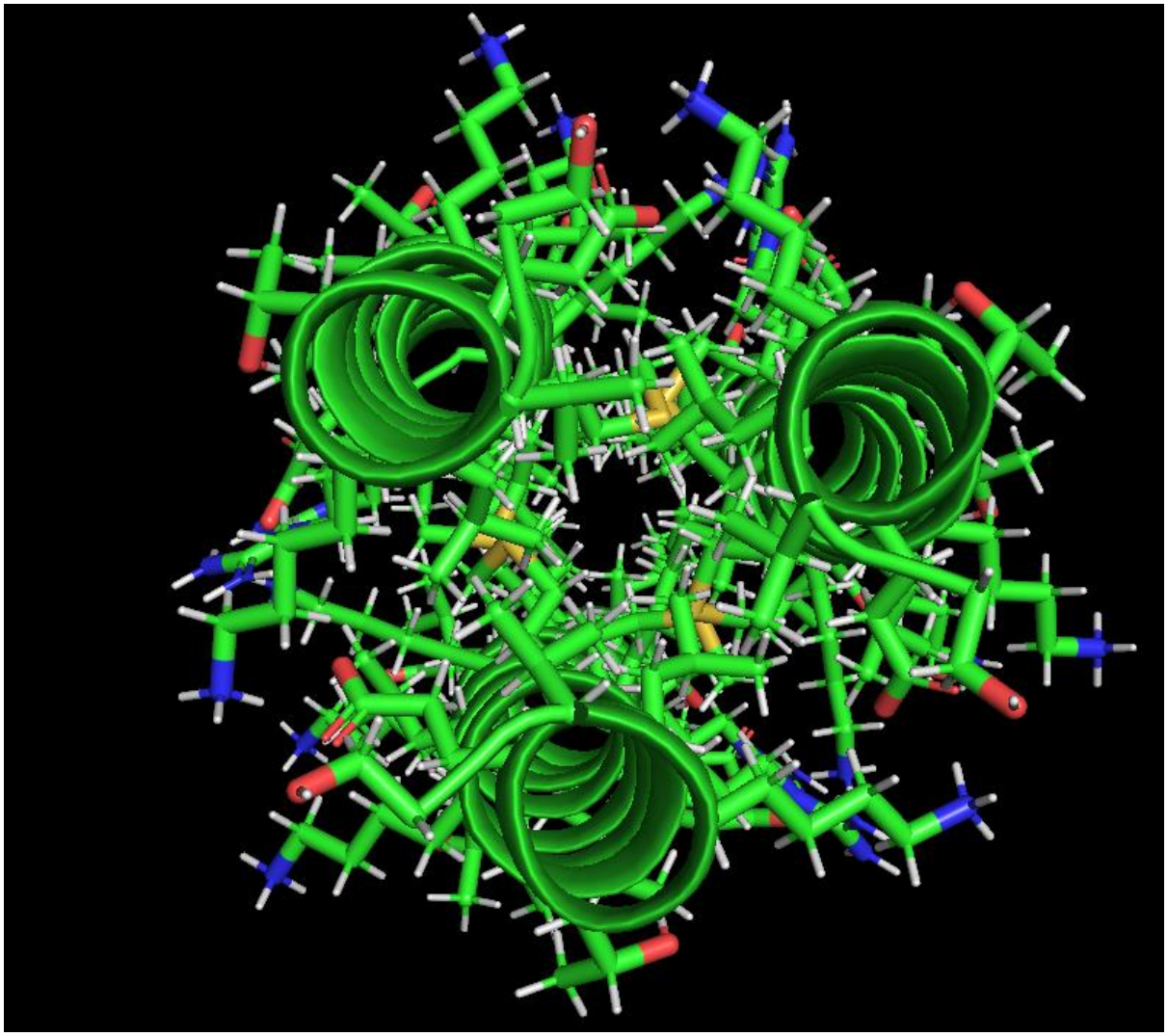
Simulating ice nucleating proteins



Joran Schoorlemmer, Jorik Schaap & Tim Woertman

Rob de Haas

22-6-2021

Introduction

Ice binding proteins have many applications, ranging from snow cannons to microvalves in microtubing. However, these proteins are currently impossible to synthesize. Specific bacteria can transport these proteins on their membrane, but creating a pure ice binding protein is impossible. This is due to the proteins having high amounts of β-helices, which are very unstable in synthesis. In this project, ice binding proteins are synthesized on a much more stable back bone, α-helices. An ice binding motif (TxxxAxxxAxx)n is bound to the α helix and is as such far easier to stabilize. Now, these proteins can be synthesized and can be used in this wide range of applications, without the hassle of dealing with living bacteria.

