Protein Secondary Structure Detection using Pattern Recognition and Geometric Modeling

Graduate Student: Tunazzina Islam
Associate Professor: Jing He
Department of Computer Science, Old Dominion University, Norfolk, VA, 23529, {tislam,jhe@cs.odu.edu}

Abstract

Electron cryo-microscopy (Cryo-EM) technique produces density maps that are 3-dimensional (3D) images of molecules. In order to derive the atomic structure of molecules, molecular features need to be identified from 3D images. Some molecular features of a protein show characteristic pattern in the image and others show a weak pattern or no pattern. We will describe an approach that uses a combination of pattern recognition and geometrical modeling to recognize protein secondary features including α-helices and β-strands. As the position of β-strands provides important constraints in backbone modeling of a protein, we will show the principle of modeling to distinguish the orientation of β-strands that are not visible in 3D images at medium resolution. In this paper, we will present the way to calculate maximum twist of β-strand traces which will help to take a decision to detect protein secondary structure when atomic structures are not available.

Introduction

Pattern recognition techniques have been successful in analysis of density maps obtained from cryo-electron microscopy (cryo-EM) technique [1, 2]. When the resolution of the density map is higher than 4 Å, the quality of the image is sufficient to distinguish the protein chain and hence the molecular structure can be derived. However, for density maps at medium resolution, such as 5-10Å, the quality of the 3D image is not sufficient to distinguish the backbone of a protein. Though detailed molecular features are not visible for medium-resolution images, rough features such as secondary structures of a protein: α-helices and β-sheets (Figure 1) can be computational identified. But it is almost impossible to detect the β-strands, the components of a β-sheet. The detection of β-strands from β-sheets in such images has been a challenging problem.

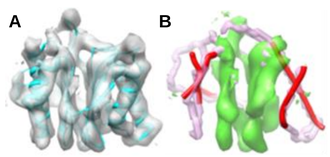

Figure 1. 3D image of cryo—electron microscopy density map, the atomic structure of a protein chain, and segmented Secondary structures based on characteristic density patterns. **A:** The density map (gray) extracted from EM Data Bank (EMDB) 1733 (6.8 Å resolution) superimposed on its atomic structure (cyan; PDB 4V68 chain BR). **B:** The helices (represented as red sticks) and two β-sheets (green density), using SSETracer [3] and the skeleton (pink) using SkelEM [4].

A helix identified from the medium resolution cryo-EM image is often represented as a line (red line in Figure 1A), referred as a α-trace that corresponds to the central axis of a helix. Location of major β-sheets can be identified from an image in low threshold value (Figure 2B). But it is not possible to identify β-strands because they don't have any fixed pattern. In low threshold, they may be visible (see Figure 2C) but in the high threshold, they are not (Figure 2D).

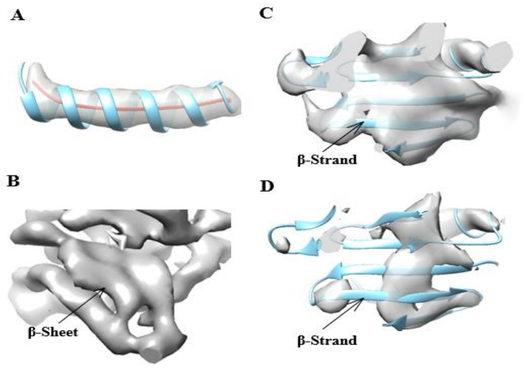

Figure 2. A: α-trace (red) generated from the 3D image (gray) with the atomic structure of the helix (blue ribbon). **B:** β-sheet. **C:** β-strands generated from the 3D image (gray) in low threshold with the atomic structure of the β-strands (blue ribbons). **D:** β-strands generated from the 3D image (gray) in high threshold with the atomic structure of the β-strands (blue ribbons).

Method

- **Patterns of helices and β-sheets and detection**

In a medium-resolution density map, a helix appears as a cylinder, and various methods exist to detect the location of these helices. We applied SSETracer [3] to detect the location of helices in a density map.

- **Modeling of β-strands**

β-sheet is composed of multiple β-strands those can be parallel, antiparallel or mix of both. We have measured the twist angle from 3D atom coordinates extracted from PDB file generated from image using StrandTwister [5], having different orientations including best case (Figure 3B) that is similar to true orientation and bad case (Figure 3B) that is far from the true orientation (Figure 3A)

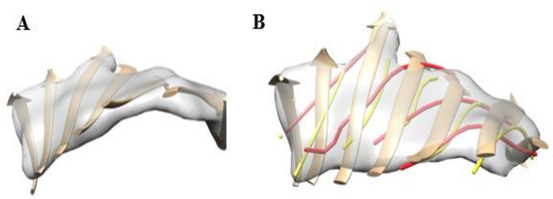

Figure 3. Twist of a β-sheet and two sets of candidates of β-strands in different orientations. Twist of the β-sheet is visualized from a side view of an extracted β-sheet region (golden gray) of the image, superimposed with the atomic structure (ribbon) of the β-sheet in (A). **A:** β-sheet (True Orientation), β-strands (True Orientation) for 1CHD_SH1 (PDB ID and Sheet ID), and **B:** two possible sets of β –strands (Yellow lines: good case, Red lines: bad case) align with the true position of β-strands (ribbon) at different orientations.

