

## Research Article

# Modeling Beta-Traces for Beta-Barrels from Cryo-EM Density Maps

Dong Si<sup>1</sup> and Jing He<sup>2</sup>

<sup>1</sup>Division of Computing and Software Systems, University of Washington Bothell, Bothell, WA 98011, USA

<sup>2</sup>Department of Computer Science, Old Dominion University, Norfolk, VA 23529, USA

Correspondence should be addressed to Jing He; [jhe@cs.odu.edu](mailto:jhe@cs.odu.edu)

Received 15 September 2016; Accepted 8 December 2016; Published 10 January 2017

Academic Editor: Slavica Jonic

Copyright © 2017 D. Si and J. He. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Cryo-electron microscopy (cryo-EM) has produced density maps of various resolutions. Although  $\alpha$ -helices can be detected from density maps at 5–8 Å resolutions,  $\beta$ -strands are challenging to detect at such density maps due to close-spacing of  $\beta$ -strands. The variety of shapes of  $\beta$ -sheets adds the complexity of  $\beta$ -strands detection from density maps. We propose a new approach to model traces of  $\beta$ -strands for  $\beta$ -barrel density regions that are extracted from cryo-EM density maps. In the test containing eight  $\beta$ -barrels extracted from experimental cryo-EM density maps at 5.5 Å–8.25 Å resolution, *StrandRoller* detected about 74.26% of the amino acids in the  $\beta$ -strands with an overall 2.05 Å 2-way distance between the detected  $\beta$ -traces and the observed ones, if the best of the fifteen detection cases is considered.

## 1. Introduction

Cryo-electron microscopy (cryo-EM) has become a major experimental technique to study structures of large protein complexes [1, 2]. Many large complexes have been resolved to about 3 Å resolution recently [3, 4], at which the position of protein backbone can be distinguished. For cryo-EM maps at lower resolutions, such as 5–8 Å (referred to as medium resolution in the paper), detailed molecular features are not resolved. It is a challenging problem to derive atomic structures from such density maps. Two types of approaches have been proposed. Fitting relies on a suitable atomic structure [5–9] and de novo modeling relies on the match of secondary structures between those in the density map and those in the protein sequence [10–17]. The most characteristic patterns in cryo-EM density maps at medium resolutions come from secondary structures of a protein chain. An  $\alpha$ -helix often appears as a cylinder or a stick and can be identified using image processing methods [14, 18–22]. A  $\beta$ -sheet consists of multiple  $\beta$ -strands (Figure 1). A  $\beta$ -sheet often appears as a thin layer of density and can be detected computationally [14, 20, 23–25]. The spacing between two neighboring  $\beta$ -strands is between 4.5 and 5 Å, and therefore  $\beta$ -strands are not resolved at medium resolutions [26, 27].

The position of  $\beta$ -strands provides important constraints in backbone modeling of a protein. We previously proposed an approach to predict the location of  $\beta$ -strands using *StrandTwister* [28]. *StrandTwister* is built on the principle of right-handed twist of  $\beta$ -strands that was discovered as early as 1970s [29]. The right-handed twist was measured along the peptide orientation as about 0° to 30° per residue [30], and an estimation of the strand orientation was proposed using the corners of a  $\beta$ -sheet [31]. *StrandTwister* was able to detect  $\beta$ -strands from single  $\beta$ -sheets. However,  $\beta$ -sheets have a variety of shapes that add complexity for  $\beta$ -strands detection. Some  $\beta$ -sheets appear as rolls and propellers, and others are  $\beta$ -barrels (Figure 2). Currently there is no computational method to detect  $\beta$ -strands from a  $\beta$ -barrel density map. In this paper, we propose a method to predict  $\beta$ -strands by utilizing prior knowledge about  $\beta$ -barrels. The proposed method *StrandRoller* is quite different from the method of *StrandTwister*.

A  $\beta$ -barrel is a large  $\beta$ -sheet in which the first  $\beta$ -strand is hydrogen-bonded with the last  $\beta$ -strand.  $\beta$ -barrels are commonly found in porins and other proteins that span cell membranes [32]. It was noticed by McLachlan in 1979 that the number of strands and their relative stagger completely determines the overall structure of a  $\beta$ -barrel [33]. The

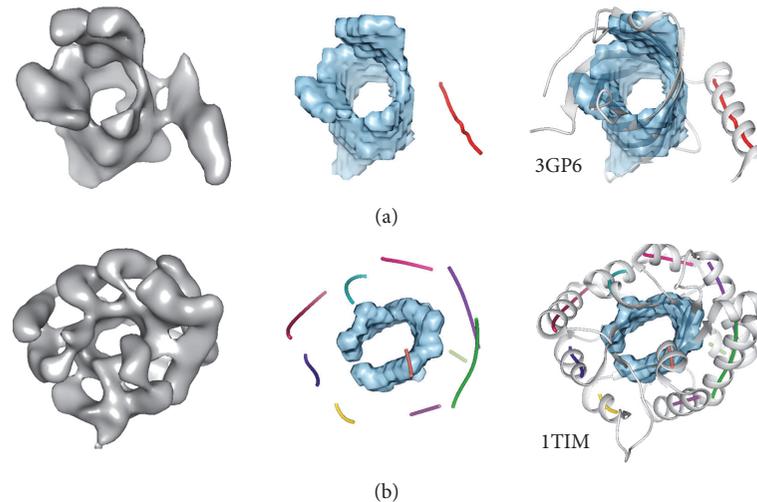


FIGURE 1: Secondary structure detection from density maps. Left: A density map (gray) was simulated at 10 Å resolution using the atomic structure of protein 3GP6 (PDB ID) in (a), 1TIM (PDB ID) in (b), and EMAN software [38]. Middle: helices (colored lines) and  $\beta$ -sheets (blue voxels) are detected from the density maps using *SSETracer* [24]. Right: The detected  $\alpha$ -helices and  $\beta$ -sheets are superimposed with the atomic structure in (a) and (b), respectively. Each atomic structure contains a  $\beta$ -barrel. Figure 1(a) is reproduced with permission from IEEE [39].

