Protein Folding: Predicting Structure from Sequence

Joe DeBartolo[1,3], Glen Hocky[2,3], Mike Wilde[4], Jinbo Xu[6], Karl F. Freed[2,3,4], and Tobin R. Sosnick[1,4,5]

[1]Department of Biochemistry and Molecular Biology, [2]Department of Chemistry, [3]The James Franck Institute, [4]Computation Institute, [5]Institute for Biophysical Dynamics, University of Chicago, [6]Toyota Technological Institute at Chicago

Naturally occurring proteins with similar sequences generally posses similar structures and often perform similar functions. Consequently, multiple sequence alignments are used to discern homologous structures for application in template-based modeling and in methods where structures are assembled using homologous fragments. While these homology-based models have some success, they require information from machine learning methods and thus provide an incomplete picture of the underlying physical principles. Our previous study avoids the use of mmachine learning and homology to develop a structure prediction method [1] based on an algorithm that mimics the folding pathway by coupling the descriptions of the formation of secondary and tertiary structure. Moves in the Monte Carlo procedure involve a change in a single pair of ϕ,ψ backbone dihedral angles that are obtained from a library that is constructed from high resolution protein structures in a manner appropriate for each amino acid, conditional on the type and conformation of the flanking residues. We now improve this method by utilizing sequence alignments to enrich the sampling library [2] in a manner that does not require structural knowledge of the homologous proteins. Effectively, this method equates to the simultaneous folding of an entire sequence alignment. In combination with other tools, including clustering and refinement, the accuracy of the predicted secondary and tertiary structures is substantially improved, and analysis of the trajectories enables the introduction of statistical measures for the accuracy of both the global and position-resolved predictions. Performance of the method in the Critical Assessment of Structure Prediction is discussed.

[1] Mimicking the Folding Pathway to Improve Homology-free Protein Structure Prediction. J. DeBartolo, A.Colubri, A. Jha, J. E. Fitzgerald, K. F. Freed, and T. R. Sosnick, Proc. Natl. Acad. Sci. **106**, 3734-9 (2009).

[2] Protein Structure Prediction Enhanced with Evolutionary Diversity: SPEED. J. DeBartolo, G. Hocky, M. Wilde, J. Xu, K. F. Freed, and T. R. Sosnick, Prot. Sci. **19**, 520 (2010).