

Research Article

Open Access

Alice Poteat and Jing He*

An Iterative Bézier Method for Fitting Beta-sheet Component of a Cryo-EM Density Map

DOI 10.1515/mlbmb-2017-0003

Received October 6, 2016; accepted April 3, 2017

Abstract: Cryo-electron microscopy (Cryo-EM) is a powerful technique to produce 3-dimensional density maps for large molecular complexes. Although many atomic structures have been solved from cryo-EM density maps, it is challenging to derive atomic structures when the resolution of density maps is not sufficiently high. Geometrical shape representation of secondary structural components in a medium-resolution density map enhances modeling of atomic structures. We compare two methods in producing surface representation of the β -sheet component of a density map. Given a 3-dimensional volume of β -sheet that is segmented from a density map, the performance of a polynomial fitting was compared with that of an iterative Bézier fitting. The results suggest that the iterative Bézier fitting is more suitable for β -sheets, since it provides more accurate representation of the corners that are naturally twisted in a β -sheet.

1 Introduction

Cryo-electron microscopy (cryo-EM) is a biophysical technique to produce electron density maps of large molecular assemblies [12, 18, 31, 33, 34]. A density map is a 3-dimensional image in which each voxel is associated with a density value that represents the local density of electrons. Many molecular complexes, such as ribosomes and viruses, have been resolved to atomic resolutions using this technique [7, 20, 32]. At high resolutions about 3Å, details of amino acids are resolved, and atomic structures can be derived. However, it is challenging to obtain high-resolution 3D images for many biological specimen due to the nature of biological samples. At medium resolutions, such as 5–10Å, amino acid details are not resolved, and it is often not possible to distinguish the chain of a protein in the image. What is possible to be distinguished from medium-resolution images includes rough location of secondary structures, such as α -helices, β -sheets [8, 14, 19, 22, 26, 27, 29, 30], and possible connections among secondary structures [4, 21]. An α -helix appears as a cylinder and a β -sheet appears as a thin layer of density image at medium resolutions. Various computational methods have been developed to segment secondary structures from cryo-EM images, such as *Helixhunter*, *SSEhunter*, *HelixTracer*, *SSElearner* *Voltrac*, and *SSETracer* [8, 14, 19, 26, 27, 29]. In addition to secondary structure detection, it is possible to match secondary structures detected from the cryo-EM density map to those predicted from protein amino acid sequence, in order to derive the backbone trace of the protein [1–3, 5, 6, 9–11, 23].

A β -sheet is composed of multiple β -strands. Figure 1 shows an example of a 7-stranded β -sheet. Knowing the position of a β -sheet without knowing the position of β -strands is often not sufficient to model the 3-dimensional structure of a protein, since each β -strand, not a β -sheet, corresponds to a segment of the protein sequence. The spacing between two neighboring β -strands is 4.5–5Å, too small to be resolved in an image with

*Corresponding Author: Jing He: Department of Computer Science, Old Dominion University, Norfolk, VA, 23529, USA, E-mail: jhe@cs.odu.edu

Alice Poteat: Department of Computer Science, Old Dominion University, Norfolk, VA, 23529, USA

 © 2017 Alice Poteat and Jing He, licensee De Gruyter Open.

This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivs 3.0 License.

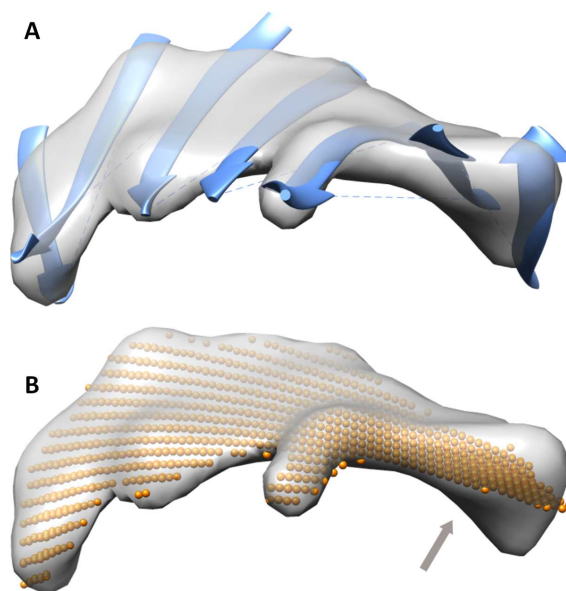


Figure 1: β -sheet, β -strands and β -sheet image. A β -sheet is composed of multiple β -strands, each shown as a ribbon in (A). An image (gray) of a β -sheet is superimposed with its atomic structure of β -sheet (ribbons) in (A) and a polynomial surface (golden dots) in (B). An example of the region where the surface does not fit well the image is marked with an arrow.

