Computationally Designed Thermostable Variant of Barley β-Glucanase

Eric Yang, Tatum Prosswimmer, and Valerie Daggett

Bioengineering 488 – Computational Protein Design (Winter 2020), University of Washington, Box 355013, Seattle, Washington 98195-5013, USA.

Introduction

Many enzymes are essential in the brewing process to convert raw ingredients such as rice, rye and wheat into beer. In the barley processing workflow, *β-Glucanase* enzymes hold an important role and are released during the malting process. Specifically, *endo-glucanase,* a derivative of the *β-Glucanase* family of enzymes, hydrolyzes β1→4 linkages next to β1→3 linkages found in barley glucans (Varghese et al, 1994). Next, in the mashing process, malted barley undergoes high temperature conditions to release critical enzymes and fermented sugars. In this step, *β-Glucanase* is needed to break down barley glucan polymers to reduce viscosity for taste and aesthetics (Bamforth et al, 1984). However, *β-Glucanase* starts to become inactive at 45C. Therefore, temperature must be carefully tuned to preserve the structural and functional integrity of the enzyme. In addition, above boiling temperatures are used in downstream processes. Due to these workflows, there is tremendous need in developing a thermostable *β-Glucanase* that can maximize enzyme efficacy at high temperatures.

There have been many 3D structures of the *β-Glucanase* enzyme studied experimentally by X-ray crystallography, nuclear magnetic resonance spectroscopy and cryo-electron microscopy, available on online protein databases. With this structural basis, there is significant opportunity to optimize the enzyme while retaining its functionality to satisfy desired performance. Here, one approach was studied, where modifications to the *β-Glucanase* enzyme were done in attempt to stabilize the structure electrostatically and thermodynamically. First, the FoldIt software was first used to make bulk optimization to the amino acids into energetically favorable residues, while maintaining the antigenicity of the epitope. The orientation of the resulting residues was balanced carefully with regards to entropy and electrostatic interactions.

After the mutations were globally identified and completed, a region of residues was manually recognized as a potential for further improvement due to its instability at higher temperature. Specifically, mutations K281N, N282Q and T289Q were made with the intention of bulkier non-polar side chain packing to contribute to the hydrophobic core. By inducing hydrophobic interactions between large nonpolar residues, it was hypothesized that the residues would pack tightly with one another, forming a more stable core structure. This study serves as a proof of concept for a significant artificially mutated *β-Glucanase* for increased thermostability.

Methods

Design and Mutations on Wild Type β-Glucanase

To achieve the final protein design of the *β-Glucanase* enzyme, the wild type structure was first modified in the FoldIt software, where non-epitope regions of the enzyme were optimized with minimization algorithms. Specifically, the side chains were roughly optimized to minimize thermodynamics energy scores, mutated for optimal amino acid interactions, and then further "fine grain" minimized for electrostatic interactions. The epitope residues that were unmodified to retain enzyme activity were Glu288, Glu232, Tyr33, Glu93, Asn92, Val134, Phe135, Asn168, Tyr170, Leu173, Phe275, Glu280, Lys283, and Trp291 (Müller et al, 1998). Then, the structure was modified in Chimera to create additional hydrophilic surface residues and hydrophobic core residues by intuition, optimized by molecular dynameomics probabilities. The unchanged active site of the wild type *β-Glucanase* enzyme and the whole extent of the sequence differences between the wild type and the final enzyme design can be found in Figure 1.

Molecular Dynamics Simulation

The two *β-Glucanase* structures were prepared with in lucem molecular mechanics for molecular dynamic simulations (Beck et. al, 2000-2018). Prior to the simulations, Hydrogen atoms were added manually to the crystal structure in all of the pdb files via FoldIt. Then, the *in lucem* algorithm further prepared the files by asserting that histidine residues are denoted by "HIE" and that all C-terminal oxygen atoms are denoted by "OT". The algorithm also performed other residue optimizations in general, although not relevant for the two *β-Glucanase* structures of interest. After the in lucem preparations, the molecular mechanics parameter library (mmpl) for the enzymes were generated, where all residues and bonds not within a single residue were listed. The proteins first underwent 500 steps of minimization. Next, through steepest descent, all atoms were minimized for 1000 steps. The worst bonded energies where confirmed to be less than 10 and the nonbonded energies confirmed to be less than 50. The enzymes were then solvated in a box contained explicitly with water. For each of the enzyme structure, three temperatures of 25C, 45C and 75C were performed to simulate temperatures of interest. Then, the prep workbook instructed further minimization of the enzymes, where solvent atoms were minimized for 1000 steps, equilibrated for 500 steps and minimized again for 500 steps. The enzyme atoms were minimized for an additional 500 steps. Finally, the molecular dynamics simulations which lasted 5 nanoseconds for each of the two structures were performed for the various temperatures.

