2018

Quantification of Twist from the Central Lines of **β**-Strands

Tunazzina Islam

Michael Poteat

Jing He
Old Dominion University

Original Publication Citation

Quantification of Twist from the Central Lines of β-Strands

TUNAZZINA ISLAM, MICHAEL POTEAT, and JING HE

ABSTRACT

Since the discovery of right-handed twist of a β-strand, many studies have been conducted to understand the twist. Given the atomic structure of a protein, twist angles have been defined using atomic positions of the backbone. However, limited study is available to characterize twist when the atomic positions are not available, but the central lines of β-strands are. Recent studies in cryoelectron microscopy show that it is possible to predict the central lines of β-strands from a medium-resolution density map. Accurate measurement of twist angles is important in identification of β-strands from such density maps. We propose an effective method to quantify twist angles from a set of splines. In a data set of 55 pairs of β-strands from 11 β-sheets of 11 proteins, the spline measurement shows comparable results as measured using the discrete method that uses atomic positions directly, particularly in capturing twist angle change along a pair, different levels of twist among different pairs, and the average of twist angles. The proposed method provides an alternative method to characterize twist using the central lines of a β-sheet.

Keywords: β-strand, cryoelectron microscopy, image, protein structure, secondary structure, spline, twist.

1. INTRODUCTION

A β-SHEET IS A MAJOR SECONDARY STRUCTURE of many proteins. Although modeling atomic structures of β-sheets is often a challenging task, recent studies show that it is possible to predict central lines of β-strands from a medium-resolution (5–10 Å) density map obtained by cryoelectron microscopy (cryo-EM) by analyzing their twist (Si and He 2014). The atomic position of the β-strands is not visible in the density maps at such resolutions. Existing measurements of twist were established on positions of β-strand atoms, and hence are not applicable to central lines of β-strands. An accurate method to measure twist of β-strands directly from central lines of β-strands is needed.

Secondary structure elements such as helices and β-sheets are the most striking structural features visible in medium-resolution density maps (three-dimensional (3D) images). Various computational methods have been developed to detect helices and β-sheets (Jiang et al., 2001; Kong et al., 2004; Dal Palu et al., 2006; Baker et al., 2007; Zeyun and Bajaj, 2008), including the recent methods *SSEhunter*, *SSELearner*, *VolTrac*, and *SSETracer* (Baker et al., 2007; Rusu and Wriggers, 2012; Si et al., 2012; Si and He, 2013). A helix

---

Department of Computer Science, Old Dominion University, Norfolk, Virginia.

detected from 3D image can be represented by its axis, and a β-sheet can be represented as a set of voxels at the β-sheet region. A few methods have been recently developed to generate candidates of central lines of β-strands from simple β-sheet images (Si and He, 2014) and β-barrel images (Si and He, 2017) that are segmented from the density map of an entire protein. An accurate measurement of β-strand twist is important to distinguish the correct set of central lines from candidate sets.

A β-sheet is composed of at least two β-strands that are stabilized by interstrand hydrogen bonds. The direction of each β-strand is the direction of protein sequence. Given the atomic positions of a protein structure, two procedures were previously defined to calculate the twist of β-strands. The former measured the twist of hydrogen bonding direction or peptide planes about the axis parallel to the β-strand (Chothia, 1973; Richardson, 1981). This measurement involves single strand demonstrating a right-handed twist of about 5° per amino acid on average (Chothia et al., 1977). The latter measured the angle formed by two vectors along two neighboring strands, respectively (Quiocho et al., 1977), and showed a left-handed twist about an axis perpendicular to the strand direction. In this article, we propose a method to measure twist using the central lines of β-strands. We compare twist angles measured using atomic positions and central lines, respectively, and show that the proposed method is sensitive enough to capture the nature of twist along a pair of central lines.

