Combining Image Processing and Modeling to Generate Traces of Beta-strands from Cryo-EM Density Images of Beta-barrels

Dong Si and Jing He

Abstract — **Electron cryo-microscopy (Cryo-EM) technique produces 3-dimensional (3D) density images of proteins. When resolution of the images is not high enough to resolve the molecular details, it is challenging for image processing methods to enhance the molecular features. β-barrel is a particular structure feature that is formed by multiple β-strands in a barrel shape. There is no existing method to derive β-strands from the 3D image of a β-barrel at medium resolutions. We propose a new method, *StrandRoller*, to generate a small set of possible β-traces from the density images at medium resolutions of 5-10Å. *StrandRoller* has been tested using eleven β-barrel images simulated to 10Å resolution and one image isolated from the experimentally derived cryo-EM density image at 6.7Å resolution. *StrandRoller* was able to detect 81.84% of the β-strands with an overall 1.5Å 2-way distance between the detected and the observed β-traces, if the best of fifteen detections is considered. Our results suggest that it is possible to derive a small set of possible β-traces from the β-barrel cryo-EM image at medium resolutions even when no separation of the β-strands is visible in the images.**

I. Introduction

Electron cryo-microscopy (Cryo-EM) has become a major experimental technique to study the structures of large protein complexes, such as ribosomes and viruses [1, 2]. It is a structure determination technique complementary to the X-ray Crystallography and Nuclear Magnetic Resonance (NMR). At medium resolutions such as 5-10Å, detailed molecular features are not resolved. However, secondary structures such as α-helices and β-sheets (Figure 1) can be computational identified. An α-helix appears as a stick (red in Figure 1A) and can be identified using image processing methods [3-6]. A β-sheet appears as a thin layer of density and can be identified computationally (blue in Figure 1B) [5-8]. Some β-sheets curve into β-barrels, which are commonly found in porins and other proteins that span cell membranes [9]. A β-barrel is composed of multiple β-strands (ribbon in Figure 1C and Figure 2) that twist and coil to form a "closed" structure in which the first strand is hydrogen bonded to the last.

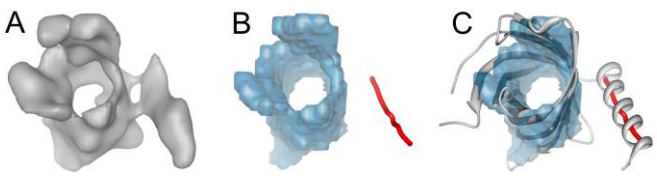

Figure 1. 3D density image, the secondary structures and a β-barrel. (A) Protein denisty image (gray) was simulated using EMAN [10] and the atomic structure of protein 3GP6 from the Protein Data Bank. (B) The location of a helix (red line) and a β-barrel region (blue voxels) was detected using *SSETracer* [8]. (C) The atomic structure of protein 3GP6 (ribbon) is overlaped with the detected position of a helix and a β-barrel.

The research is partially supported by the MSF and M&S scholarship of Old Dominion University and NSF-1356621.

Dong Si and Jing He are with the Computer Science Department, Old Dominion University, Norfolk, VA, USA, 23529. (Correspondence to: Jing He, jhe@cs.odu.edu).

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 [11, 12]. The detection of β-strands from β-sheets in such images has been a challenging problem since it was first attempted in 2004 [13]. The detection of β-strands from β-barrels is perhaps even more challenging. β-barrel is one of many types of β-structures, on which the β-strands appear to roll over to form a barrel. There is no existing method to detect β-strands from the β-barrel cryo-EM images at medium resolutions. In this paper, we propose a new approach that combines image processing and modeling to address this problem. Although exact β-strands are impossible to be detected directly from such images, we show that it is possible to generate a small set of possible β-strand traces. Our novel method, *StrandRoller*, is designed based on the intrinsic nature and feature of β-barrel structures.

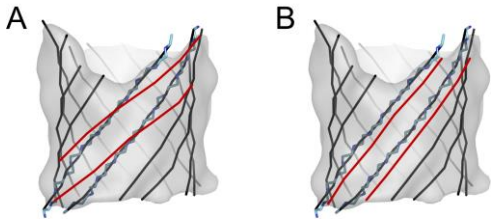

Figure 2. β-strands of a β-barrel image. A set of β-traces is shown in black lines at the front and gray lines at the back in (A) and (B). Two possible sets of β-traces (represented by two black lines and two red lines) may have different orientations shown in (A) or locations/shifts shown in (B). The atomic structure of the two β-strands is superimposed on the two representative β-traces in (A) and (B).

