Project Year
2013-2014

Project Title
PyRosetta: Advanced Biomolecular Modeling for Undergraduates

Project Team
Jeffrey Gray, Whiting School of Engineering, Chemical and Biomolecular Engineering, Faculty
Jason Labonte, Whiting School of Engineering, Chemical and Biomolecular Engineering, Post Doc
Alex Mathews, Whiting School of Engineering, Biomedical Engineering, Fellow

Audience
The project will serve the course, (ChemBE) 540.414 *Computational Protein Structure Prediction and Design*, offered annually or biannually, with the next offering in Fall 2014. New and updated teaching modules and sample Python scripts developed under this fellowship will be integrated into class. The class enrolls 15–20 undergraduates and graduate students from the departments of ChemBE, BME, Biophysics, Chemistry, Computer Science, and Applied Math. Past offerings (Spring 2012, 2010, 2009) served approximately 50 students. Students who took prior courses have enrolled in top computational biology graduate programs (NYU, CMU, MIT).

In addition to the JHU audience, students at other universities benefit since our modules are used at U-Kansas, NYU, UNC, Stanford, and MIT. The workshops and updated code are distributed freely to all users on pyrosetta.org. To date, there are over 2,200 no-cost academic licensees and 8 commercial licensees.

Pedagogical Challenge
Our challenge is to teach modern protein structure modeling and design approaches to JHU undergraduates. Hands-on use of protein modeling software provides a much deeper learning experience than lectures alone. However, the typical software package for predicting and designing structures is complex and difficult to learn, often requiring a background in programming. Moreover, the algorithms involved are often hard to modify or adapt to specific tasks.

With prior support from the Center for Educational Resources (CER), we have created a workbook of pedagogical modules that uses PyRosetta to introduce structure prediction and design applications. Our previous proposals have focused exclusively on protein models. However, Rosetta has grown to now include methods for handling nucleic acids, NCAAs, PTMs, carbohydrates, surfaces, and other molecules. Students frequently request access to these other biomolecules for class projects.

Solution
We propose writing new chapters of our PyRosetta workbook to include modules and sample scripts for teaching modeling of DNA, post-translationally modified protein residues, sugar molecules and glycosylations, and bound ligands. We will provide sample scripts and step-by-step tutorial modules that lead students through the creation and manipulation of these biomolecules in silico. In addition to the educational workbook, we will modify the code directly to allow students to understand the calculations by providing print access to relevant data structures (sugar branching trees, non-canonical amino acid parameters) and by providing informative error messages for common misuse cases.

Faculty Statement

Computational modeling of biomacromolecules is an important, growing field, yet a need is present to make its applications accessible to those with a variety of backgrounds. The PyRosetta platform provides a means of modeling macromolecules in an interactive way and serves as a practical tool in the teaching of current computational methods. We currently have nine modules to teach topics in the field from protein structural analysis and visualization to folding and docking. The modules are bound in a workbook available through Amazon.com (in print and Kindle versions) and online at pyrosetta.org. I use the workbook as the basis of once/week interactive learning sessions in ChemBE 414.

In 2008, the CER supported two Tech Fellows who drafted initial modules for teaching concepts of protein structure prediction and design in ChemBE 540:460, an undergraduate course co-taught by Professor Marc Ostermeier and me. The computational portion of these modules was later expanded into a more comprehensive, computationally focused class, ChemBE 540:414 *Protein Structure Prediction and Design*, which has been taught Spring 2009, 2010, and 2012. In 2012–2013, a CER-supported Tech Fellow: (1) resolved errors in all sample Python scripts and the workbook modules caused by updates to the underlying C++ code; (2) added seven sample scripts including demonstrations of loop predictions, RNA modeling, and non-canonical amino acid modeling; (3) added output functionality to 30 data classes referenced in the second edition of the workbook; and (4) published an updated edition of the workbook to Amazon.com.

While our educational materials are now the leading standard in the field, there are significant gaps in the treatment of biomolecular modeling. Students need the ability to model post-translational modifications, bound ligands, DNA, and sugars. In fact, every year students propose applications for class projects that require these common features in protein systems. Past students have been forced to make approximations or remove these key biophysical features from their models.

We propose to add new workbook modules for nucleic acids, sugars, and non-canonical and post-translationally modified amino acids, so that we can teach these emerging areas of interest with an active-learning component. In addition, we will continue to improve the helpfulness and user-friendliness of the PyRosetta platform by adding or rewriting helpful error and warning messages.

Updates and improvements to code will be immediately available for download from on pyrosetta.org, as the code is automatically compiled and posted daily. Each new workshop will be posted on pyrosetta.org, and at the end of the year we will publish the full, expanded workbook on Amazon.com. The new modules and code modifications will be tested during the Fall 2014 offering of ChemBE 540:414.