doi:10.1017/S1431927617001891
Microsc. Microanal. 23 (Suppl 1), 2017

Detection of Protein Secondary Structure Patterns from 3D Cryo-TEM Maps at Medium Resolution – Combining the Best of SSETracer and VolTrac

Christopher Spillers[1,2], Willy Wriggers[2], and Jing He[1]

[1] Dept. of Computer Science, Old Dominion University, Norfolk, VA 23529
[2] Dept. of Mechanical & Aerospace Engineering and Institute of Biomedical Engineering, Old Dominion University, Norfolk, VA 23529

Cryo-electron microscopy (Cryo-TEM) has been successfully used to derive the atomic structures of macromolecular assemblies. Structure details are often sufficient for deriving the atomic structure when a 3D density map exhibits a resolution better than 4Å. However, it is still challenging to derive atomic structures from medium-resolution (5–10Å) density maps. At medium resolution it is important to know the location of the major components of a protein structure, such as α-helices and β-sheets, in the density map. Although the backbone of a protein structure is often not visible at medium resolution, secondary structures, such as α-helices and β-sheets, can be computationally identified [1-4]. In addition, the location of β-strands can be derived from the density region of a β-sheet once it is segmented out from the surrounding density [5].

Over the years, various pattern recognition methods have been developed for the identification of α-helices [1-4, 6, 7] and β-sheets [1, 2, 4, 8]. More recently, machine learning techniques, such as Support Vector Machine (SVM) and Convolutional Neural Networks (CNN), have been applied to this problem [2, 9]. Figure 1 shows an example of α-helices, β-sheet, and β-strands detected in a cryo-TEM density map at 5.5Å resolution. The location of secondary structures in the density map can be utilized for prediction of the tertiary structure of the protein. This goal is particularly feasible when the location of the β-strands is predictable from the density map. We have developed a computational method, StrandTwister, to derive the orientation of β-strands from the β-sheet region in the density map. Although β-strands are not distinguishable in a density map at medium resolution, we have shown that their locations are predictable utilizing the twist of a β-sheet. The detected traces of α-helices and β-strands were combined with multiple secondary structure predictions from the protein sequence to score candidate topologies of the secondary structure traces. The true topological order of the traces was ranked the 2nd highest on the list of possible topologies [10] (Figure 1B). Although major secondary structures can be identified, the precise detection of secondary structures remains a challenging problem [11]. For example, a predicted helix can be longer or shorter than a known benchmark. Small helices and β-sheets are particularly challenging to detect.

We have taken advantage of two independently developed methods for secondary structure detection. SSETracer [4] utilizes a feature voting technique to identify the spatial characteristics of a cylinder. VolTrac [3] uses a genetic algorithm for placing seeds, followed by a bidirectional optimization of cross-correlation between cylinder templates and the density map. The two α-helix detection methods have complementary strengths. SSETracer is faster than VolTrac, but it is not as accurate in determining the end position of a helix. Our new hybrid method is built on a quick scan of initial positions of α-helices using SSETracer. A fine process for determination of end points is carried out using the extension step in VolTrac. The merging of the two methods shows an improved accuracy and good efficiency of helix detection. In our presentation, we will discuss our exploration of effective methods

for secondary structure detection, including the recent development and implementation of the hybrid method, as well as the use of such information in deriving the protein backbone [13].

References:

[1] ML Baker *et al*, Structure **15** (2007), p. 7.
[2] D Si *et al*, Biopolymers **97** (2012), p. 698.
[3] M Rusu and W Wriggers, J Struct Biol **177** (2012), p. 410.
[4] D Si and J He, BCB'13: Proceedings of ACM Conference on Bioinformatics (2013) p. 764.
[5] D Si and J He, Structure **22** (2014), p. 1665.
[6] W Jiang *et al*, J Mol Biol **308** (2001), p. 1033.
[7] A Dal Palu *et al*, Proceeding of Computational Systems Bioinformatics Conference (CSB) (2006) p. 89.
[8] Y Kong and J Ma, J Mol Biol **332** (2003), p. 399.
[9] R Li *et al*, 2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) (2016) p. 41.
[10] A Biswas *et al*, IEEE/ACM Transactions on Computational Biology and Bioinformatics, **PP** (2016), p. 1.
[11] S Zeil *et al*, Journal of Computational Biology **24** (2016), p. 52.
[12] A Biswas *et al*, Journal of Computational Biology **22** (2015), p. 837.
[13] The work in this paper was supported in part by NSF DBI-1356621 and by NIH R01-GM062968.

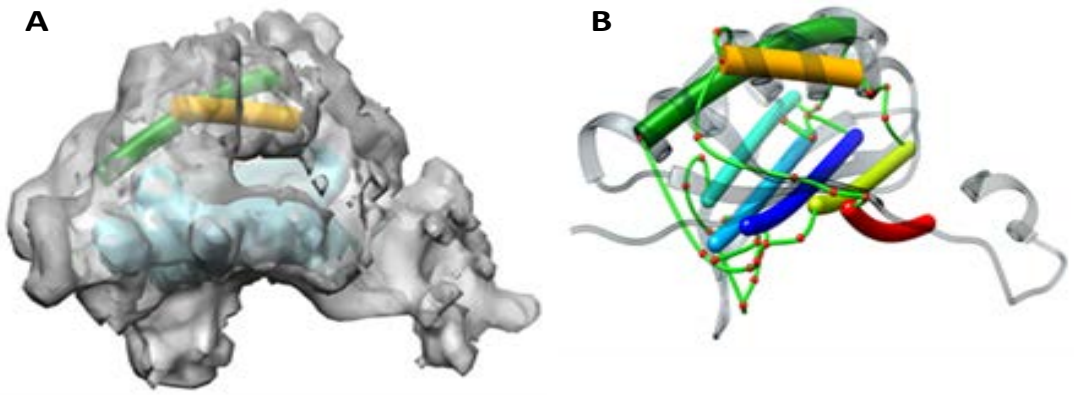

Figure 1. α-helices, β-sheet, and β-strands detected in a medium resolution cryo-TEM density map. (A) The density region of EMD-1780 (5.5Å resolution) corresponding to chain K of PDB ID 3IZ6 is superimposed with the helices (colored cylinders) and β-sheet (blue) detected by SSETracer [4]. (B) The atomic structure of PDB ID 3IZ6 chain K (ribbon) is superimposed with the detected helices (thick cylinders) and β-strand traces (thin sticks) using StrandTwister [5]. The detected helices and strands are colored from N-terminal to C-terminal using rainbow colors. Using Multi-DP-TOSS [10, 12], the true topology of secondary structure traces (rainbow color from N to C terminal) was ranked the 2nd highest among possible topologies.

https://doi.org/10.1017/S1431927617001891 Published online by Cambridge University Press