Proving competing pathways in Protein-B for analysis in variable water mass simulations

Brett Velasquez, Margaret S. Cheung, Andrei G. Gasic, Fabio C. Zegarra

UNIVERSITY of HOUSTON — CTBP — NSF — RICE

Abstract

The immediate goal of the project is performing simulations of Protein-B trajectories at varying water mass to observe whether the varied water mass would affect appearance rate of various, competing, folding pathways. Progress unto this point was focused on proving competing pathways within the Protein-B trajectories acquired from the D.E. Shaw research group [1]. Fifty-five trajectories numbering 10,000 frames each were all sections of a single simulation of Protein-B and were the only sets of trajectories used in-so-far. The trajectories were used in conjunction with the VMD software [4] and Python scripting language and MDAnalysis [2, 3] Python libraries to produce all evidence seen here. Proof for competing pathways was provided by fraction of native contact (Q) calculations, non-native contact (Z) calculations and contact exclusivity heatmaps.

Background

Protein folding is the process by which a protein attempts to reach its 'native structure', the structure that allows the protein to achieve its biological function. The process of protein folding itself is still being investigated with varying topics of interest to be researched. Due to protein folding research being done via computer simulation large amounts of time can be between meaningful results. One such topic to be researched is if varying the mass of water within simulations will affect the appearance rate of protein folding pathways compared to one another.

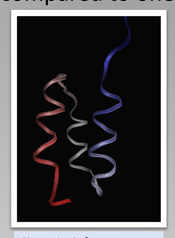


Figure 1 – Reference structure of Protein-B, fully folded

Method

After the possibility of competing folding pathways were substantiated by contact maps and visual analysis of Protein-B models in VMD we decided to calculate fraction of native contacts and non-native contacts to see if substantially different data existed to prove existence of competing pathways. Contacts were determined between each of the three helix structures of a folded Protein-B. Fraction of native contact and non-native contacts were determined by a distance cutoff between backbone atoms of each residue at a cutoff of 6 Angstroms.

Results

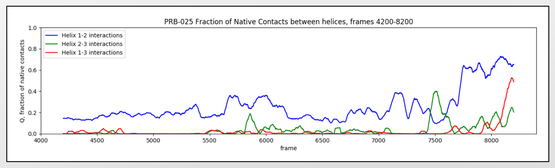


Figure 2a(above)- Fraction of native contact(Q) values vs. time(frames) for a folding event during trajectory 025

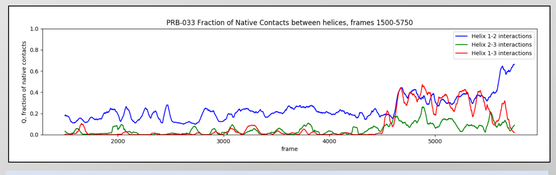


Figure 2b(left)- Fraction of native contact values vs. time during trajectory 033, characterized by Helix 1-2 and 1-3 development in tandem
Figure 2c(below)- Q vs. t of trajectory 037, characterized by development of Helix 1-2 and 2-3 interactions in tandem

The fraction of native contact(Q) vs time(frame) values gives information on the order in which certain Helixes of Protein-B developed during the folding event. Through the three events we see Helixes 1, 2 and 3 develop in varying orders across all trajectories.

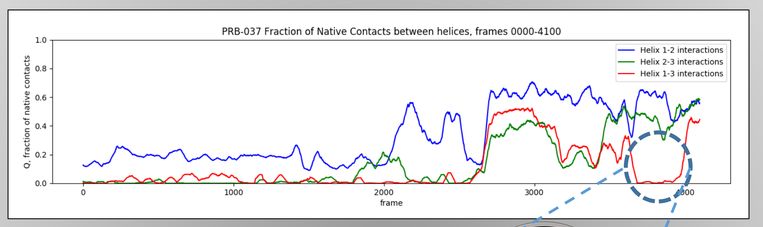


Figure 3a- During trajectory 037 (*Fig.2c*) we had a large depression in Helix 1-3 interactions before the protein finished folding, indicating the Helix 3 would be the last to fold. The model shows fully folded Helices, however, the helixes 1 and 3 are separate from one another by Helix 2.

Figure 3b- During trajectory 025 (*Fig.4a*) we had a peak in non-native interactions between Helix 1 and 3, as shown by the model with Helix 1 and 3 close to one another in a 'unique' fold over the course of the proteins development.

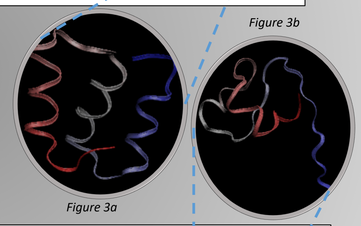
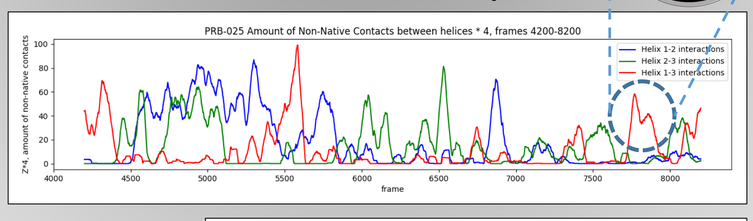
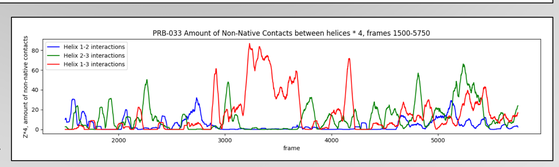


