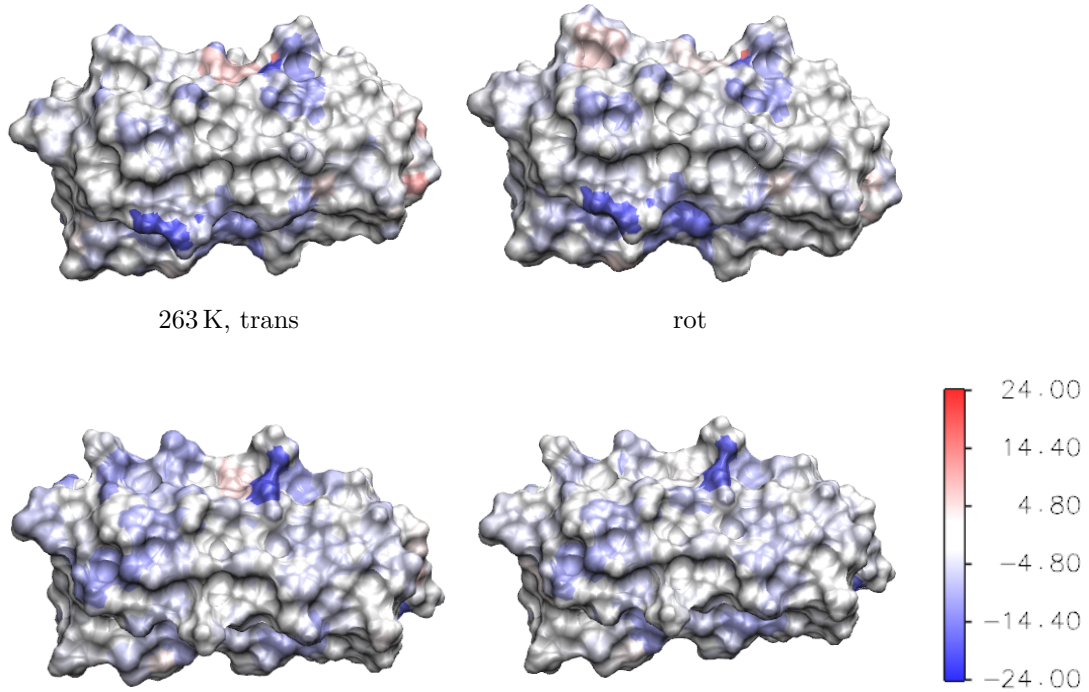


Figure C.3.: 283 K.

C.2. Entropy differences

The entropy is given in $\mathrm{Jmol^{-1}K^{-1}}$.



263 K, trans | rot

273 K, trans | rot

283 K, trans | rot

C.3. Rotational and tranlsational entropies at the IBS and non-IBS of the TmAFP

The following images show the entropy of the water surrounding the protein mapped onto the surface of the protein, as described in section 5.3.1.The left and right side show the IBS and non-IBS, respectively. The top four images show the entropy of the water molecules surrounding the protein mapped onto the surface of the protein. The first row shows the translational entropy, the second row the rotational entropy. The colors go from blue ($0\,\mathrm{Jmol^{-1}K^{-1}}$) to red ($70\,\mathrm{Jmol^{-1}K^{-1}}$). In the last row images of the proteins are added to demonstrate the orientation.