Workflow

First, a stable backbone had to be generated. For this, atomistic simulations were used using the Pyrosetta software. This was done on one peptide and multiplied using C3 symmetry to decrease computational workload.

Secondly, α-helices were designed using the generated backbones. The goal of this step is to do a broad scan on the influence of the input parameters on the stability of the protein, in order to find a range of optimal values which can be used in the next step. The input parameters are the coil radius, the twist of the whole protein and the phase of the individual chains. To influence the process of Pyrosetta designing an optimal amino-acid sequence, an RES-file is necessary to exclude or include certain amino-acids in general or at specific positions. Cysteines have the tendency to promote β-sheet formation and are therefore excluded from the usable amino-acid set for this broad scan. After this designing step, the proteins are visualized in PyMOL. By visual inspection of the protein, the amino-acid positions suitable for the ice-binding motive were determined.

In the next step, Pyrosetta was once again used, with the optimal parameter range as input, to find the optimal amino-acid sequence this time including the ice-binding amino-acids at fixed positions at the outside of the α-helices. The best scoring proteins were selected and relaxed. The relaxing step is necessary to assess the stability of the designed proteins.

In the last step, the three most optimal amino-acid sequences were analysed by an ab initio, fold and dock algorithm. The algorithm works in the following way: an amino-acid sequence is fragmentized into 3-mers and 9-mers and is folded into a new configuration by Monte-Carlo moves. This process is repeated multiple times. The resulting score values are stored, together with the RMSD value relative to the input model. To decrease computation time, these 9-mer and 3-mer fragments are structured according to values found in literature for the specific amino acids. For example into α helices.

Results

The results of multiple parameter sampling runs are displayed below, the three parameters sampled were the radius of the designed helices packed together, the twist of the helices and the phase of the helices.

Results radius

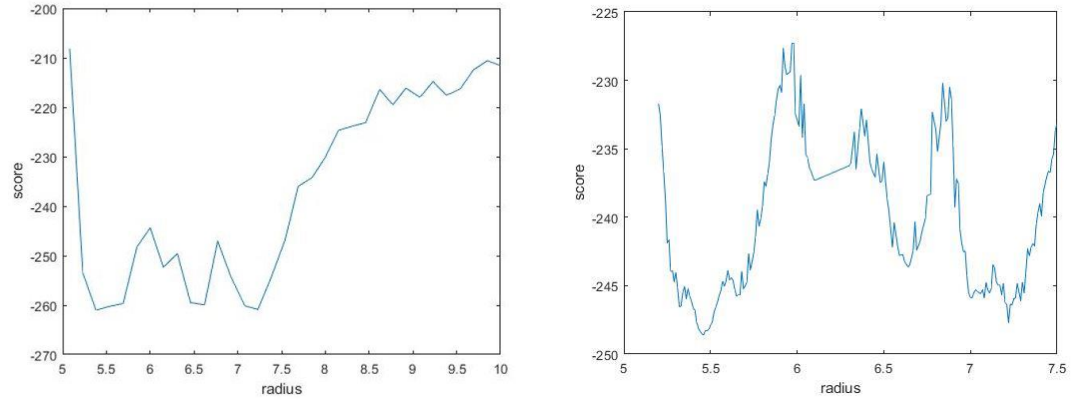


Figure 1: Minimum score of designed peptides per radii. Left) Broad scan between 5 and 10 Angstrom. Right) detailed scan of the three local minima between 5.2 and 7.7 angstrom.