5–10Å resolution. In spite of the difficulty in distinguishing individual β -strands from a medium-resolution map, we previously showed that it is possible to predict the position of the β -strands of a β -sheet through the analysis of β -sheet twist [28]. A unique nature of a β -sheet is that each one is right-handed twisted [13]. *StrandTwister* utilizes this special character of a β -sheet to model the orientation of β -strands [28].

Given an atomic structure of a protein, a twist angle can be calculated along the peptide orientation and has been shown to have right-handed twist on a β -strand [13]. For a given atomic model of a β -sheet, the twist angle is often less than 40° [17, 25]. However, when precise location of atoms is not available, as for a β -sheet image, precise calculation of a twist angle is challenging. We measured a twist angle that is represented by two neighboring β -traces, each of which is a central line of the β -strand [28]. In such a case, accurate calculation of twist angles depends on accurate modeling of a β -sheet.

Since the density image of a β -sheet appears as a thin layer, two previous methods exist to represent a β -sheet from its 3D image. Thinning and pruning is an image processing technique, and has been used to produce a thin surface from a 3D image at the β -sheet region [21]. This method is an effective way to detect a β -sheet, but the resulting surface is not smooth enough for direct calculation of twist angles. A polynomial fitting method was used in *StrandTwister* [27, 28]. The polynomial was selected using least square fitting of the voxels in the β -sheet image. Although the surface appears to fit well in most part of the image, it does not fit accurately for a large β -sheet, particularly at the corner regions (arrow in Figure 1). It is possible that a polynomial is not a sufficient model to represent a twisted surface.

In this paper, we explore a general mathematical model to represent the surface of any given β -sheet, large or small. We propose to use an iterative Bézier surface to fit a β -sheet image. We show that the proposed model improves fitting accuracy, particularly for the corners of a β -sheet.

2 Method

The surface model is derived iteratively using a Bézier model. At each iteration, a set of control points are identified and used to guide the surface model. The overall process includes four major steps: (1) initial plane

fitting; (2) generating initial bounding box of the β -sheet image; (3) iterative fitting of Bézier surfaces; (4) definition of a polygonal boundary of the surface.

2.1 Pre-processing

Before iterative fitting of Bézier surfaces, initial control points need to be identified and the initial boundary of the image need to be outlined. Since a β -sheet is a thin layer of density, a simple geometrical representation of a β -sheet is a plane. We utilized a plane to assist the initialization of the boundary and control points. To search for the best-fit plane of the β -sheet image, we define the plane in (1) and (2) and search for the best fit plane using (3).

$$\begin{aligned} A &= \sin \alpha \cos \beta \\ B &= \sin \alpha \sin \beta \\ C &= \cos \alpha \\ D &= \Delta \end{aligned} \quad (1)$$

$$Ax + By + Cz + D = 0 \quad (2)$$

Where $p = (x, y, z)$ is a point in 3-dimensional space, α and β are angles variable between 0 and 2π . Δ is a real number. P is the set of all points in the β -sheet image.

$$\min_{\alpha, \beta, \Delta} \sum_{p \in P} (Ax + By + cz + D)^2 \quad (3)$$

In principle a least square fitting of a plane will determine the best-fit plane of the β -sheet image, but we implemented an approach simply varying α , β , and Δ based on the feedback of the fitting error in (3). It appears to work in all test cases. Note that the accuracy of the initial plane is not critical. The optimization stops when a desired threshold of error is reached.

Once a plane was found, the initial boundary and control points were defined by projecting the β -sheet image to the plane to create a 2-dimensional image. A rectangle was identified on the 2-dimensional image using Deming regression [15], a heuristic method to approximate a bounding box. This heuristic method is in $O(n)$ where n is the number of points to be bounded. Once the bounding box is defined, the four corner points are projected back to the β -sheet image to define the initial four control points in 3-dimensional space.