Analysis of Molecular Dynamics Simulation

The four main metrics that were calculated by algorithms during the simulation were Cα root-mean-squared-deviation (RMSD) values, residue contact plots, root-mean-squared-fluctuation (RMSF) values, and DSSP secondary structures, with RMSF not reported in this work. Each of these metrics were included inside of module files after molecular dynamic

simulations. The workbooks and enzyme structures were transferred to the local machine for analysis, see figures. The Cα RMSD values were calculated based on the structural variance of the models during timepoints of the simulations compared to a reference model. They were analyzed with respect to simulation timepoints as well as individual residues. The contact plots were generated based on varying degree of native and nonnative interactions between and within main chain and side chain residues for the wild type and final design of *β-Glucanase* enzymes. The native and nonnative structures were determined based on the degree of residue interactions pre and post simulation. These interactions include hydrogen bonding, hydrophobic-hydrophilic interactions and bond angles among others. Finally, the DSSP secondary structures were generated based on structural characteristics such as helix, loop, bridge, and sheet phenotypes throughout the whole simulation with respect to individual residues.

Statistics and Plotting

All statistical analyses and figures were generated with the simulation algorithms, Chimera or manually with Microsoft Excel.

Results

Rosetta energy scores and sequence identity

After the modifications of the wild type *β-Glucanase*, the modified design compared to the wild type had a sequence identity of 65.6%. While the majority of the residues were preserved, a significant number of residues were mutated both in the core and on the surface. Comparing the Rosetta energy scores, which accounts for both electrostatic and hydrophobic interactions, the designed enzyme had a score of -328.019 while the wild type *β-Glucanase* had a score of 456.72. Although the scores are not all-encompassing, this tremendous difference in

energy scores suggests that the designed structure with the optimized electrostatic interactions exhibits and fosters much more stable and lower energy interactions between residues.

Protein backbone deviations

The enzyme structural deviations were calculated and assessed primarily with RMSD values. Alpha carbons on the main chain were compared to a reference structure with each residue's carbon's distances normalized. Over the course of MD simulations at 25, 45 and 75C, the wild type enzyme resulted with RMSD averages of 2.21, 2.74, 3.91 Angstroms and a standard deviation of 0.25, 0.35, 0.78 respectively. At the end of the simulations, the RMSD values all reached a plateau, at about 2.5, 3, and 4.5 Angstroms at 25, 45, and 75C after a steady increase over the time course. For the same temperatures, the designed enzyme resulted with RMSD averages of 2.63, 2.47 and 3.07 significantly. At the end of the simulations, the RMSD values reached a plateau at about 2.5, and 3.5 Angstroms at 45, and 75C after a steady increase over the time course. For the 25C designed enzyme simulation, the RMSD value seems to still be fluctuating at the end of the simulation slightly above 3 Angstroms. Of note is the comparison between the two structures at 45 and 75C, with the designed enzyme holding much lower RMSD values, suggesting a more stable and less deviant main chain structures.

Protein deviations by residue

The RMSD values by residue were also calculated for the wild type and the designed *β-Glucanase* enzyme. Here, instead of assessing RMSD values at different timepoints over the course of the simulation, the reference point was shifted to individual residues, where the average for each residue over the full time course was assessed. For both structures at all temperatures, at the ends of the protein where there are much more degrees of freedom for residues to conform, the RMSD values were high. In addition, as temperatures increase, both

enzymes exhibited higher overall RMSD values, suggesting higher entropy and instability as one would expect. At higher temperatures of 45 and 75C, the wild type structure's residues had higher RMSD averages compared to the mutated design. Specifically, at regions of the enzyme that were heavily mutated (around residues 100, 200, and 250), the mutated design's RMSD values dropped well below those of the wild type structure. Most of the mutated residues had lower RMSD values compared to the wild type residues at the same position, suggesting minimal deviations from the reference structure especially at higher temperatures.

Core contact analysis

For each of the structure and at various temperatures, the residue interactions were assessed via contact plots (Figure 5). Over the course of simulation, the blue data points represent native contacts that exist in starting structures while the red data points represent nonnative contacts resulting from thermodynamics and electrostatics over the course of the simulation. Specifically, the alpha helix of the wild type and the beta sheets for both structures are highlighted in the diagonal, with native contacts present all throughout. The inter-sheet and inter-helix structures were represented by the off diagonal. Comparing the contact plots between the wild type and mutated design, both structures had fairly consistent native and nonnative contacts over time and compared to each other. For both enzymes, the structures were well retained throughout the simulation, even as temperatures increased significantly. In addition to the optimization of electrostatic interactions, the minor core deviations also contributed to the improvement in stability as the plots suggest.

