

Workshop #3: PyRosetta Scoring

Scoring Poses

A basic function of Rosetta is calculating the energy or *score* of a biomolecule. Rosetta has a standard energy function for all-atom calculations as well as several scoring functions for low-resolution protein representations. In addition, you can tailor an energy function by including scoring terms of your choice with custom weights.

For these exercises, use the protein ras (PDB ID 6Q21) and load it into a pose called `ras`. Be sure to clean the PDB file and use only one chain.

1. To score a protein, you must first define a scoring function:

```
scorefxn = create_score_function('standard')
```

The option for `standard` tells Rosetta to load the standard all-atom energy terms. To see these terms, you can print the score function:

```
print scorefxn
```

What terms are in the score function, and what are their relative weights?

2. Set up your own custom score function which includes just van der Waals, solvation, and hydrogen bonding terms, all with weights of 1.0. Use the following to start:

```
scorefxn2=ScoreFunction()  
scorefxn2.set_weight(fa_atr,1.0)  
scorefxn2.set_weight(fa_rep,1.0)
```

Enter the other weights and then confirm that the weights are set correctly.

3. Evaluate the energy of `ras` with the standard score function:

```
scorefxn(ras)
```

What is the total energy of `ras`?

Break the energy down into its individual pieces:

```
scorefxn.show(ras)
```

Which are the three most dominant contributions, and what are their values? Is this what you would have expected? Why?

4. Break the energy down into each residue's contribution:

```
print ras.energies().show(residue_number)
```

What are the van der Waals, solvation, and hydrogen bonding contributions of residue 24?

5. Analyze the energy between residues Y102 and Q408 in Cetuximab (1YY9). You'll need to load in that structure.
 - a. Internally, a pose object has a list of residues, numbered starting from 1. To find the residue number of Y102 and Q408 of chain A, use the residue chain identifier and the PDB residue number to convert to the pose numbering:

```
pose.pdb_info().pdb2pose("A",102)
```

- b. Score the pose and determine the Van der Waals energy and solvation energy between these two residues. Use the following commands to isolate contributions from particular pairs of residues, where `rsd1` and `rsd2` are the two residue objects of interest (not the residue number – use `pose.residue(resnum)` to access the object):

```
emap = core.scoring.TwoBodyEMapVector()  
scorefxn.eval_ci_2b( rsd1, rsd2, pose, emap )  
print emap[fa_atr]  
print emap[fa_rep]  
print emap[fa_sol]
```

- c. How do Rosetta's calculations compare to the ones you completed by hand (see references)?

- d. Create a new PDB file containing coordinates for just the two residues Y102 and Q408. Repeat the above calculations. Which energies change? Why?

Programming exercises

1. *Interface energy.* Write a program that can calculate the binding energy of EGFR to Cetuximab. You will need to make separate PDB files for the antigen, antibody, and complex. In PyMOL, select one of these peptides, then use File→Save Molecule.

Use the following formula for binding energy:

$$E_{\text{binding}} = E_{\text{complex}} - E_{\text{antibody}} - E_{\text{antigen}}$$

Submit your script along with output values for the total binding energy, Van der Waals, hydrogen bonding and solvation energies.

What does your result suggest about these two proteins *in vitro*? What are some inaccuracies in the way you've calculated the binding energy?

2. *Statistical energy functions.* Write a program to create a histogram of the C-N bond lengths observed in a set of ten high-resolution x-ray protein structures. (One source of curated structures is the WHATIF sets at <http://swift.cmbi.kun.nl/swift/whatif/select>).

 - a. Plot the data as a histogram of probability versus bond length, and also as a statistical free energy versus bond length. Try a bin size of 0.5 Å.
 - b. Look up the CHARMM potential for this bond stretch, and plot a curve over each of your figures to show the CHARMM model for this motion.
 - c. Do the statistics match that which would be produced by a harmonic oscillator under the CHARMM potential? Specifically, is the average bond length and the CHARMM spring constant K correct? If not, what should it be? You may need to fit a parabola to your data to find the average bond length and the spring constant K .

References

1. E. Neria, S. Fischer & M. Karplus, "Simulation of activation free energies in molecular systems," *J. Chem. Phys.* **105**, 1902-1921 (1996).
2. T. Kortemme, A. V. Morozov & D. Baker, "An orientation-dependent hydrogen bonding potential improves prediction of specificity and structure for proteins and protein-protein complexes," *J. Mol. Biol.* **326**, 1239-1259 (2003).
3. D. Eisenberg & A. D. McLachlan, "Solvation energy in protein folding and binding," *Nature* **319**, 199-203 (1986).
4. T. Lazaridis & M. Karplus, "Effective energy function for proteins in solution," *Proteins* **35**, 133-152 (1999).