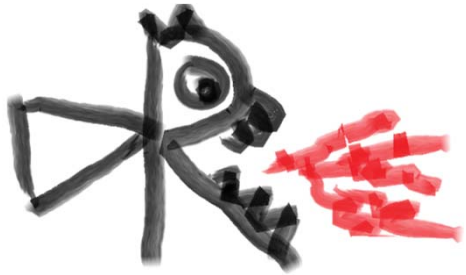


So…You Want to Teach Rosetta?
Or,
How to Bring Biomolecular Modeling to the World

Includes Time-saving Tips for Those on Track to Tenure!

Jeffrey J. Gray
Johns Hopkins University, Baltimore, MD
http://graylab.jhu.edu/courses/540.414

RosettaCON 2010
August 5, 2010

Course Objectives

Students should be able to:

1. Explain, interpret or modify classic algorithms in structure prediction and design
2. Use standard tools to model biomolecules *de novo* or by homology, dock biomolecules, and design biomolecules
3. Create new custom methods and algorithms for specific problems

http://graylab.jhu.edu/courses/540.414/docs/Learning_Objectives.pdf

ChemBE 414/614: Computational Protein Structure Prediction and Design (Spring 08, 09, 10)

- 15-25 Students
  (limited by size of computer classroom)

- About 1/3 Graduate, 2/3 Undergraduate
- ChemBE, Biophysics, BME, Biology, Chemistry
- Prereqs: familiarity with computers, biochemistry, thermodynamics
- Each week:
  - 75 minute lecture
  - 75 minute hands-on workshop
  - 1-2 literature articles to read, also reference data

Topics I: Fundamentals

1. Why do we want to engineer proteins?
2. Protein structure and geometry / PyMol
3. Molecular energies and forces / PyRosetta
4. Ab-initio folding / Monte Carlo
5. Refinement / Monte Carlo minimization
6. Side-chain prediction / Packing & design

Homework Week 1

- *Torsion space calculations*
  - Calculate the coordinates of the N atom of residue *i*+1, given that residue *i* has φ = -126.1° and ψ = 154.4°. Use bond lengths and angles from Engh & Huber *Acta Cryst.* (1991). To make the calculation simple, place the carbonyl C atom at the origin, the C_α on the negative *x*-axis, and the first N in the *xy* plane with a positive *y* coordinate. For completeness, also give the coordinates of the C_α and N atoms. [Ref: Parsons *et al.* 2005 *J. Comp. Chem.*]

Homework Week 2

- *Molecular energy calculations*
  - Use the **Lennard-Jones potential** to calculate the van der Waals energy between the side chains of Cetuximab heavy chain residue Y102 and EGFR residue Q408. Use the polar-hydrogen model of Neria *et al.* (CHARMm param19), and only consider the two closest atom pairs.
  - Calculate the **hydrogen bond energy** between the same two side chains, using the statistical potential of Kortemme *et al.*
  - Calculate the **solvation potential** energy between the same two side chains using the pair-wise model of Lazaridis & Karplus. Again consider only the closest three atoms.
  - Which type of energy contributes most to the Y102-Q408 interaction, van der Waals, solvation or hydrogen bonding? Is this a valid question? What is the reference state?

Homework Week 5

- *Ab initio* folding algorithm. Using your Monte Carlo energy optimization algorithm from Workshop 4, write a complete program that will fold a protein. A suggested algorithm involves low-resolution fragment insertion (first 9-mers, then 3-mers), followed by high-resolution refinement using small, shear, and minimization movers, as well as side-chain packing.

  Test your code by attempting to fold a zinc finger. How do your results compare with the crystal structure? If your lowest-energy conformation is different than the native structure, explain why this is so in terms of the limitations of the computational approach.

Topics II: Applications

7. Computational design of structure
8. Computational design of function
9. Protein-protein docking
10. Loop modeling & fold trees
11. Homology modeling
12. RNA structure prediction
13. Small-molecule docking & screening

Reading Rubric, Week 8

Choose one of the articles below and then answer the following questions. Be prepared to discuss your article in class on Tuesday.

1. What is (are) the authors' design goal(s)?
2. What is the scope of the sequence/conformation space searched?
3. What strategy do they use to search this space?
4. What energy function do they use?
5. What other tricks are necessary (*e.g.* additional filtering or steps for efficiency)?
6. How do they validate their design experimentally?
7. Did they meet their design goal?
8. Do *you* believe they met their design goal? Consider carefully the experimental data presented.
9. How reliable is the technique? *I.e.*, do all the designs work?
10. How is this work important in the field? You may need to learn about the other articles in class to answer this.

Seminal articles in the design of protein structure

Dahiyat & Mayo 1997.
Harbury, Plecs, Tidor, Alber & Kim 1998.
Kuhlman & Baker 2000.
Kuhlman, Dantas, Ireton, Varani, Stoddard & Baker 2003.
Hu, Wang, Ke & Kuhlman 2007.