Secondary structure analysis

For all the MD simulations at various temperatures, the Define Secondary Structure of Proteins (DSSP) plots were generated for analysis. Hydrogen bonds between backbones were

evaluated to infer different secondary structures in the algorithm (Kabsch & Sander, 1983). Consistent with visual inspection, all structures were largely composed of alpha helices, which maintained as temperatures increased. Comparing the two structures, the designed enzyme developed more 3/10 loops and had more residues participate in beta sheet interactions while the wild type enzyme did not. This can be seen especially in the 75C condition. For both structures at all temperatures, the number of residues participating in any form of secondary structure increased as the enzymes are given time to optimize. Inspecting the DSSP plots macroscopically, at all temperatures, the designed structure had more residues involved in secondary structures both in the core and at the N and C termini.

Discussion

With all the metrics available, it is clear that the modified protein structure did in fact result in higher stability thermodynamically and electrostatically at higher temperatures. An overhaul in residue optimization resulted to a more structurally stable protein at 75C without reduced functionality at the active site. At a lower temperature of 25C, the wild type does seem to be slightly more stable compared to the designed enzyme. However, this may have a minimal effect with the high operation temperatures of the brewing processes.

Looking at the RMSD values, for a protein with around 300 residues, the protein's structural deviation was reasonably low across the whole simulation as well as for some key active site residues. Stronger residue-residue interactions were evident in the mutated protein's contact plots that contributed to further electrostatic stability. Observing the evolution of secondary structures, it seems that these mutations to the primary structure significantly impacted the residues' access and ability to form more stable loop and helical phenotypes locally. Even with such a drastic change in residue identity, most of the wild type secondary

structures were preserved and additional favorable interactions were formed over the course of the simulations.

From introducing metal particles to optimizing protein's loop structures, there have been many approaches demonstrated that lead to increased protein stabilization (Bier et. al 2016). With the ends shown to be more "floppy", protective capping of N and C termini residues also hold promise in preserving the antigenicity of the protein function in the context of this *β-Glucanase* structure. Ultimately, a stable core and favorable surfaces are critical for a protein's stability. Whether through favorable hydrophobic, electrostatic, enthalpic, entropic, solvent accessibility approaches, it is essential to create a protein interior that is not negatively exposed to the environment as well as an exterior that can foster stable interactions under stress.

Conclusions

We have successfully engineered a thermostable *β-Glucanase* enzyme structure while preserving its hydrolyzing functionalities. Through significant mutation both in the protein core and surface, it is clear that such an approach holds promise in preserving enzyme activity at higher temperatures. Longer simulations and experimental verifications need to be performed in order to confirm the computational design.

Acknowledgements

Protein modifications, images and optimization were completed via UCSF Chimera and the FoldIt software, developed by the Baker Lab at the University of Washington.

References

Bamforth, C. W. "Barley β-glucans: their role in malting and brewing." *Brewers Digest* 57.1 (1982): 22-35.

Beck, DAC, McCully, ME, Alonso, DOV, Daggett, V. in lucem molecular mechanics, University of Washington, Seattle, (2000-2018).

Beck, D. a C. & Daggett, V. Methods for molecular dynamics simulations of protein folding/unfolding in solution. Methods 34, 112–20 (2004).

Bier D., M. Bartel, K. Sies, S. Halbach, Y. Higuchi, Y. Haranosono, T. Brummer, N. Kato, C. Ottmann, ChemMedChem 11, 911 (2016).

Kabsch, W. & Sander, C., Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features, Biopolymers, Issue 22, pp. 2577-2637 (1983).

Müller, J., Thomsen, K. & Heinemann, U. Crystal Structure of Barley 1,3–1,4-β-Glucanase at 2.0-Å Resolution and Comparison with Bacillus1,3–1,4-β-Glucanase, J. Biol. Chem. 273: 3438-(1998).

Varghese, Joseph N., et al. "Three-dimensional structures of two plant beta-glucan endohydrolases with distinct substrate specificities." Proceedings of the National Academy of Sciences 91.7 (1994): 2785-2789.

Figure Legends

Figure 1. Ribbon structure and sequence comparisons of wild type and designed *β-Glucanase.* (Left) The wild type *β-Glucanase* 1ghr sequence and structure was obtained from the Protein Data Bank. Regions highlighted in magenta represent the residues part of the enzyme active site that were unchanged in the final design. The residues are listed in the methods section. (Right) The final design's sequence (Automated.pdb) and structure was mutated through FoldIt automatically with several manual modifications. (Bottom) Sequence comparison between wild type and final design of *β-Glucanase* enzyme. Residues highlighted in green indicate mutation from the wild type enzyme.

Figure 2. Trends of overall RMSD over 5 ns simulation at 25, 45 and 75C. The C-alpha root-mean-squared-deviations for both the wild type and designed *β-Glucanase* over 5 nanoseconds of MD simulation at 25, 45 and 75C are shown in the connected scatterplots.