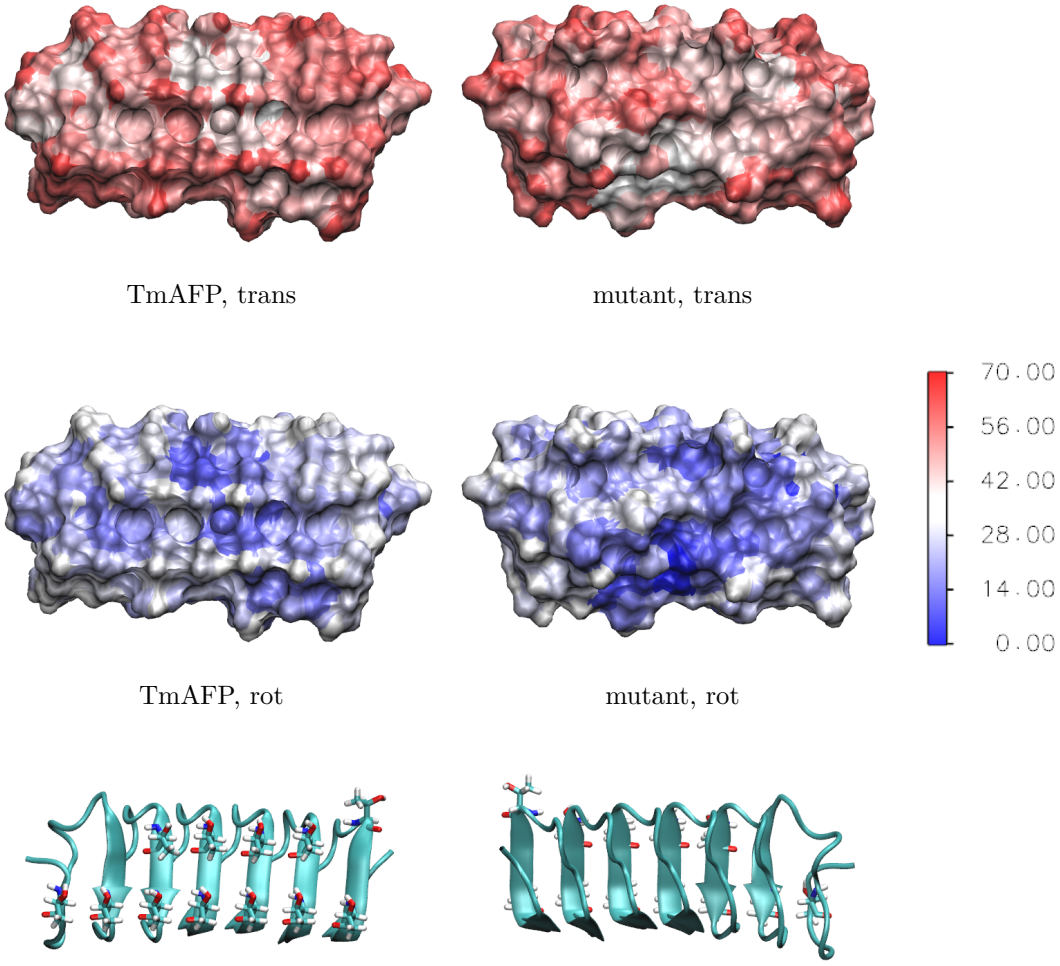


Figure C.4.: 263 K

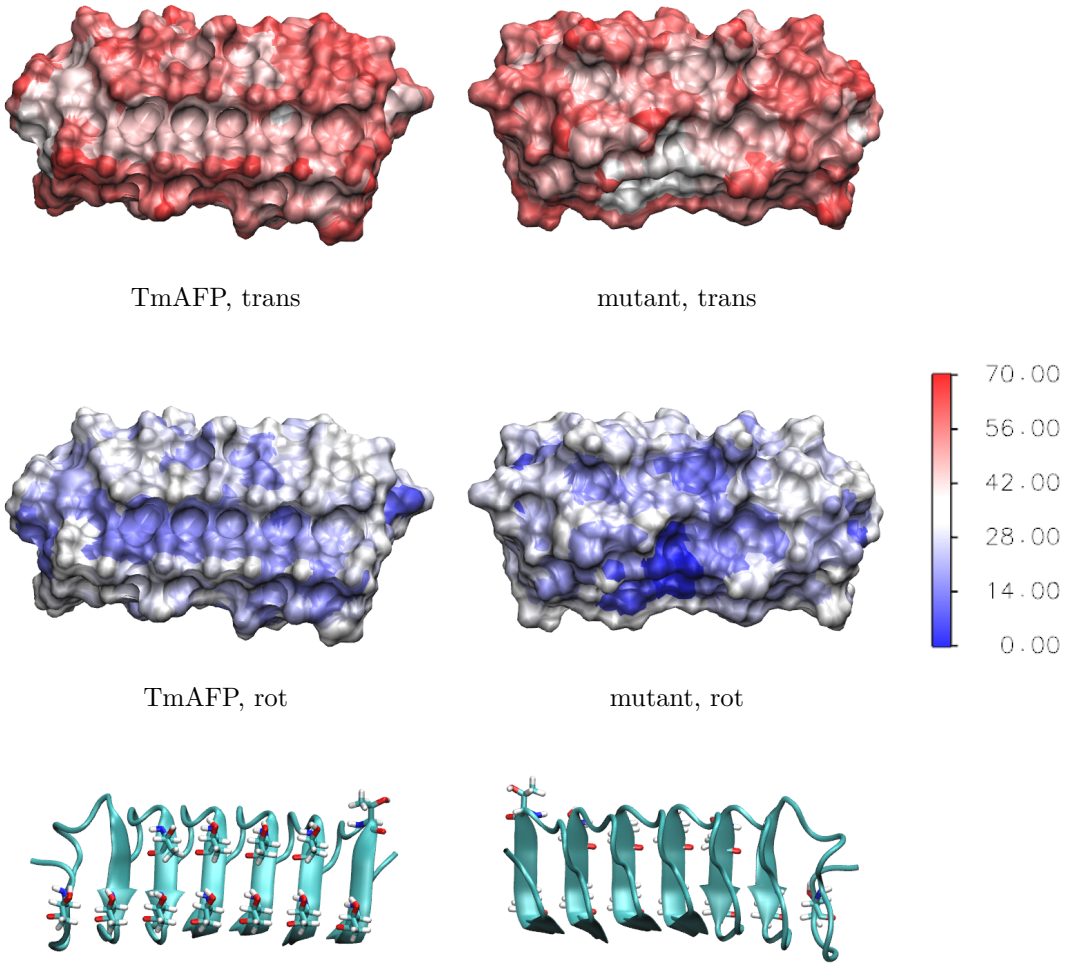


Figure C.5.: 273 K

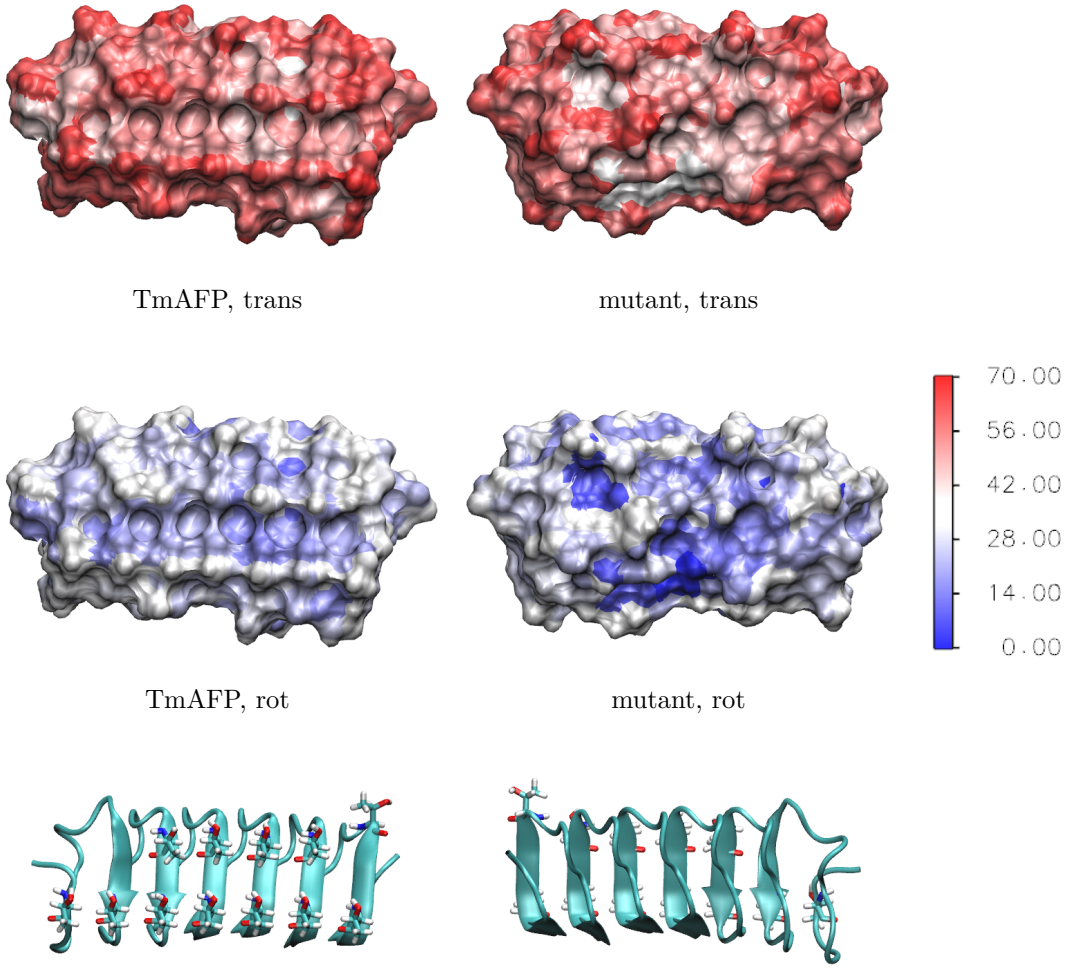


Figure C.6.: 283 K

D. Entropy as a function of distance

D.1. Total entropies

In this appendix, the entropy as a function of distance is shown, as discussed in sections 5.5.1 and 5.5.2. The figures show the entropy for four different sides of the protein, one of them being the IBS. Each figure corresponds to the total entropy of the TmAFP or the mutant or the entropy difference between the two for on temperature.