2. METHODOLOGY

### 2.1. Measurement of a twist angle from atomic positions of two neighboring β-strands

We measured twist angles directly from atomic positions of the backbone described previously (Ho and Curmi, 2002). Let v_1 be a vector formed by two consecutive midpoints on β-strand β_1. A midpoint is the middle point of the peptide bond, the C–N bond, on the backbone. Let v_2 be a similar vector formed by consecutive midpoints of a neighboring β-strand β_2, that is the closest vector to v_1 in β_2. The twist angle is the acute angle formed by v_1 and v_2 (Ho and Curmi, 2002). Note that each angle involves three consecutive amino acids. Once the backbone of a protein is given, twist angles can be calculated for the hydrogen-bonded portion of a pair of β-strands (Murzin et al., 1994). We refer this measurement as the discrete (sparse) method in comparison with the method using splines (details in Section 2.2).

2.2. Measurement of twist angles from a pair of central lines of β-strands

When the 3D image of a β-sheet is available, central lines of β-strands can be derived from the image using computational methods such as StrandTwister (Si and He, 2014) and StrandRoller (Si and He, 2017). For the purpose of characterization of twist, a cubic spline generated from the midpoints used in the sparse method in Section 2.1 was used as a central line, instead of splines derived from the image of a β-sheet. Let S_1, S_2, ..., S_m, be the central lines of β-strands, and S_i and S_{i+1} are hydrogen bonded for $i = 1, ..., m-1$. Without the knowledge of hydrogen bond locations, a pairing segment needs to be identified for each pair of neighboring strands (Fig. 1A). This was done by taking each of the two ends of the splines and identifying the nearest point on the pairing spline. For simplicity, the shorter strand of the pair was used as the reference in dividing into equal-length vectors. We investigated two methods of selecting vectors of the same length in the pairing strand. Suppose that the pairing segment of the reference spline S_i is divided into n_i vectors of a specific length $(v_1^i, v_2^i, ..., v_{n_i}^i)$. Let $(v_1^{i+1}, v_2^{i+1}, ..., v_{n_i}^{i+1})$ be a tuple of vectors of the same length on the neighboring spline S_{i+1}, and let twist θ_j^i be the acute angle formed by v_j^i and v_j^{i+1} for $j = 1, ..., n_i$. For nearest-vector measurement, v_j^{i+1} is the nearest vector of v_j^i. For sequential-vector measurement, the beginning point of v_{j+1}^{i+1} is the ending point of v_j^{i+1}. Note that nearest-vector measurement allows two consecutive vectors to overlap as long as the nearest-vector principle is maintained. We investigated vector length of 0.7, 1.1, 1.5, 2.5, and 3.2 Å.

2.3. Averaged minimum twist

Both discrete and spline methods measure twist locally for individual β-strands. The level of twist is often different at different locations of the strand and for different β-strands. To characterize the level of twist for different β-sheets, we define in Equation (1) an average minimum twist (AMT), λ, for a set of splines. The idea of AMT is to characterize the region with the minimum twist for the two longest pairs of β-strands. In principle, all twist angles can be included in characterization, but we observed, in practice,
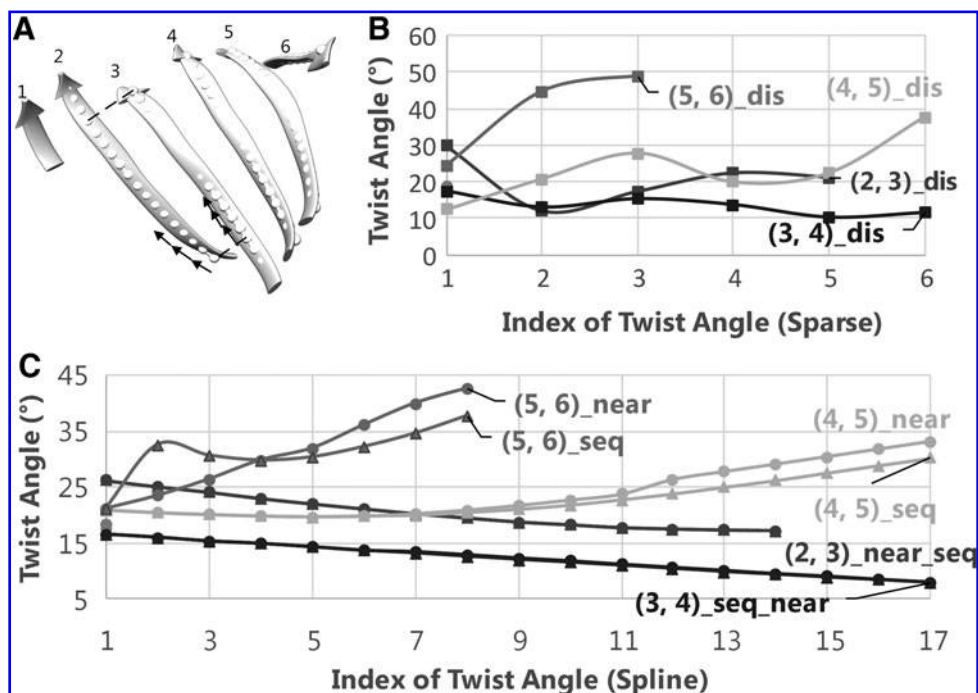
FIG. 1. Twist angles calculated from atomic positions and central lines of β-strand in protein 1ATZ (PDB ID) sheet A. **(A)** The atomic structure of β-strands shown as ribbons is superimposed with the splines (dots) of the β-strands that are sampled with a 1.1 Å vector length. The pairing segment (marked with dashed lines) and vectors are marked (black arrows). **(B)** Twist angles measured using the discrete method. **(C)** Twist angles measured from splines. Those angles measured using nearest-vector and sequential-vector methods are marked with circles and triangles, respectively. Pair (1,2) is colored dark in **(B, C)**. PDB, Protein Data Bank.