2.2 Iterative Bézier Surface Fitting

Bézier surface is a parametric surface S defined as in (4).

$$S(u, v) = \sum_{i=0}^n \sum_{j=0}^m B_i^n(u) B_j^m(v) \mathbf{C}_{i,j} \quad (4)$$

In which $\mathbf{C}_{i,j}$ is a control point in 3-dimensional space, $i = 0, \dots, n; j = 0, \dots, m$. A control point is the position of a selected voxel of the β -sheet image through a pre-processing step. $B_i^n(u)$ is defined as in (5).

$$B_i^n(u) = \binom{n}{i} u^i (1-u)^{n-i} \quad (5)$$

Using the initial four control points derived in the pre-processing step, an initial surface is decided by searching for a Bézier surface that generates the smallest fitting error evaluated as Least Squared Orthogonal Projection distance in (6).

$$\text{RMSE}(P, S) = \sqrt{\frac{1}{|P|} \sum_{p \in P} (d(p, \bar{p}))^2} \quad (6)$$

p is a point on the β -sheet image, and $\bar{p}$ is a point on surface S with the shortest distance from p . $d(p, \bar{p})$ is the distance between point p and $\bar{p}$. For a general surface, calculating point projection onto a surface is a non-trivial optimization problem. We approximate the closest point using a fast local optimization. Initially a set of points were uniformly distributed on the surface. Given a voxel in the β -sheet image, its closest point on the surface is searched in the local neighborhood of an initial point using gradient descent method. To search for the surface with the minimal orthogonal distance, the position of control points were searched in a local neighborhood.

The iterative step of fitting Bézier surfaces consists of adding more control points and regenerating approximately best-fit surfaces. The number of control points changed from 2×2 to 3×3 and 4×4 , as examples. The number of control points nm is chosen to balance between precision and performance. Figure 2A was generated with $n = 6$ and $m = 10$. Because each point projection takes a constant amount of time, our heuristic allows us to compute RMSE of orthogonal in $O(n)$ time.

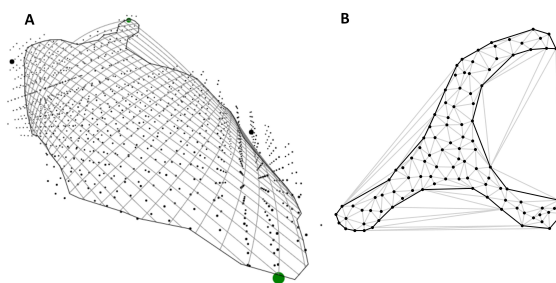


Figure 2: Polygonal boundary and Delaunay triangulation. A: An example of the polygonal boundary (thick lines) of the surface for protein 1CHD (PDB ID) β -sheet 1. Larger dots are control points of Bézier surface; fine dots are voxels of the β -sheet image. B: An example of the long edges (gray lines) that are pruned after Delaunay triangulation using χ algorithm [16]. Dark lines are the resulting boundary of the point set.

2.3 Polygonal Boundary of the β -sheet Surface

A Bézier surface is traditionally defined as a patch, with surface variables u and v ranging from 0 to 1. However, a β -sheet will not have the shape of a square patch. Thus, a way to outline the β -sheet region on the Bézier surface is needed. We first projected the voxels of the β -sheet image onto the optimized Bézier surface. We then calculated the Delaunay triangulation of the point set using the code provided at <https://github.com/ironwallaby/delaunay>.

After the triangles are generated using Delaunay, long edges at the boundary are iteratively pruned using χ algorithm until a desired maximum edge length is reached [16] (Figure 2). A polygonal boundary is generated this way on the Bézier surface to outline the boundary of the β -sheet. Note that the Delaunay triangulation and the pruning using χ algorithm has $O(n \lg n)$ time, since the algorithm uses a relational map to represent and access the triangulation efficiently. An example of the surface boundary derived for a β -sheet of protein 1CHD (PDB ID) is shown in Figure 2.

2.4 Polynomial Fitting of a β -sheet Image

Previous studies have shown that an order 3 polynomial can be an approximation of a beta-sheet. We summarize the polynomial fitting here. More details can be found in [27, 28]. Given an isolated 3D image of a beta-sheet, it was first aligned with the Z-axis of the coordinate system using the normal vector that passes through the center of the image. Least square fitting of the image using (7) was performed to determine parameters A to J. Finally, the beta-sheet image and the polynomial surface were rotated back to the original

position.

$$z = Ax^3 + By^3 + Cx^2 + Dy^2 + Ex^2y + Fy^2x + Gxy + Hx + Iy + J \quad (7)$$

Fitting error was estimated using the RMSE between the image and the surface in (8).

$$RMSE(P, S) = \sqrt{\frac{1}{|P|} \sum_{p \in P} (z_i - \bar{z}_i)^2} \quad (8)$$

Where z_i is the Z coordinate of i -th voxel of the β -sheet image P , and $\bar{z}_i$ is the Z coordinate of its corresponding point on surface S . Note that the image and the surface are registered using the normal vector of image P and the Z-axis of surface S .