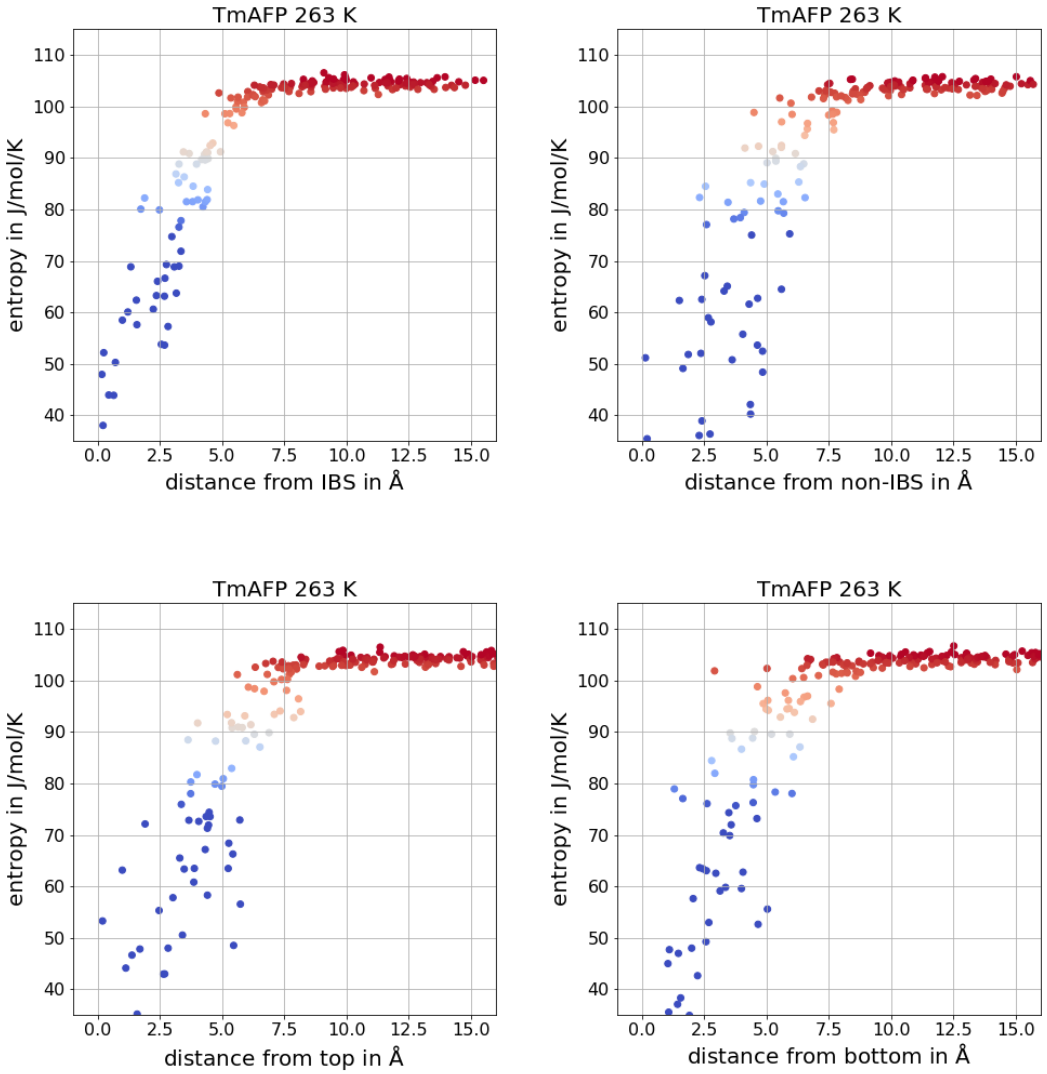


Figure D.1.: TmAFP 263 K

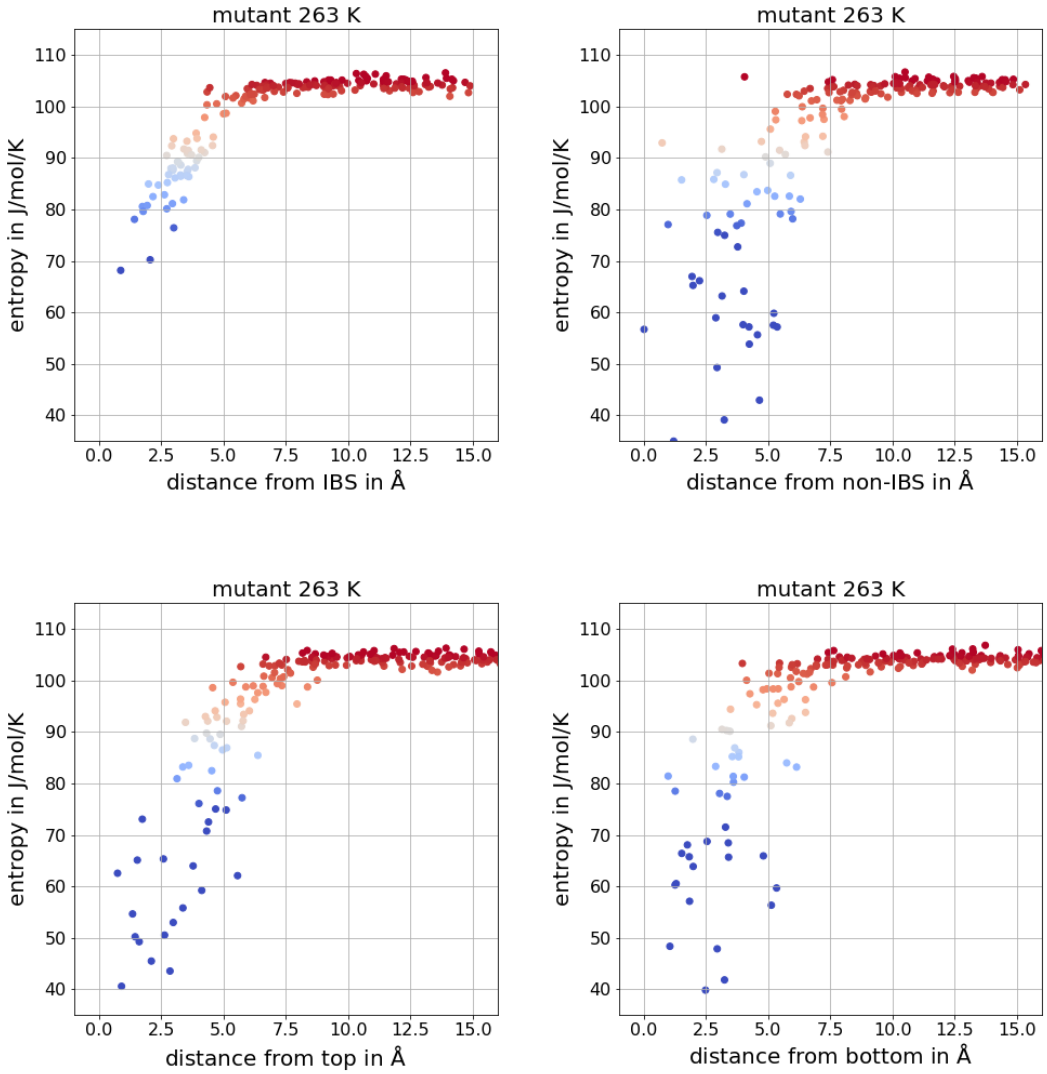


Figure D.2.: mutant 263 K

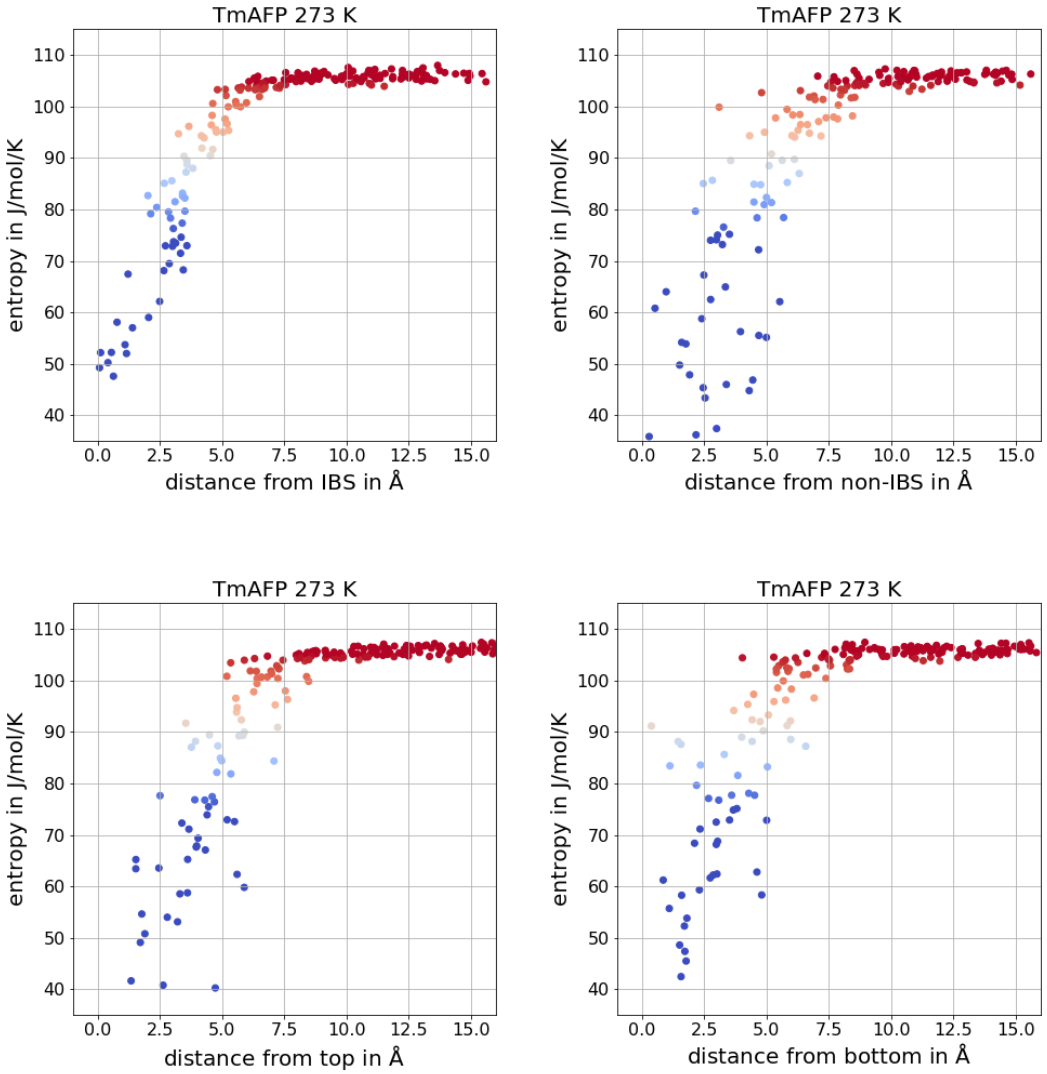


Figure D.3.: TmAFP 273 K

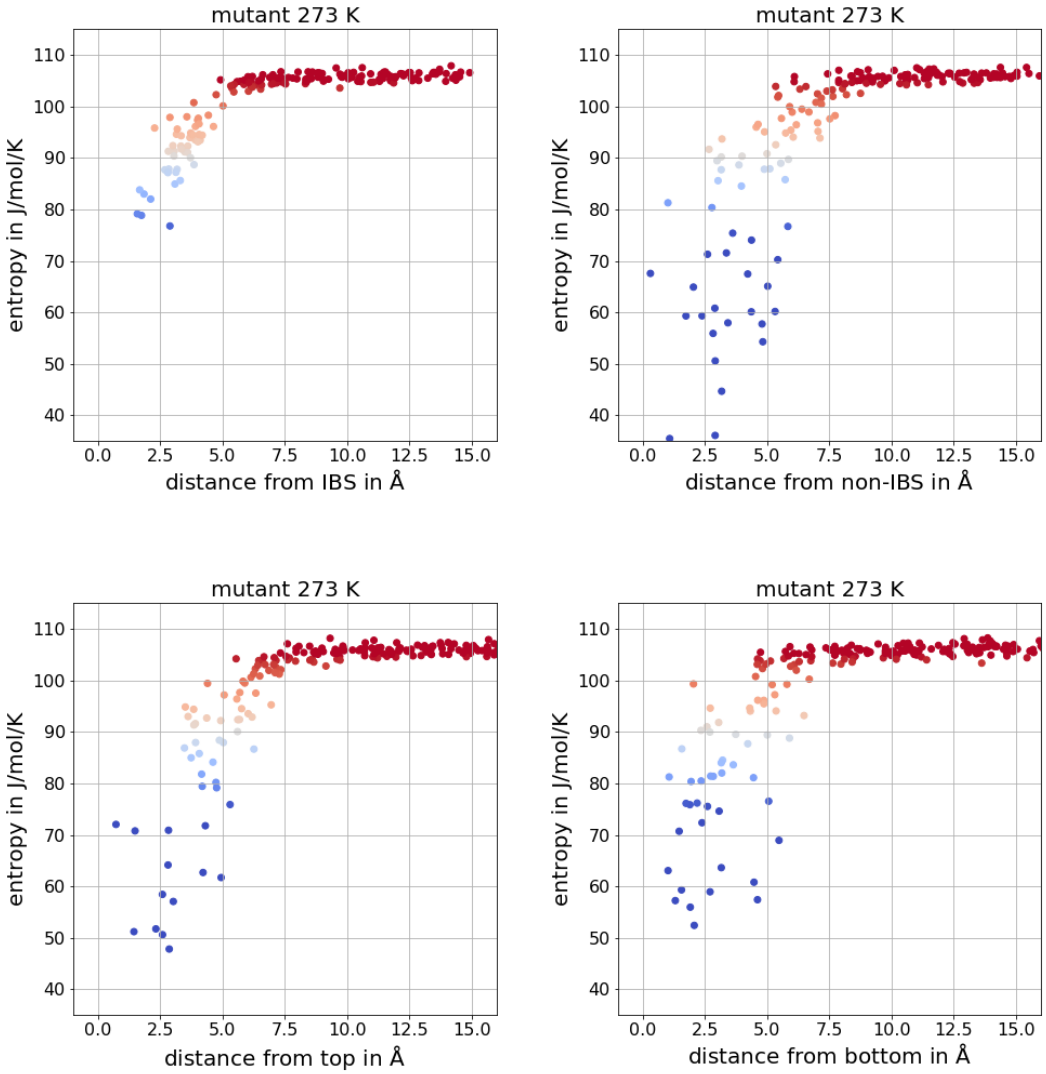


Figure D.4.: mutant 273 K

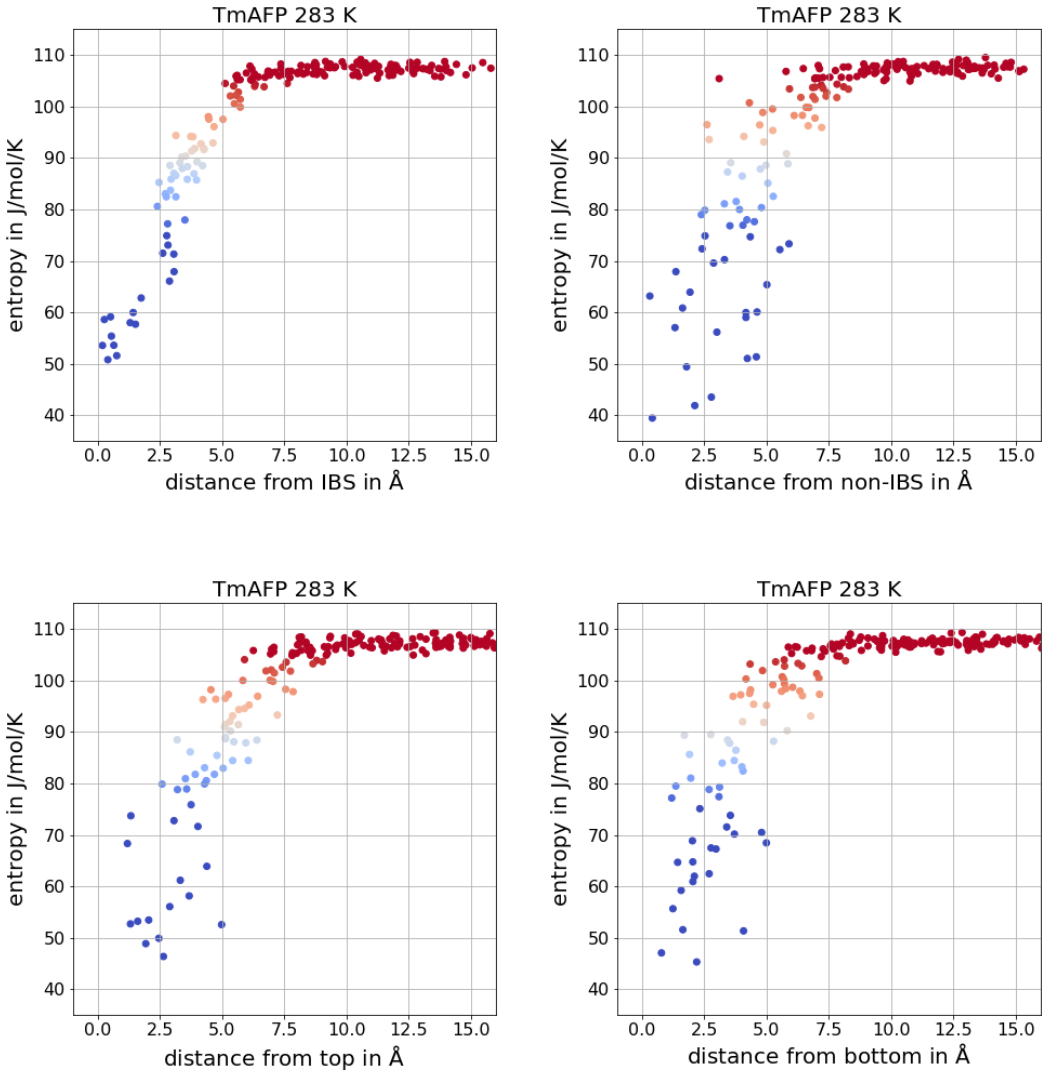


Figure D.5.: TmAFP 283 K

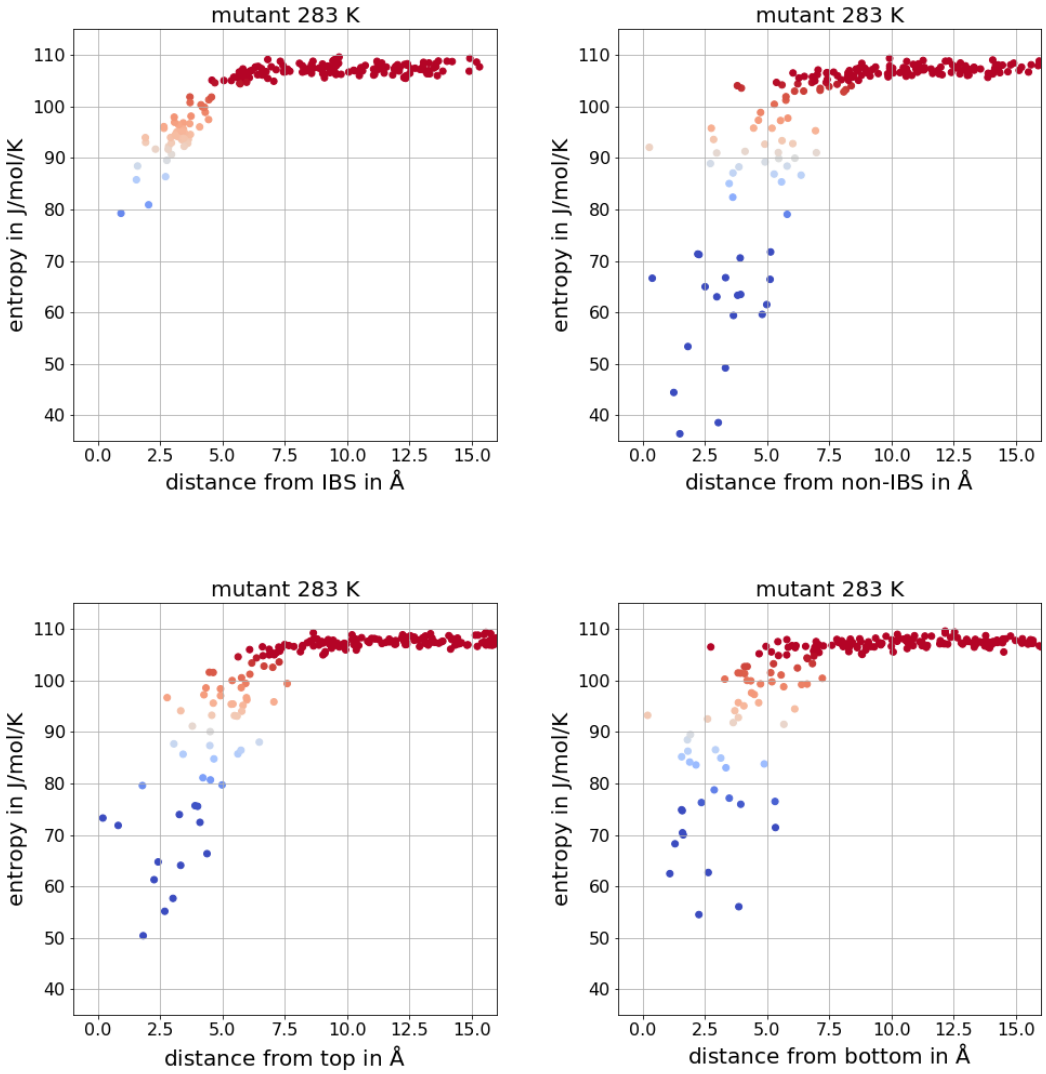


Figure D.6.: mutant 283 K

D.2. Entropy differences

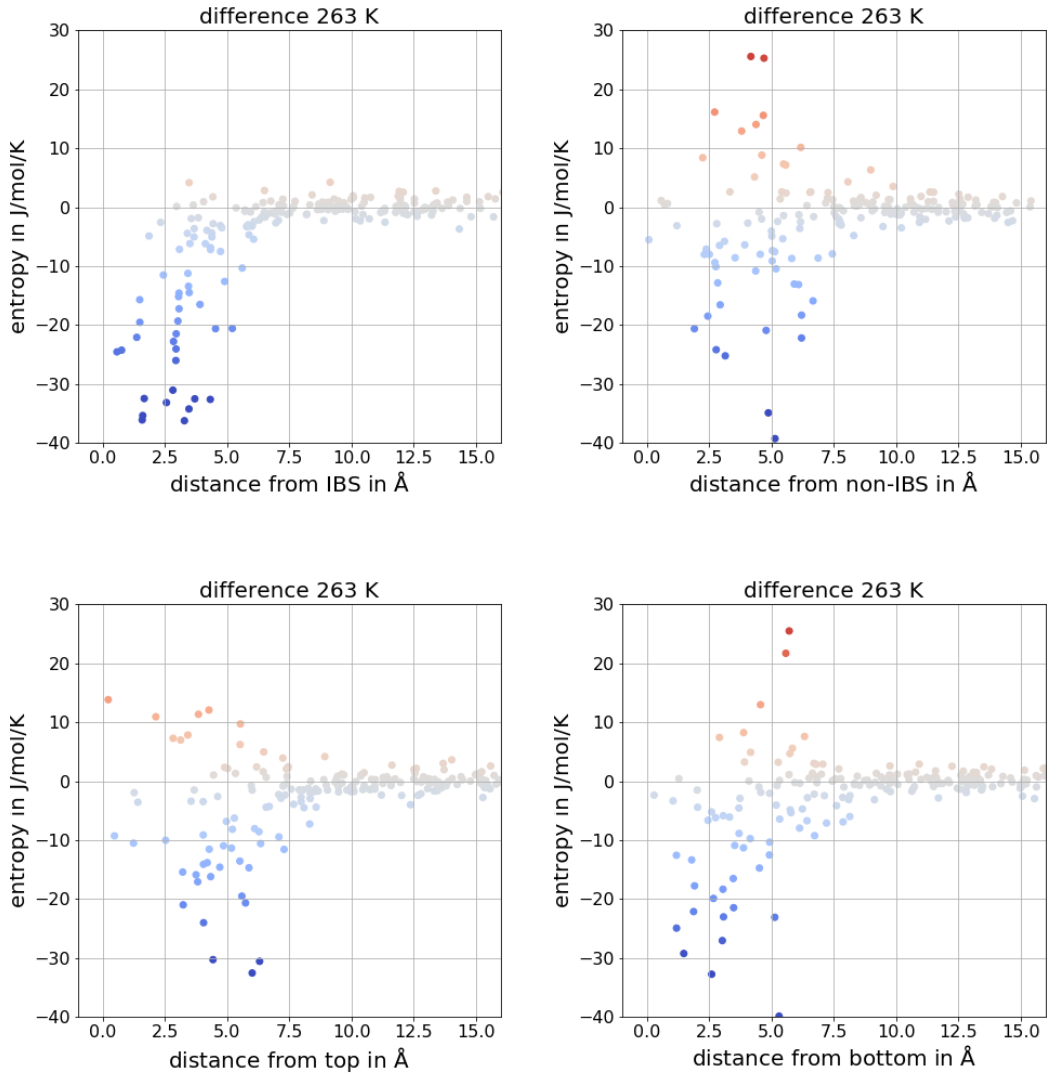


Figure D.7.: Difference between TmAFP and mutant 263 K

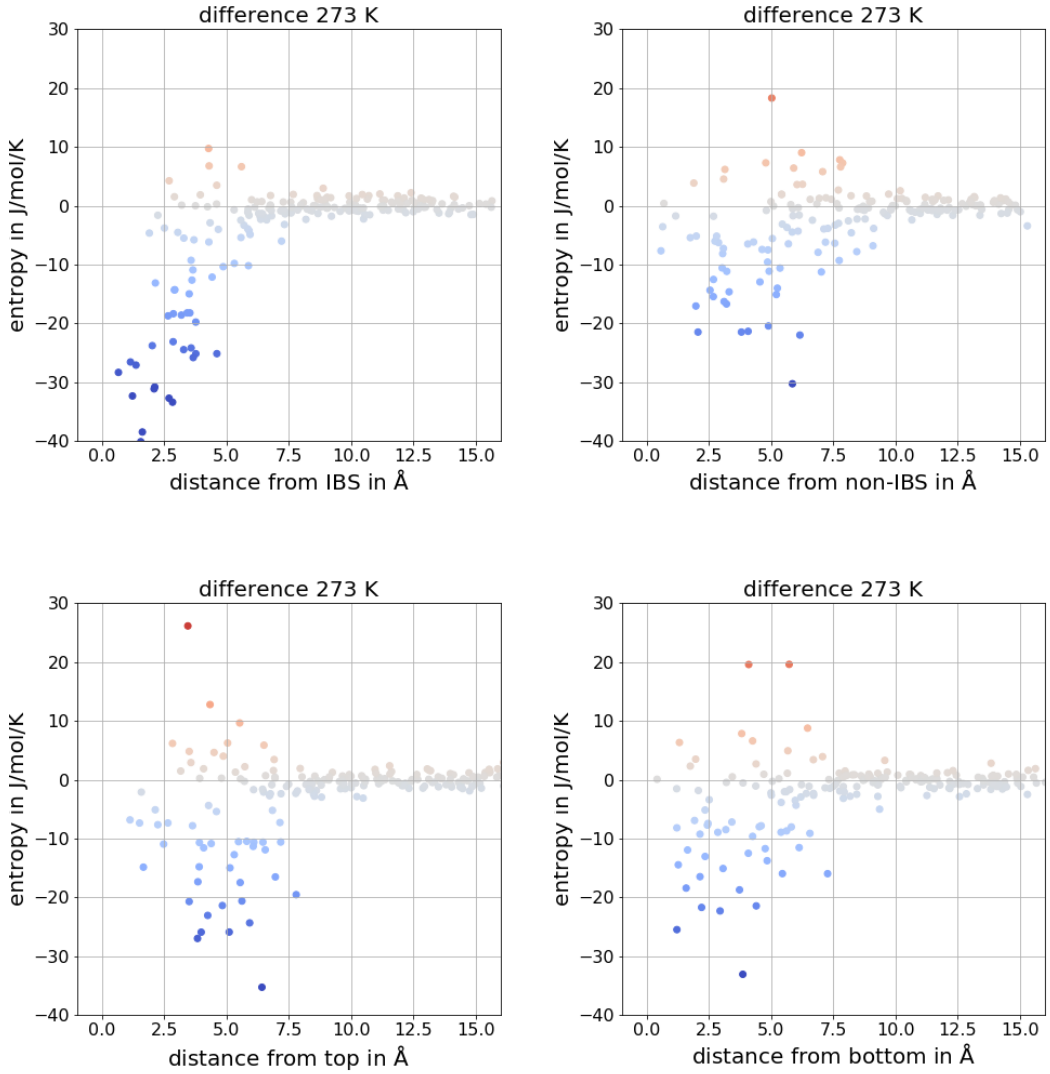


Figure D.8.: Difference between TmAFP and mutant 273 K

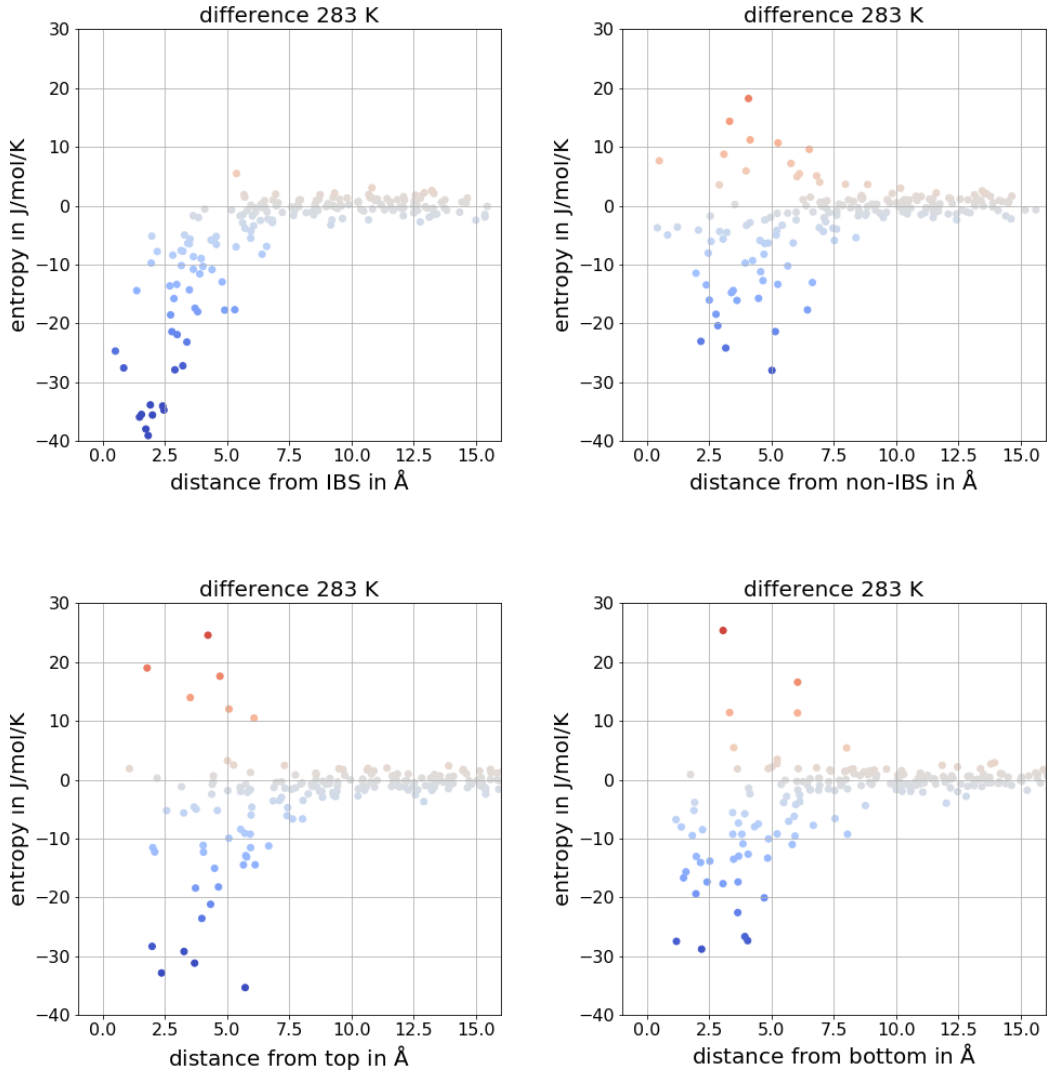


Figure D.9.: Difference between TmAFP and mutant 283 K

Bibliography

[1] "Rcsb protein data bank, 1ezg." [Online]. Available: https://www.rcsb.org/structure/1ezg