*Note: The MD simulation of wild type *β-Glucanase* enzyme at 75C was given at the start of the project. This simulation only ran for the duration of 2 nanoseconds. To preserve computational resources, the full 5 nanosecond MD simulation for that structure at 75C was not re-run.

Figure 3. End of simulation RMSD by individual residue at 25, 45 and 75C. The C-alpha RMSD with respect to individual residues for both the wild type and designed *β-Glucanase* at the end of the MD simulations at 25, 45 and 75C are shown in the connected scatterplots. The residue indices are the reverse of the sequences shown in figure 1.

Figure 4. Average and standard deviation of overall RMSD over 5 ns simulation at 25, 45 and 75C. The average C-alpha RMSD for both the wild type and designed *β-Glucanase* over the duration of 5 nanoseconds of MD simulation at 25, 45 and 75C are displayed in bar-graph format. The error bars represent +/- one standard deviation.

Figure 5. Native and non-native contact plots for wild type and designed *β-Glucanase at* 25, 45 and 75C. Plots displaying the varying degree of native and nonnative contacts between main chain and side chain residues for the designed (Left) and wild type (Right) of *β-Glucanase* at 25, 45 and 75C of MD simulation are shown. The blue data points represent native contacts that exist in starting structures while the red data points represent nonnative contacts resulting from thermodynamics and electrostatics over the course of the simulation.

Figure 6. Secondary structure analyses for wild type and designed *β-Glucanase at* 25, 45 and 75C. The DSSP structural plots with respect to individual residues for both the wild type and designed *β-Glucanase* over the course of 5 nanoseconds MD simulations at 25, 45 and 75C are shown. The legend on the bottom highlights loop, sheet, helix, bridge and other structures as well as their color correlations. The residue indices are the reverse of the sequences shown in figure 1.

Figures

Figure 1

Wild Type	Designed (Automated.pdb)

