

Analysis of β -strand Twist from the 3-dimensional Image of a Protein

Tunazzina Islam

Department of Computer Science
Old Dominion University
Norfolk, VA 23529
tislam@cs.odu.edu

Alice Poteat

Old Dominion University
Norfolk, VA 23529

Jing He

Department of Computer Science
Old Dominion University
Norfolk, VA 23529
jhe@cs.odu.edu

ABSTRACT

Electron cryo-microscopy (Cryo-EM) technique produces density maps that are 3-dimensional (3D) images of molecules. It is challenging to derive atomic structures of proteins from 3D images of medium resolutions. Twist of a β -strand has been studied extensively while little of the known information has been directly obtained from the 3D image of a β -sheet. We describe a method to characterize the twist of β -strands from the 3D image of a protein. An analysis of 11 β -sheet images shows that the Averaged Minimum Twist (AMT) angle is larger for a close set than for a far set of β -traces.

CCS CONCEPTS

- **Pattern Recognition** $\rightarrow$ **Models** $\rightarrow$ Geometric; Structural;
- **Simulation and Modeling** $\rightarrow$ General

KEYWORDS

Protein, β -strand; secondary structure; image; pattern, twist

1 INTRODUCTION

3-dimensional (3D) molecular images of large biological assemblies are produced and archived in the Electron Microscopy Data Bank (EMDB) [12]. As of April 2017, EMDB contains 4730 entries of 3D images (density maps) of molecules. Most of the 3D images with a higher than 10Å resolution are produced using cryo-electron microscopy (cryo-EM). This technique is particularly suitable for structure determination of large molecular assemblies that are often challenging for traditional methods such as X-ray crystallography and Nuclear Magnetic Resonance (NMR). Some recently solved atomic structures from high resolution images include the complex of β -galactosidase and inhibitor (2.2Å resolution) [5], bacteriophage Sf6 (2.9Å resolution)

[26], and P22 bacteriophage (3.3Å resolution) [10]. EMDB is linked with Protein Data Bank (PDB). It is possible to obtain a data set that contains both 3D images (from the EMDB) and their corresponding atomic models (from the PDB).

Although atomic structures are being routinely solved from density maps with sufficient quality, it is computationally challenging to derive atomic structures from density maps at medium resolutions (5-10 Å). Existing approaches require fitting a known atomic structure into the Cryo-EM density map [18; 24; 25]. De novo methods do not rely on template structures; rather, rely on the intrinsic relationship among secondary structures to construct the backbone of the protein [1-3; 6].

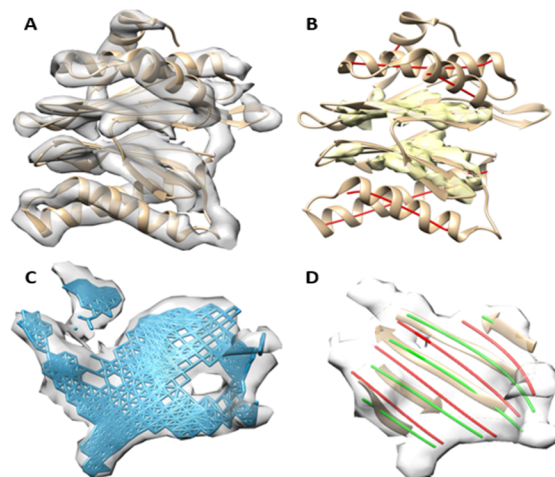


Figure 1: 3D image of cryo-electron microscopy density map, the atomic structure of a protein chain, and segmented secondary structures based on characteristic density patterns. (A): The density map (gray) extracted from Electron microscopy Data Bank 1733 (6.8 Å resolution) superimposed on its atomic structure (ribbon; PDB 3C91 chain H). (B): Segmented helices (represented as red sticks) and two β -sheets (yellow density), using SSETracer [19]. (C): The iteratively derived Bézier surface (blue) from β -sheet image (gray) that corresponds to sheet Q. (D): Two sets of β -traces (green and red lines) that were derived from the Bézier surface are superimposed on the atomic structure of β -strands (ribbon) and the image.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
ACM-BCB '17, August 20-23, 2017, Boston, MA, USA
© 2017 Association for Computing Machinery.
ACM ISBN 978-1-4503-4722-8/17/08...\$15.00
<http://dx.doi.org/10.1145/3107411.3107507>