3 Results

Iterative Bézier surface fitting was tested using two sets of β -sheet images. The first set contains eight simulated β -sheet images (see 3.1) and the second contains five β -sheet images extracted from cryo-EM density map EMD1237 (see 3.2). We compared two surface fitting methods – the polynomial fitting and the iterative Bézier fitting. In evaluation of the performance, we compared the root-mean-square-error (RMSE) between the β -sheet image and the surface model. Since the polynomial fitting was implemented in *StrandTwister*, the RMSE calculation used in *StrandTwister*, as in (8), was directly used in comparison. The RMSE between the β -sheet image and the Bézier surface was calculated after each voxel in the image is registered with its closest point on the surface model. We also performed visual inspection by superposition of the atomic structure, the image, and the surface model of the β -sheet.

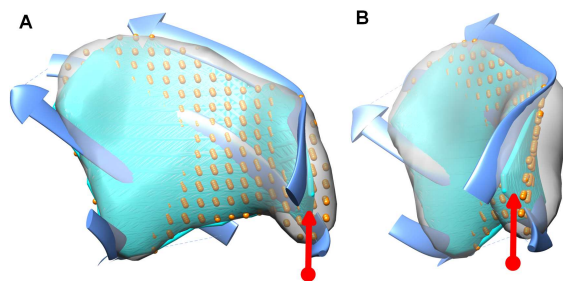


Figure 3: Bézier surface and polynomial surface fitted in the β -sheet image of 1A12 (PDB ID) chain A sheet A. The image (gray), atomic structure (blue ribbon), the Bézier surface (cyan), and the polynomial surface (golden dots) are superimposed. The most prominent visible difference is seen at the lower-right corner, at which the Bézier surface is seen to fit the β -strands more closely (arrow). (A) and (B) are two views of the same object.

3.1 Fitting Simulated β -sheet Images

For each of the eight simulated test cases, the atomic structure of a β -sheet was used to simulate the β -sheet image. The atomic structures were downloaded from the Protein Data Bank, and all atoms on a β -sheet were used to simulate the 3-dimensional image of the β -sheet to 10 Å resolution using *molmap* function in Chimera [24]. To guide the surface fitting towards the dense points of the image, a density threshold was used to generate the β -sheet image. The same threshold was used in each test case for both the polynomial and Bézier fitting. An example of the simulated β -sheet image (gray region in Figure 4) is superimposed with the atomic structure (ribbon in Figure 4). Although most of the polynomial surface (golden dots) is fairly close to the Bézier surface (blue), differences are seen mostly at the corner or the edge of the image (Figure 4). Bézier

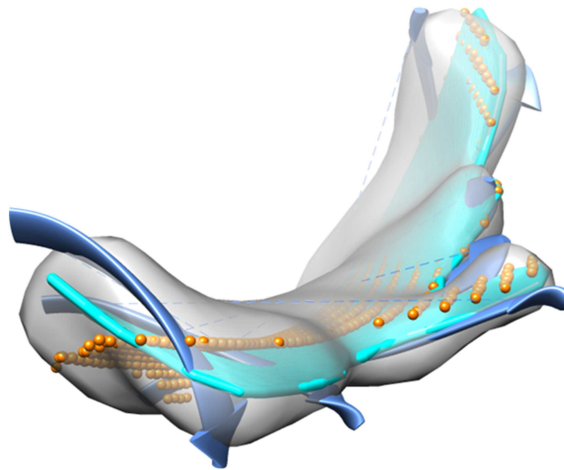


Figure 4: Bézier surface and the polynomial surface for β -sheet 1 of protein 1CHD (PDB ID). Figure 4 and Figure 3 use the same color annotation.

surface appears to be closer to the atomic structure (ribbon) at the corner in this case than the polynomial surface.