that those angles at the edge of a β-sheet are often not as accurate as those at the central region if splines are derived from cryo-EM images rather than the atomic structure. If (S_{i_1}, S_{i_1+1}) and (S_{i_2}, S_{i_2+1}) are the two longest pairs of splines, then the AMT λ is defined in Equation (1). p and q are parameters, for which we used 4 in this article.

$$\lambda = \frac{1}{2} \ (min_{k_1} \sum_{j_1=k_1}^{k_1+p} \frac{\theta_{j_1}^{i_1,}}{p} + min_{k_2} \sum_{j_2=k_2}^{k_2+q} \frac{\theta_{j_2}^{i_2.}}{q}) \qquad (1)$$

3. RESULTS AND DISCUSSION

Eleven β-sheets, each from a different protein, were used in analysis of twist angles. The atomic structure of each protein was downloaded from Protein Data Bank (PDB). Twist angles were measured both sparsely (using the discrete method) and finely (using splines). The splines (Figs. 1A and 2A) appear to align well with the ribbons that represent the central lines of β-strands in Chimera (Pettersen et al., 2004). Since the distance between two consecutive midpoints is about 3.3 Å, the number of twist angles measured from a pair of splines is three times of twist angles measured sparsely if 1.1 Å vector length is used for splines.

3.1. Twist angles measured using the discrete method and the spline method

The discrete measurement shows that the smallest twist angle is 10.21° that is the fifth angle of pair (β_3, β_4) (Fig. 1B) formed by β-strand 3 and β-strand 4 (Fig. 1A). The largest twist angle is 48.92°, the third angle of pair (β_5, β_6) for 1ATZ sheet A (Fig. 1B). A typical range of twist angles of proteins was previously reported (Ho and Curmi, 2002). Since twist angles are often different depending on where they are measured, we calculated an average of twist angles for each pair. The pair of β-strands that has the smallest
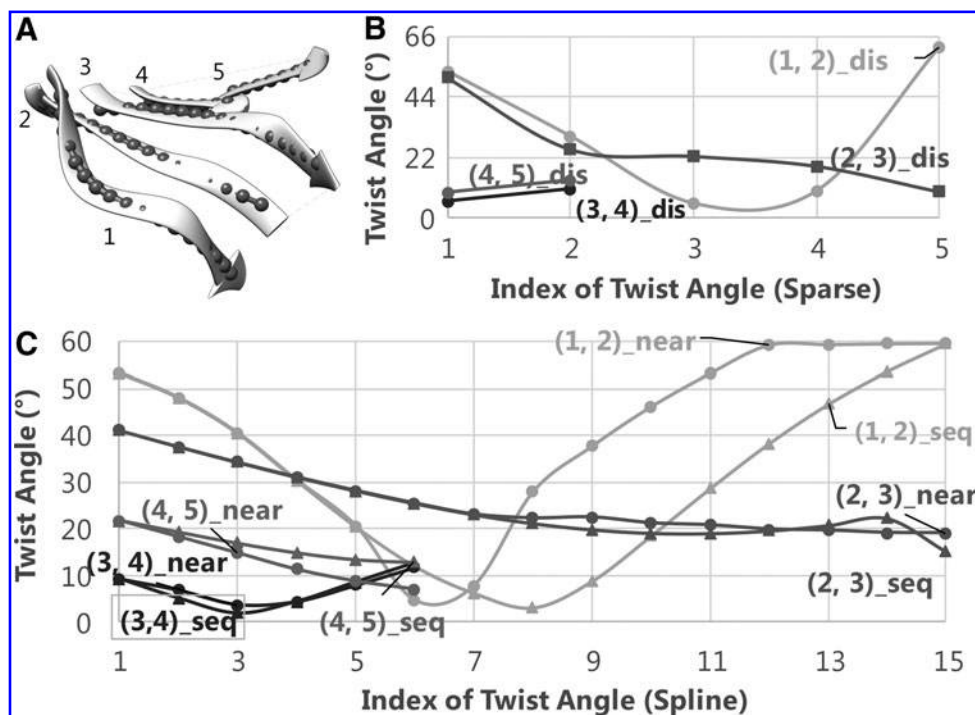
FIG. 2. Twist angles calculated from atomic positions and central lines of β-strands in protein 1AOP (PDB ID). See Figure 1 caption for more details in labels and legends.