[2] W. Humphrey, A. Dalke, and K. Schulten, "VMD – Visual Molecular Dynamics," *Journal of Molecular Graphics*, vol. 14, pp. 33–38, 1996.

[3] Y.-C. Liou, A. Tocilj, P. L. Davies, and Z. Jia, "Mimicry of ice structure by surface hydroxyls and water of a β-helix antifreeze protein," *Nature*, vol. 406, no. 6793, pp. 322–324, 2000. [Online]. Available: https://doi.org/10.1038/35018604

[4] G. L. Fletcher, C. L. Hew, and P. L. Davies, "Antifreeze proteins of teleost fishes," *Annu Rev Physiol*, vol. 63, pp. 359–90, 2001.

[5] M. Griffith and M. Yaish, "Antifreeze proteins in overwintering plants: a tale of two activities." *Trends in plant science*, vol. 9, pp. 399–405, 09 2004.

[6] J. Gilbert, P. Hill, C. Dodd, and J. Laybourn-Parry, "Demonstration of antifreeze protein activity in antarctic lake bacteria," *Microbiology Society*, vol. 150, pp. 171–180, 2004.

[7] M. Kuiper, C. Morton, S. Abraham, and A. Gray-Weale, "The biological function of an insect antifreeze protein simulated by molecular dynamics," *eLife*, 2015. [Online]. Available: https://doi.org/10.7554/eLife.05142

[8] S. P. Graether, M. J. Kuiper, S. M. Gagné, V. K. Walker, Z. Jia, B. D. Sykes, and P. L. Davies, "β-Helix structure and ice-binding properties of a hyperactive antifreeze protein from an insect," *Nature*, vol. 406, no. 6793, pp. 325–328, 2000. [Online]. Available: https://doi.org/10.1038/35018610

[9] J. Ramos-Elorduy, E. A. González, A. R. Hernández, and J. M. Pino, "Use of Tenebrio molitor (Coleoptera: Tenebrionidae) to Recycle

Organic Wastes and as Feed for Broiler Chickens," *Journal of Economic Entomology*, vol. 95, no. 1, pp. 214–220, 02 2002. [Online]. Available: https://doi.org/10.1603/0022-0493-95.1.214

[10] E. Leinala, P. Davies, D. Doucet, M. Tyshenko, V. Walker, and Z. Jia, "A β-helical antifreeze protein isoform with increased