Each of the eight β -sheets contains three to ten β -strands. We observed that the RMSE error between the image and the Bézier surface (column 3 of Table 1) is consistently smaller than that for the polynomial surface (column 4 of Table 1). As an example, β -sheet 1A12_A has four strands. The RMSE between the image and Bézier surface is 1.01 Å (Table 1). The RMSE between the β -sheet image and the polynomial surface is 1.15 Å. The major difference in fitting appears to be located at the corner region (Figure 3). Bézier surface represents more accurately at the corner of a β -sheet, and that is expected since the shape of the corners will be mostly amplified by the twist of a β -sheet. Our test suggests that the iteratively fitted Bézier surface is more accurate than the polynomial surface. Although this case involves a β -sheet of only four strands, one of the β -strands has a significant kink, as seen for some β -sheets. It is possible that a polynomial is not the best way to fit such a case. We noticed that a Bézier surface is more flexible to represent such a shape.

Table 1: Comparison of surface fitting error.

PDB ID ^a	#Strd ^b	BzFit ^c	PolyFit ^d	Thrsh ^e	#Voxels ^f
1IG0_B	10	1.19	2.00	0.33	1868
1QNA_C	9	1.07	1.45	0.33	1764
1DTD_A	8	1.64	2.13	0.29	2621
1CHD_SH1	7	0.99	1.53	0.33	1106
1ATZ_A	6	0.99	1.21	0.35	1117
1AKY_A	5	1.37	1.49	0.29	1102
1A12_A	4	1.01	1.15	0.34	644
1A12_B	3	0.87	1.13	0.34	327

a. The PDB and sheet ID;

b. The number of β -strands;

c. The RMSE (root-mean-square-error, in Å) for the iterative Bézier surface fitting;

d. The RMSE for the polynomial fitting;

e. The density cut-off threshold;

f. The total number of voxels included in each fit;

Table 2: Fitting Errors of β -sheets in Cryo-EM density maps.

PDB ID ^a	#Strd ^b	BzFit ^c	PolyFit ^d	Thrsh ^e	#Voxel ^f
2GSY_A	4	1.29	2.34	0.18	362
2GSY_B	5	1.61	2.19	0.16	863
2GSY_C	6	1.65	2.34	0.15	1239
2GSY_E	4	1.49	1.80	0.12	1181
2GSY_G	4	1.45	2.21	0.14	1036

a. The PDB and sheet ID;

b. The number of β -strands;

c. The RMSE (root-mean-square-error, in Å) for the spline fitting;

d. The RMSE for the polynomial fitting;

e. The density cut-off threshold;

f. The total number of voxels included in each fit;

Similar result is observed for 1CHD-SH1 that has seven β -strands. In this case, the β -sheet is larger than previous case 1A12_A. We observe a difference of 0.54 Å RMSE when two different surfaces are used in calculating RMSE. Although some of the difference might be resulted from two slightly different ways to register a voxel with its closest point on the surface, some of the RMSE difference is expected to be from the different fitting at the corners, since visual inspection clearly shows difference at the corners (Figure 4). It is possible that a flexible surface is needed to fit a larger β -sheet image, particularly when the corners of a β -sheet appear highly-twisted, presumably from the accumulation of twist from every pair of neighboring β -strands of the β -sheet. We observe that the Bézier surface is better in representing the twist of a β -sheet and provides a more accurate fitting of the β -sheet image (Figure 4).

3.2 Fitting Cryo-EM β -sheet Images

The performance of iterative Bézier surface fitting was further tested using five β -sheet volumes of an experimentally derived cryo-EM density map. Cryo-EM density map (EMD-1237, 7.2 Å resolution) was aligned with atomic structure 2GSY (PDB ID) and was downloaded in September 2011. Chain A of 2GSY was used to mask the density region corresponding to chain A. *SSETracer* was then used to identify α -helices and β -sheets. Chain A of 2GSY has seven β -sheets, two of them are small β -sheets with 2 of 3 β -strands. The five larger ones (with four to six β -strands) were segmented out using *SSETracer* and were used in the test. We observed similar results as for the test using simulated density maps. The RMSE fitting error of iterative Bézier surface is consistently lower than that of the polynomial surface (Table 2). We also observed that the fitting error is generally slightly higher for cryo-EM volumes than for simulated volumes (Table 1 and Table 2). A possible reason is that cryo-EM maps generally have lower quality than simulated maps.