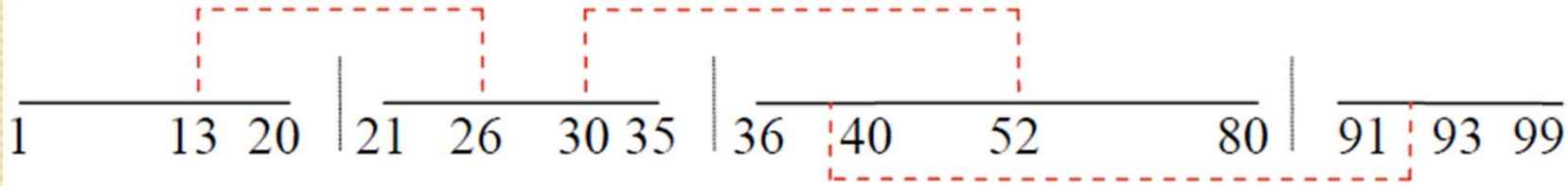
Homework Week 9

- Explain the difference between conformer selection and induced fit binding.

- How would *peptide*-protein docking need to differ from protein-protein docking?

Homework Week 10

- *Fold Tree*
  - Given the fold tree below, which residues would move (relative to the rest of the protein) if:

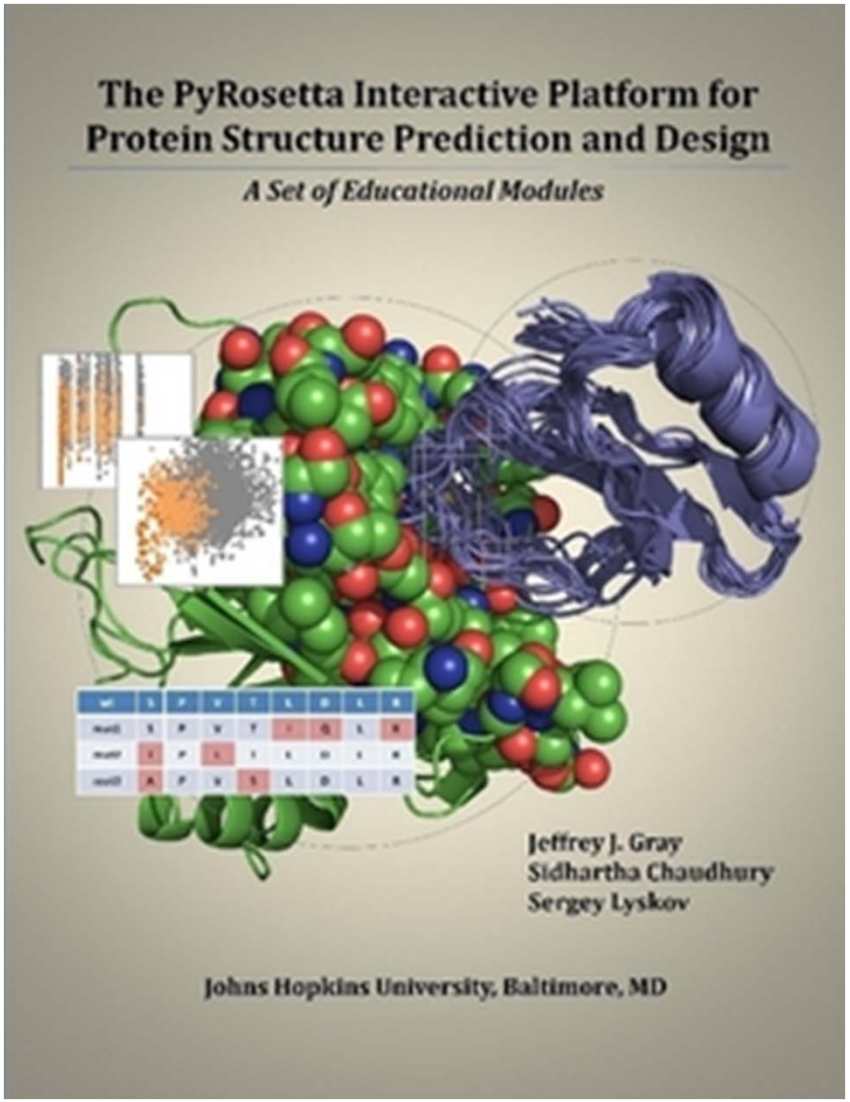


  - ϕ of residue 15 was changed
  - χ_1 of residue 61 was changed
  - ψ of residue 38 was changed
  - ω of residue 45 was changed