Figure 3b

Figure 3a



Contacts made that do not align with the reference structure are known as non-native contacts, also called Z-values. Z-values provide a means to find unique folding events and helix interactions that the reference structure does not account for as well as folding values to supplement those seen in the Q vs. t graphs



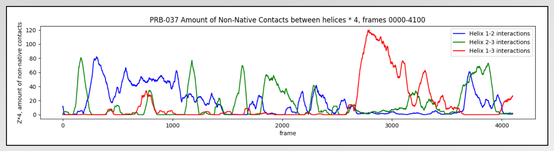


Figure 4a(top)- Trajectory 025, Helix 1-3 interaction peak to supplement rapid growth in Q vs. t graph for folding of Helix 3
Figure 4b(middle)- Trajectory 033, peaks for Helix 2-3 and 1-3 interactions to supplement low Q values from Q vs. t
Figure 4c(bottom)- Trajectory 037, peak for Helix 1-2 & 2-3 interactions alongside already high interactions in Q vs. t graph is valuable proof towards independent interactions verifying competing pathways

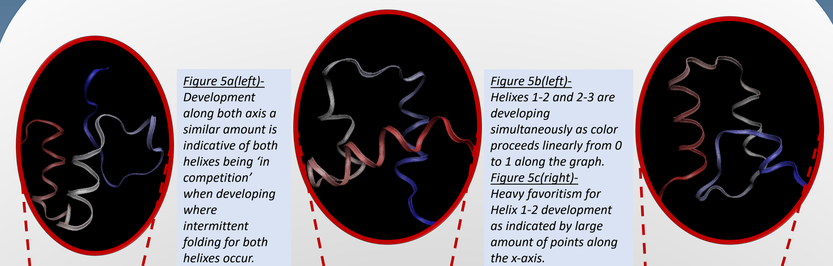
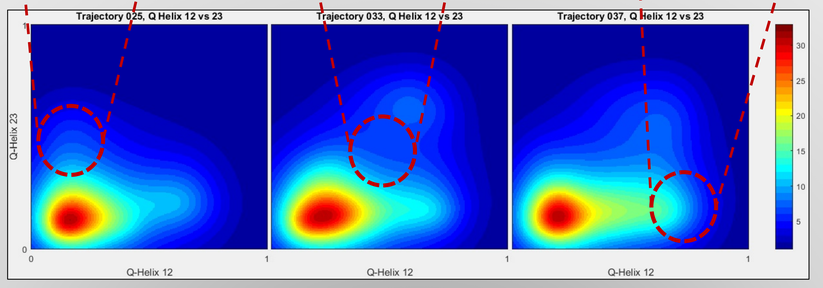


Figure 5a(left)- Development along both axis a similar amount is indicative of both helixes being 'in competition' when developing where intermittent folding for both helixes occur.

Figure 5b(left)- Helixes 1-2 and 2-3 are developing simultaneously as color proceeds linearly from 0 to 1 along the graph.
Figure 5c(right)- Heavy favoritism for Helix 1-2 development as indicated by large amount of points along the x-axis.



The heat maps above are indicative of every Q value graphed against one another between two helices to compare the development of both over the time frame of the trajectory. Color is based on clustering of points, wherein points where the protein are unfolded will be close to 0 and folded at 1. Growth along one axis of the heat map indicates the development of those Helix interactions takes priority/ folds first in a folding event, whereas a linear increase in Q along both axis implies development in tandem of values. Herein we see trajectory 033 favor simultaneous Helix interaction development, whereas 025 and 037 favor a single axis.

Conclusions

Fraction of native contact graphs per helix interaction provided visuals to compare the folding of helixes and thus their order between trajectories, where one could make assumptions about whether the protein is folding differently compared to other folding events. Z values for folding events are used for supplementing the fraction of native contact Q graphs for contacts that may not have been present in the reference structure, so high values of Z during periods of low values of Q could be indicative of alternate folding pathways. The exclusivity heatmaps provide a means to determine exclusivity of contact formation over the proteins folding trajectories, allowing us to draw the conclusion that the protein pathways vary enough to be competing.

Future Plans

A large portion of the projects development time went into creating and using the methods seen, so along the current slew of evidence toward competing pathways additional data will be scrounged via using unfolding events as another source of folding events. Compiling of all folding events and describing them using specific contact formation will as well be used for a metric when doing water mass variation simulations. After data compilation the group will move toward water mass variant MD simulations.

Acknowledgements

This project and the results were supported by resources of the FiS funding grant and uHPC funding grant:
FiS funding provided by grant NSF PHY-1427654
uHPC funding provided by grant NSF ACI-1531814

References

[1] Lindorff-Larsen K, Piana S, Dror RO, Shaw DE. How Fast-Folding Proteins Fold. Science. 2011;334:517–520.[2] N. Michaud-Agrawal, E. J. Denning, T. B. Woolf, and O. Beckstein. MDAnalysis: A Toolkit for the Analysis of Molecular Dynamics Simulations. *J. Comput. Chem.* **32**(2011), 2319-2327, doi:10.1002/jcc.21787. PMCID:PMC3144279
[3] R. J. Gowers, M. Linke, J. Barnoud, T. J. E. Reddy, M. N. Melo, S. L. Seyler, D. L. Dotson, J. Domanski, S. Buchoux, I. M. Kenney, and O. Beckstein. MDAnalysis: A Python package for the rapid analysis of molecular dynamics simulations. In S. Benthall and S. Rostrup, editors, *Proceedings of the 15th Python in Science Conference*, pages 98-105, Austin, TX, 2016. SciPy.
[4] Humphrey, W., Dalke, A. and Schulten, K., "VMD - Visual Molecular Dynamics", J. Molec. Graphics, 1996, vol. 14, pp. 33-38