Major secondary structures such as helices and β -sheets are visible in density maps of medium resolutions. An α -helix often appears as a cylinder in a density map of medium resolutions. Various methods have been developed to evaluate cylindrical character in the 3D image [4; 8; 11; 13; 17; 19; 23]. Recently a deep learning method was developed using a convolutional neural network [13]. A β -sheet may appear as a thin layer of density (Fig. 1A). Each β -sheet is composed of multiple β -strands with an inter-strand distance of 4.5Å–5Å. A β -strand possesses a right-handed twist [7], resulting in a non-flat structure of a β -sheet [9; 16]. This property may be visible in a medium-resolution image. Although a β -sheet is not as characteristic as a helix in an image, computational methods exist to detect major region of a β -sheet [4; 19; 23]. Fig. 1B shows an example of detected β -sheet region (yellow) using SSETracer [19].

Due to the spacing of β -strands (about 4.5–5Å), it is almost impossible to distinguish the location of β -strands from a density map with lower than 5Å resolution. We previously showed that right-handed twist of β -sheet is effective in elimination of wrong candidate sets of β -strands [21]. We recently have showed that it is also possible to predict β -strand traces from β -barrel images utilizing the nature of barrels [22]. However, it is still an open question as to how to distinguish the best candidate among a pool of candidate β -strands that are all right-handed twisted to slightly different levels. An important question is how to measure twist angles accurately for a set of lines. Twist angles are slightly different when measured at different locations of the lines. We previously used one angle at the central region of a pair of lines [20]. In this paper, we sample many angles along a pair of lines and analyze the behavior of twist angles for two types of sets: sets close and far from the realistic positions of the β -strands.

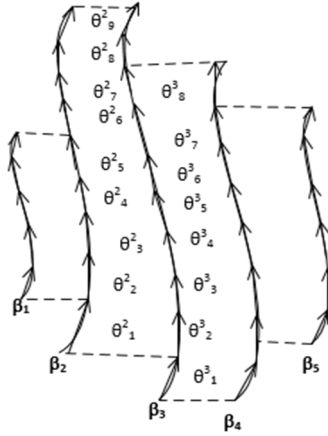


Figure 2: Twist angle calculation. The pairing length of two neighboring β -traces are marked with dashed lines.

2 METHOD

2.1 Computing twist angles from 3D image of a β -sheet

In order to represent twist angles precisely, a 3D image of the β -sheet was first converted to a surface using iterative Bézier fitting [15] or a polynomial fitting [21]. Once the surface is optimized, a

set of lines was generated from the surface to mimic traces of β -strands. We refer to the set of such generated lines as β -traces. One set of β -traces may differ from another set by orientation or by translation. It is important to distinguish the set that is closest to the actual set representing the atomic structure of β -strands.

Let β -strand traces be $\beta_1, \beta_2, \dots, \beta_m$, and let θ_j^i be the angle formed by β_i and β_{i+1} at location j of β_i , $1 \leq i < m$, $1 \leq j \leq n_i$. As an example, the angles formed by β_1 and β_2 along the strands are $(\theta_1^1, \theta_2^1, \dots, \theta_{n_1}^1)$. We divided a line of β -trace into consecutive vectors of certain length, and calculated the angle formed by two vectors from adjacent traces, as illustrated in Fig. 2. The longest pairing length of a pair of β -traces was determined for each two neighboring traces. Note that the longest pairing length is often different from the individual length of each trace due to the relative position of the two lines. Since longer pairs of β -strands are major components of a β -sheet, we used the two longest pairs of β -traces to calculate an overall twist angle of a set of β -traces. Note that since the level of twist is often different at different locations of a β -sheet, a consistent measure of twist is needed to compare the level of twist created by different sets of β -traces. We define an averaged minimum twist (AMT), τ , for each set of β -traces as in (1). The idea of AMT is that the region with the minimum twist is the most stable spot for a pair of β -traces, and the two longest pairs formed by β_{i_1} , β_{i_1+1} , and β_{i_2} , β_{i_2+1} respectively were used for representing the major region of a β -sheet.