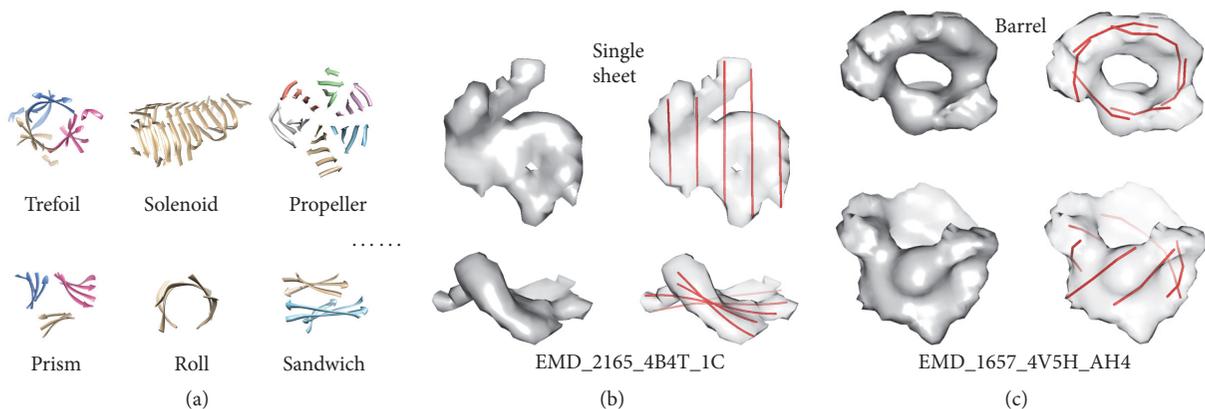


FIGURE 2: Challenges of  $\beta$ -structure modeling from cryo-EM density maps. (a) Various  $\beta$ -structures classified in CATH database. (b) An example of an isolated  $\beta$ -sheet density map and the  $\beta$ -traces predicted using *StrandTwister*. (c) An example of an isolated  $\beta$ -barrel density map and the  $\beta$ -traces detected using the method proposed in this paper. The top view (upper row) and the side view (bottom row) are shown in (b) and (c).

main structural characteristics of an ideal  $\beta$ -barrel have been discussed based on a cylindrical barrel [33–35]. Studies have shown that tilt angle and interstrand distance for all  $\beta$ -barrel structures vary within a fairly small range [35–37]. Our method, *StrandRoller*, is designed to utilize such characteristics of tilt angles and interstrand distance.

A helix identified from the medium resolution cryo-EM map is often represented as a line (colored line in Figure 1), referred to as an  $\alpha$ -trace that corresponds to the central axis of a helix. We define a  $\beta$ -trace as the central line along a  $\beta$ -strand. In particular, the observed  $\beta$ -trace is the line interpolating all geometrical centers of three consecutive  $C\alpha$  atoms on a  $\beta$ -strand plus two  $C\alpha$  atoms at the ends of the  $\beta$ -strand (black line in Figure 3). An observed  $\beta$ -trace represents the line along the atomic structure of a  $\beta$ -strand (Figure 3). Given the  $\beta$ -barrel density voxels, the problem of

$\beta$ -strands detection is to find the orientation (Figure 3(a)) and location (Figure 3(b)) of  $\beta$ -traces from the three-dimensional cryo-EM map.

Preliminary result of *StrandRoller* was shown in [39]. Details of the method and more thorough tests on different sizes of  $\beta$ -barrels are shown in this paper. In addition, we tested our method on eight pieces of experimental cryo-EM data that were downloaded from EM DataBank. The result suggests that *StrandRoller* can be used for the prediction of  $\beta$ -traces from medium resolution cryo-EM density maps of  $\beta$ -barrel when the rough barrel region is segmented.

## 2. Method

$\beta$ -Barrels have a characteristic shape. Atomic structures of  $\beta$ -barrels have been modeled as hyperboloid surfaces [40–42]

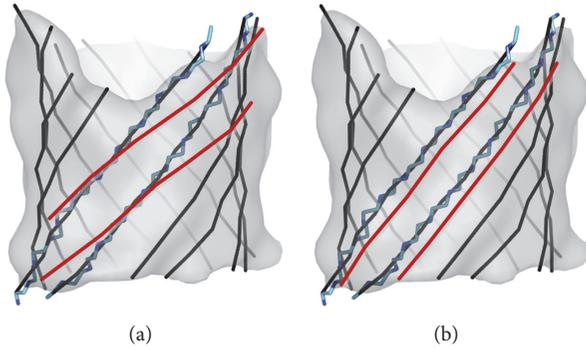


FIGURE 3: The problem of  $\beta$ -barrel modeling from medium resolution density maps. A set of  $\beta$ -traces is shown in black lines at the front and gray lines at the back. Two possible sets of  $\beta$ -traces (represented by two black lines and two red lines) may have different orientations shown in (a) or locations/shifts shown in (b). The atomic structure of the two  $\beta$ -strands is superimposed on the two representative  $\beta$ -traces in (a) and (b). Figure 3 is reproduced with permission from IEEE [39].

or catenoid surfaces [43]. Although atomic structures of  $\beta$ -barrels are different from  $\beta$ -barrel density maps, many small  $\beta$ -barrels visually appear as cylinders with nonuniform ends. Although various models can be used to approximate the major area of a  $\beta$ -barrel,  $\beta$ -barrel density maps often deviate from the mathematical models at certain regions. In order to create a surface well representing the density map, we used a simple elliptical model initially and adjusted the model at those regions that do not fit. Strand generation was then performed on the adjusted barrel surface model (Figure 4). We assume that the  $\beta$ -barrel density has been segmented from the entire cryo-EM density map.

**2.1.  $\beta$ -Barrel Surface Model from Cryo-EM Density Voxels.** In order to represent the shape of a  $\beta$ -barrel, we reduced the  $\beta$ -barrel density map to a surface. Two steps are involved in creating the surface model. The first step involves identification of the axis of the barrel density map (Figure 5(a)). The barrel density map was first translated to global origin (0, 0, 0) based on its geometric center. An elliptical cylinder (1) was then utilized to search for the orientation of central axis. The orientation was selected using exhaustive search and least square fitting to the cylinder. The entire barrel density was then rotated such that the central axis aligns with the  $z$ -axis (Figure 5(b)).

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

Instead of a mathematical formula, our  $\beta$ -barrel model consists of a thin layer of density voxels that closely represents the morphed barrel shape and outline of barrel at two ends (yellow points in Figure 5(e)). The barrel model was generated using cross-sections from bottom to top of the volume. The density voxels on each cross-section of  $Z$ -axis (Figure 5(c), gray) appear to be nearby the ideal model of ellipse. The voxels that are closest to the ideal ellipse were

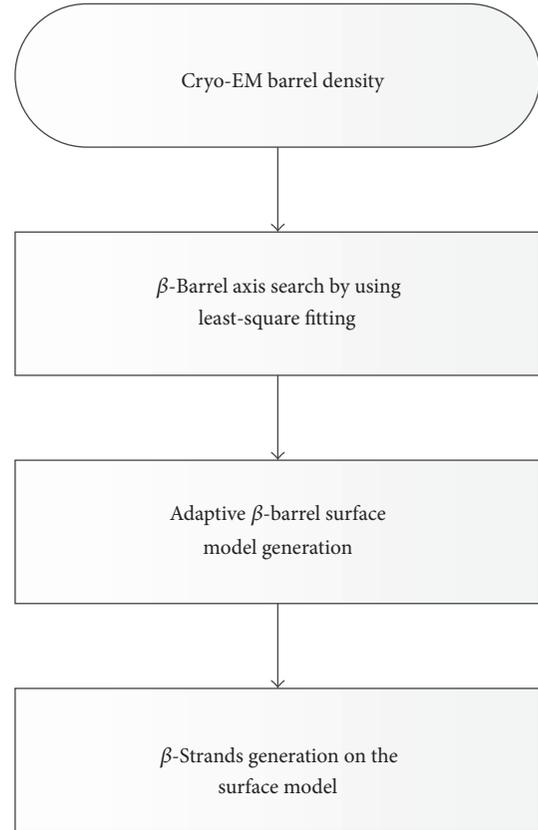


FIGURE 4: The flow chart of *StrandRoller*.

selected as the points on barrel model (Figure 5(c), yellow). Note that such a discrete model closely represents the three-dimensional distribution of the voxels. For example, when the fitted ideal elliptical cylinder is outside the density (arrows in Figure 5(c)), the voxels on the density map were used to adjust the barrel model. It appears that the resulting barrel model clearly represents the morphed regions, especially at the two ends of barrel (arrows in Figure 5(e)). We find that it is important to have an accurate barrel surface to model the  $\beta$ -traces accurately.

