Preface

Structures of proteins and protein complexes help explain biomolecular function. Computational methods provide an inexpensive way to predict unknown structures, manipulate behavior, and design new proteins or functions. The protein structure prediction program Rosetta, developed by a consortium of laboratories in the Rosetta Commons, has an unmatched variety of functionalities and is one of the most accurate protein structure prediction and design approaches (Das & Baker, *Ann Rev Biochem* 2008; Gray, *Curr Op Struct Biol* 2006). To make the Rosetta approaches broadly accessible to biologists and biomolecular engineers with varied backgrounds, we developed PyRosetta, a Python-based interactive platform for accessing the objects and algorithms within the Rosetta protein structure prediction suite. In PyRosetta, users can measure and manipulate protein conformations, calculate energies in low- and high-resolution representations, fold proteins from sequence, model variable regions of proteins (loops), dock proteins or small molecules, and design protein sequences. Furthermore, with access to the primary Rosetta optimization objects, users can build custom protocols for operations tailored to particular biomolecular applications. Since the program can communicate with the visualization software PyMOL, search algorithms can be viewed on-screen in real time.

In this book, we have compiled a set of workshops to teach both the fundamentals and the practical applications of protein structure prediction and design. The workshops assume basic knowledge of protein structure and familiarity with computers and the Python programming language. Readings and references are provided in each chapter for more in-depth study. Each workshop covers a single topic in the field and walks the reader through the basic operations in a one- to two-hour session. Interactive exercises are incorporated so that the reader gains hands-on experience using the variety of commands available in the toolkit. The text is arranged progressively, beginning with an introduction to the PyMOL visualization package, proceeding through the fundamentals of protein structure and energetics, and then progressing through the applications of protein folding, refinement, packing, design, docking, and loop modeling. A set of tables is provided at the end of the book as a reference of the available commands.

Additional resources on the Rosetta program are available online. The PyRosetta web site, pyrosetta.org, includes additional example and application scripts. At the web-based user forum www.rosettacommons.org/forum, the PyRosetta community shares questions, answers, and useful scripts. Documentation on Rosetta in general including expert details on the underlying C++ code is available at www.rosettacommons.org/support. PyRosetta is built upon the Rosetta 3 platform, so objects available in PyRosetta will have the same underlying data structures and functionality.

The bulk of these modules were created at the Homewood campus of Johns Hopkins University over the course of two semesters, Spring 2008 and Spring 2009, for the Chemical & Biomolecular Engineering class "Computational Protein Structure Prediction and Design," and further developed in 2010, 2012, 2014, and 2016. We acknowledge the contributions of the

many developers of the Rosetta community (see www.rosettacommons.org/about) for their creation of the Rosetta protein structure prediction suite, upon which PyRosetta is built. Julian Rosenberg and J. D. Bagert, former students of the class before PyRosetta, pioneered early drafts of the workshops in 2008 through a Technology Fellowship from the JHU Center for Educational Resources. We thank Richard Shingles of the Center for Educational Resources for assistance in the workshop conception and in the formal assessment over multiple CER grants. Brian Weitzner, Justin Porter, and Liza Lee identified corrections in the first printing. Evan H. Baugh extensively tested the modules for the PyRosetta 2.011 and 2.012 releases, and Jason W. Labonte made corresponding improvements in the text's second edition, respectively. E.H.B. and S.L. developed the PyMOL Mover. J.W.L. contributed Workshop #9. Boonsom Uranukul extensively tested the additions to PyRosetta 3.4, wrote additional sample scripts, and improved the code usability under a Technology Fellowship from the JHU Center for Educational Resources. Alex Mathews, another Technology Fellow, updated the workshops to be current with the newest Talaris2013 scoring function. Rebecca F. Alford updated the workshops to be current with PyRosetta 4.0. The National Institutes of Health supported J.J.G., S.C., and S.L. through grant numbers GM-078221 and GM-073151, and the National Science Foundation supported J.J.G. through CAREER award number 0846324. Finally, we thank the wonderful JHU students in all semesters for their help, feedback, patience, and fun times writing code in the lab.

To complete these modules, you will need:

- PyMOL – www.pymol.org
- Python – www.python.org
- IPython – ipython.scipy.org
- PyRosetta – www.pyrosetta.org

All packages are free and available for Mac, Linux, or Windows platforms.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation or the National Institutes of Health.