We calculated twist angles along lines that are interpolated from β-strands. We divided a line of β-trace into consecutive vectors of a certain length and calculated the angle formed by two vectors from adjacent traces respectively. Since twist angles change depending on where they are measured, we measured using a unit vector of adjustable length along two neighboring lines. For measuring twist angles between two neighboring β-strands we noticed the length of two neighboring β-strands and their overlapping characteristics. Here four cases can be happened (Figure 4).

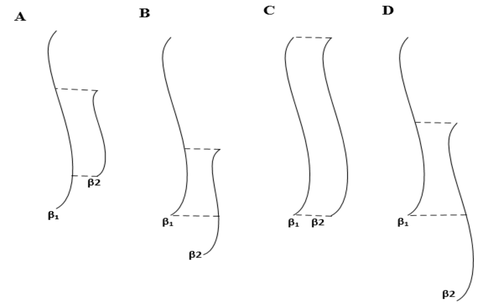

Figure 4. Various length of two neighboring β-strands traces (curved lines named β1, β2) and their overlapping characteristics (dotted lines are showing overlapping points). **A:** two neighboring β-strands having different length with full overlapping. **B:** two neighboring β-strands having different length with partial overlapping. **C:** two neighboring β-strands having equal length with full overlapping. **D:** two neighboring β-strands having equal length with partial overlapping.

Such collected twist angles are used to calculate the overall twist angle between two lines. Since the levels of twist are often different at different locations of the β-sheet, we used an average at two most stable spots to represent the twist. A twist angle is calculated as follows.

$$A_T = \frac{min_{k_1} \sum_{j_1=k_1}^{k_1+p} \frac{\theta_{j_1}^{i_1,}}{p} + min_{k_2} \sum_{j_2=k_2}^{k_2+q} \frac{\theta_{j_2}^{i_2,}}{q}}{2}$$

Let, β-strand traces be $\beta_1, \beta_2, \beta_3, \ldots\ldots, \beta_m$, and θ_j^i be the angle formed by β_i and β_{i+1} at location j of β_i.
i_1, and i_2 are two longest pairs of β-traces. For each β-sheet, two longest pairs of β-strand lines are used to represent the twist of the β-sheet.

Results

We used ten proteins, for which the atomic structures were downloaded as PDB file from The RCSB/PDB Database, and their corresponding 3D density maps were simulated at 10 A˚ resolution using Chimera-1.10.2. We generated different orientations of β-traces from StrandTwister [5]. For good orientation, we have larger twist (see 3rd column of Table 1) than bad orientation (see 4th column of Table 1) because good orientation matches more with the true atomic structure (see Figure 3B).

Table 1. The maximum twist for the set of β-strand traces

PDB ID_SHEET ID	No. of β-Strands in a β-Sheet	Max twist (good case)	Max twist (bad case)
1B3A_B	3	5.4545	1.6145
1A12_A	4	9.38	3.5405
1AOP_A	5	2.6175	0.921
1AKY_A	5	6.3305	0.9875
1ATZ_A	6	12.842	5.387
1CHD_SH1	7	2.908	1.291
1ELU_B	7	7.642	0.416
1D5T_A	7	10.009	4.466
1DTD_A	8	10.3985	5.8475
1QNA_C	9	15.7895	5.191

Conclusion

The position of β-strands is critical for modeling the atomic structure of the entire protein. However, it has been a challenging problem when no separation of the β-strands is visible in the images. We proposed a novel approach using image processing and modeling to generate a small set of possible β-strand traces from the images and find out the maximum twist. This maximum twist represents the most stable region of β-strand traces as well as the twist of the β-sheet.
For the society in general, this research is a step forward in detection of protein secondary structure accurately which is an essential pre-requisite for structure-based drug design.

Acknowledgements & References

I would like to thank Dr. Mohammad Zubair and Dr. Desh Ranjan for allowing me to do this work.
This work was partially supported by NSF DBI-1356621, NIH R01-GM062968 and MSVE department of Old Dominion University.

[1] W. Chiu, M. L. Baker, W. Jiang, M. Dougherty, and M. F. Schmid, "Electron cryomicroscopy of biological machines at subnanometer resolution," Structure, vol. 13, pp. 363-72, Mar 2005.
[2] Z. H. Zhou, "Atomic resolution cryo electron microscopy of macromolecular complexes," Adv Protein Chem Struct Biol, vol. 82, pp. 1-35, 2011.
[3] K. Al Nasr, C. Liu, M. Rwebangira, L. Burge, and J. He, "Intensity-based skeletonization of CryoEM gray-scale images using a true segmentation-free algorithm," IEEE/ACM Trans Comput Biol Bioinform, vol. 10, pp. 1289-98, Sep-Oct 2013.
[4] M. L. Baker, M. R. Baker, C. F. Hryc, T. Ju, and W. Chiu, "Gorgon and pathwalking: macromolecular modeling tools for subnanometer resolution density maps," Biopolymers, vol. 97, pp. 655-68, Sep 2012.
[5] D. Si, and J. He, "Tracing beta-strands using strandtwister from cryo-EM density maps at medium resolutions," *Structure*, vol. 22, no. 11, pp. 22(11):1665-76, 2014.