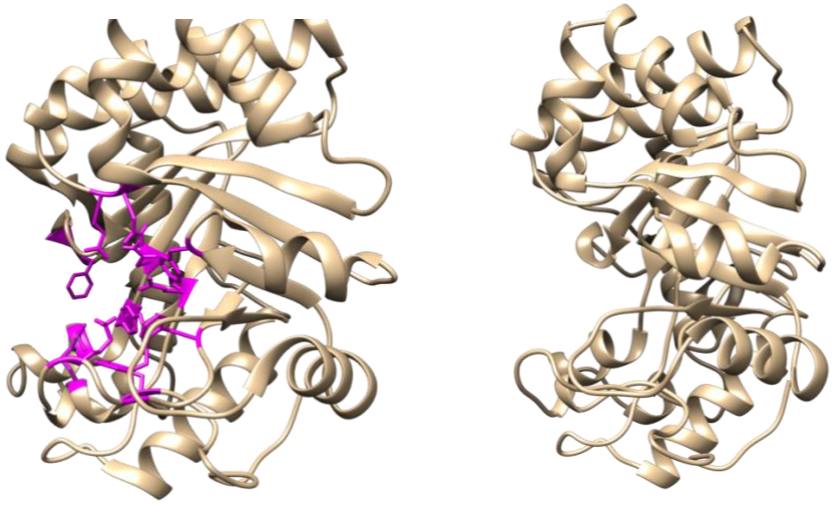


Figure 2

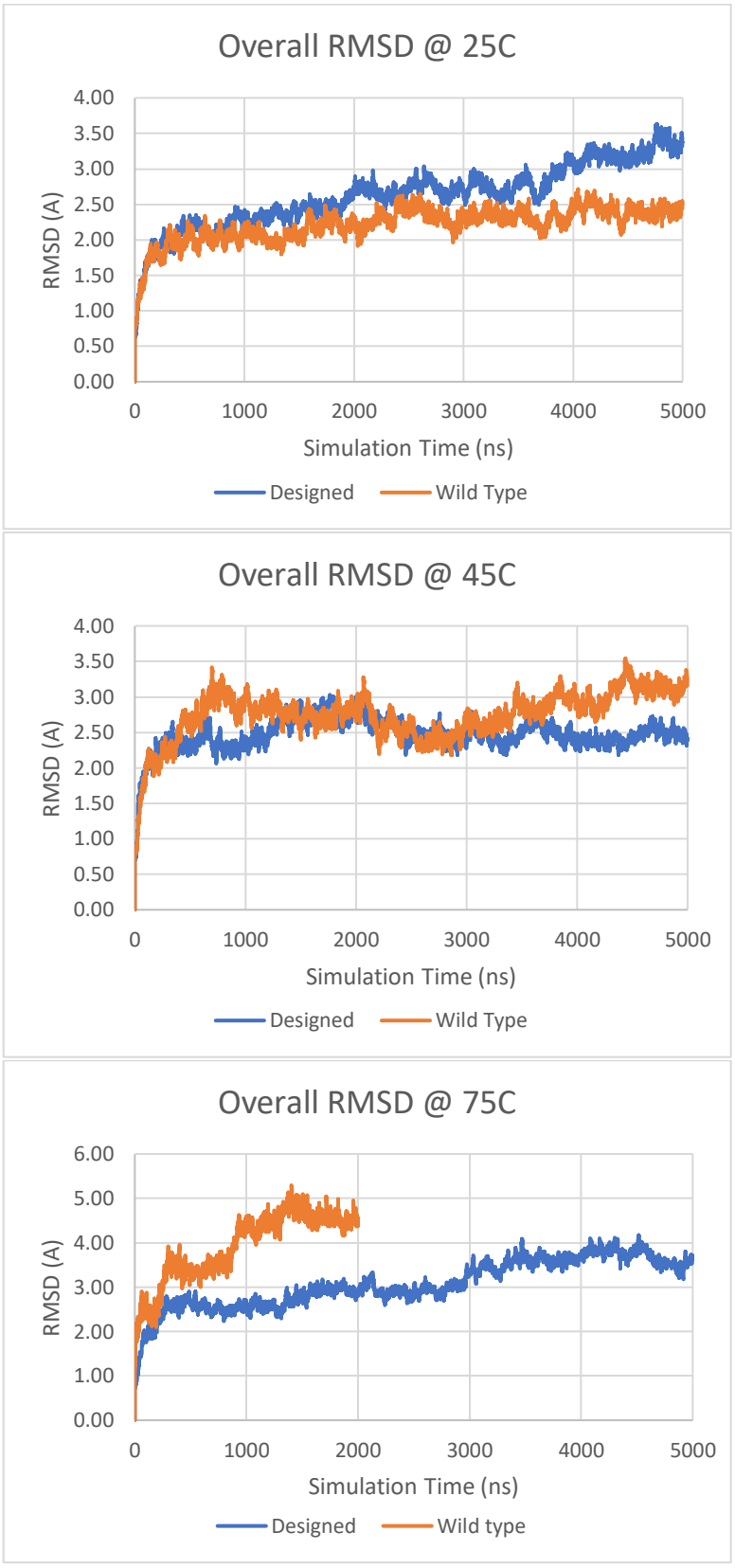
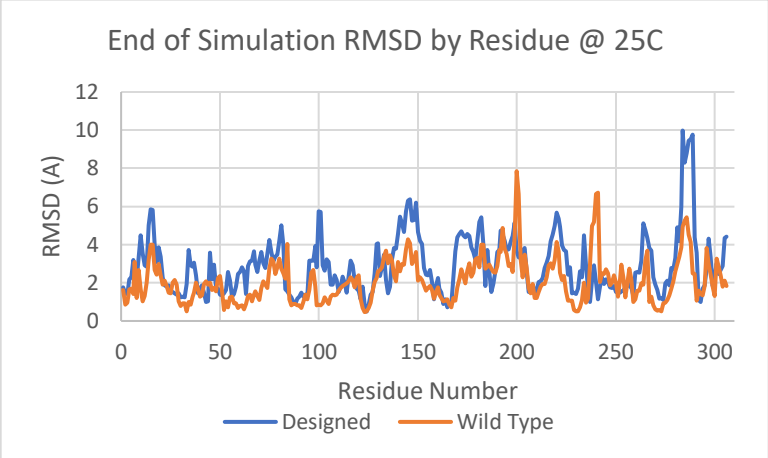
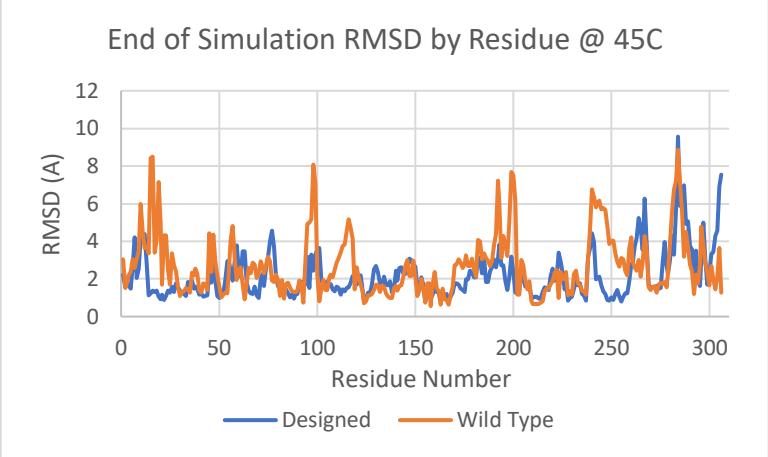