4 Discussions

Deriving β -strands from β -sheet density volume requires accurate representation of the β -sheet surface. It was shown in *StrandTwister* method that candidate sets of β -strand traces can be drawn on the surface, and the correct orientation of β -strand traces can be identified based on maximum twist angle measurement. Although a polynomial surface fit well in the central region of a β -sheet, it does not fit well at the corners. The inaccurate surface representation at the corners may result in inaccuracy of β -traces at the corner regions. We show that an iterative fitting of a Bézier surface provides more flexible surface representation for the twisted corners. We expect that the resulting β -traces on the Bézier surface are more accurately predicted than those on a polynomial surface.

5 Conclusions

Cryo-EM technique has recently been used to produce 3-dimensional images for many large molecular complexes. When the resolution of 3-dimensional images is not sufficient to derive atomic structure directly from the image, it is still challenging to interpret the image. At medium resolution of 5–8 Å, secondary structures such as α -helices and β -sheets show their characteristic shape. In order to model β -strands from a β -sheet image, we propose an iterative fitting method using Bézier surface. This method is compared to the polynomial fitting that we previously used in *StrandTwister* [28]. A test involving eight simulated β -sheet images of various sizes shows that the fitting error RMSE is reduced for all cases. Iterative Bézier surfaces is capable of representing twisted corners of a β -sheet image more precisely. Our future work will involve the development of a method to predict β -strand with the enhanced accuracy than *StrandTwister*.

Acknowledgement: This work is supported in part by NSF Bio-DBI-1356621, NIH R01 GM062968, and Undergraduate Research Scholarship (to Alice Poteat) of Honors College of Old Dominion University.

References

- [1] Abeysinghe, S., Ju, T., Baker, M.L., Chiu, W., 2008. Shape modeling and matching in identifying 3D protein structures. *Computer Aided-design* 40, 708-720.
- [2] Al Nasr, K., He, J., 2016. Constrained cyclic coordinate descent for cryo-EM images at medium resolutions: beyond the protein loop closure problem. *Robotica* 34, 1777-1790.
- [3] Al Nasr, K., Ranjan, D., Zubair, M., He, J., 2011. Ranking valid topologies of the secondary structure elements using a constraint graph. *Journal of bioinformatics and computational biology* 9, 415-430.
- [4] Al Nasr, K., Liu, C., Rwebangira, M., Burge, L., He, J., 2013. Intensity-based skeletonization of CryoEM gray-scale images using a true segmentation-free algorithm. *IEEE/ACM transactions on computational biology and bioinformatics* / IEEE, ACM 10, 1289-1298.
- [5] Al Nasr, K., Ranjan, D., Zubair, M., Chen, L., 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 transactions on computational biology and bioinformatics* / IEEE, ACM 11, 419-429.
- [6] Al Nasr, K., Chen, L., Si, D., Ranjan, D., Zubair, M., He, J., 2012. Building the initial chain of the proteins through de novo modeling of the cryo-electron microscopy volume data at the medium resolutions, *Proceedings of the ACM Conference on Bioinformatics, Computational Biology and Biomedicine*. ACM, Orlando, Florida, 490-497.
- [7] Anger, A.M., Armache, J.P., Berninghausen, O., Habeck, M., Subklewe, M., Wilson, D.N., Beckmann, R., 2013. Structures of the human and Drosophila 80S ribosome. *Nature* 497, 80-85.
- [8] Baker, M.L., Ju, T., Chiu, W., 2007. Identification of secondary structure elements in intermediate-resolution density maps. *Structure* 15, 7-19.
- [9] Baker, M.L., Abeysinghe, S.S., Schuh, S., Coleman, R.A., Abrams, A., Marsh, M.P., Hryc, C.F., Ruths, T., Chiu, W., Ju, T., 2011. Modeling protein structure at near atomic resolutions with Gorgon. *Journal of structural biology* 174, 360-373.
- [10] Biswas, A., Ranjan, D., Zubair, M., He, J., 2015. A Dynamic Programming Algorithm for Finding the Optimal Placement of a Secondary Structure Topology in Cryo-EM Data. *Journal of Computational Biology* 22, 837-843.
- [11] Biswas, A., Ranjan, D., Zubair, M., Zeil, S., Nasr, K.A., 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* PP, 1-1.
- [12] Chiu, W., Baker, M.L., Jiang, W., Dougherty, M., Schmid, M.F., 2005. Electron cryomicroscopy of biological machines at sub-nanometer resolution. *Structure* 13, 363-372.
- [13] Chothia, C., 1973. Conformation of twisted beta-pleated sheets in proteins. *J Mol Biol* 75, 295-302.
- [14] Dal Palu, A., He, J., Pontelli, E., Lu, Y., 2006. Identification of Alpha-Helices from Low Resolution Protein Density Maps. *Proceeding of Computational Systems Bioinformatics Conference(CSB)*, 89-98.
- [15] Deming, W.E., 1980. The statistical procedure in the SENIC Project. *American journal of epidemiology* 111, 470-471.
- [16] Duckham, M., Kulik, L., Worboys, M., Galton, A., 2008. Efficient generation of simple polygons for characterizing the shape of a set of points in the plane. *Pattern Recogn.* 41, 3224-3236.
- [17] Ho, B.K., Curmi, P.M., 2002. Twist and shear in beta-sheets and beta-ribbons. *J Mol Biol* 317, 291-308.
- [18] Hryc, C.F., Chen, D.H., Chiu, W., 2011. Near-Atomic-Resolution Cryo-EM for Molecular Virology. *Curr Opin Virol* 1, 110-117.

