2016

Geometric algorithms for modeling protein structures

Dan Zeng

Follow this and additional works at: https://openscholarship.wustl.edu/undergrad_research

Part of the Other Computer Sciences Commons

Recommended Citation

Zeng, Dan, "Geometric algorithms for modeling protein structures" (2016). Undergraduate Research Symposium Posters. 84.
https://openscholarship.wustl.edu/undergrad_research/84

This Unrestricted is brought to you for free and open access by the Undergraduate Research at Washington University Open Scholarship. It has been accepted for inclusion in Undergraduate Research Symposium Posters by an authorized administrator of Washington University Open Scholarship. For more information, please contact digital@wumail.wustl.edu.
Geometric algorithms for modeling protein structures
Dan Zeng, Tao Ju, Ph. D.
Department of Computer Science and Engineering, Washington University in St. Louis

Introduction
Macromolecular assemblies drive nearly all cellular events. The assemblies' structures, which consist of tens to hundreds of interlocking proteins, are critical to understanding how biological systems work\(^1\). Our work addresses the computing challenges of generating structural models of macromolecular assemblies from Cryo-EM (cryo-electron microscopy) density maps.

Purpose
Develop geometric algorithms for modeling protein structures from Cryo-EM density maps of macromolecular assemblies by implementing geometric algorithms in Gorgon, a molecular modeling software.

Gorgon
Gorgon is an interactive molecular modeling system collaboratively developed between Washington University and Baylor College of Medicine. It is geared towards cryo-EM structures of macromolecular complexes.

Algorithms
Pathwalking
Pathwalking traces the primary backbone of a protein\(^2\):
1. Pseudo-atom generation: Use k-means clustering to generate vertices at locations that likely make up a protein’s backbone.
2. Pathwalking step: Fit a model through the pseudo-atoms such that every atom is connected to two others and deviation from the expected bond distance is minimized. Analogous to the Traveling Salesman Problem (Shortest path visiting each node).

Pathwalking in Gorgon
Our user interface for Pathwalking allows the user to set constraints and manipulate the predicted model both before and after Pathwalking:

Extremal Curve Skeletonization
This technique identifies locally maxima:
- Points with density values that are local maxima on a single axis are parts of max surfaces. These may be part of beta sheets.
- Maxima on two axes are part of max curves. These may be part of alpha helices.

Results
Pathwalking
Shown below are sample results for Pathwalking on a density map of rotavirus vp6. In the right figure, thin and blue ribbons correspond to no error, white and medium-thickness ribbons correspond to average error, and red and thick ribbons correspond to maximum error.

Extremal Curve Skeletonization
As evidenced by the significant presence of thin blue ribbons, Pathwalking produces topologically correct models of protein backbones on both simulated and authentic density maps without user intervention.

Evaluation
- At lower resolutions (> 5 Å), Pathwalking provides a clearer picture of the protein backbone.
- At higher resolutions (<4 Å), the shapes of secondary structures are more distinguishable in max curves and surfaces produced by skeletonization than in protein backbones generated with Pathwalking.

Next Steps
Our results suggest that resolution-aware algorithms would optimize both the speed and accuracy of the modeling process. We will create algorithms that will detect the resolution of the input density map, then reap the benefits of both Pathwalking and Extremal curve skeletonization. Skeletonization would first identify secondary structure structures. These structures will be used as anchors to find the rest of the protein structure using Pathwalking.

References

I would like to thank Dr. Tao Ju for his support and advice. This project was funded by National Science Foundation Grants IIS-1319573 and DBI-1356388.