

Building and Validating Atomic Models for Cryo-EM Density Maps

Matthew L. Baker¹, Muyuan Chen^{1,2}, Tunay Durmaz¹, Phil Baldwin¹, Tao Ju³ and Steven J. Ludtke¹

¹ Verna and Marrs McLean Department of Biochemistry and Molecular Biology, Baylor College of Medicine, Houston TX USA.

² Program in Structural and Computational Biology and Molecular Biophysics, Baylor College of Medicine, Houston TX USA.

³ Department of Computer Science and Engineering, Washington University in St. Louis, St. Louis MO USA.

While it is now possible to produce structures for large macromolecular complexes at better than 3Å resolution electron cryo-microscopy (cryoEM), the vast majority of structures are determined to lower resolutions (3-10Å). As atomic details are often lacking in this resolution range, accurate interpretation of macromolecular structure is often challenging and somewhat limited [1]. To this end we have developed and applied a number of *de novo* model building and validation tools [2-4].

Designed specifically for subnanometer resolution cryo-EM density maps, Gorgon (<http://gorgon.wustl.edu>) offers an integrated set of utilities including feature recognition, rigid-body and flexible fitting and *de novo* modeling [5,6]. With Gorgon, modeling a single protein structure has been reduced from a manual, weeks-long project to a computationally guided process often requiring less than a day to complete a structural model for a single protein. In the latest version of our Gorgon molecular modeling software, we have improved upon a number of our core algorithms allowing for faster and more accurate modeling of entire macromolecular assemblies. Additionally, these core features, plus a host of new functionality, will be made available through a stand-alone toolkit, The Gorgon Toolkit, which will be compatible with our own molecular viewer, Gorgon Explorer, as well as other visualization software such as UCSF's Chimera. Additionally, design considerations of The Gorgon Toolkit will also provide enhanced interoperability with various image processing and modeling software.

While Gorgon provides an interactive modeling environment for subnanometer resolution cryo-EM density maps, Pathwalking offers a more automated solution for model building in density maps better than 7Å resolution [7]. Recent improvements in our Pathwalking utilities, available in EMAN2.12, nearly completely automate the model building process, as well as incorporate a number of new density, geometrical and biophysical constraints to insure optimal structural modeling. Additionally, new features in Pathwalking also allow for simultaneous modeling of multiple protein structures in one density map. In a recent example of Pathwalking on a ~4.5Å resolution cryo-EM density map of a ribosome, we were able to automatically compute Cα backbone models for every protein component in the ribosome in less than a day. While some registration errors were present, the unsupervised Pathwalking produced models with the correct folds. As an extension to Pathwalking, we are adapting our Pathwalking tools to generate reference models in successive map-refinement iterations, consequently improving 2D image classification and 3D reconstruction. By combining this with a deep learning approach, we can monitor structural variations in the 2D particle images, allowing us to 1) identify the most homogenous group of particles and thereby producing the highest resolution density maps possible with the least amount of noise and 2) characterize particle motions.

In addition to cryo-EM modeling tools, we have implemented a Fitting Validation and Evaluation Server (FiVES). FiVES uses a number of statistical measures to evaluate the fit an atomic model to a cryo-EM density map, including the comparison of fittings from multiple software packages, evaluation of fit as compared to proteins of similar size, quality of fit and uniqueness of fit. FiVES is available as web-service and only requires the user to provide their map, model and map resolution. Results of the statistical analysis are made available through an interactive webpage, as well as downloadable file containing all of the raw data and statistical analysis. FiVES is currently available for beta-testing at <http://ncmi.bcm.edu/FiVES>.

References:

- [1] Baker, M. L., Baker, M. R., Hryc, C. F., and Dimaio, F. *Methods Enzymol* **483** (2010), p. 1-29.
- [2] Baker, M. L., Baker, M. R., Hryc, C. F., Ju, T., and Chiu, W. (2012). *Biopolymers* **97** (2012), p. 655-68.
- [3] Wang, Z., Hryc, C.F., Bammes, B., Afonine, P.V., Jakana, J., Chen, D.C., Liu, X., Baker, M.L., Kao, C., Ludtke, S.J., Schmid, M.F., Adams, P.A., and Chiu, W. *Nat Commun* **5** (2014), p. 4808.
- [4] Baker, M. L., Hryc, C. F., Zhang, Q., Wu, W., Jakana, J., Haase-Pettingell, C., Afonine, P. V., Adams, P. D., King, J. A., Jiang, W., and Chiu, W. (2013). *Proc Natl Acad Sci USA* **110** (2013), p. 12301-12306.
- [5] Baker, M. L., Abeysinghe, S. S., Schuh, S., Coleman, R. A., Abrams, A., Marsh, M. P., Hryc, C. F., Ruths, T., Chiu, W., and Ju, T. *J Struct Biol* **174** (2010), p. 360-73.
- [6] Dao, H., **Baker, M. L.**, Ju, T. (2015). *The Visual Computer* **31** (2015), p. 961-977.
- [7] Baker, M. R., Rees, I., Ludtke, S. J., Chiu, W., and Baker, M. L. *Structure* **20** (2012), p. 450-463.
- [8] The authors acknowledge funding from NIH (1R21GM100229, 5R01GM079429 and 8P41GM103832) and NSF (DBI-1408887).

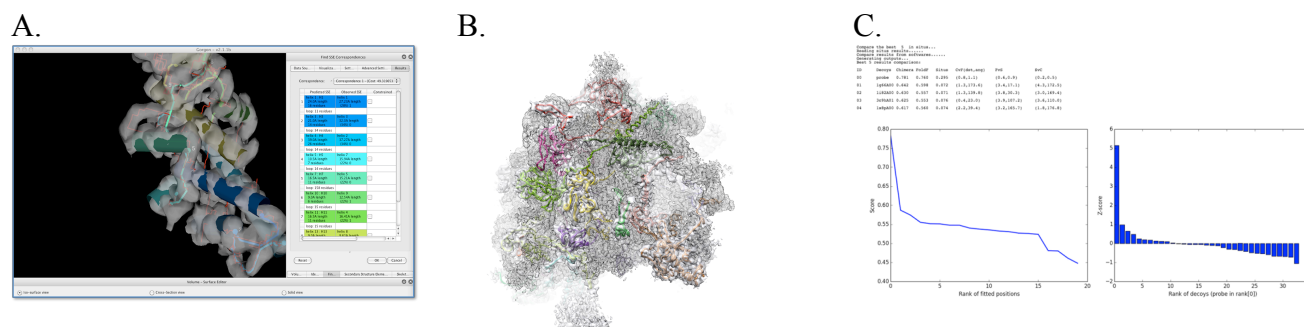


Figure 1. Examples of (A) *de novo* modeling in Gorgon, (B) Pathwalking on the 4.5Å resolution structure of a ribosome and (C) FiVES output.