Protein Secondary Structure Detection using Pattern Recognition and Modeling

Tunazzina Islam
Ph.D. Student
Department of Computer Science
Old Dominion University
+17572648410
tislam@cs.odu.edu

Jing He
Associate Professor
Department of Computer Science
Old Dominion University
+17576837716
jhe@cs.odu.edu

1. INTRODUCTION

Pattern recognition techniques have been successful in analysis of density maps obtained from cryo-electron microscopy (cryo-EM) technique. Cryo-EM is a biophysical technique to determine 3-dimensional structures of molecules [1, 2]. This technique is particularly suitable for large molecular assemblies that are often challenging for traditional techniques such as X-ray Crystallography and Nuclear Magnetic Resonance (NMR). A density map of molecules is a 3-dimensional image. 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. It is challenging to derive atomic structure from such images.

Though detailed molecular features are not visible for medium-resolution images, rough features such as secondary structures of a protein is visible. Secondary structure of a protein such as α-helices and β-sheets (Figure 1) can be computational identified. An α-helix often appears as a cylinder (represented as a red stick in Figure 1A) and can be identified using image processing methods [3-6]. A β-sheet may appear as a thin layer of density (green region in Figure 1B) and can be identified computationally [5-8].

Although β-sheets can be identified from cryo-EM density images at 5-10Å, it is almost impossible to detect the β-strands, the components of a β-sheet. The spacing between two neighboring β-strands is between 4.5 and 5Å, and therefore they are not visible when the resolution is at 5-10 Å. The detection of β-strands from β-sheets in such images has been a challenging problem since it was first attempted in 2004 [11].

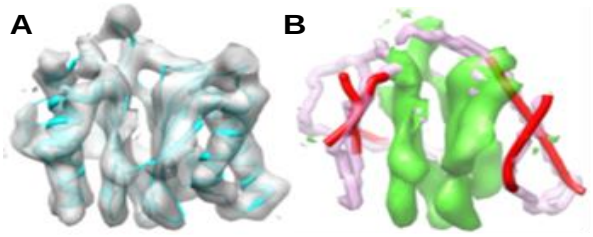

Figure 1. 3D image of cryo—electron microscopy density map, the atomic structure of a protein chain, and segmented Secondary structures based on characteristic density patterns. A: The density map (gray) extracted from EM Data Bank (EMDB) 1733 (6.8 Å resolution) superimposed on its atomic structure (cyan; PDB 4V68 chain BR). B: The helices (represented as red sticks) and two β-sheets (green density), using *SSETracer* [10] and the skeleton (pink) using *SkelEM* [11].

A helix identified from the medium resolution cryo-EM image is often represented as a line (red line in Figure 1A), referred as an α-trace that corresponds to the central axis of a helix. Location of major β-sheets can be identified from 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 (Figure 2C) but in high threshold they are not (Figure 2D).

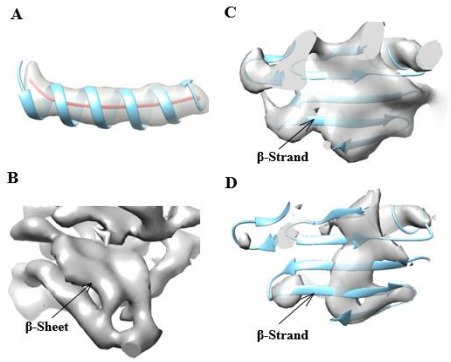

Figure 2. A: α-trace (red) generated from the 3D image (gray) with the atomic structure of the helix (blue ribbon). B: β-sheet. C: β-strands generated from the 3D image (gray) in low threshold with the atomic structure of the β-strands (blue ribbons). D: β-strands generated from the 3D image (gray) in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

2. METHOD

2.1 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 [10] to detect the location of helices in a density map. SSETracer detects helices (and β-sheets) based on a characterization of local density features such as local structure tensor, local thickness, continuity of the skeleton, and density value. A detected helix is represented by a set of points located along the central axis of the helix. The current implementation of SSETracer contains a modified step in the axis extension to enhance the geometric characterization of the helix

2.2 Modeling of β-strands

A β-sheet is composed of multiple β-strands those can be parallel, antiparallel or mix of both. The sheet twist is defined as the angle between the backbone vectors of the two residues in the pair [12]. We have measured the twist angle from 3D atom coordinates extracted from PDB file generated from image having different

orientations including best case (Figure 3B) that is similar to true orientation and bad case (Figure 3B) that is far from true orientation.