PyRosetta Workshop Book

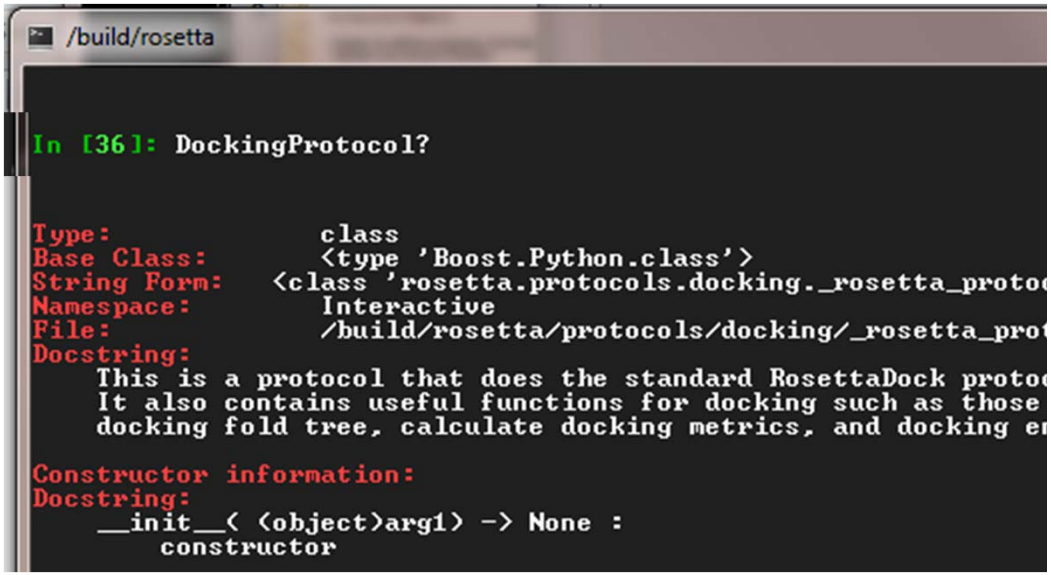
Paperback (lulu.com) or free PDF download (www.pyrosetta.org)

Table of Contents



Getting Help

- iPython
  - **`help(command)`**
  - **`command?`**
  - **`command??`**
- PyRosetta website
  - User's Manual
  - Sample Scripts
- Class GoogleGroup!
- RosettaCommons
  - User Forums
  - Doxygen

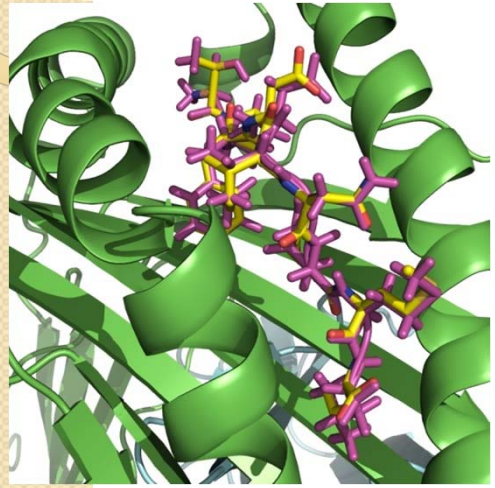


Hello everyone,
Maybe someone can help me out with this but I am trying
jd=JobDistributor("output",10,scorefxn_low) and I keep o
error saying that JobDistributor is not defined.
-Ben

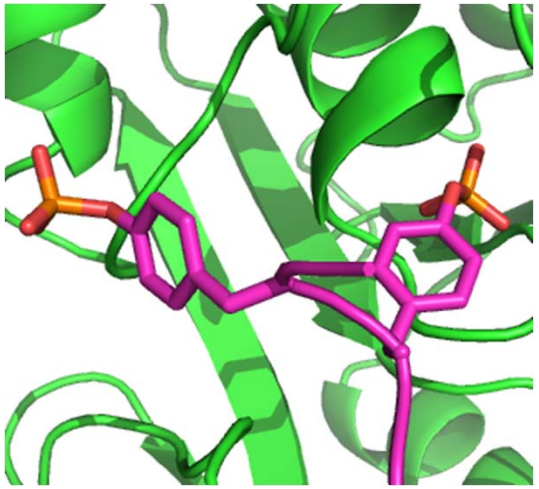
Hi Ben,
Even I am getting the same error. Also when I try tab comp
JobDistributor option. This indicates that there is probably
called JobDistributor.
Moiz

PyJobDistributor
-- Brian Weitzner

Example Projects



MHC peptide specificity and design (Y. Zhang and Y.C. Law)



Phosphotyrosine modeling (A. Chen)

Statistically biased loop-modeling (N. Tippens)

Graduate students present to class in the last week

PyRosetta
Interactive Molecular Modeling for Proteins

Issues

- Varying levels of programming skills
  - Increase prerequisites?
  - Give short intro to programming?
- Student workload is front-heavy (energy and by-hand structure calculations)
- PyRosetta hurdles…

Future Directions

- PyRosetta fixes
  - Cleaner exits on errors
  - More complete access to data / state
    - HBond energies, rotamer states, etc.
  - More coverage of namespaces
    - KIC, EnzymeDesign
  - PyMOL-PyRosetta integration
- Split into intro/advanced courses?
- Writing a full textbook?

Acknowledgements

PyRosetta β-Testers ("Students") of ChemBE 414/614

TAs:

- Sid Chaudhury
- Brian Weitzner

Support:

- Sergey Lyskov
- Evan Baugh
- J.D. Bagert & Julian Rosenberg

Lecture Material

- Brian Kuhlman, Carol Rohl, Jens Meiler, Phil Bradley
- Rosetta Scientists!


Pravin Muthu


Liza Lee


Justin Porter


Sid Chaudhury


Brian Weitzner


Evan Baugh


Sergey Lyskov

http://graylab.jhu.edu/courses/540.414
http://pyrosetta.org