**2.2. Strand Generation on the Barrel Model.** McLachlan noticed in 1979 that the number of strands and their relative stagger completely determines the overall structure of a  $\beta$ -barrel [33]. The main structural characteristics of ideal  $\beta$ -barrel have been discussed based on a cylindrical barrel [33–35]. Studies have shown that tilt angle  $\alpha$  (Figure 5(f)) of a  $\beta$ -strand can vary between  $30^\circ$  and  $60^\circ$ , as reflected in the known structures of membrane proteins [35–37]. The tilt angle  $\alpha$  may vary by  $\pm 15^\circ$  for different strands in the same  $\beta$ -barrel [37]. However, the interstrand distance  $d$  remains to be  $4.5\text{--}5 \text{ \AA}$  due to hydrogen bonds between two neighboring  $\beta$ -strands. *StrandRoller* uses previous knowledge about the tilt angle  $\alpha$  and the interstrand distance  $d$  in the modeling of  $\beta$ -strands.

An initial  $\beta$ -trace (blue in Figure 5(f)) was produced by tilting the barrel axis with  $\alpha$  and projecting it onto the barrel surface model. The second  $\beta$ -trace was then generated from

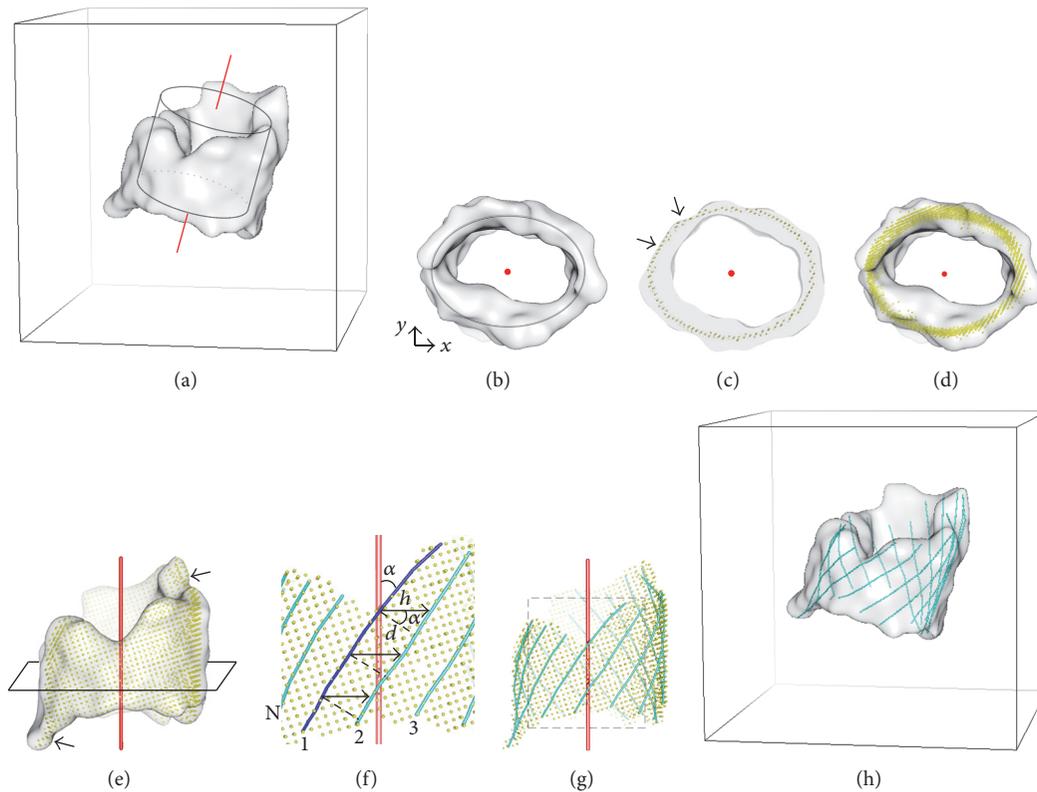


FIGURE 5: Generation of  $\beta$ -barrel model and  $\beta$ -traces from a  $\beta$ -barrel density map. (a-b) The barrel axis (red) was identified by fitting an elliptical cylinder to the density map (gray). (c) A cross-section of the barrel shows adjustment (arrow) of the ideal elliptical cylinder model at morphed regions. Top view (d) and the side view (e) of the modeled barrel surface (yellow) show the barrel axis (red) and a cross-section of the  $\beta$ -barrel. The zoom-in view (f) and the entire-view (g) of the  $\beta$ -traces generated recursively on the surface model using tilt angle  $\alpha$  and interstrand distance  $d$  of  $\beta$ -strands.  $\beta$ -Traces are superimposed with the  $\beta$ -barrel density map in (h). Figure 5 is reproduced with permission from IEEE [39].

the previous one by traveling a horizontal distance  $h$  (2) on the barrel surface (Figure 5(f)).

$$h = \frac{d}{\cos \alpha}, \quad d = 4.8 \text{ \AA}. \quad (2)$$

Given a tilt angle, the entire set of  $\beta$ -traces can be built iteratively on the barrel surface (Figure 5(g)) until the last  $\beta$ -trace is generated (Figure 5(f)). The tilt angle was sampled every  $5^\circ$  between  $35^\circ$  and  $55^\circ$ , and three translations were sampled at each tilt angle. There are fifteen sets of  $\beta$ -traces generated for one barrel volume. Note that the fifteen sets are within a very small range of tilt angle ( $20^\circ$ ) and translation distance ( $4.8 \text{ \AA}$ ). The barrel density along with the detected  $\beta$ -strands was eventually translated and rotated back to the original position in the map after the detection is done (Figure 5(h)).