A helix identified from the medium resolution cryo-EM image is often represented as a line (red line in Fig 1B and C), referred as an α-trace that corresponds to the central axis of a helix. We define the β-trace (black line in Figure 2) as the central line along a β-strand. In particular, the observed β-trace is the line interpolating all geometrical centers of three consecutive Cα atoms on a β-strand plus two Cα atoms at the end of the β-strand. An observed β-trace represents the line along the atomic structure of a β-strand (Figure 2). Given the β-barrel image voxels, the problem of β-strands detection is to find the orientation (Figure 2A) and location (Figure 2B) of β-traces from the three-dimensional cryo-EM image.

978-1-4244-7929-0/14/$26.00 ©2014 IEEE

II. METHOD

A. β-barrel Surface Modeling from Cryo-EM Image

β-barrels have characteristic shapes and have been modeled mathematically in previous studies. The atomic structure of a β-barrel has been modeled as hyperboloid surfaces [14-16] and catenoid surfaces [17]. All these methods concentrated on the fitting of a particular mathematical model to the β-barrel structures by using linear or non-linear fitting procedure. Although these models can approximate the major area of a β-barrel, cryo-EM images of β-barrels often deviate from the rigid mathematical models in certain regions. We present an adaptive method to generate a surface that fit in the three-dimensional image of a β-barrel. The idea is to use a rigid model at the regions it fit well and to adjust the model at those regions it does not fit.

Given the density voxels of a β-barrel, least-square fitting of an elliptical cylinder (1) was performed to find the central axis, the Z-axis, of the barrel (Figure 3A).

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1 \quad (1)$$

The surface model of the β-barrel was then built from bottom cross-sections to those at the top. The density voxels on each cross-section of Z axis (Figure 3B, gray) appear to be nearby the ideal model of ellipse. The voxels that are closest to the ideal ellipse was selected as the points in surface model (Figure 3B, yellow). Note that such a discrete model closely represents the 3D image. For example, when the ideal elliptical cylinder is outside the density (arrow in Figure 3B), the voxels on the image was used to adjust the surface model. It appears that the resulting surface model clearly follows to the morphed regions (arrows in Figure 3D), and we find that it is important to have an accurate barrel surface to model the β-traces accurately.

B. Strand Walking on the Surface Model

It was first noticed by MacLachlan in 1979 that the number of strands and their relative stagger completely determines the overall structure of a β-barrel [18]. The main structural characteristics of ideal β-barrel have been discussed based on a cylindrical barrel [18-20]. In short, β-barrel forms a closed cylindrical barrel. In all known β-barrel structures, the β-strands are right-twisted, and in order to satisfy hydrogen bonding, each β-strand is right-tilted with respect to the barrel axis. Studies have shown that tilt angle α (Figure 3E) of a β-strand can vary between 30° and 60°, as reflected in the known structures of membrane proteins [20-22]. Note the α may vary by $\pm 15°$ for different strands in the same β-barrel [22]. However, the inter-strand distance d remains the same due to the hydrogen bonding pattern of β-strands. In general, the inter-strand distance d is roughly between 4.5 and 5Å. *StrandRoller* is based on previous knowledge about the tilt angle α and the inter-strand distance d.

An initial β-trace (blue in Figure 3E) was produced by tilting the barrel axis with α and projecting it onto the surface model. The second β-trace was then generated by traveling a horizontal distance h (2) on the surface (Figure 3E).