- [19] Jiang, W., Baker, M.L., Ludtke, S.J., Chiu, W., 2001. Bridging the information gap: computational tools for intermediate resolution structure interpretation. *J Mol Biol* 308, 1033-1044.
- [20] Jiang, W., Baker, M.L., Jakana, J., Weigele, P.R., King, J., Chiu, W., 2008. Backbone structure of the infectious epsilon15 virus capsid revealed by electron cryomicroscopy. *Nature* 451, 1130-1134.
- [21] Ju, T., Baker, M.L., Chiu, W., 2007. Computing a family of skeletons of volumetric models for shape description. *Comput Aided Des* 39, 352-360.
- [22] Kong, Y., Ma, J., 2003. A structural-informatics approach for mining beta-sheets: locating sheets in intermediate-resolution density maps. *J Mol Biol* 332, 399-413.
- [23] Lindert, S., Alexander, N., Wotzel, N., Karaka, M., Stewart, P.L., Meiler, J., 2012. EM-Fold: De Novo Atomic-Detail Protein Structure Determination from Medium-Resolution Density Maps. *Structure* 20, 464-478.
- [24] Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E., 2004. UCSF Chimera—A visualization system for exploratory research and analysis. *Journal of Computational Chemistry* 25, 1605-1612.
- [25] Richardson, J.S., 1981. The anatomy and taxonomy of protein structure. *Advances in protein chemistry* 34, 167-339.
- [26] Rusu, M., Wriggers, W., 2012. Evolutionary bidirectional expansion for the tracing of alpha helices in cryo-electron microscopy reconstructions. *Journal of structural biology* 177, 410-419.
- [27] Si, D., 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*, 764-770.
- [28] Si, D., He, J., 2014. Tracing beta-strands using strandtwister from cryo-EM density maps at medium resolutions. *Structure* 22, 22(11):1665-1676.
- [29] Si, D., Ji, S., Al Nasr, K., He, J., 2012. A machine learning approach for the identification of protein secondary structure elements from electron cryo-microscopy density maps. *Biopolymers* 97, 698-708.
- [30] Zeyun, Y., Bajaj, C., 2008. Computational Approaches for Automatic Structural Analysis of Large Biomolecular Complexes. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 5, 568-582.
- [31] Zhang, X., Jin, L., Fang, Q., Hui, W.H., Zhou, Z.H., 2010. 3.3 angstrom Cryo-EM Structure of a Nonenveloped Virus Reveals a Priming Mechanism for Cell Entry. *Cell* 141, 472-482.
- [32] Zhang, X., Ge, P., Yu, X., Brannan, J.M., Bi, G., Zhang, Q., Schein, S., Zhou, Z.H., 2013. Cryo-EM structure of the mature dengue virus at 3.5-A resolution. *Nature structural & molecular biology* 20, 105-110.
- [33] Zhou, Z.H., 2008. Towards atomic resolution structural determination by single-particle cryo-electron microscopy. *Current Opinion in Structural Biology* 18, 218-228.
- [34] Zhou, Z.H., Dougherty, M., Jakana, J., He, J., Rixon, F.J., Chiu, W., 2000. Seeing the herpesvirus capsid at 8.5 Å. *Science* 288, 877-880.