$$\tau = \frac{1}{2} \left(\min_{k_1} \sum_{j_1=k_1}^{k_1+p} \frac{\theta_{j_1}^{i_1}}{p} + \min_{k_2} \sum_{j_2=k_2}^{k_2+q} \frac{\theta_{j_2}^{i_2}}{q} \right) \quad (1)$$

Here, p and q are parameters, for which we used 4 in this paper. The choice of p and q is related to the length of the vector in angle calculation.

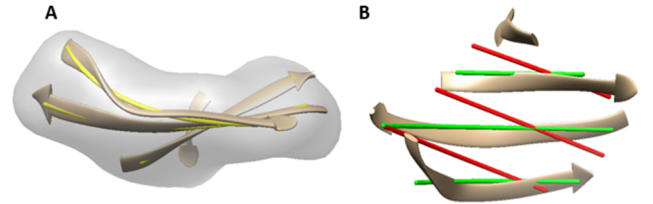


Figure 3: Twist analysis for β -strand splines, the close set, and the far set of β -traces. (A): The 3D image (gray) of β -sheet A of 1A12 (PDB ID), the atomic structure of β -strands (ribbon), and the β -strand splines (yellow lines) derived from the atomic structure. (B): The atomic structure (ribbon), the close set (green) and the far set of β -traces (red) derived from the 3D image of the β -sheet. Two different views of the same β -sheet are shown in (A) and (B) respectively.

2.2 Twist calculation using splines of β -strands

Given the atomic structure of a protein, one way to calculate the twist angle is to use two vectors, one from each of two neighboring β -strands [9]. The vector is defined by the two mid-points of consecutive C-N bonds along the β -strand. In order to

characterize the AMT of a set of lines that precisely represent β -strands, we derived splines for β -strands. A cubic spline of a β -strand was calculated using mid-points of C-N bonds on the β -strand. Once the splines were generated, AMT was calculated. Note that the process of generating β -strand splines is different from the process of generating β -traces. One uses the atomic structure of β -strands, while the other uses 3D image of the β -sheet. In order to generate a robust AMT for β -strand splines, we measured the AMT three times and took an average. Each pass of the AMT was calculated by shifting the beginning of the line by one, two, or three points on the line.

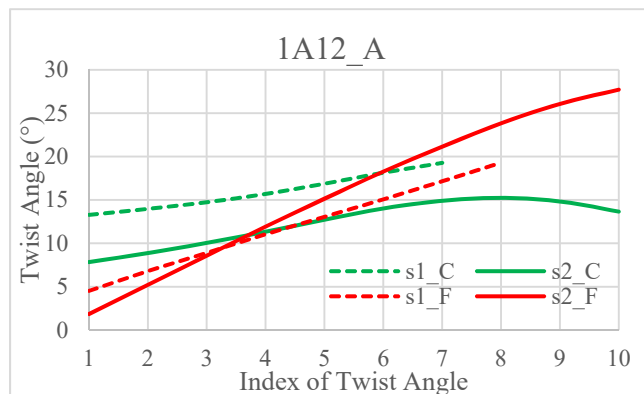


Figure 4: Twist angles along a pair of β -traces. Twist angles of the two longest pairs of β -traces in 1A12 (PDB ID) sheet A. The longest pair (S2) and the second longest pair (S1) of β -traces are shown for the close set (green) and far set (red).

3 RESULTS

Ten β -sheet images were generated using atomic structures of the β -sheets and Chimera [14] to a 10Å resolution. The data set represents small to large β -sheets, each of which has three to nine β -strands. StrandTwister was used to generate candidate sets of β -traces. StrandTwister only returns β -traces that have right-handed twist. The close set of β -traces aligns better with the atomic structure of β -strands than the far set, since it is selected based on the smallest overall distance between atoms of β -strands and the β -traces [21]. As an example, the close set of 1A12 sheet A appears to align much better than the far set with a visual inspection in Fig. 3.

