Abstract

Structural characterization of the protein native state is often the key to better understanding protein function. The conformations that comprise the native state reside in the lowest-energy basin of the protein’s energy landscape. Discovering and populating this basin with native conformations in silico demands powerful search algorithms that can navigate high-dimensional conformational spaces. Even short protein chains have many degrees of freedom. Additionally, semi-empirical energy functions employed to guide the conformational search may introduce significant biases to the energy landscape, often leading to rough landscapes rich in local minima. A successful search algorithm must balance the competing goals of sampling a diverse representation of the conformational space, maintaining a representative bias in the fragment library.

We present a novel algorithm which effectively balances the goals of exploration and exploitation through the use of projection layers. A geometric layer keeps track of the geometric grid cells of the protein, maintaining a representative bias in the fragment library. A probabilistic search of a coarse-grained conformational landscape with the need to further populate promising energy minima.

The algorithm framework proposed in [1,2] takes a protein sequence σ as input and outputs an ensemble Qε, where the lowest-energy backbone conformations are sufficiently close to the native state that they can be further refined to recover this state in an all-atom detail. The framework explores the conformational landscape through a Rapidly Exploring Random Tree (RRT). Each iteration of the search is guided by the selection of an existing protein conformation (tree node) and the expansion of this node to sample a new conformation from the energy landscape.

Methodology

Algorithm Framework

Input: a protein sequence σ
Output: ensemble Qε of conformations

1. Extend coarse-grained fragment con g
2. Add all atoms to existing conformation
3. While Qε not exceed limits do
4. SelectEnergyLevel(Layerε);
5. cell = SelectCell(σ); cell = cell at Lε
6. Cε = SelectConf(cell.conf);
7. ExpandConform(Cε);
8. Add all at atoms
9. ε = ε + 1.

Selection

A conformation is selected with some probability P. First, an energy level ε is selected over a 1d energy grid. A weighting function w(ε) bias this selection towards lower-energy levels. Next, an implicit 3d grid is associated with each conformation and a geometric grid cell is assigned to each conformation. A second weighting function selects geometric conformations with fewer collisions. There are many degrees of freedom. Additionally, semi-empirical energy functions employed to guide the conformational search may introduce significant biases to the energy landscape, often leading to rough landscapes rich in local minima. A successful search algorithm must balance the competing goals of sampling a diverse representation of the conformational space, maintaining a representative bias in the fragment library.

We compile an extensive list of structurally diverse proteins on which we apply our algorithm. Our results show that the algorithm efficiently yields native-like conformations in conformational energy landscapes. Geometric layers keep track of the lowest-energy conformations and are able to guide the search in the conformational space by balancing coverage of conformational space with population of lower-energy levels.

Conclusions

We present an algorithmic framework for successfully navigating the high-dimensional protein conformational space. We employ energetic and geometric projection layers to create a low-dimensional mapping of this search space, allowing our algorithm to effectively balance exploration of diverse structures with exploitation of promising local minima.

The addition of a granularity reduction technique allows the algorithm to run for extended periods of time necessary for uncovering near-native conformations.

Our method employs a sample of the art fragment library which maximizes the structural diversity explored during the search. Future work will focus on enhanced secondary structure prediction methodologies to improve the sampling bias in the fragment library.

Acknowledgements

Granularity Reduction

Calculate intra-cell least Root Mean Square Deviation (lRMSD) only retain conformations with lRMSD > 1. (1 - 2 Å)

Fragment Library

Select position Select new fragment Replace fragment

All-Atom Refinement

The conformation obtained from an all-atom refinement of the lowest lRMSD conformation obtained by our method is colored red and superimposed over the known native structure in transparent blue. The refined backbone RMSD is given for each structure with the non-refined IRMSD in parentheses. The refinement is carried out with the Rosetta software package.

References

3. B. Olson, K. Molloy, and A. Shehu. Toward high-resolution de novo structure prediction methodologies to improve the sampling bias in the fragment library.

Mapping the Protein Conformational Landscape with Adaptive Probabilistic Search

Brian Olson1, Kevin Molloy1, Amarda Shehu1,2

1Department of Computer Science 2Department of Bioinformatics and Computational Biology

George Mason University, Fairfax, VA, 22030

Shehu amarda@gmu.edu

Results

Backbone Structure Prediction

Eleven medium length protein sequences are selected to test our structure prediction method.

The backbone structure with the minimum Root Mean Square Deviation (IRMSD) to the native structure predicted by our method (Shehu) is compared to results published by the Sosnick and Baker research groups. Our algorithm samples conformations closer to the native structure for four out of the eleven target proteins evaluated. The four cases in which our method performs poorly are explained by inferior secondary structure prediction, shown by the Q3 scores in columns 4-6 (The Q3 score measures the percentage of amino acids in correctly predicted secondary structures).

Fragment Library

Filter based on predicted secondary structure

Predicted Secondary Structure (SS)

Filter based on predicted secondary structure

Multiple sequence alignment

Pre-extracted Fragments

Pre-extracted Fragments

Pre-extracted Fragments

Pre-extracted Fragments

Pre-extracted Fragments

Pre-extracted Fragments