overall twist angles and smallest variation is pair (β_3, β_4), as seen from the average of 13.44° and standard deviation of 2.32° [row of 1ATZ_(3,4) with columns 4 and 5 in Table 1]. The plot of twist angles shows that there is smaller twist along the pair formed by β_3 and β_4 (brown curve in Fig. 1B) when they are compared with other pairs. However, there is much larger twist for pair (β_5, β_6) (purple curve in Fig. 1B), with an average of 39.38° (Table 1). We observe a similar trend of overall twist among different pairs for both the discrete method and the spline method for the 1ATZ sheet A. For example, both methods rank overall amount of twist from the highest to the lowest as pairs (5,6), (4,5), (2,3), and (3,4) (Fig. 1B, C). This order is also shown in the average of twist angles of 31.38°, 23.99°, 20.53°, and 12.16° for those pairs measured using the spline method and 39.38°, 23.45°, 20.55°, and 13.44° measured, respectively, using the discrete method (rows of 1ATZ in Table 1). Although the two methods produce different twist angles, they produce the same result in distinguishing the overall level of twist among different pairs.

The change among twist angles along a pair is shown similarly from both the spline and the discrete measurements. As an example, twist angles of pair (2,3) of 1AOP are monotonically decreasing shown from the discrete measurement (red curve in Fig. 2B). They are also monotonically decreasing in the spline measurement if the nearest vector was used (red curve in Fig. 2C). The twist curve appears as a well for pair (1,2) of 1AOP in both measurements (cyan curves in Fig. 2B, C). Although the overall nature of twist change along a pair can be reflected from both methods, the spline measurement shows smoother transition than the discrete measurement, as expected from more sampling using splines. We observed this property from both 1ATZ, 1AOP, and the rest of the nine β-sheets in the test (data not shown). The comparison of twist angles using the two methods is summarized in Table 1. We observed that the average twist angles among 55 pairs of β-strands are comparable between the two methods (Column 2 vs. Column 4 and Column 7 vs. Column 9 of Table 1). The average twist angles of all 55 pairs are 22.25° for discrete measurement, very close to 22.06° for the spline measurement.

