doi:10.1017/S1431927615003505 Paper No. 0271 Microsc. Microanal. 21 (Suppl 3), 2015

Modeling Protein Structure in Macromolecular Assemblies at Near Atomic Resolutions

Matthew L. Baker[1], Muyuan Chen[1,2], Philip R. Baldwin[1], Steven J. Ludtke[1], Wah Chiu[1]

[1.] National Center for Macromolecular Imaging, Verna and Marrs McLean Department of Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, TX USA
[2.] Program in Structural and Computational Biology, Baylor College of Medicine, Houston, TX USA

In nearly every cellular process, macromolecular machines play critical roles. As such, understanding the structure of these complexes is critical in preventing disease and developing efficacious treatments. However, structural studies of such large complexes, typically by electron cryo-microscopy or X-ray crystallography are often difficult and result in structures with non-atomic resolutions. Limited resolvability and noise in the density map can complicate direct interpretation, and as such, model construction at near-atomic resolutions is generally not automated and often results in only Cα only models[1].

Previously, we developed the *Pathwalking* protocol, which semi-automatically enumerates putative configurations of protein models from a density map[2,3]. *Pathwaking* is based on the Traveling Salesman Problem (TSP), in which possible cyclical paths (i.e. protein fold) are calculated through a density map without using any sequence or structure constraints. A TSP solver is used to find a path through a set of pseudoatoms by optimizing the spatial distance between the pseudoatoms such that they are representative of Cα-Cα distances in consecutive amino acids in the protein structure. Here, the only required inputs are a density map better than 6Å resolution and the number of amino acids in the protein of interest, which is used for the initial seeding of pseudoatoms in the density map.

In the initial testing of *Pathwalking*, reasonable first-approach models, models with the correct overall folds, were derived directly from the density map with limited user intervention. However, *Pathwalking* was not designed to directly consider protein chemistry or density map constraint, and as such, non-protein like connections were sometimes observed and required the user to correct.

Building on the success of our first implementation of *Pathwalking* and the rapidly growing number of near atomic resolution structures, we developed an enhanced version of our original protocol capable of producing more accurate models with reduced user interaction. In the new version of *Pathwalking*, all of the interactive steps in the original version, including identifying secondary structures assignment, pseudoatom placement and path evaluation, have been optimized for nearly automated usage. Additionally, *Pathwalking* improvements in the implementation of our TSP-based search now allow for modeling multiple chains in a density map simultaneously. In testing of our new *Pathwalking* protocol, we have not only improved the ease of use but have also increased the accuracy of our models. In our benchmark of 20 authentic density maps between 3.2Å and 7Å resolution, *Pathwalking* models averaged ~69% structural overlap, 2.35Å RMSD and ~67% correctly registered Cαs when compared to the known structure. Errors in modeling were generally restricted to register shifts and improperly placed pseudoatoms due to map noise/resolution, though these errors generally did not affect the overall model topology. Three examples of *Pathwalking* on cryo-EM and X-ray crystallographic density maps from 3-4Å resolution are in Figure 1.

References:
[1] Baker ML, Zhang J, Ludtke SJ, Chiu W. Cryo-EM of macromolecular assemblies at near-atomic resolution. Nat Protoc. 2010 Sep;5(10):1697-708. PMCID: PMC3107675.
[2] Baker MR, Rees I, Ludtke SJ, Chiu W, Baker ML. Constructing and validating initial Cα models from subnanometer resolution density maps with pathwalking. Structure. 2012 Mar 7;20(3):450-63. PMCID: PMC3307788.
[3] Baker ML, Baker MR, Hryc CF, Ju T, Chiu W. Gorgon and pathwalking: macromolecular modeling tools for subnanometer resolution density maps. Biopolymers. 2012 Sep;97(9):655-68 PMCID: PMC3899894.
[4] Wang Z, Hryc CF, Bammes B, Afonine PV, Jakana J, Chen DH, Liu X, Baker ML, Kao C, Ludtke SJ, Schmid MF, Adams PD, Chiu W. An atomic model of brome mosaic virus using direct electron detection and real-space optimization. Nat Commun. 2014 Sep 4;5:4808. PMCID: PMC4155512.

