doi:10.1016/j.jmb.2009.07.008 J. Mol. Biol. (2009) 392, 181–190

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Refinement of Protein Structures into Low-Resolution Density Maps Using Rosetta

Frank DiMaio[1]*, Michael D. Tyka[1], Matthew L. Baker[2], Wah Chiu[2] and David Baker[1]*

[1]*Department of Biochemistry, University of Washington, Box 357350, Seattle, WA 98195, USA*

[2]*National Center for Macromolecular Imaging, Verna and Marrs McLean Department of Biochemistry and Molecular Biology, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA*

Received 15 April 2009; received in revised form 2 July 2009; accepted 2 July 2009
Available online 8 July 2009

Edited by R. Huber

We describe a method based on Rosetta structure refinement for generating high-resolution, all-atom protein models from electron cryomicroscopy density maps. A local measure of the fit of a model to the density is used to directly guide structure refinement and to identify regions incompatible with the density that are then targeted for extensive rebuilding. Over a range of test cases using both simulated and experimentally generated data, the method consistently increases the accuracy of starting models generated either by comparative modeling or by hand-tracing the density. The method can achieve near-atomic resolution starting from density maps at 4–6 Å resolution.

© 2009 Elsevier Ltd. All rights reserved.

Keywords: cryoEM; density fitting; structure prediction; Rosetta; comparative modeling

Introduction

Electron cryomicroscopy (cryoEM) has matured to the point that density maps can regularly be obtained at 4–8 Å resolution. Methods have been developed to fit solved structures into such maps, to find locations of secondary-structure elements[1,2] and determine the topology of these elements,[3] to select threaded homology models using density data,[4] and to flexibly fit models into density.[5–11] These methods generally start with complete all-atom models, rather than the C^{α}-only models that are often traced through low-resolution density.

The Rosetta structure prediction methodology[12] has been successful at predicting structures *de novo* for small proteins and for refining comparative models to higher resolution. Rosetta uses Monte Carlo sampling to search for the lowest-energy structure of the polypeptide chain according to a detailed all-atom force field. For small proteins (less than 100 amino acids), Rosetta can, in some cases, generate atomic-accuracy models with no experimental data. The bottleneck to more consistent *de novo* prediction is conformational sampling: conformations within 1.5–2 Å RMSD of the native structure generally have much lower energies than nonnative models, but for larger proteins, such models are generated extremely rarely. With even a small amount of data (e.g., NMR chemical shift data[13]) to guide conformational sampling, Rosetta can consistently build atomic-level models for proteins of 120 amino acids or less. Rosetta's rebuild-and-refinement protocol often improves the accuracy of comparative models, especially distant homologues (<30% sequence identity).

In this article, we adapt Rosetta to refine comparative models and low-resolution C^{α} traces using density maps as a guide. A local measure of the fit to density is used to identify regions incompatible with the density that are targeted for extensive rebuilding, and the whole structure is then refined using this measure as a guide. The new method generates models that fit the density, are low in energy, and can have near atomic resolution starting from 4-8 Å density maps.

**Corresponding authors.* E-mail address: dimaio@u.washington.edu.

Abbreviations used: cryoEM, electron cryomicroscopy; RDV, rice dwarf virus; PDB, Protein Data Bank.

0022-2836/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.

Results and Discussion

The adaptation of Rosetta to utilize input density maps is described in Materials and Methods. We have developed two protocols: the first starts with an alignment to a homologous protein of known structure, and the second starts with a low-resolution C^{α} trace through the density. In this section, we describe application of the two methods to a variety of structure modeling problems using both synthetic and experimentally determined density maps.

Comparative modeling using synthesized density

This test involves the refinement of a set of models built from distant homologues into synthesized low-resolution cryoEM density maps at 5 and 10 Å resolution. For each of eight structures, noise-free maps were constructed using EMAN's mrc2pdb,[1] at both 5 and 10 Å resolution. The starting models are based on Moulder reference alignments.[14] Moulder uses a genetic algorithm that simultaneously optimizes a sequence-alignment potential and a potential on the threaded model implied by a particular sequence alignment. The top 300 threaded models according to Moulder's fitness function were refined into density using the protocol outlined in Fig. 1 (see Materials and Methods and Supplementary Materials for more details).