3.2. Measurement using nearest vectors and sequential vectors

Unlike discrete measurement in which vectors are defined by atomic positions of the backbone, twist measurement using splines needs to have well-defined vectors. We investigated two ways to define vectors. The nearest-vector method emphasizes local twist angles more than the sequential method. Although similar

TABLE 1. AVERAGE AND STANDARD DEVIATION OF TWIST ANGLES MEASURED USING THE SPLINE
METHOD AND THE DISCRETE METHOD FOR EACH PAIR OF β-STRANDS

| Index[a]	A_s[b]	S_s[c]	A_d[d]	S_d[e]	Index[a]	A_s[b]	S_s[c]	A_d[d]	S_d[e]
1B3A_(1,2)	25.85	6.65	21.64	6.85	1ELU_(5,6)	20.04	7.43	16.41	0.97
1B3A_(2,3)	38.14	1.86	35.35	3.74	1ELU_(6,7)	29.80	9.51	30.10	8.10
1A12_(1,2)	51.03	N/A	52.27	N/A	1D5T_(1,2)	38.28	3.53	39.00	6.37
1A12_(2,3)	24.71	6.97	24.19	7.68	1D5T_(2,3)	15.82	5.51	15.22	8.07
1A12_(3,4)	23.34	10.74	26.00	15.03	1D5T_(3,4)	8.43	0.93	10.39	6.12
1AKY_(1,2)	26.57	3.33	27.51	4.60	1D5T_(4,5)	17.65	2.63	18.08	2.57
1AKY_(2,3)	20.08	0.20	19.26	7.54	1D5T_(5,6)	16.01	0.72	18.37	9.17
1AKY_(3,4)	21.12	0.90	20.33	8.41	1D5T_(6,7)	23.04	10.15	25.77	14.14
1AKY_(4,5)	16.72	1.40	16.49	1.04	1DTD_(1,2)	20.37	5.51	19.92	3.88
1AOP_(1,2)	40.51	18.04	31.88	22.68	1DTD_(2,3)	22.41	13.51	21.14	4.50
1AOP_(2,3)	25.66	6.94	25.25	13.81	1DTD_(3,4)	12.11	6.71	11.10	6.29
1AOP_(3,4)	7.29	2.76	8.26	2.29	1DTD_(4,5)	15.76	4.92	9.38	1.67
1AOP_(4,5)	13.67	5.16	9.13	N/A	1DTD_(5,6)	15.94	2.69	15.80	3.03
1ATZ_(1,2)	18.42	N/A	18.54	N/A	1DTD_(6,7)	16.20	4.62	24.07	13.26
1ATZ_(2,3)	20.53	3.00	20.55	5.86	1DTD_(7,8)	35.09	3.73	38.36	7.56
1ATZ_(3,4)	12.16	2.60	13.44	2.32	1QNA_(1,2)	24.73	N/A	25.50	N/A
1ATZ_(4,5)	23.99	4.56	23.45	7.78	1QNA_(2,3)	26.01	1.99	28.00	3.28
1ATZ_(5,6)	31.38	7.23	39.38	10.65	1QNA_(3,4)	14.33	1.46	16.33	8.86
1CHD_(1,2)	39.97	7.10	43.32	11.34	1QNA_(4,5)	18.00	1.55	22.51	13.95
1CHD_(2,3)	15.44	2.07	13.27	1.95	1QNA_(5,6)	25.80	5.30	24.34	18.15
1CHD_(3,4)	15.95	1.66	16.26	1.95	1QNA_(6,7)	17.42	1.12	12.93	4.76
1CHD_(4,5)	15.83	4.54	15.31	6.05	1QNA_(7,8)	16.75	0.16	19.06	9.73
1CHD_(5,6)	17.13	6.83	16.69	5.47	1QNA_(8,9)	22.75	0.86	24.00	2.89
1CHD_(6,7)	36.28	17.89	30.90	16.66	1733_(1,2)	20.26	6.60	15.71	6.17
1ELU_(1,2)	29.82	6.27	26.56	11.31	1733_(2,3)	13.77	4.36	14.77	4.88
1ELU_(2,3)	10.43	3.77	9.71	4.42	1733_(3,4)	16.69	5.97	18.76	0.44
1ELU_(3,4)	20.61	0.71	14.40	2.18	1733_(4,5)	33.93	10.53	34.12	12.58
1ELU_(4,5)	23.95	3.78	24.84	4.49	Avg_all[f]	22.25		22.06	

[a]The PDB ID and the pair ID; [b]the average and [c]standard deviation of twist angles of the pair measured using the spline method with nearest vectors; [d]the average and [e]standard deviation of twist angles measured by the discrete method; [f]Avg_all: the average of 55 pairs of average twist angles in the column.