$$h = d/\cos\alpha, d = 4.8\,\text{Å} \quad (2)$$

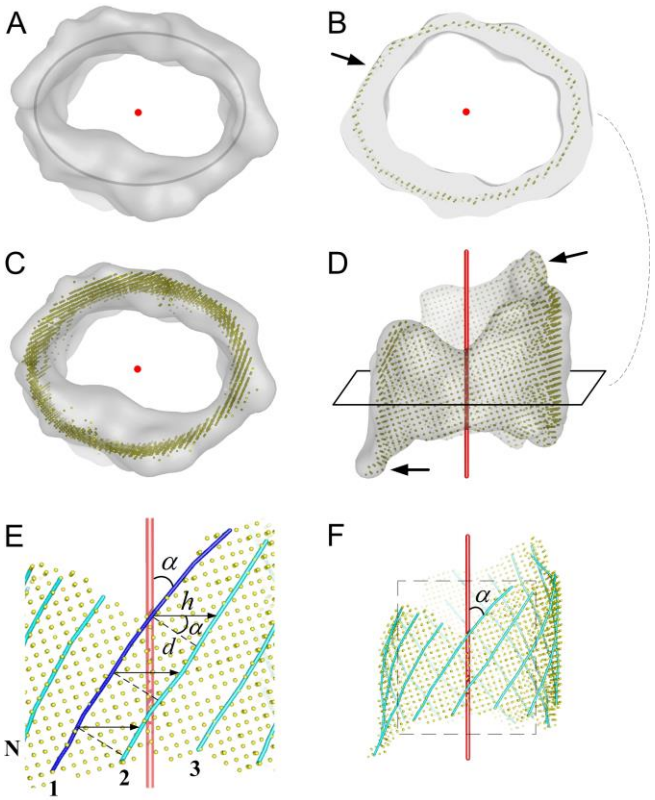

Figure 3. Modeling barrel surface and building β-traces from 3D image of a β-barrel. (A) The barrel axis (red) was identified by fitting an elliptical cylinder (line) to the density image (gray), as shown in the top view. (B) A cross-section of the barrel shows adjustment (arrow) of the ideal surface model at the morphed region. Top view (C) and the side view (D) of the modeled barrel surface (yellow) show the barrel axis (red) and a cross-section of the β-barrel. The zoom-in view (E) and the entire-view (F) of the β-traces generated recursively using tilt angle α and inter-strand distance d of β-strands.

Given a tilt angle, the entire set of β-traces was built iteratively on the barrel surface (Figure 3F) until the last β-trace (Figure 3E). The tilt angle was sampled every 5° between 35° to 55°, and three translations were sampled at each tilt angle. There are fifteen sets of β-traces generated for one β-barrel.

III. RESULT

StrandRoller was tested on eleven β-barrel images that are simulated to 10Å resolution and one experimentally derived cryo-EM image from EMDB (http://www.emdatabank.org/) at 6.7Å resolution. The accuracy of detection was calculated using 2-way distance between the set of detected β-traces and the set of observed β-traces. The observed β-trace is the line interpolating all geometrical centers of three consecutive Cα atoms on a β-strand plus the two Cα atoms at the end of the β-strand, as shown in Figure 2. In order to calculate the 2-way distance, one-to-one correspondence between the β-traces in the detected set and those in the observed set was first determined based on the overall smallest distance. This ensures that the same number of detected β-traces $(S_1, S_2, \ldots, S_T)$ are compared to the same number of observed traces $(S'_1, S'_2, \ldots, S'_T)$ in which S_k is compared with S'_k for $k = 1, \ldots, T$. The number of miss-detected (or wrongly detected) β-strands can be inferred from the difference

between the total number of the observed and that of the detected β-strands. D_k, the 2-way distance of a β-strand k, was calculated for each pair of lines S_k and S'_k. The overall 2-way distance D reflects the quality of the detected β-traces that are corresponded to their observed ones.

$$D_k = \left(\sum_{i=1}^{N} D_i^{ss'}/N + \sum_{j=1}^{M} D_j^{s's}/M\right)/2 \quad (3)$$

$$D = \left(\sum_{s=1}^{T} D_k\right)/T \quad (4)$$

In formula (3), N and M are the numbers of points on the detected and the observed β-traces S_k and S'_k respectively. i and j are the indexes of the point along line S_k and S'_k respectively. $D_i^{ss'}$ is the projected distance from point i of line S_k to line S'_k. We require the projection of point i to be within line S'_k. In case it is outside, the distance between point i and the end of line S'_k was used to approximate the error.

A. Performance on the Simulated Images

The purpose of this test is to investigate if the traces of β-strands can be modeled from β-barrel images simulated to 10Å resolution, at which the separation of β-strands is not visible. The proteins were collected from the β-barrel transmembrane super family of Orientations of Proteins in Membranes (OPM) database [23] with less than 40% sequence similarity. The atomic structures of β-barrels were used to generate β-barrel density images at 10Å resolution using EMAN [10], a popular software to produce simulated density, with a sampling of 1Å/pixel.