[5] This work was supported by grants from the National Institutes of Health (P41GM103832, R01GM079429, R21GM100229), and National Science Foundation (DBI-1356306).

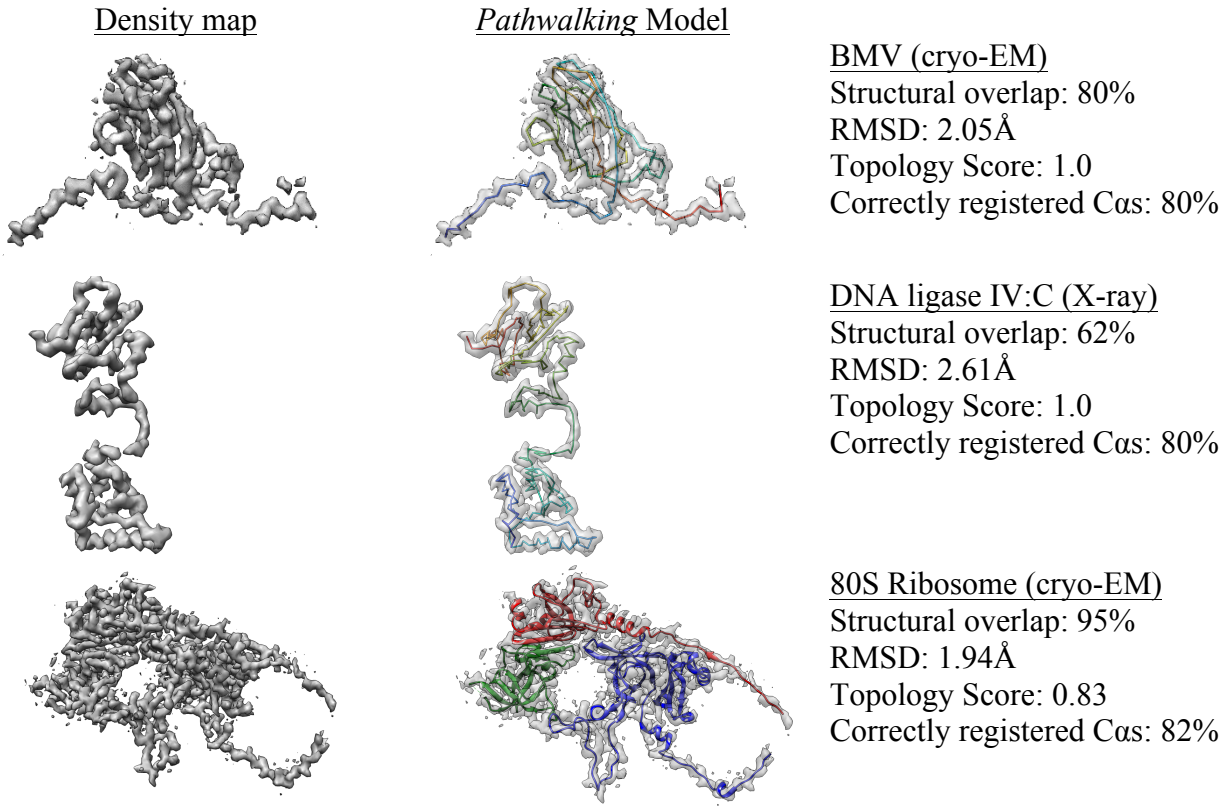

Figure 1. *Pathwalking* at near atomic resolutions. Shown are two examples of the new *Pathwalking* protocol. In the left panels, the density maps for the structural protein from Brome Mosiac Virus [4] (top row, EMDB ID:6000), subunit C from the DNA ligase IV complex (PDB ID: 1Z56, middle row) and the 80S ribosome (chains C,I and M only) (bottom row, EMDB ID:2566). In the middle panel, the *Pathwalking* model is shown overlaid on the density map. Model quality statistics are shown in the right panel. For the 80S ribosome example, a portion of the map was segmented and contained density for only 3 chains. *Pathwalking* results for the 80S ribosome data are averaged over all three subunits.

https://doi.org/10.1017/S1431927615003505 Published online by Cambridge University Press