Overall RMSD @ 25C
RMSD (A)
4.00
3.50
3.00
2.50
2.00
1.50
1.00
0.50
0.00
0
1000
2000
3000
4000
5000
Simulation Time (ns)
Designed
Wild Type
Overall RMSD @ 45C
RMSD (A)
4.00
3.50
3.00
2.50
2.00
1.50
1.00
0.50
0.00
0
1000
2000
3000
4000
5000
Simulation Time (ns)
Designed
Wild Type
Overall RMSD @ 75C
RMSD (A)
6.00
5.00
4.00
3.00
2.00
1.00
0.00
0
1000
2000
3000
4000
5000
Simulation Time (ns)
Designed
Wild type

Figure 3


End of Simulation RMSD by Residue @ 25C
RMSD (A)
0
2
4
6
8
10
12
0
50
100
150
200
250
300
Residue Number
Designed
Wild Type


End of Simulation RMSD by Residue @ 45C
RMSD (A)
0
2
4
6
8
10
12
0
50
100
150
200
250
300
Residue Number
Designed
Wild Type

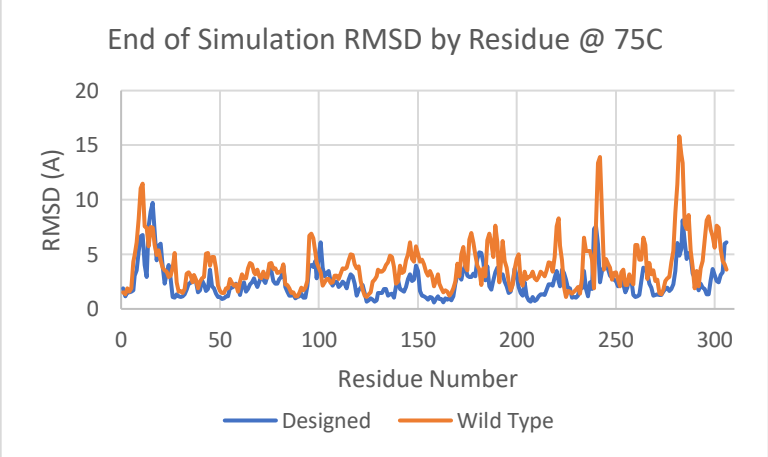

End of Simulation RMSD by Residue @ 75C
RMSD (A)
0
5
10
15
20
0
50
100
150
200
250
300
Residue Number
Designed
Wild Type