3.1 Twist angles along a pair of β -traces

The magnitude of twist is often different at different locations of a β -strand. We investigated twist angles measured along a pair of β -traces for both the close set and the far set. Since most short pairs are near the edge, the most challenging region of the β -sheet, we focused on the longest pairs. For each set, the twist angles of two longest pairs are calculated. In the case of 1A12 sheet A, as shown in Fig. 4, the longest pair has twist angles between 7.8° and 15.3° for the close set, and between 1.9° and 27.7° for the far set. Note that both pairs are on the same surface, but they show

different levels of twist depending on where they are located on the surface due to twist of the surface. In this case, the region with the smallest twist is on the left side of the curve for the close set (solid green curve). This is the similar region with the smallest twist in the far set (solid red curve) (Fig. 4). If we compare the smallest twist of the longest pair between the close set and the far set, the close set would have bigger twist.

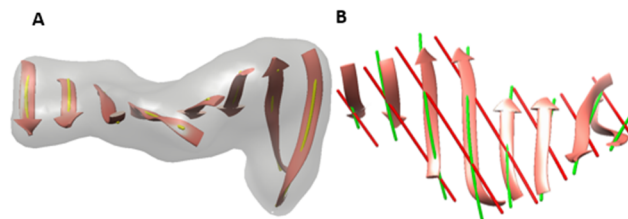


Figure 5: Twist analysis of β -strand splines, the close set, and the far set of β -traces 1DTD (PDB ID) sheet A. Colors and lines are annotated as in Fig. 3.

1DTD sheet A has eight β -strands. For the close set of 1DTD, as shown as green lines in Fig. 5B, the twist angles of the longest pair of β -traces (longest blue line in Fig. 6) are between 7° and 14°. The second longest pair (blue line in Fig. 6) has twist angles between 11° and 18°. Twist angles along a pair of β -traces show the challenge of twist analysis. The longest pair in a close set does not always have larger twist angles than the longest pair in a far case. In fact the two lines may cross (longest solid line and longest dashed line) indicating that in certain regions one has bigger twist than the other. However, if both the longest and the second longest pairs are considered, solid lines in Fig. 6 appear to out-twist corresponding dashed lines if AMT is calculated. Fig. 7 shows two sets of β -traces for 1ATZ sheet A.

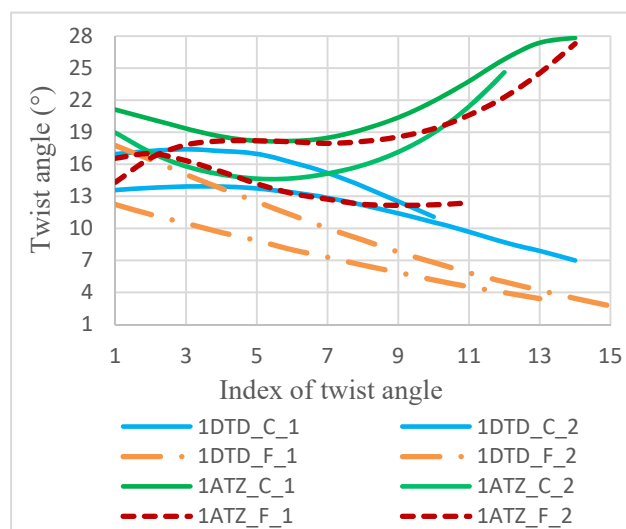


Figure 6: Twist angles along a pair of β -traces. Twist angles were calculated for the longest and the second longest pair of β -traces for a close set (solid lines) and a far set (dashed lines) in two cases: 1DTD sheet A (green lines) and 1ATZ sheet A (blue lines).

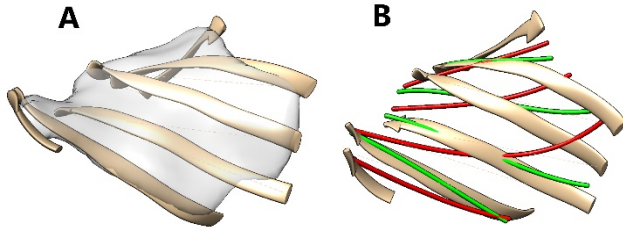


Figure 7: The close set and the far set of β -traces for 1ATZ (PDB ID) sheet A. Annotation method for color and lines is as in Fig. 3.