From figure 1, it follows that a change in radius has a major impact on the score, the optimal range for the peptide trimer to form seems to be between 5.2 and 7.7 angstrom where a global minimum is observed. Within in the global minimum 3 distinctive local minima are present at approximately 5.5, 6.6 and 7.3 angstrom. These local minima arise from the fact that for larger radii, larger amino acids can be used to construct the helix. This results in the helices around 5.5 angstrom to consist mainly out of alanine and the chain at 7.3 angstrom out of larger hydrophobic residues like leucine.

The presence of alanine generally results in very stable helices, but larger hydrophobic residues are desired to increase the amount of hydrophobic interactions between the helices and thus potentially yielding a more stable trimer.

Results twist and phase

To investigate the effect of the twist angle and the phase of the helix, these parameters were sampled as well. The sampling of the phase for a fixed radius and twist angle, yielded very similar score values ranging within 8 points, which is not very significant.

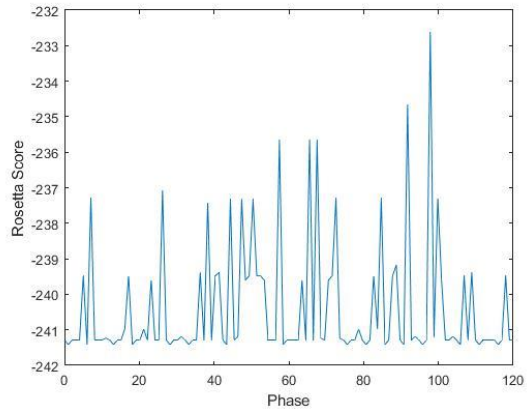


Figure 2: sampled phase angles, for a twist of degrees and a radius of 5.8 Angstrom

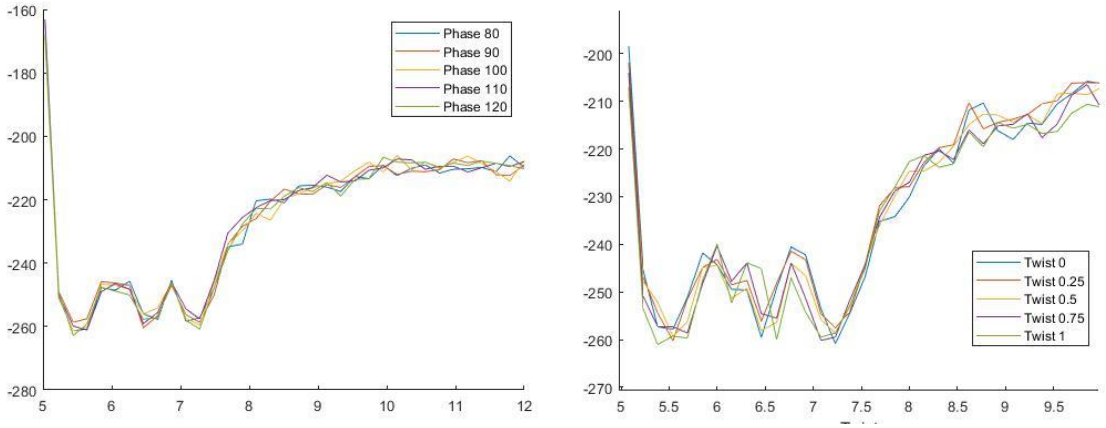


Figure 3: The minimum score values of the sampled phase and twist angles for a certain radius.

From figure 3 it again follows that the parameters of phase and twist do not lead to significant changes the difference in score is never larger than 10 points and exactly the same pattern is followed for all values of phase and twist.

Results fold and dock

Three sequences found in the design step were used in the fold an dock. These input models have the following set parameters:

Model 1: twist =1, radius = 5.5 Å and phase = 30
Sequence: AEVEAALTKALAAVKAALTKALAAVKAALTKAS

Model 2: twist =0.5, radius = 6.61 Å and phase = 80
Sequence: AELEATITEMLARMKALETKVLALIKALETKVS

Model 3: twist = 0.5, radius = 6.65 Å and phase = 80
Sequence: AELEATITEVLARIKALKTEMLARMKALETKVS