Figure 4

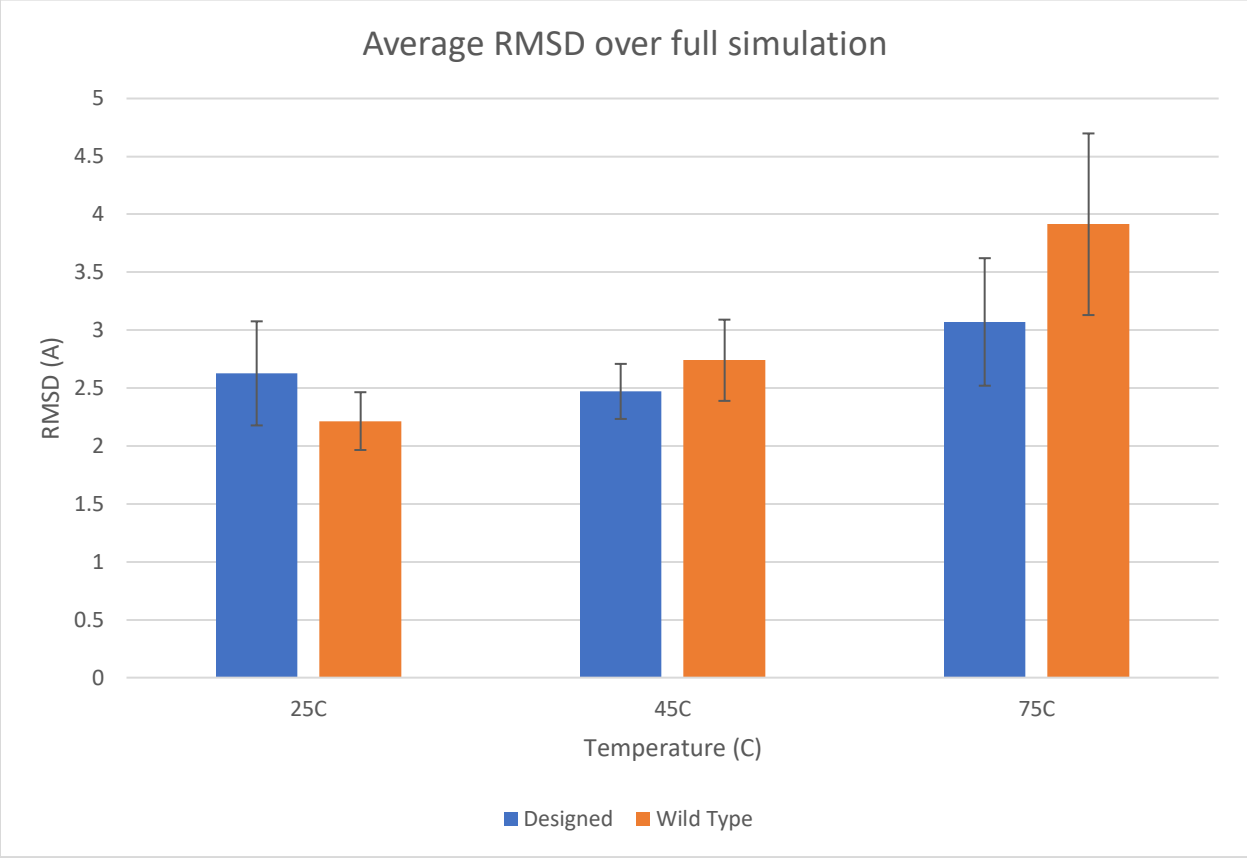

Average RMSD over full simulation
5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
RMSD (A)
25C
45C
75C
Temperature (C)
Designed
Wild Type