3.2 Twist characterization for simulated β -sheet images

In order to distinguish two sets of β -traces, we defined AMT to characterize the regional twist of a β -sheet. As an example for 1A12 sheet A, The AMT shows 11.98° for the close set, larger than that of the far set of 7.36° in Table 1. We observed similar pattern for β -sheet A of 1DTD that has eight β -strands. The close set (green lines in Fig. 5B) has an AMT of 10.75° , and the far set (red lines in Fig. 5B) has an AMT of 4.07° in Table 1. In all the eleven cases, we observed that the close set of β -traces shows larger AMT than the far set (Table 1 column 4 and 5).

Table 1: Average Minimum Twist angle of three sets of lines: the set of β -strand splines, two sets of β -traces.

ID ^a	Str ^b	AMT std ^c	AMT close ^d	AMT far ^e	AMT diff ^f
1B3A_B	3	27.22	9.76	4.69	5.07
1A12_A	4	27.24	11.98	7.36	4.62
1AOP_A	5	24.12	2.81	1.28	1.53
1AKY_A	5	23.40	7.54	6.16	1.38
1ATZ_A	6	16.70	16.62	12.37	4.25
1CHD_SH1	7	12.97	6.40	1.79	4.61
1ELU_B	7	21.02	7.76	3.88	3.88
1D5T_A	7	16.34	10.88	8.09	2.79
1DTD_A	8	13.21	10.75	4.07	6.68
1QNA_C	9	22.99	16.62	8.12	8.50
EMD1733-3C91_H_Q	5	21.69	9.70	8.97	0.73

a. the PDB/EMDB and sheet ID of the protein;

b. number of strands in a β -sheet;

c. average of minimum twist angle (in degrees) for splines c β -strands;

d. average minimum twist angle (in degrees) for a close set;

e. average minimum twist angle (in degrees) for a far set;

f. difference (in degrees) between d and e.

The difference of AMT between a close set and a far set is shown in column 6 of Table 1. We analyze three cases that are shown in Fig. 1, Fig. 3, and Fig. 5. The difference of AMT shows 0.73° (EMD1733 chain H Sheet Q), 4.62° (1A12_A), and 6.68° (1DTD_A).

A visual examination of the three cases show that the smallest difference in AMT of 0.73° corresponds to similar orientation and position between the close set and far set (Fig. 1D). In this case, the two sets are mostly different by a translation with the orientations being similar. The largest difference in AMT of 6.68° correlates with the large difference between the close set and the far set, particularly at major region where longest lines are located (Fig. 5B). By relating the difference of AMT (column 6 of Table 1) with visual examination for the three cases (shown in Fig. 1, Fig. 3, and Fig. 5), we observed that the measurement of AMT was able to distinguish the level of closeness between sets across different cases.

3.3 Twist characterization for a β -sheet image of a cryo-EM density map

The cryo-EM density map (EMD-1733) was downloaded from EMDB, and the corresponding structure 3C91 (PDB ID) was downloaded from PDB. Chain H of 3C91 was used as a mask to isolate the density region corresponding to the chain in Fig. 1A. SSETracer [19] was used to detect the location of α -helices (red lines in Fig. 1B) and β -sheet (yellow in Fig. 1B). Since the two β -sheet regions detected by SSETracer are naturally separate, we used one of them (sheet Q of chain H) in this case study. Iterative Bézier fitting was performed to derive a surface (blue in Fig. 1C) for the 3D image of sheet Q [15]. Two sets of β -traces were constructed on the Bézier surface. One set of β -traces (green lines in Fig. 1D), referred to the close set, appears to align better with the atomic structure (ribbon in Fig. 1D) than the other set (red lines in Fig. 1D), referred to the far set. We tested if the two sets can be distinguished using the estimation we defined in (3). We observed that by quantifying twist at the stable region of a pair of β -traces, a close set of β -traces shows bigger twist than a far set. The close set has an AMT of 9.70° , slightly larger than 8.97° for the far set Table 1. A visual examination between the two sets show similar orientation between the two sets in Fig. 1D, yet AMT detected slight difference between the two. Since β -strand splines were derived directly from atomic structure of β -strands, they are expected to align with β -strands quite accurately. A measure of the AMT for β -strand splines shows 21.69° in Table 1. We observe that the AMT of β -strand splines is even larger than that of the close set. This case study suggests that AMT might be suitable for distinguishing a set of β -traces that aligns well with the atomic structure of β -strands.