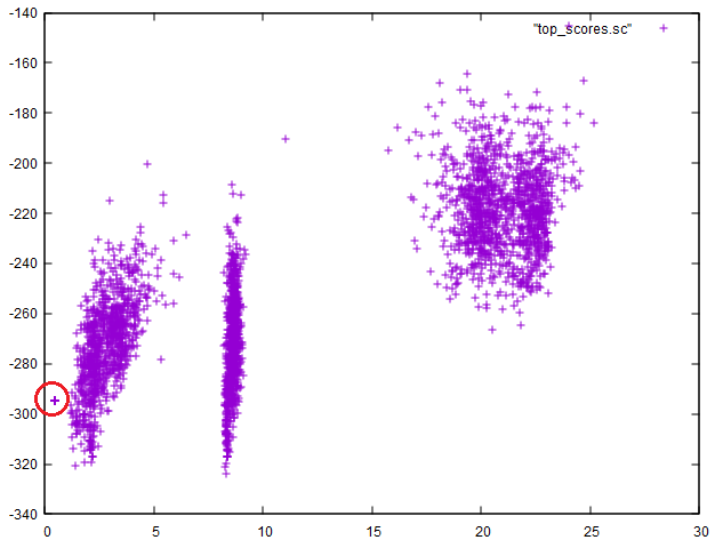


Figure 4: The results from the fold and dock step of sequence 1. The score is depicted on the y-axis and the RMSD in Å on the x-axis. The results from the relax step of a structure with twist =1, radius = 5.5 and phase = 30 are depicted with the purple crosses in the red circle.

For sequence 1, an unexpected result appears. Configurations are found which are much lower in energy relative to the expected structure with score values of -295. This indicates a different configuration has formed in the monte carlo folds. Also, a completely different structure with RMSD = 8 has formed. This could be disastrous to the formation of the original structure because multiple energy minima of completely different configurations are possible.

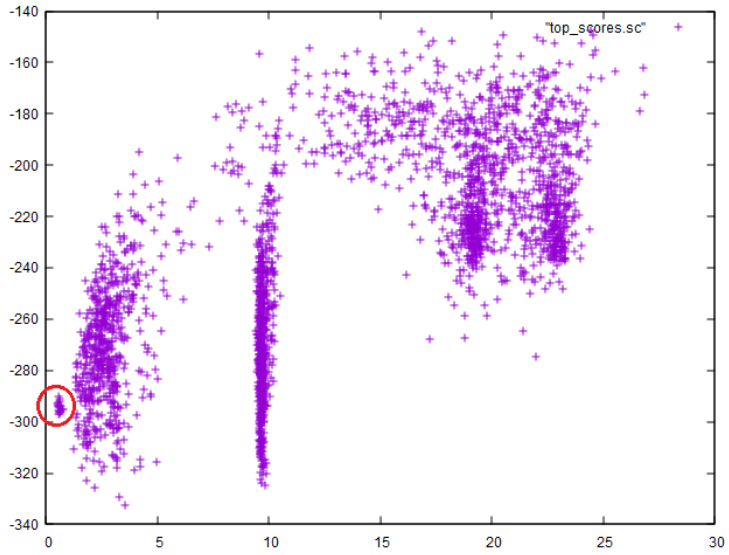


Figure 5: The results from the fold and dock step of sequence 2. The score is depicted on the y-axis and the RMSD in Å on the x-axis. The results from the relax step of a structure with twist =0.5, radius = 6.61 and phase = 80 are depicted with the purple crosses in the red circle.