### 3. Result

*StrandRoller* was tested using three sets of  $\beta$ -barrel density maps: eighteen small simulated maps, fourteen large simulated maps, and eight experimental cryo-EM  $\beta$ -barrel maps. The proteins used in the simulated test set were collected from the  $\beta$ -barrel transmembrane superfamily of Orientations of

Proteins in Membranes (OPM) database [44] with less than 40% sequence similarity. The atomic structures of  $\beta$ -barrels were used to generate  $\beta$ -barrel density maps at  $10 \text{ \AA}$  resolution using the `pdb2mrc` function in EMAN [38], with a sampling of  $1 \text{ \AA}/\text{pixel}$ . The experimental cryo-EM density maps were downloaded from EMDB (<http://www.emdatabank.org/>). Since atomic structures are available for such cryo-EM maps, the density region that corresponds to one chain of the protein was first segmented using the atomic structure as an envelope. The  $\beta$ -sheet voxels were then manually outlined based on the atomic structure of the  $\beta$ -barrel. Such segmented testing maps bare the characteristic of a  $\beta$ -sheet and have an outline of a  $\beta$ -barrel. The accuracy of  $\beta$ -strand detection was evaluated using two parameters as previously implemented [28]: 2-way distance between the set of detected  $\beta$ -traces and the set of observed  $\beta$ -traces and number of amino acids covered in the detected  $\beta$ -trace. The observed  $\beta$ -trace is the line interpolating all geometrical centers of three consecutive  $C\alpha$  atoms on a  $\beta$ -strand plus the two  $C\alpha$  atoms at the ends of  $\beta$ -strand, as shown in Figure 3.

In order to estimate how much of a  $\beta$ -strand was detected, the percentage of the detected  $C\alpha$  atoms of an observed  $\beta$ -strand was calculated. An amino acid of a  $\beta$ -strand is considered detected if the projection distance from its  $C\alpha$

atom to the corresponding detected  $\beta$ -trace is less than 2.5 Å, which is about half  $\beta$ -strand spacing. Since the number of detected  $\beta$ -strands may be different from the number of observed  $\beta$ -strands, one-to-one correspondence needs to be established between subsets of the  $\beta$ -traces. For example, if detected set contains five  $\beta$ -traces while observed set contains six  $\beta$ -traces, five out of the six observed  $\beta$ -traces which have the overall smallest 2-way distance with the five detected  $\beta$ -traces will be selected for the calculation of 2-way distance. This ensures that the same number of detected  $\beta$ -traces ( $S_1, S_2, \dots, S_T$ ) is compared to the same number of observed  $\beta$ -traces ( $S'_1, S'_2, \dots, S'_T$ ) in which  $S_k$  is compared with  $S'_k$  for  $k = 1, \dots, T$ . The number of misdetected (and/or wrongly detected)  $\beta$ -strands can be inferred from the difference between the total number of the observed and that of the detected  $\beta$ -traces. The 2-way distance of a  $\beta$ -strand  $k$ ,  $D_k$  was calculated for each pair of lines  $S_k$  and  $S'_k$ . The overall 2-way distance  $D$  reflects how far the two sets of  $\beta$ -traces (detected and observed) are from each other.

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

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

In formula (3),  $N$  and  $M$  are the numbers of points on detected  $\beta$ -traces ( $S_k$ ) and observed  $\beta$ -traces ( $S'_k$ ), respectively.  $i$  and  $j$  are the indices of a point along lines  $S_k$  and  $S'_k$ , respectively.  $D_i^{ss'}$  is the projection distance from point  $i$  of  $S_k$  to  $S'_k$ . The projection of point  $i$  is required to be within line  $S'_k$ . In case it is outside, the distance between  $i$  and an end of  $S'_k$  was used as an approximate distance.

**3.1. Performance on the Simulated Density Data.** The purpose of this test is to investigate if traces of  $\beta$ -strands can be modeled from  $\beta$ -barrel density maps simulated to 10 Å resolution, at which the separation of  $\beta$ -strands is not visible. To discuss the ability of our  $\beta$ -trace detection, we use the best of fifteen sampled sets. The best set is the one that is closest to the observed set in terms of 2-way distance.

**3.1.1. Small-Medium Barrels.** Small-medium  $\beta$ -barrels refer to those with less than 15  $\beta$ -strands in each. The test of eighteen simulated small-medium sized  $\beta$ -barrel density maps shows that one of the fifteen sets of  $\beta$ -traces aligns very well with the observed set of  $\beta$ -traces, with an overall 2-way distance of 1.61 Å for the detected  $\beta$ -traces (Table 1). In the case of sheet A13 of PDB structure 1G7K, the detected set of  $\beta$ -traces appears to align with the  $\beta$ -strands very well (Figure 6(a)). In this case, all the eleven strands were detected with a small 2-way distance of 1.8 Å (Table 1 row 1).

To analyze the sensitivity of the detection, we calculated the percentage of the detected C $\alpha$  atoms of an observed  $\beta$ -strand. For example, ITX2\_B has all eight  $\beta$ -strands detected (Table 1 row 5). It missed three amino acids. For the eight detected  $\beta$ -strands, the 2-way distance is only 0.92 Å. Among the eighteen test cases, *StrandRoller* appears to be able to detect 78.26% of the  $\beta$ -strands fairly accurately in one of

TABLE 1: Accuracy of medium-small size (less than fifteen strands) beta-barrel modeling from simulated density maps at 10 Å resolution.

PDB ID <sup>a</sup>	#Det./#Obs. Strd <sup>b</sup>	2-w Dist. <sup>c</sup>	#Det./#Obs. AA <sup>d</sup>
1G7K_A13	11/11	1.80	85/124
1QD5_A17	12/12	1.93	99/140
1QJP_A	8/8	1.78	70/107
1RRX_A12	12/11	1.81	86/118
1TX2_B	8/8	0.92	31/34
1UYN_XA12	12/12	1.83	144/189
2ERV_A10	8/8	1.17	74/94
2K0L_A	8/8	1.96	57/79
2LHF_A	8/8	1.44	61/79
2QOM_C	12/12	1.71	153/190
2VDF_AA28	10/10	1.61	144/172
2WJR_AA15	12/12	1.50	114/130
2X9K_AA15	14/14	1.63	151/188
3AEH_A15	12/12	1.85	141/194
3FID_A14	12/12	1.42	131/155
3GP6_A	8/8	1.94	60/88
4EIS_A13	12/12	1.39	118/136
4FQE_A	12/12	1.28	121/134
Average		1.61	1840/2351 = 78.26%
Standard deviation		0.29	

<sup>a</sup>PDB.Sheet ID.

<sup>b</sup>The number of  $\beta$ -traces in the best of the fifteen possible sets/the number of  $\beta$ -strands in the  $\beta$ -sheet of the PDB structure.

<sup>c</sup>The 2-way distance (in Å) between observed  $\beta$ -traces and modeled  $\beta$ -traces for the best of the fifteen possible sets.

<sup>d</sup>The number of detected/the total number of amino acids in the  $\beta$ -barrel.

the fifteen sampled sets of  $\beta$ -traces (Table 1). Seventeen test cases have the number of  $\beta$ -strands detected the same as observed. The number of detected amino acids and 2-way distance are two parameters that have been used previously in accuracy measurement. Length-association method was proposed recently and can be a potentially more sensitive method to evaluate secondary structure detection [45].