4 CONCLUSIONS

β -strands are naturally twisted. Given the 3D image of a β -sheet, there are many ways to generate β -traces from the image to satisfy the inter-trace distance of 4.5 to 5Å. Multiple sets of β -traces need to be generated using a reference surface that captures the overall shape of the 3D image. One set of β -traces may differ from another set mostly by orientation or by translation. We investigated a method to distinguish a close set from a far set, and observed from 11 test cases that the AMT is larger for a close set of β -traces than for a far set. This suggests that AMT can be a potential measure to identify most accurate β -traces from a pool

of candidates. Additional measures may be combined for accurate detection of β -strands from medium-resolution images of β -sheets.

ACKNOWLEDGMENTS

The work in this paper was supported in part by NSF DBI-1356621, NIH R01-GM062968, Modeling and Simulation Scholarship (to TI) and Honors College Undergraduate Research Scholarship (to MP) of Old Dominion University.

REFERENCES

- [1] ABEYSINGHE, S., JU, T., BAKER, M.L., and CHIU, W., 2008. Shape modeling and matching in identifying 3D protein structures. *Computer Aided-design* 40, 708-720.
- [2] AL NASR, K., RANJAN, D., ZUBAIR, M., CHEN, L., and HE, J., 2014. Solving the secondary structure matching problem in cryo-EM de novo modeling using a constrained K-shortest path graph algorithm. *IEEE/ACM Trans Comput Biol Bioinform* 11, 2, 419-429.
- [3] AL NASR, K., RANJAN, D., ZUBAIR, M., and HE, J., 2011. Ranking valid topologies of the secondary structure elements using a constraint graph. *J Bioinform Comput Biol* 9, 3 (Jun), 415-430.
- [4] BAKER, M.L., JU, T., and CHIU, W., 2007. Identification of secondary structure elements in intermediate-resolution density maps. *Structure* 15, 1 (Jan), 7-19.
- [5] BARTESAGHI, A., MERK, A., BANERJEE, S., MATTHIES, D., WU, X., MILNE, J.L.S., and SUBRAMANIAM, S., 2015. 2.2 Å resolution cryo-EM structure of β -galactosidase in complex with a cell-permeant inhibitor. *Science* 348, 6239, 1147-1151.
- [6] BISWAS, A., RANJAN, D., ZUBAIR, M., ZEIL, S., NASR, K.A., and HE, J., 2016. An Effective Computational Method Incorporating Multiple Secondary Structure Predictions in Topology Determination for Cryo-EM Images. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 14, 3, 578-586.
- [7] CHOTHIA, C., 1973. Conformation of twisted beta-pleated sheets in proteins. *J Mol Biol* 75, 2 (Apr 5), 295-302.
- [8] DAL PALU, A., HE, J., PONTELLI, E., and LU, Y., 2006. Identification of Alpha-Helices from Low Resolution Protein Density Maps. *Proceeding of Computational Systems Bioinformatics Conference(CSB)*, 89-98.
- [9] HO, B.K. and CURMI, P.M., 2002. Twist and shear in beta-sheets and beta-ribbons. *J Mol Biol* 317, 2 (Mar 22), 291-308.
- [10] HRYC, C.F., CHEN, D.-H., AFONINE, P.V., JAKANA, J., WANG, Z., HAASE-PETTINGELL, C., JIANG, W., ADAMS, P.D., KING, J.A., SCHMID, M.F., and CHIU, W., 2017. Accurate model annotation of a near-atomic resolution cryo-EM map. *Proceedings of the National Academy of Sciences* 114, 12 (March 21, 2017), 3103-3108.