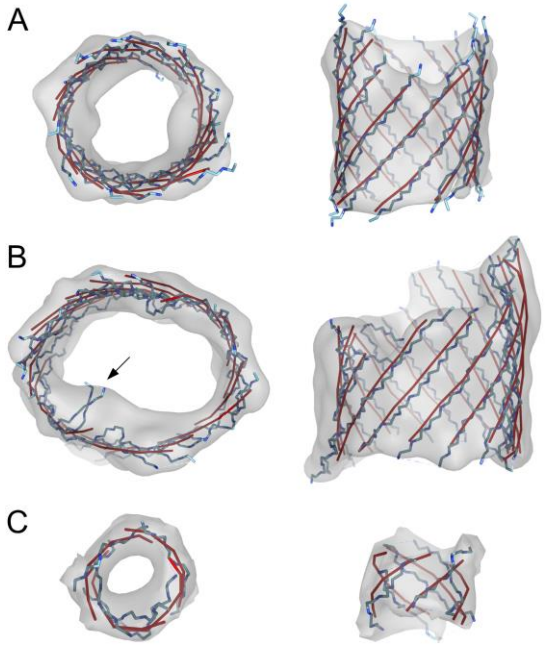

Figure 4. β-strands modeling from the simulated density image at 10Å and an experimentally derived cryo-EM image. The best of the fifteen sets of modeled β-traces (red) are superimposed with the back-bone of the β-strands (blue) and the density images (gray) for β-barrels 4FQE_A in (A) and 2J1N_AA18 in (B). The top view (left) and the side view (right) are shown in each case. An experimental derived cryo-EM image of β-barrel (EMD_5036, sheet H of protein 3FIK) is shown in (C).

To discuss the ability of our β-trace detection, we use the best of the fifteen sampled sets in the Result section. The best set is the one that is closest to the observed set in terms of 2-way distance. Figure 4 shows the best of the fifteen modeled β-traces (red) for three cases with 12, 16 and 5 β-strands respectively. In the case of sheet A of PDB structure 4FQE, the detected set of β-traces appears to align with the β-strands very well (Figure 4A). In this case, all the twelve strands were detected with a small 2-way distance of 1.25Å (Table I). The test shows that *StrandRoller* is able to model various sizes of β-barrels ranging from 5 strands to 16 strands. Although the detection is slightly better for the smaller β-barrels, some large β-barrels, such as the 16-stranded β-barrel 2J1N_AA18, were still well detected, (Figure 4B). The 2-way distance is only 1.67Å in this case (Table I). The error appears to be at the edge of the β-barrels (arrows in Figure 4B), where the β-strands tend to be more flexible.

TABLE I. Accuracy of β-barrel modeling from simulated density images at 10Å resolution.

PDB ID[a]	#Det./#Obs. Strd[b]	2-w Dist.[c]	#Det./#Obs. AA[d]
1G7K_A13	11 / 11	1.53	103 / 124
1QJP_A	8 / 8	1.81	74 / 107
1RRX_A12	11 / 11	1.71	95 / 118
1TX2_B	7 / 8	1.30	30 / 34
2ERV_A10	8 / 8	1.11	79 / 94
2J1N_AA18	16 / 16	1.67	148 / 181
2QDZ_C17	15 / 16	1.50	165 / 198
2QOM_C	12 / 12	1.71	153 / 190
2WJR_AA15	11 / 12	1.59	103 / 130
3FID_A14	12 / 12	1.36	127 / 155
4FQE_A	12 / 12	1.25	122 / 134
Average		**1.50**	**1199/1465 = 81.84%**
Standard deviation		**0.22**	

a. PDB_Sheet ID;
b. The number of detected β-traces in the best of the fifteen modeled sets / the number of observed β-strands in the true β-barrel;
c. The 2-way distance (in Å) between the observed β-traces and the modeled β-traces for the best of the fifteen modeled sets;
d. The number of detected / total number of amino acids in the β-barrel.

The test of 11 simulated β-barrel density images shows that one of the fifteen sets of β-traces aligns very well with the observed set of β-traces, with an overall 2-way distance of 1.5Å for the detected β-traces (Table I). To analyze the sensitivity of the detection, we measured the number of amino acids that were missed in the detection. An amino acid was considered detected if its Cα atom is within 2.5Å from the detected β-trace that corresponds to the strand where the Cα resides. For example, 1TX2_B has seven of the eight β-strands detected (Table I row 5). It missed four amino acids. For the seven detected strands, the 2-way distance is 1.3Å. Among the 11 test cases, *StrandRoller* appears to be able to detect 81.84% of the β-strands fairly accurately in one of the fifteen sampled sets of β-traces (Table I).

Figure 4C shows the density region of a β-barrel (3FIK_H) from the experimentally derived cryo-EM image at 6.7Å resolution (EMD_5036). At this resolution, single β-strands are not visible. *StrandRoller* was able to detect all five strands, and they align fairly well with the observed β-traces. In this case, the 2-way distance for the five β-strands is only 1.6Å, and it detected 27 of 29 amino acids on the β-barrel.