**3.1.2. Large Barrels.** Large barrels in this paper refer to those with more than 15  $\beta$ -strands. Large barrels appear to be more challenging. Some extremely large  $\beta$ -barrels, such as the 22-stranded  $\beta$ -barrels 2GUF\_D23 and 2HDI\_D23 (Table 2), were still well detected. The 2-way distance is 2.03 Å in the case of 2GUF\_D23 (Table 2 row 1) and 1.93 Å in the case of 2HDI\_D23 (Table 2 row 2).  $\beta$ -Barrel 2GUF\_D23 has twenty-one of twenty-two  $\beta$ -strands detected (Table 2 row 1). It missed sixty-six amino acids in which most are at the edge (arrows in Figure 6(b)). For the twenty-one detected  $\beta$ -strands, the 2-way distance is 2.03 Å. Among the fourteen test cases of large sized  $\beta$ -barrels, *StrandRoller* appears to be able to detect 69.46% of the  $\beta$ -strands fairly accurately in one of the fifteen possible sets of  $\beta$ -traces, with an overall 2-way distance of 2.12 Å for the detected  $\beta$ -traces (Table 2).

It is noticed that the performance of *StrandRoller* is better on small-medium sized barrel than on the large sized barrel.

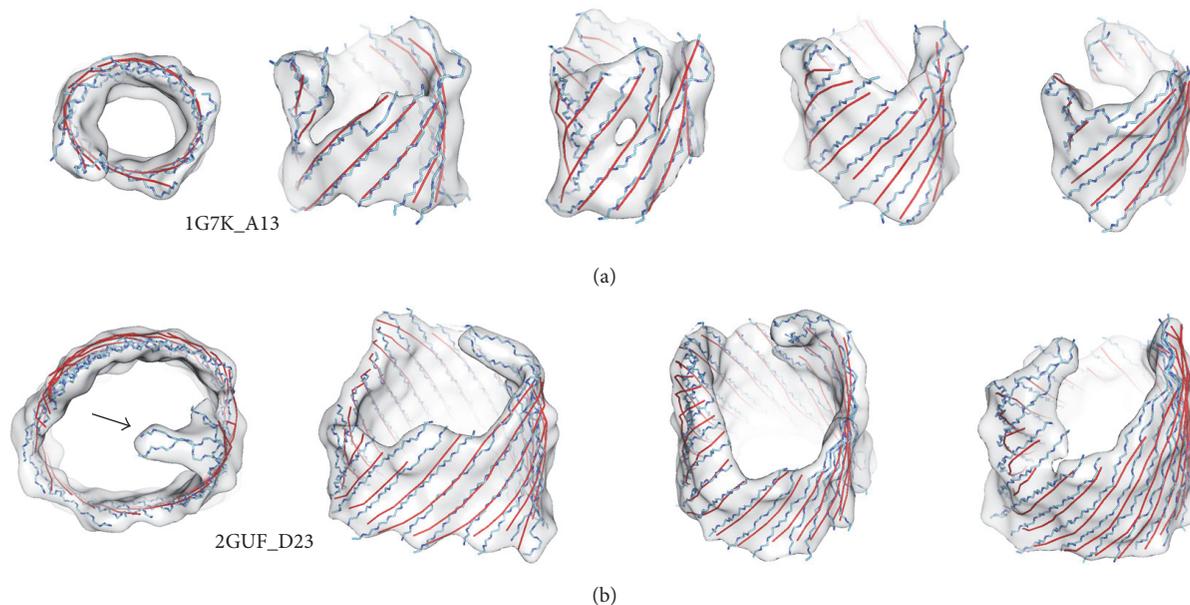


FIGURE 6:  $\beta$ -Strands modeling from the simulated density map at 10 Å. The best of the fifteen sets of modeled  $\beta$ -traces (red) are superimposed with the backbone of  $\beta$ -strands (blue) and the density maps (gray) for  $\beta$ -barrels 1G7K\_A13 in (a) and 2GUF\_D23 in (b). The top view (left) and four side views are shown in each case.

TABLE 2: Accuracy of large size beta-barrel (more than fifteen strands) modeling from simulated density maps at 10 Å resolution.

PDB ID <sup>a</sup>	#Det./#Obs. Strd <sup>b</sup>	2-w Dist. <sup>c</sup>	#Det./#Obs. AA <sup>d</sup>
2GUF_D23	21/22	2.03	230/296
2HDI_D23	22/22	1.93	226/296
2JIN_AA18	16/16	1.77	146/181
2MPR_A19	19/19	2.49	134/224
2POR_S117	12/16	2.53	94/178
2QDZ_C17	16/16	1.32	177/198
2YNK_AC20	20/18	2.26	133/194
3CSL_C23	23/22	2.59	171/300
3EMN_A20	20/19	1.83	135/183
3PRN_A17	15/17	1.67	131/162
3SYB_A19	19/18	2.12	155/220
4C00_AE17	17/16	2.37	121/202
4GEY_A18	19/16	2.33	116/189
4K3C_C18	18/16	2.43	126/193
<i>Average</i>		2.12	2095/3016 = 69.46%
<i>Standard deviation</i>		0.38	

<sup>a</sup>PDB\_Sheet ID.

<sup>b</sup>The number of  $\beta$ -traces in the best of the fifteen possible sets/the number of  $\beta$ -strands in the  $\beta$ -sheet of the PDB structure.

<sup>c</sup>The 2-way distance (in Å) between the observed  $\beta$ -traces and the modeled  $\beta$ -traces for the best of the fifteen possible sets.

<sup>d</sup>The number of detected/the total number of amino acids in the  $\beta$ -barrel.

Large  $\beta$ -barrels are more likely to adopt flexible shapes. The missing detection appears to be more at the edge of large  $\beta$ -barrels, where the  $\beta$ -strands tend to be more flexible (arrows in Figure 6(b)). The number of  $\beta$ -strands in large  $\beta$ -barrels also tends to be hard to detect due to the error accumulated during strand generation step. Since each  $\beta$ -trace is deducted

from the previous generated one, error could be propagated while traveling around the barrel.

3.2. *Performance on Experimental Cryo-EM Data. StrandRoller* was tested using eight  $\beta$ -barrels obtained from experimental cryo-EM density maps. The eight test cases are small ribosomal proteins in which the first  $\beta$ -strand is hydrogen-bonded with the last  $\beta$ -strand. Experimental data are often more challenging to analyze due to the noise and missing density. Figure 7 shows three density regions that were segmented from cryo-EM maps at 5.8 Å, 5.5 Å, and 6.7 Å resolutions, respectively. At these resolutions,  $\beta$ -strands are not visible in density maps. *StrandRoller* was able to detect all  $\beta$ -strands on the barrels and they align fairly well with the observed  $\beta$ -traces. In the case of 70S ribosome EMD\_1657 (sheet AH4 in protein 4V5H), the 2-way distance for the five-stranded barrel is 1.94 Å, and it detected 25 of 30 amino acids on the  $\beta$ -barrel (Figure 7(a)). In the case of 80S Ribosome EMD\_1780 (sheet AH4 in protein 4V7E), the 2-way distance for the six-stranded barrel is 1.88 Å, and it detected 24 of 28 amino acids on the  $\beta$ -barrel (Figure 7(b)). We noticed that the eight  $\beta$ -barrels in the cryo-EM maps are all small barrels with less than nine  $\beta$ -strands. *StrandRoller* appears to be fairly accurate in detection of  $\beta$ -strands from such cryo-EM maps with an overall 2.05 Å 2-way distance and 74.26% of amino acids detected (Table 3).