N/A, not available when the number of data points is <3; PDB, Protein Data Bank.

twist angles are produced by the two definitions of vectors for most of the angles in the test, different values are observed. For example, the second angle of pair (S_5, S_6) is 23.4° using the nearest vectors, but it is 32.5° using sequential vector for 1ATZ (Fig. 1C). Given that the vector length is only 1.1 Å in this case, it is not expected to have a large twist change for the backbone in such a small distance. More ''jumps'' of twist angles were observed in the use of sequential vectors than the use of nearest vectors. We show that using nearest vectors produces more robust measurement of twist angles.

3.3. Vector length

Twist angles are generally different when measured with different vector lengths. We sampled vector length of 0.7, 1.1, 1.5, 2.5, and 3.2 Å for the data set of 11 proteins. To examine the overall effect of twist for the comparable regions among the 11 β-sheets, we monitored the AMT that is the average of twist angles nearby the minimum twist regions of two longest pairs for a β-sheet. We observe that AMT stays similar when different vector lengths are used for some β-sheets (1ATZ, 1AKY, and 1B3A). However, it is quite different for some other β-sheets (1D5T, 1A12, and 1QNA), and thus further investigation is needed. The nature of twist produced by different pairs is different. We observed that most pairs show increasing or decreasing twist angles along the pair, but some pairs show odd fluctuations. Those β-sheets that show less stable AMT between 0.7 and 1.5 Å vector length tend to have ''unusual'' pairs, such as 1AOP (Figs. 2 and 3). It appears that AMT is more stable for vector length <1.5 Å for most of the cases. The twist angles shown in Figures 1, 2, and Table 1 were calculated using 1.1 Å that is about one-third of the distance between two
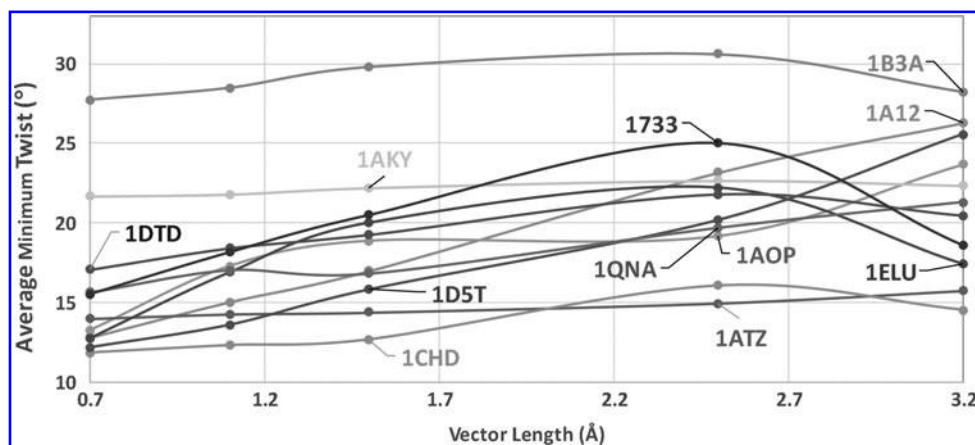
FIG. 3. Effect of vector length shown as the averaged minimum twist for 11 β-sheets. Nearest vectors were used in the measurement.

consecutive midpoints of the backbone. It is notable that the average twist angles for the spline method are also comparable with those generated by the discrete methods when 1.1 Å was used as the vector length (Table 1).

4. CONCLUSION

Twist of β-strands has been analyzed extensively in the literature. However, most studies are based on the discrete measurement directly using atomic positions of the protein backbones. We propose a new method to measure twist angles using a pair of splines of β-strands. In this method, a pairing segment is first determined and the angles are measured using the nearest-vector method with a 1.1 Å vector length. In a data set of 55 pairs of β-strands from 11 β-sheets of 11 proteins, the spline measurement shows comparable results as measured using the discrete method, particularly in terms of capturing twist angle change along a pair, overall amount of twist among different pairs, and the average of twist angles. The proposed method provides an alternative method to characterize twist angles at different regions of the β-sheet. It can be used to characterize twist when the atomic structure is not available but the central lines of β-strands are available, a situation in analysis of 3D image data at medium resolutions obtained from cryo-EM.