IV. 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. There is no existing method to detect the location of β-strands from β-barrel images at the medium resolutions. We propose a novel approach using image processing and modeling to generate a small set of possible β-strand traces from the β-barrel images. Our preliminary results show that it is possible to derive such a small sets. Each set of detected β-traces can be further evaluated when more atomic details are added in the protein modeling.

StrandRoller does not require the resolution of cryo-EM image to be high enough (<5Å) to resolve the separation of β-strands in a β-barrel image. It applies to the images with lower resolutions. In the test containing eleven β-barrel images, *StrandRoller* detected about 81.84% of the amino acids in the β-strands with an overall 1.5Å 2-way distance between the detected β-traces and the observed ones, if the best of the fifteen detections is considered. Our results suggest that β-strands can be generated from the medium resolution cryo-EM images of β-barrel proteins. To our knowledge, this is the first method to address the problem of β-strands detection from medium resolution β-barrel images.

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] V. Cherezov, W. Liu, J. P. Derrick, B. Luan, A. Aksimentiev, V. Katritch, and M. Caffrey, "In meso crystal structure and docking simulations suggest an alternative proteoglycan binding site in the OpcA outer membrane adhesin," *Proteins,* vol. 71, pp. 24-34, Apr 2008.

[10] S. J. Ludtke, P. R. Baldwin, and W. Chiu, "EMAN: semiautomated software for high-resolution single-particle reconstructions," *J Struct Biol,* vol. 128, pp. 82-97, Dec 1 1999.

[11] Z. H. Zhou, "Towards atomic resolution structural determination by single-particle cryo-electron microscopy," *Current Opinion in Structural Biology,* vol. 18, pp. 218-228, Apr 2008.

[12] 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.

[13] Y. Kong, X. Zhang, T. S. Baker, and J. Ma, "A Structural-informatics approach for tracing beta-sheets: building pseudo-C(alpha) traces for beta-strands in intermediate-resolution density maps," *J Mol Biol,* vol. 339, pp. 117-30, May 21 2004.

[14] J. Novotny, R. E. Bruccoleri, and J. Newell, "Twisted hyperboloid (Strophoid) as a model of beta-barrels in proteins," *J Mol Biol,* vol. 177, pp. 567-73, Aug 15 1984.

[15] E. Tolonen, B. Bueno, S. Kulshreshta, P. Cieplak, M. Argaez, L. Velazquez, and B. Stec, "Allosteric transition and binding of small molecule effectors causes curvature change in central beta-sheets of selected enzymes," *J Mol Model,* vol. 17, pp. 899-911, Apr 2011.

[16] I. Lasters, S. J. Wodak, P. Alard, and E. van Cutsem, "Structural principles of parallel beta-barrels in proteins," *Proc Natl Acad Sci U S A,* vol. 85, pp. 3338-42, May 1988.

[17] E. Koh and T. Kim, "Minimal surface as a model of beta-sheets," *Proteins,* vol. 61, pp. 559-69, Nov 15 2005.

[18] A. D. McLachlan, "Gene duplications in the structural evolution of chymotrypsin," *J Mol Biol,* vol. 128, pp. 49-79, Feb 15 1979.

[19] A. G. Murzin, A. M. Lesk, and C. Chothia, "Principles determining the structure of beta-sheet barrels in proteins. I. A theoretical analysis," *J Mol Biol,* vol. 236, pp. 1369-81, Mar 11 1994.

[20] A. G. Murzin, A. M. Lesk, and C. Chothia, "Principles determining the structure of beta-sheet barrels in proteins. II. The observed structures," *J Mol Biol,* vol. 236, pp. 1382-400, Mar 11 1994.

[21] T. Pali and D. Marsh, "Tilt, twist, and coiling in beta-barrel membrane proteins: relation to infrared dichroism," *Biophys J,* vol. 80, pp. 2789-97, Jun 2001.

[22] G. E. Schulz, "The structure of bacterial outer membrane proteins," *Biochimica Et Biophysica Acta-Biomembranes,* vol. 1565, pp. 308-317, Oct 11 2002.

[23] M. A. Lomize, A. L. Lomize, I. D. Pogozheva, and H. I. Mosberg, "OPM: orientations of proteins in membranes database," *Bioinformatics,* vol. 22, pp. 623-5, Mar 1 2006.