## 4. Discussion

We previously showed that the accuracy of  $\beta$ -strand detection is affected by the accuracy of  $\beta$ -sheet detection [28]. This is also true in the context of  $\beta$ -barrels. A  $\beta$ -barrel is a closed structure, and the number of  $\beta$ -strands may be estimated

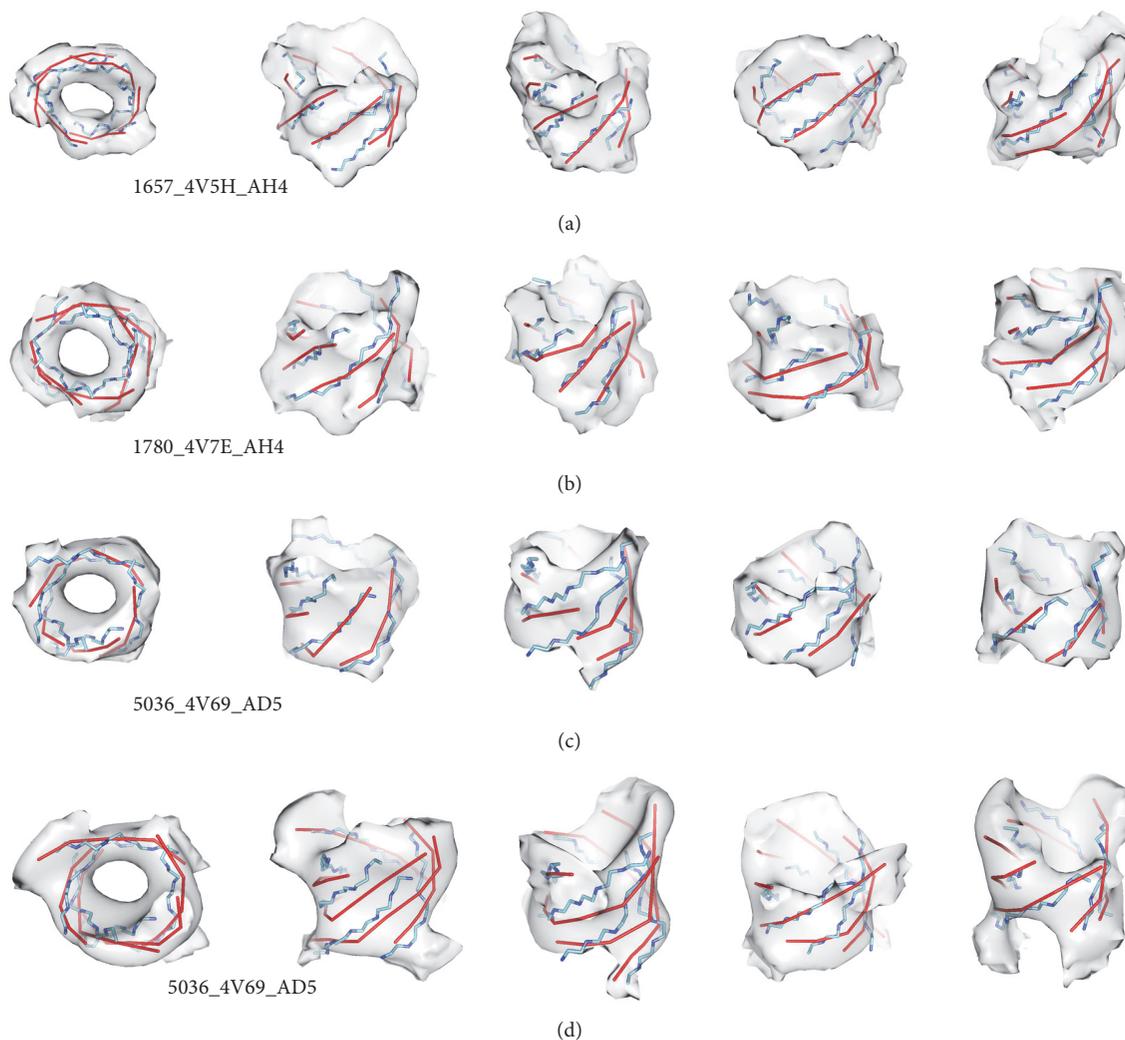


FIGURE 7:  $\beta$ -Traces detected from  $\beta$ -barrel density maps obtained from experimentally derived cryo-EM density maps. The best of the fifteen sets of modeled  $\beta$ -traces (red) are superimposed with the backbone of the  $\beta$ -strands (blue) and the density maps (gray) for  $\beta$ -barrel 1657\_4V5H\_AH4 (EMD\_1657, sheet AH4 of protein 4V5H) in (a) and 1780\_4V7E\_AH4 (EMD\_1780, sheet AH4 of protein 4V7E) in (b). The top view (left) and four side views are shown in each case. (c) A conservative segmentation for  $\beta$ -barrel 5036\_4V69\_AD5 (EMD\_5036, sheet AD5 of protein 4V69) and its corresponding  $\beta$ -strands modeling result. (d) A relaxed segmentation of the same  $\beta$ -barrel as in (c) superimposed with the backbone and the detected  $\beta$ -traces.  $\beta$ -Barrels in (c) and (d) are shown using the same density threshold in similar view points.

TABLE 3: Accuracy of beta-barrel modeling from cryo-EM experimental maps at medium resolutions.

EMD_PDB ID <sup>a</sup>	Res. <sup>b</sup>	#Det./#Obs. Strd <sup>c</sup>	2-w Dist. <sup>d</sup>	#Det./#Obs. AA <sup>e</sup>
1657_4V5H_AH4	5.8 Å	6/6	1.94	25/30
1780_4V7E_AH4	5.5 Å	5/5	1.88	24/28
1829_4V5H_AH4	5.6 Å	8/6	2.11	25/30
1849_4V6K_AE6		7/6	2.34	24/40
1849_4V6K_AC6	8.25 Å	6/5	1.90	26/31
1849_4V6K_AG2		6/6	2.74	31/52
5030_4V68_AF8	6.4 Å	4/5	1.72	24/32
5036_4V69_AD5	6.7 Å	5/5	1.73	23/29
Average			2.05	202/272 = 74.26%
Standard deviation			0.35	

<sup>a</sup>EMDB\_PDB\_Sheet ID.

<sup>b</sup>Resolution of the cryo-EM density map.

<sup>c</sup>The number of  $\beta$ -traces detected in the best of the fifteen possible sets/the number of  $\beta$ -strands in the  $\beta$ -sheet of the PDB structure.