Figure 5

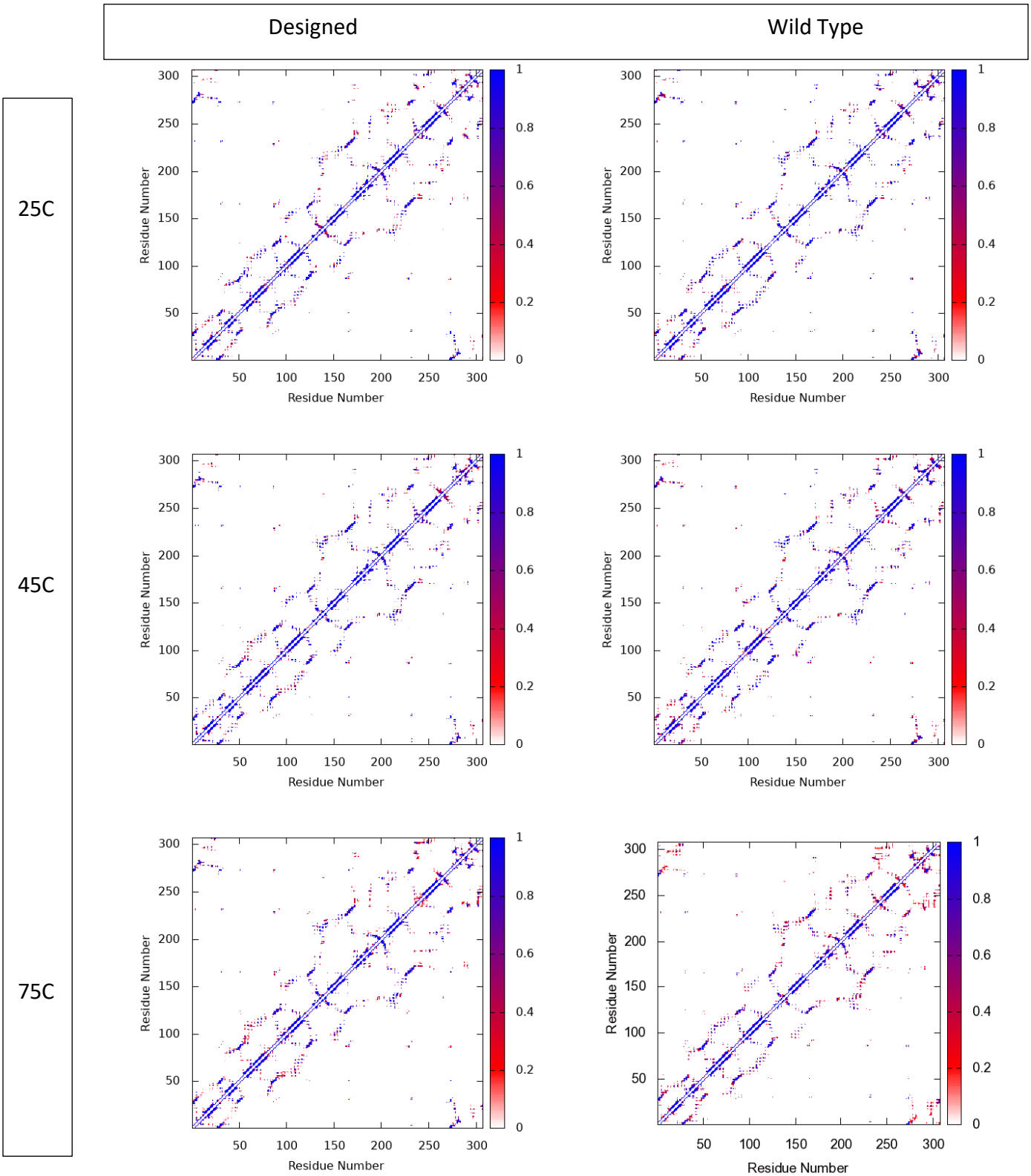
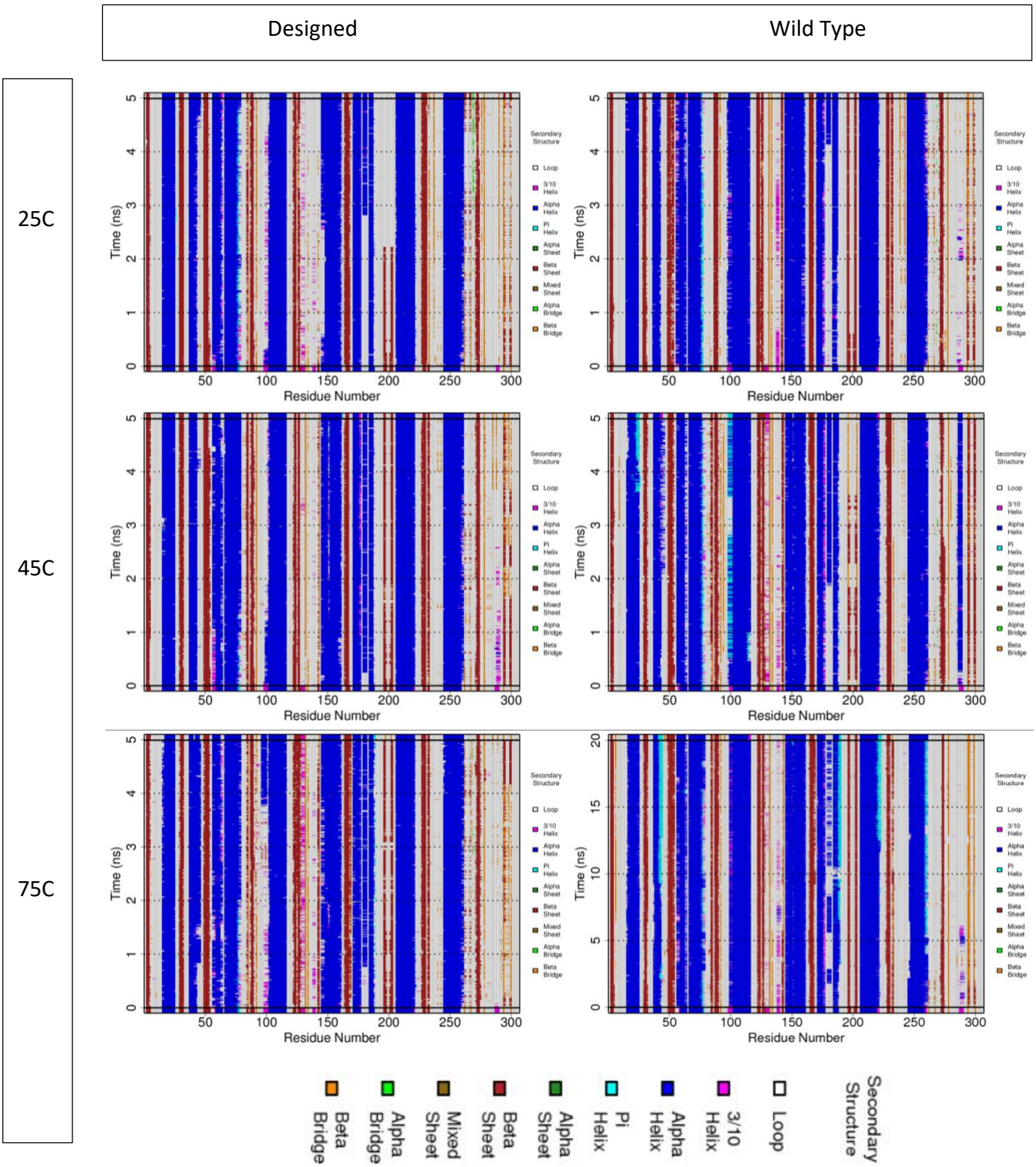

Designed
Wild Type
25C
45C
75C
Residue Number
Residue Number

Figure 6


Designed
Wild Type
25C
45C
75C
Time (ns)
Residue Number
Secondary Structure
Loop
3/10 Helix
Alpha Helix
Pi Helix
Alpha Sheet
Beta Sheet
Mixed Sheet
Alpha Bridge
Beta Bridge