ACKNOWLEDGMENTS

AUTHORS' CONTRIBUTIONS

All authors participated in the design of the method. T.I. implemented the method and conducted the test. M.P. assisted the organization of data and results. J.H. wrote the article.

AUTHOR DISCLOSURE STATEMENT

No competing financial interests exist.

REFERENCES

Baker, M.L., Ju, T., and Chiu, W. 2007. Identification of secondary structure elements in intermediate-resolution density maps. *Structure* 15:7–19.

Chothia, C. 1973. Conformation of twisted beta-pleated sheets in proteins. *J Mol Biol* 75:295–302.

Chothia, C., Levitt, M., and Richardson, D. 1977. Structure of proteins: Packing of alpha-helices and pleated sheets. *Proc Natl Acad Sci U S A* 74:4130–4134.

Dal Palu, A., He, J., Pontelli, E., et al. 2006. Identification of alpha-helices from low resolution protein density maps. *Proc Comput Syst Bioinformatics Conf (CSB)* 89–98.

Ho, B.K., and Curmi, P.M. 2002. Twist and shear in beta-sheets and beta-ribbons. *J Mol Biol* 317:291–308.

Jiang, W., Baker, M.L., Ludtke, S.J., et al. 2001. Bridging the information gap: Computational tools for intermediate resolution structure interpretation. *J Mol Biol* 308:1033–1044.

Kong, Y., Zhang, X., Baker, T.S., et al. 2004. A Structural-informatics approach for tracing beta-sheets: Building pseudo-C(alpha) traces for beta-strands in intermediate-resolution density maps. *J Mol Biol* 339:117–130.

Murzin, A.G., Lesk, A.M., and Chothia, C. 1994. Principles determining the structure of beta-sheet barrels in proteins. I. A theoretical analysis. *J Mol Biol* 236:1369–1381.

Pettersen, E.F., Goddard, T.D., Huang, C.C., et al. 2004. UCSF Chimera—A visualization system for exploratory research and analysis. *J Comput Chem* 25:1605–1612.

Quiocho, F.A., Gilliland, G.L., and Phillips, G.N., Jr. 1977. The 2.8-A resolution structure of the L-arabinose-binding protein from *Escherichia coli*. Polypeptide chain folding, domain similarity, and probable location of sugar-binding site. *J Biol Chem* 252:5142–5149.

Richardson, J.S. 1981. The anatomy and taxonomy of protein structure. *Adv Protein Chem* 34:167–339.

Rusu, M., and Wriggers, W. 2012. Evolutionary bidirectional expansion for the tracing of alpha helices in cryo-electron microscopy reconstructions. *J Struct Biol* 177:410–419.

Si, D., and He, J. 2013. Beta-sheet detection and representation from medium resolution cryo-EM density maps. *BCB'13: Proc ACM Conf Bioinform Comput Biol Biomed Inform* 764–770.

Si, D., and He J. 2014. Tracing beta-strands using strandtwister from cryo-EM density maps at medium resolutions. *Structure* 22:1665–1676.

Si, D., and He, J. 2017. Modeling beta-traces for beta-barrels from cryo-EM density maps. *Biomed Res Int* 2017:9.

Si, D., Ji, S., Al Nasr, K., et al. 2012. A machine learning approach for the identification of protein secondary structure elements from electron cryo-microscopy density maps. *Biopolymers* 97:698–708.

Zeyun, Y., and Bajaj, C. 2008. Computational approaches for automatic structural analysis of large biomolecular complexes. *IEEE/ACM Trans Comput Biol Bioinform* 5:568–582.

Address correspondence to:
Dr. Jing He
Department of Computer Science
Old Dominion University
Norfolk, VA 23529

E-mail: jhe@cs.odu.edu