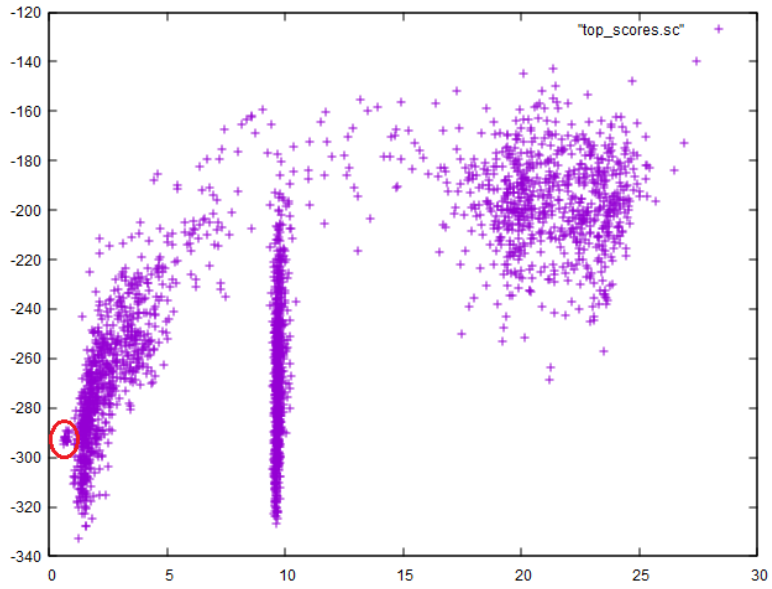


Figure 6: The results from the fold and dock step of sequence 3. The score is depicted on the y-axis and the RMSD in Å on the x-axis. The results from the relax step of a structure with twist = 0.5, radius = 6.65 and phase = 80 are depicted with the purple crosses in the red circle.

Very similar results are found for sequence 2 and 3. They still funnel towards a different protein configuration and show a similar structure at RMSD = 10. These also indicate the resulting structure to be different from the desired structure, like sequence 1.

The models with the lowest energy are further analysed in PyMol.

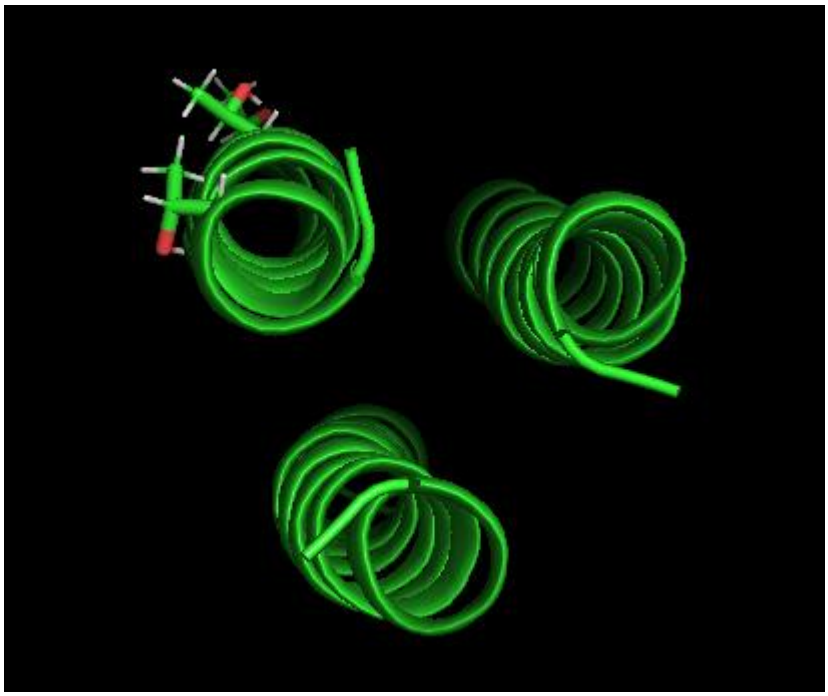


Figure 7: A picture of the model with the lowest energy, found in the fold and dock step for sequence 3.

As is visible in Figure 7, the Threonine groups are not aligned. This is a major problem for the effectivity of the ice binding motif.

Conclusion

A stable structure containing the ice binding motif was found during the design phase. Multiple configurations were found with low score values. The coil radius was found to be the most impactful parameter, while the twist and the phase of the protein seem to have less effect. Three local minima were found for the radius, these are 5.2-6 Å, 6.3-6.9 Å, 6.9-7.5 Å. The effect of the larger radii was a decrease in attraction between the three chains. Meanwhile, larger, more hydrophobic amino acids could be sampled in the core, resulting in a lower score.

However, when looking at the fold and dock step, the likelihood of creating this structure from scratch is very low. Three different sequences were looked at, these all resulted in a slightly different structure with RMSD = 1.5 Å. For all 3, another completely different structure was present at RMSD = 8 or 10 Å. This indicates that the probability of the desired structure forming in the lab is very low.

In the future, a broader range of input parameters must be taken in account. This could be a higher number of chains, fixing hydrogen bonds in the core or other parameters.