- [11] JIANG, W., BAKER, M.L., LUDTKE, S.J., and CHIU, W., 2001. Bridging the information gap: computational tools for intermediate resolution structure interpretation. *J Mol Biol* 308, 5 (May), 1033-1044.
- [12] LAWSON, C.L., BAKER, M.L., BEST, C., BI, C., DOUGHERTY, M., FENG, P., VAN GINKEL, G., DEVKOTA, B., LAGERSTEDT, I., LUDTKE, S.J., NEWMAN, R.H., OLDFIELD, T.J., REES, I., SAHNI, G., SALA, R., VELANKAR, S., WARREN, J., WESTBROOK, J.D., HENRICK, K., KLEYWEGT, G.J., BERMAN, H.M., and CHIU, W., 2011. EMDatabank.org: unified data resource for CryoEM. *Nucleic Acids Res* 39, suppl 1 (Jan), D456-464.
- [13] LI, R., SI, D., ZENG, T., JI, S., and HE, J., 2016. Deep convolutional neural networks for detecting secondary structures in protein density maps from cryo-electron microscopy. In *2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 41-46.
- [14] PETTERSEN, E.F., GODDARD, T.D., HUANG, C.C., COUCH, G.S., GREENBLATT, D.M., MENG, E.C., and FERRIN, T.E., 2004. UCSF Chimera—A visualization system for exploratory research and analysis. *Journal of Computational Chemistry* 25, 13, 1605-1612.
- [15] POTEAT, M. and HE, J., 2017. An Iterative Bezier Method for Fitting Beta-sheet Component of a Cryo-EM Density Map. *Mol. Based Math. Biol* 5, 31-39.
- [16] RICHARDSON, J.S. and RICHARDSON, D.C., 1985. Interpretation of electron density maps. *Methods Enzymol* 115, 189-206.
- [17] RUSU, M. and WRIGGERS, W., 2012. Evolutionary bidirectional expansion for the tracing of alpha helices in cryo-electron microscopy reconstructions. *J Struct Biol* 177, 2 (Feb), 410-419.
- [18] SCHRÖDER, G.F., BRUNGER, A.T., and LEVITT, M., 2007. Combining Efficient Conformational Sampling with a Deformable Elastic Network Model Facilitates Structure Refinement at Low Resolution. *Structure* 15, 12, 1630-1641.
- [19] SI, D. and HE, J., 2013. 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*(September 22-25), 764-770.
- [20] SI, D. and HE, J., 2014. Orientations of beta-strand traces and near maximum twist. In *Proceedings of the Proceedings of the 5th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics* (Newport Beach, California 2014), ACM, 2660835, 690-694.
- [21] SI, D. and HE, J., 2014. Tracing beta-strands using strandtwister from cryo-EM density maps at medium resolutions. *Structure* 22, 11, 22(11):1665-1676.
- [22] SI, D. and HE, J., 2017. Modeling Beta-Traces for Beta-Barrels from Cryo-EM Density Maps. *Biomed Res Int* 2017, 1793213.
- [23] SI, D., JI, S., AL NASR, K., and HE, J., 2012. A machine learning approach for the identification of protein secondary structure elements from electron cryo-microscopy density maps. *Biopolymers* 97, 9 (Sep), 698-708.
- [24] TRABUCO, L.G., VILLA, E., MITRA, K., FRANK, J., and SCHULTEN, K., 2008. Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. *Structure* 16, 5 (May), 673-683.
- [25] WRIGGERS, W. and BIRMANNS, S., 2001. Using situs for flexible and rigid-body fitting of multiresolution single-molecule data. *J Struct Biol* 133, 2-3 (Feb-Mar), 193-202.
- [26] ZHAO, H., LI, K., LYNN, A.Y., ARON, K.E., YU, G., JIANG, W., and TANG, L., 2017. Structure of a headful DNA-packaging bacterial virus at 2.9 Å resolution by electron cryo-microscopy. *Proceedings of the National Academy of Sciences* 114, 14 (April 4, 2017), 3601-3606.