activity, structural and functional insights," *Journal of Biological Chemistry*, vol. 277, no. 36, pp. 33 349–33 352, 2002. [Online]. Available: https://doi.org/10.1074/jbc.M205575200

[11] U. S. Midya and S. Bandyopadhyay, "Interfacial water arrangement in the ice-bound state of an antifreeze protein: A molecular dynamics simulation study," *Langmuir*, vol. 33, no. 22, pp. 5499–5510, 2017. [Online]. Available: https://doi.org/10.1021/acs.langmuir.7b01206

[12] ——, "Operation of kelvin effect in the activities of an antifreeze protein: A molecular dynamics simulation study," *The Journal of Physical Chemistry B*, vol. 122, no. 12, pp. 3079–3087, 2018, pMID: 29488381. [Online]. Available: https://doi.org/10.1021/acs.jpcb.8b00846

[13] A. Hudait, D. R. Moberg, Y. Qiu, N. Odendahl, F. Paesani, and V. Molinero, "Preordering of water is not needed for ice recognition by hyperactive antifreeze proteins," *Proceedings of the National Academy of Sciences*, vol. 115, no. 33, pp. 8266–8271, 2018. [Online]. Available: https://www.pnas.org/content/115/33/8266

[14] U. S. Miya and S. Bandyopadhyay, "Role of polar and nonpolar groups in the activity of antifreeze proteins: A molecular dynamics simulation study," *The Journal of Physical Chemistry B*, vol. 122, no. 40, pp. 9389–9398, 2018, pMID: 30222341. [Online]. Available: https://doi.org/10.1021/acs.jpcb.8b08506

[15] Y. Qiu, E. Hudait, and V. Molinero, "How size and aggregation of ice-binding proteins control their ice nucleation efficiency," *Journal of the American Chemical Society*, vol. 141, pp. 7439 – 7452, 2019.

[16] D. R. Nutt and J. C. Smith, "Dual function of the hydration layer around an antifreeze protein revealed by atomistic molecular dynamics simulations," *J. Am. Chem. Soc.*, vol. 130, no. 39, pp. 13 066–13 073, 2008.