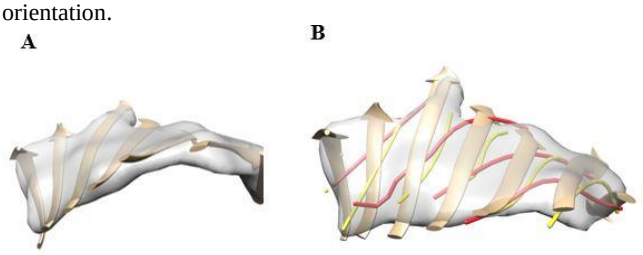

Figure 3. A: β-sheet, β-strands, and B: two sets of β –traces (Yellow: good case, Red: bad case)

Let β-strand traces be $\beta_1, \beta_2, \ldots, \beta_m$, and let θ_j^ibe the angle(Figure 4) formed by β_i and β_{i+1} at location j of β_i. As an example, the angles formed by β_1 and $\beta_2$along the strands are$(\theta_1^1, \theta_2^1, \ldots, \theta_{n_1}^1)$. We divide a line of β-trace into consecutive vectors of certain length, and calculated the angle formed by two vectors from adjacent traces respectively. Since the level 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.

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

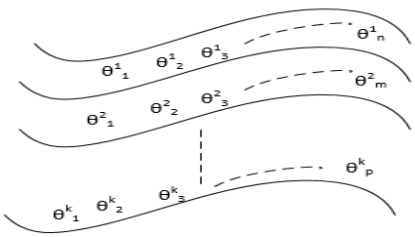

Figure 4. Twist angles between two consecutive β-strands

3. RESULT

We used six proteins, for which the atomic structures were downloaded from the PDB, and their corresponding 3D density maps were simulated at 10 A˚ resolution using Chimera. For good orientation we have larger minimum twist (3rd column of Table 1) than bad orientation (4th column of Table 1) because good orientation matches more with the atomic structure (Figure 3B).

Table 1. The minimum twist angle for the set of β-traces

PDB ID_SHEET ID	No. of β-Strands in a β-Sheet	Minimum twist (good case)	Minimum twist (bad case)
1A12_A	4	13.526	8.954
1AKY_A	5	15.285	13.068
1ATZ_A	6	13.071	9.674
1CHD_SH1	7	10.93	5.736
1DTD_A	8	12.524	6.664
1QNA_C	9	13.415	3.501

4. 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 propose 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 minimum twist.

5. REFERENCES

[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] W. Jiang, M. L. Baker, S. J. Ludtke, and W. Chiu, "Bridging the information gap: computational tools for intermediate resolution structure interpretation," *J Mol Biol*, vol. 308, pp. 1033-44, May 2001.

[4] A. Del Palu, J. He, E. Pontelli, and Y. Lu, "Identification of Alpha-Helices from Low Resolution Protein Density Maps," presented at the Proceeding of Computational Systems Bioinformatics Conference(CSB), 2006.

[5] M. L. Baker, T. Ju, and W. Chiu, "Identification of secondary structure elements in intermediate-resolution density maps," *Structure*, vol. 15, pp. 7-19, Jan 2007.

[6] D. Si, S. Ji, K. A. Nasr, and J. He, "A machine learning approach for the identification of protein secondary structure elements from electron cryo-microscopy density maps," *Biopolymers*, vol. 97, pp. 698-708, Sep 2012.

[7] Y. Kong and J. Ma, "A structural-informatics approach for mining beta-sheets: locating sheets in intermediate-resolution density maps," *J Mol Biol*, vol. 332, pp. 399-413, Sep 12 2003.

[8] S. Dong and H. Jing, "Beta-sheet Detection and Representation from Medium Resolution Cryo-EM Density Maps," in *Proceedings of the International Conference on Bioinformatics, Computational Biology and Biomedical Informatics %@ 978-1-4503-2434-2*, ed. Wshington DC, USA: ACM, 2013, pp. 764-770.

[9] D. Si and J. He, "Beta-sheet Detection and Representation from Medium Resolution Cryo-EM Density Maps.," *BCB'13: Proceedings of ACM Conference on Bioinformatics, Computational Biology and Biomedical Informatics*, vol. Washington, D.C., pp. 764-70, 2013.

[10] 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.

[11] 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.

[12] B. K. Ho and P. M. G. Curmi, "Twist and Shear in β-Sheets and β-Ribbons," DOI= 10.1006/jmbi.2001.5385