

Project Year

2012-2013

Project Title

PyRosetta: Advanced Biomolecular Modeling for Undergraduates

Project Team

Jeff Gray, Whiting School of Engineering, Department of Chemical and Biomolecular Engineering, Faculty
Jason Labonte, Whiting School of Engineering, Department of Chemical and Biomolecular Engineering, Post-Doctoral Fellow

Boon Uranukul, Whiting School of Engineering, Department of Chemical and Biomolecular Engineering, Fellow

Audience

The project will serve the course, (ChemBE) 540.414 *Computational Protein Structure Prediction and Design*, offered annually or biannually, enrolling 10-20 undergraduates from the departments of ChemBE, BME, Biophysics, Chemistry, Computer Science, and Applied Math. The improved educational materials will also benefit other universities that use our materials (including MIT, NYU and U-Kansas), and other users of the PyRosetta code (over 750 academic no-cost licensees).

Pedagogical Issue

PyRosetta is a Python library of the Rosetta computational methods for protein structure prediction and design originally developed in a 2008 CER Technology Fellows project, and currently used in the course 540.414. We propose to update and expand the package, sample scripts, and tutorial modules. These improvements will help Hopkins undergraduates master a broader scope of topics with a deeper level of comprehension of the underlying computer science behind the protein structure prediction and design calculations.

Solution

Upon completion of the project, we will achieve the following objectives:

1. Resolve errors in sample Python scripts on pyrosetta.org and build automated testing framework.
2. Add new sample scripts to pyrosetta.org for (i) modeling and designing nucleic acids, (ii) using non-canonical and post-translationally modified amino acids, (iii) custom score functions and protein manipulations, and (iv) loop prediction.
3. Add functionality to print internal data structures for several PyRosetta objects (MoveMap, PackTask, ConformerEnsemble).
4. Add new modules (matching sample scripts in goal 2) to the 2nd edition of The PyRosetta Interactive Platform for Protein Structure Prediction and Design workbook. Publish and distribute on Amazon.com.

Technologies Used

Python
Rosetta
PyRosetta

C++

Publishing through CreateSpace.com

Faculty Statement

PyRosetta is a Python library of the Rosetta computational methods for protein structure prediction and design. In 2008, the CER supported two Tech Fellows who drafted initial modules for teaching concepts of protein structure prediction and design in ChemBE 460, an undergraduate class. With this basis, I expanded the computational portion of the course into a more comprehensive computationally-focused class, ChemBE 414, *Protein Structure Prediction and Design*, taught in Spring 2009, 2010, and 2012. We currently have seven modules to teach topics in the field from protein structural analysis and visualization to protein folding and docking. The modules are bound in a workbook available through lulu.com, and the materials have been used at other universities including the University of Kansas and MIT. Students who have taken the prior courses have enrolled in graduate programs in computational biology at leading institutions including NYU and MIT.

While our PyRosetta modules have been very successful, there is a need to update them. The underlying Rosetta code has progressed rapidly, with new abilities to model nucleic acids, non-canonical amino acids, and custom scoring functions. Some of the prior modules have lost the ability to teach some tasks (e.g. flexible docking) because the functionality of the original code was expanded and changed. Additionally, there is an opportunity to 'go to the next level' with these modules by revealing more sophisticated data structures in the code. These improvements will help Hopkins undergraduates master a broader range of topics with a deeper level of comprehension of the underlying computer science behind the protein structure prediction and design calculations. In addition, by giving students the ability to create new PyRosetta objects, we will be able to teach more creative learning objectives.

The new computer functionality will be incorporated continuously into emerging releases of the PyRosetta package (automatically compiled and posted to pyrosetta.org nightly), which will be used in future sections of ChemBE 414. The modules will be used in class and shared more broadly by self-publishing on CreateSpace.com and distribution through Amazon.com.