[17] J. A. Raymond and A. L. DeVries, "Adsorption inhibition as a mechanism of freezing resistance in polar fishes," *PNAS*, vol. 74, pp. 2589–2593, 1977. [Online]. Available: https://doi.org/10.1073/pnas.74.6.2589

[18] C. A. KNIGHT and A. L. DEVRIES, "Melting inhibition and superheating of ice by an antifreeze glycopeptide," *American Association for the Advancement of Science*, vol. 245, pp. 505–507, 1989. [Online]. Available: https://doi.org/10.1126/science.245.4917.505

[19] A. J. Scotter, C. B. Marshall, L. A. Graham, J. A. Gilbert, C. P. Garnham, and P. L. Davies, "The basis for hyperactivity of antifreeze proteins," *Cryobiology*, vol. 53, no. 2, pp. 229–239, 2006. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0011224006001106

[20] J. Baardsnes, L. H. Kondejewski, R. S. Hodges, H. Chao, C. Kay, and P. L. Davies, "New ice-binding face for type i antifreeze protein," *FEBS*, vol. 463, pp. 87–91, 1999. [Online]. Available: https://doi.org/10.1016/S0014-5793(99)01588-4

[21] Y. Kobashigawa, Y. Nishimiya, K. Miura, S. Ohgiya, A. Miura, and S. Tsuda, "A part of ice nucleation protein exhibits the ice-binding ability," *FEBS Letters*, vol. 579, no. 6, pp. 1493–1497, 2005. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0014579305001596

[22] R. Hockney and J. Eastwood, "Computer simulation usingparticle," *McGraw-Hill, New York*, 1981.

[23] L. P. Heinz and H. Grubmüller, "Per|mut: Spatially resolved hydration entropies from atomistic simulations," *Journal of Chemical Theory and Computation*, vol. 17, no. 4, pp. 2090–2098, 2021, pMID: 33710881. [Online]. Available: https://doi.org/10.1021/acs.jctc.0c00961

[24] L. P. Heinz, "Spatially resolved hydration statistical mechanics at biomolecular surfaces from atomistic simulations," dissertation, Georg-August-Universität Göttingen, 2020.

[25] F. Reinhard and H. Grubmüller, "Estimation of absolute solvent and solvation shell entropies via permutation reduction," *The Journal of*

Chemical Physics, vol. 126, no. 1, p. 014102, 2007. [Online]. Available: https://doi.org/10.1063/1.2400220

[26] L. P. Heinz and H. Grubmüller, "Computing spatially resolved rotational hydration entropies from atomistic simulations," *Journal of Chemical Theory and Computation*, vol. 16, no. 1, pp. 108–118, 2020, pMID: 31822062. [Online]. Available: https://doi.org/10.1021/acs.jctc.9b00926

[27] A. Kraskov, H. Stögbauer, and P. Grassberger, "Estimating mutual information," *Physical Review E*, vol. 69, 2004.

[28] W. Zhang and R. A. Laursen, "Structure-function relationships in a type i antifreeze polypeptide the role of threonine methyl and hydroxyl groups in antifreeze activity," *J. Biol. Chem.*, vol. 273, p. 3480634812, 1998.

[29] H. Bekker, H. Berendsen, E. Dijkstra, S. Achterop, R. Vondrumen, D. Vanderspoel, A. Sijbers, H. Keegstra, and M. Renardus, "Gromacs - a parallel computer for molecular-dynamics simulations," in *PHYSICS COMPUTING '92*, R. DeGroot and J. Nadrchal, Eds. World Scientific Publishing, 1993, pp. 252–256, 4th International Conference on Computational Physics (PC 92) ; Conference date: 24-08-1992 Through 28-08-1992.

[30] W. Jorgensen, J. Chandrasekhar, J. Madura, R. Impey, and M. Klein, "Comparison of simple potential functions for simulating liquid water," *The Journal of Chemical Physics*, vol. 79, no. 2, p. 926–935, 1983.

[31] J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B. L. de Groot, H. Grubmüller, and A. D. MacKerell, "Charmm36m: an improved force field for folded and intrinsically disordered proteins," *Nature Methods*, vol. 14, no. 1, pp. 71–73, 2017. [Online]. Available: https://doi.org/10.1038/nmeth.4067

[32] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, "Molecular dynamics with coupling to an external bath," *The Journal of Chemical Physics*, vol. 81, no. 8, pp. 3684–3690, 1984. [Online]. Available: https://doi.org/10.1063/1.448118

[33] G. Bussi, D. Donadio, and M. Parrinello, "Canonical sampling through velocity rescaling," *The Journal of Chemical Physics*, vol. 126, no. 1, p. 014101, 2007. [Online]. Available: https://doi.org/10.1063/1.2408420

[34] M. Parrinello and A. Rahman, "Polymorphic transitions in single crystals: A new molecular dynamics method," *Journal of Applied Physics*, vol. 52, no. 12, pp. 7182–7190, 1981. [Online]. Available: https://doi.org/10.1063/1.328693

[35] F. Reinhard, O. Lange, J. Hub, J. Haas, and H. Grubmüller, "g_permute: Permutation-reduced phase space density compaction," *Computer Physics Communications*, vol. 180, no. 3, pp. 455–458, 2009. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010465508003743

[36] B. Naidan, L. Boytsov, Y. Malkov, and D. Novak, "Non-metric space library manual," 2019.

[37] D. Q. Huynh, "Metrics for 3d rotations: Comparison and analysis," *Journal of Mathematical Imaging and Vision*, vol. 35, p. 155–164, 2009.

[38] K. Röttger, A. Endriss, J. Ihringer, S. Doyle, and W. Kuhs, "Lattice constants and thermal expansion of h2o and d2o ice ih between 10 and 265 k (vol 50, pg 644, 1994)," *Acta crystallographica. Section B, Structural science*, vol. 68, p. 91, 02 2012.

[39] M. Bar, Y. Celik, D. Fass, and I. Braslavsky, "Interactions of - helical antifreeze protein mutants with ice." *Cryst. Growth Des.*, vol. 8, p. 29542963, 2008.

Acknowledgements

I would like to acknowledge everyone who supported me while I wrote my bachelor thesis.

First of all, I want to thank Hon.-Prof. Dr. Karl Helmut Grubmüller, who proposed this interesting topic to me. Due to his feedback, I did not only learn a lot about AFPs and entropy calculation, but also about scientific writing.

A big thank you goes to Leonard Heinz, who supervised me and was always available for questions. Many thanks also for the introduction to molecular dynamics simulations and entropy calculation, as well as for proofreading.

I would also like to thank Marie Kleis and Jonas Willms for further proofreading.

Laura Müller

Erklärung Ich versichere hiermit, dass ich die vorliegende Arbeit ohne fremde Hilfe selbstständig verfasst und nur die von mir angegebenen Quellen und Hilfsmittel verwendet habe. Wörtlich oder sinngemäß aus anderen Werken entnommene Stellen habe ich unter Angabe der Quellen kenntlich gemacht. Die Richtlinien zur Sicherung der guten wissenschaftlichen Praxis an der Universität Göttingen wurden von mir beachtet. Eine gegebenenfalls eingereichte digitale Version stimmt mit der schriftlichen Fassung überein. Mir ist bewusst, dass bei Verstoß gegen diese Grundsätze die Prüfung mit nicht bestanden bewertet wird.

Göttingen, den September 4, 2021

(Laura Müller)