The results of this refinement are shown in Table 1, and two examples are illustrated in Fig. 2. For each of the eight structures, the refined model is closer to the native structure (in terms of C^{α} and all-atom RMSD) than the best initial model. In some cases, the initial model that was closest to native was not the one highest ranked by Moulder; in some cases, it was not even in the top 20. In six of the eight cases at 5 Å and four of the eight cases at 10 Å, the lowest-energy model was closer than 2 Å to the crystal structure. Refinement improved individual starting models from 1 to 3 Å (see Supplementary Fig. 1). Several structures at 5 Å resolution refined from 2 to 4 Å RMSD to sub-1 Å accuracy. These results show that the Rosetta refinement procedure—restricted by a low-resolution density map to focus sampling in relevant regions—can improve homology models, even those that are already quite close to the native.

Benchmark tests on real data

Refining the upper domain of RDV

The rebuilding and refinement into density protocol illustrated in Fig. 1 was applied to the upper domain (residues 173–292) of the rice dwarf virus (RDV) capsid protein P8. A 6.8-Å-resolution cryoEM map of this structure has been determined.[15] The crystal structure of this protein has also been solved [Protein Data Bank (PDB) code: 1uf2],[16] giving a standard against which to compare. A starting model was generated from an alignment to a structural homologue from bluetongue virus[17] (coat protein vp7, PDB code: 1bvp) produced by GenTHREADER.[18] Details of this alignment are shown in Supplementary Fig. 2. The standard Rosetta rebuild-and-relax protocol (without density data) was used to create an initial 10,000 models, which were then refined into density as described in Materials and Methods.

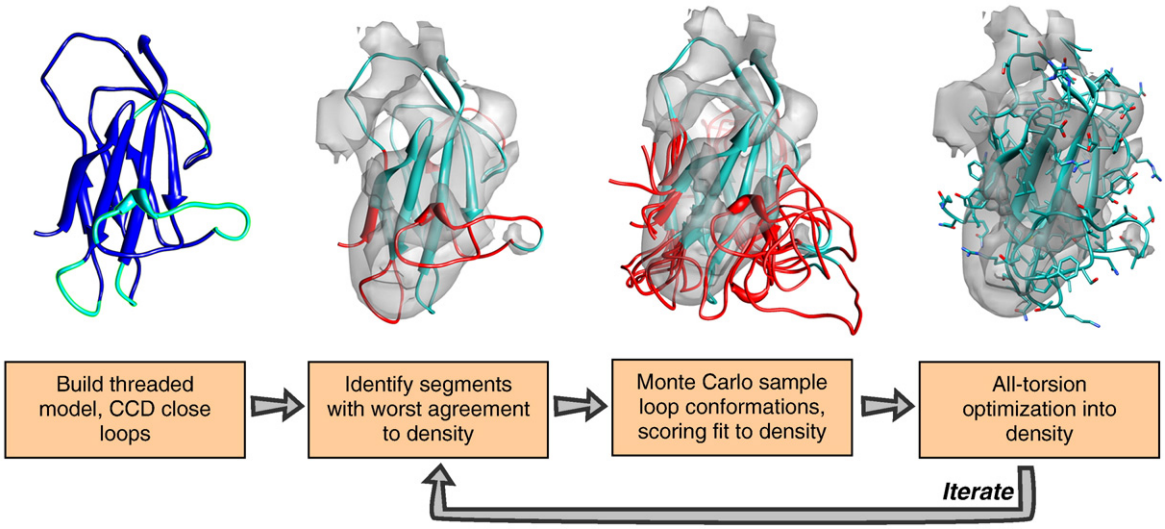

Fig. 1. The comparative modeling into density protocol. We initially build a threaded model from some alignment (blue), using fragment assembly to model insertions (cyan) and cyclic coordinate descent to close gaps. We then dock this threaded model into density and identify regions that have a poor local agreement with the density data (red). We aggressively resample the conformations in these regions, scoring each potential conformation with Rosetta's low-resolution energy function together with an agreement-to-density score. Finally, we optimize side-chain rotamers and minimize all backbone and side-chain torsions using Rosetta's high-resolution potential, also augmented with this agreement-to-density score. We iterate over these final three steps until the lowest-energy models converge, at each iteration enriching our population for those models with both favorable Rosetta energy and good fit to density.

Download English Version:

https://daneshyari.com/en/article/2186352

Download Persian Version:

https://daneshyari.com/article/2186352

Daneshyari.com