<sup>d</sup>The 2-way distance (in Å) between the observed  $\beta$ -traces and the modeled  $\beta$ -traces for the best of the fifteen possible sets.

<sup>e</sup>The number of detected/the total number of amino acids in the  $\beta$ -barrel.

from the diameter of the barrel. Figures 7(c) and 7(d) show the  $\beta$ -traces detected from  $\beta$ -barrel 5036\_4V69\_AD5 that were manually extracted from the cryo-EM density map using two different segmentations. The segmentation in Figure 7(c) is more conservative than that in Figure 7(d). The relaxed segmentation of the same barrel includes more density volume at the edge of the barrel. Although the 2-way distance is 1.73 Å in the more conservative segmentation versus 1.83 Å in the other, the main difference in the resulting  $\beta$ -traces appears to be the length difference. Both detected the same number of  $\beta$ -strands with similar orientation and position. Our result suggests that the number of detected  $\beta$ -strands is not sensitive to the density segmentation errors at the two ends of  $\beta$ -barrel.

## 5. Conclusion

The position of  $\beta$ -strands is critical for modeling atomic structures of proteins. However, it has been a challenging problem to detect  $\beta$ -strands when no separation of the  $\beta$ -strands is visible from the density maps. The variety of shapes of  $\beta$ -sheets adds the complexity of this problem. We previously proposed *StrandTwister* to detect  $\beta$ -strands from single  $\beta$ -sheet using right-handed twist [28]. We propose a new method to predict  $\beta$ -strands from a  $\beta$ -barrel density map directly using the characteristic tilt angles of the  $\beta$ -barrel. This approach bypasses the need to measure twist angles. Our results show that this approach is feasible. As long as the rough density region of a  $\beta$ -barrel is isolated from the entire density map, location of  $\beta$ -strands can be modeled. However, current limiting factor is the lack of automatic detection methods of  $\beta$ -barrels from a cryo-EM density map. In fact a  $\beta$ -barrel has a fundamental shape character in which a hole is surrounded by a  $\beta$ -sheet. However, accurate detection of  $\beta$ -barrels needs to consider different characters of the hole depending on different sizes of  $\beta$ -barrels. We are hopeful that such a detection tool will be available in the near future.

*StrandRoller* does not require the resolution of cryo-EM density map to be higher than 5 Å to resolve the separation of  $\beta$ -strands. It applies to the maps with lower resolutions. In the test containing eight experimental cryo-EM  $\beta$ -barrel maps between 5.5 Å and 8.25 Å, *StrandRoller* detected about 74.26% of the amino acids in the  $\beta$ -strands in one of the fifteen sets of predicted traces. We demonstrate again that it is possible to derive  $\beta$ -strands from density maps at medium resolutions. To our knowledge, *StrandRoller* is the first method that attempts to address the problem of  $\beta$ -strands detection from medium resolution  $\beta$ -barrel maps. Future work includes developing more accurate methods in identification of  $\beta$ -traces and generating alternative  $\beta$ -traces for further evaluation in modeling.

## Competing Interests

The authors declare that they have no competing interests.

## Authors' Contributions

Dong Si and Jing He developed the method. Dong Si implemented the method and conducted the tests. Dong Si and Jing He wrote the manuscript.

## Acknowledgments

The work in this paper is partially supported by NSF DBI-1356621, NIH R01-GM062968, and the FP3 fund of the Old Dominion University.

## 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, no. 3, pp. 363–372, 2005.
- [2] Z. H. Zhou, "Atomic resolution cryo electron microscopy of macromolecular complexes," *Advances in Protein Chemistry and Structural Biology*, vol. 82, pp. 1–35, 2011.
- [3] Z. Liu, F. Guo, F. Wang, T.-C. Li, and W. Jiang, "2.9 Å resolution cryo-EM 3D reconstruction of close-packed virus particles," *Structure*, vol. 24, no. 2, pp. 319–328, 2016.
- [4] A. Bartesaghi, A. Merk, S. Banerjee et al., "2.2 Å resolution cryo-EM structure of  $\beta$ -galactosidase in complex with a cell-permeant inhibitor," *Science*, vol. 348, no. 6239, pp. 1147–1151, 2015.
- [5] M. G. Rossmann, "Fitting atomic models into electron-microscopy maps," *Acta Crystallographica Section D: Biological Crystallography*, vol. 56, no. 10, pp. 1341–1349, 2000.
- [6] L. G. Trabuco, E. Villa, K. Mitra, J. Frank, and K. Schulten, "Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics," *Structure*, vol. 16, no. 5, pp. 673–683, 2008.
- [7] G. F. Schröder, A. T. Brunger, and M. Levitt, "Combining efficient conformational sampling with a deformable elastic network model facilitates structure refinement at low resolution," *Structure*, vol. 15, no. 12, pp. 1630–1641, 2007.
- [8] K.-Y. Chan, L. G. Trabuco, E. Schreiner, and K. Schulten, "Cryo-electron microscopy modeling by the molecular dynamics flexible fitting method," *Biopolymers*, vol. 97, no. 9, pp. 678–686, 2012.
- [9] M. Topf, K. Lasker, B. Webb, H. Wolfson, W. Chiu, and A. Sali, "Protein structure fitting and refinement guided by cryo-EM density," *Structure*, vol. 16, no. 2, pp. 295–307, 2008.
- [10] K. Al Nasr, D. Ranjan, M. Zubair, L. Chen, and J. He, "Solving the secondary structure matching problem in de novo modeling using a constrained K-shortest path graph algorithm," *IEEE Transaction of Computational Biology and Bioinformatics*, vol. 11, no. 2, pp. 419–430, 2014.
- [11] M. L. Baker, S. S. Abeysinghe, S. Schuh et al., "Modeling protein structure at near atomic resolutions with Gorgon," *Journal of Structural Biology*, vol. 174, no. 2, pp. 360–373, 2011.
- [12] S. Lindert, N. Alexander, N. Wötzel, M. Karakaş, P. L. Stewart, and J. Meiler, "EM-Fold: de novo atomic-detail protein structure determination from medium-resolution density maps," *Structure*, vol. 20, no. 3, pp. 464–478, 2012.
- [13] K. Al Nasr and J. He, "Constrained cyclic coordinate descent—beyond the protein loop closure problem," *Robotica*, vol. 34, no. 8, pp. 1777–1790, 2016.
- [14] 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, no. 9, pp. 698–708, 2012.
- [15] A. Biswas, D. Si, K. Al Nasr, D. Ranjan, M. Zubair, and J. He, "Improved efficiency in Cryo-Em secondary structure topology

