Map segmentation, automated model-building and their application to the Cryo-EM Model Challenge

Thomas C. Terwilliger, Paul D. Adams, Pavel V. Afonine, Oleg V. Sobolev

PII: S1047-8477(18)30193-X
DOI: https://doi.org/10.1016/j.jsb.2018.07.016
Reference: YJSBI 7227

To appear in: *Journal of Structural Biology*

Received Date: 4 April 2018
Revised Date: 11 July 2018
Accepted Date: 27 July 2018

Please cite this article as: Terwilliger, T.C., Adams, P.D., Afonine, P.V., Sobolev, O.V., Map segmentation, automated model-building and their application to the Cryo-EM Model Challenge, *Journal of Structural Biology* (2018), doi: https://doi.org/10.1016/j.jsb.2018.07.016

Map segmentation, automated model-building and their application to the Cryo-EM Model Challenge

Thomas C. Terwilliger[1,2], Paul D. Adams[3,4], Pavel V. Afonine[3,5], Oleg V. Sobolev[3]

[1]Los Alamos National Laboratory, Los Alamos NM 87545 USA
[2]New Mexico Consortium, Los Alamos NM 87544 USA
[3]Molecular Biophysics & Integrated Bioimaging Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720-8235, USA
[4]Department of Bioengineering, University of California Berkeley, Berkeley, CA, USA
[5]Department of Physics and International Centre for Quantum and Molecular Structures, Shanghai University, Shanghai, 200444, People's Republic of China

E-mail: tterwilliger@newmexicoconsortium.org

Abstract

A recently-developed method for identifying a compact, contiguous region representing the unique part of a density map was applied to 218 cryo-EM maps with resolutions of 4.5 Å or better. The key elements of the segmentation procedure are (1) identification of all regions of density above a threshold and (2) choice of a unique set of these regions, taking symmetry into consideration, that maximize connectivity and compactness. This segmentation approach was then combined with tools for automated map sharpening and model-building to generate models for the 12 maps in the 2016 cryo-EM model challenge in a fully automated manner. The resulting models have completeness from 24% to 82% and RMS distances from reference interpretations of 0.6 Å to 2.1 Å.

Introduction

In the 2016 Cryo-EM Modeling Challenge (see http://challenges.emdatabank.org/?q=model_challenge; accessed 2017-11-19), a total of 12 maps were supplied to contestants along with reconstruction symmetry and the sequences of the molecules present. One of the goals of the Challenge was to fully interpret such a map given only the map, the symmetry and the sequence information. There are a number of tools being developed by several groups for automated interpretation of cryo-EM maps (DiMaio and Chiu, 2016). These include methods for identification of secondary structure (Jiang et al., 2001; Kong and Ma, 2003; Kong et al., 2004; Baker, Ju and Chiu, 2007), methods for combination of structure-modeling tools such as Rosetta with cryo-EM model-building (Lindert et al., 2012; Wang et al., 2015; Frenz et al., 2017), semi-automated tools for full map interpretation (Baker et al., 2011), and automated tools based on chain-tracing (Chen et al., 2016; Collins and Si, 2017) and template-matching approaches (Zhou, Wang and Wang, 2017).