- determination from inaccurate data,” *Journal of Bioinformatics and Computational Biology*, vol. 10, no. 3, Article ID 1242006, 2012.
- [16] A. Biswas, D. Ranjan, M. Zubair, S. Zeil, K. Al Nasr, and J. He, “An effective computational method incorporating multiple secondary structure predictions in topology determination for cryo-EM images,” in *Proceedings of the IEEE/ACM Transactions on Computational Biology and Bioinformatics*, March 2016.
- [17] K. Al Nasr, D. Ranjan, M. Zubair, and J. He, “Ranking valid topologies of the secondary structure elements using a constraint graph,” *Journal of Bioinformatics and Computational Biology*, vol. 9, no. 3, pp. 415–430, 2011.
- [18] W. Jiang, M. L. Baker, S. J. Ludtke, and W. Chiu, “Bridging the information gap: computational tools for intermediate resolution structure interpretation,” *Journal of Molecular Biology*, vol. 308, no. 5, pp. 1033–1044, 2001.
- [19] A. Dal Palu, J. He, E. Pontelli, and Y. Lu, “Identification of  $\alpha$ -helices from low resolution protein density maps,” in *Proceeding of the Computational Systems Bioinformatics Conference (CSB '06)*, pp. 89–98, 2006.
- [20] M. L. Baker, T. Ju, and W. Chiu, “Identification of secondary structure elements in intermediate-resolution density maps,” *Structure*, vol. 15, no. 1, pp. 7–19, 2007.
- [21] M. Rusu and W. Wriggers, “Evolutionary bidirectional expansion for the tracing of alpha helices in cryo-electron microscopy reconstructions,” *Journal of Structural Biology*, vol. 177, no. 2, pp. 410–419, 2012.
- [22] Z. Yu and C. Bajaj, “Computational approaches for automatic structural analysis of large biomolecular complexes,” *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 5, no. 4, pp. 568–582, 2008.
- [23] Y. Kong and J. Ma, “A structural-informatics approach for mining  $\beta$ -sheets: locating sheets in intermediate-resolution density maps,” *Journal of Molecular Biology*, vol. 332, no. 2, pp. 399–413, 2003.
- [24] D. Si and J. He, “Beta-sheet detection and representation from medium resolution cryo-EM density maps,” in *Proceedings of the 4th ACM Conference on Bioinformatics, Computational Biology and Biomedical Informatics (BCB '13)*, pp. 764–770, ACM, September 2013.
- [25] J. He, D. Si, and M. Arab, “Detection of secondary structures from 3D protein images of medium resolutions and its challenges,” in *Image and Graphics: 8th International Conference, ICIG 2015, Tianjin, China, August 13–16, 2015, Proceedings, Part II*, vol. 9218 of *Lecture Notes in Computer Science*, pp. 147–155, Springer, Berlin, Germany, 2015.
- [26] Z. H. Zhou, “Towards atomic resolution structural determination by single-particle cryo-electron microscopy,” *Current Opinion in Structural Biology*, vol. 18, no. 2, pp. 218–228, 2008.
- [27] 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, no. 9, pp. 655–668, 2012.
- [28] D. Si and J. He, “Tracing beta strands using strandtwister from cryo-EM density maps at medium resolutions,” *Structure*, vol. 22, no. 11, pp. 1665–1676, 2014.
- [29] C. Chothia, “Conformation of twisted  $\beta$ -pleated sheets in proteins,” *Journal of Molecular Biology*, vol. 75, no. 2, pp. 295–302, 1973.
- [30] J. S. Richardson, “The anatomy and taxonomy of protein structure,” *Advances in Protein Chemistry*, vol. 34, pp. 167–339, 1981.
- [31] J. S. Richardson and D. C. Richardson, “Interpretation of electron density maps,” *Methods in Enzymology*, vol. 115, pp. 189–206, 1985.
- [32] V. Cherezov, W. Liu, J. P. Derrick et al., “In meso crystal structure and docking simulations suggest an alternative proteoglycan binding site in the OpcA outer membrane adhesin,” *Proteins: Structure, Function and Genetics*, vol. 71, no. 1, pp. 24–34, 2008.
- [33] A. D. McLachlan, “Gene duplications in the structural evolution of chymotrypsin,” *Journal of Molecular Biology*, vol. 128, no. 1, pp. 49–79, 1979.
- [34] A. G. Murzin, A. M. Lesk, and C. Chothia, “Principles determining the structure of  $\beta$ -sheet barrels in proteins I. A theoretical analysis,” *Journal of Molecular Biology*, vol. 236, no. 5, pp. 1369–1381, 1994.
- [35] A. G. Murzin, A. M. Lesk, and C. Chothia, “Principles determining the structure of  $\beta$ -sheet barrels in proteins II. The observed structures,” *Journal of Molecular Biology*, vol. 236, no. 5, pp. 1382–1400, 1994.
- [36] T. Páli and D. Marsh, “Tilt, twist, and coiling in  $\beta$ -barrel membrane proteins: relation to infrared dichroism,” *Biophysical Journal*, vol. 80, no. 6, pp. 2789–2797, 2001.
- [37] G. E. Schulz, “The structure of bacterial outer membrane proteins,” *Biochimica et Biophysica Acta (BBA)—Biomembranes*, vol. 1565, no. 2, pp. 308–317, 2002.
- [38] S. J. Ludtke, P. R. Baldwin, and W. Chiu, “EMAN: semiautomated software for high-resolution single-particle reconstructions,” *Journal of Structural Biology*, vol. 128, no. 1, pp. 82–97, 1999.
- [39] D. Si and J. He, “Combining image processing and modeling to generate traces of beta-strands from cryo-EM density images of beta-barrels,” in *Proceedings of the 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC '14)*, pp. 3941–3944, Chicago, Ill, USA., August 2014.
- [40] J. Novotný, R. E. Bruccoleri, and J. Newell, “Twisted hyperboloid (*strophoid*) as a model of  $\beta$ -barrels in proteins,” *Journal of Molecular Biology*, vol. 177, no. 3, pp. 567–573, 1984.
- [41] E. Tolonen, B. Bueno, S. Kulshreshtha et al., “Allosteric transition and binding of small molecule effectors causes curvature change in central  $\beta$ -sheets of selected enzymes,” *Journal of Molecular Modeling*, vol. 17, no. 4, pp. 899–911, 2011.
- [42] I. Lasters, S. J. Wodak, P. Alard, and E. van Cutsem, “Structural principles of parallel beta-barrels in proteins,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 85, no. 10, pp. 3338–3342, 1988.
- [43] E. Koh and T. Kim, “Minimal surface as a model of  $\beta$ -sheets,” *Proteins: Structure, Function and Genetics*, vol. 61, no. 3, pp. 559–569, 2005.
- [44] M. A. Lomize, A. L. Lomize, I. D. Pogozheva, and H. I. Mosberg, “OPM: orientations of proteins in membranes database,” *Bioinformatics*, vol. 22, no. 5, pp. 623–625, 2006.
- [45] S. Zeil, J. Kovacs, W. Wriggers, and J. He, “A method to compare an atomic model or a structure and its Cryo-EM image at the central axis of a helix,” *Journal of Computational Biology*, vol. 24, no. 1, pp. 52–67, 2017.



**Hindawi**

Submit your manuscripts at  
<https://www.hindawi.com>

