Molecular Modeling 2014. e-Review session

Final exam will be held Tuesday May 13, 2014 in rm 2C13

Structure of the final exam will be..

30% will be multiple choice questions. These will test concepts and terminology.
40% will be a MOE exercise that requires you to run one or more of the tools that we
used in class, in homework assignments, or in your term project.
30% will be short essays, graded based on (content + clarity)/length.

Tasks you should be able to do, and questions you should be able to answer, either by
multiple choice or short essay:

First half of course
- Use SCOP-based terminology to describe a given protein.
- Identify contacts in 3D space based on a 2D contact map.
- Identify 2D contacts in a contact map based on a TOPS diagram.
- Draw a TOPS diagram based on 3D coordinates.
- What does resolution mean? R-factor?
- What are examples of internal coordinates?
- Describe the amphipathic (hydrophobic/hydrophilic) patterns found in alpha
  helices and beta strands and explain why they exist.
- Describe the database loop search method for modeling indels.
- Build a small molecule based on a drawing.
- Make and break bonds.
- Create and use distance restraints.
- Edit a sequence.
- Edit an alignment.
- Run the Homology model script.
- Assess the quality of a homology model.
- Understand the statistical and physical bases of confidence assessment.
- Point out errors in a model.
- Find void spaces.
- Add water molecules to a model.

Second half of course

Electrostatic surfaces
- Define the charge distribution.
- Define the electrostatic potential
- What is the "dielectric"?
- Describe the assumptions built into the Poisson-Boltzmann equation.
- Describe the Poisson-Boltzman grid-based approach for solving the electrostatic
  potential.
Identify and understand terms in the Poisson-Boltmann equation.
Understand the entropy/enthalpy differences between water molecules on a hydrophobic surface and waters on a hydrophilic surface
Understand the basis of the hydrophobic effect.
Why does Nature abhor a void?

**Force fields and energy minimization**
- Why is modeling solvation important?
- What is "implicit" versus "explicit" solvation?
- What atom-based parameters are used in the Generalized Born model?
- How does the Monte Carlo algorithm work?
- What is the energy gradient?
- How is the gradient used in energy minimization?
- What is a "local minimum"?
- What is "simulated annealing"? How does it overcome the local minimum problem?

**Protein folding**
- What is a protein folding pathway? Initiation site? Nucleation site?
- What did Anfinsen do?
- What did Levinthal prove?
- How do we know early folding units exist?
- There are two approaches to protein structure prediction.
- What is a knowledge-based potential?
- What assumptions do we make in converting probability to free energy?
- What is the equation for free energy as a function of probability?
- How does Rosetta work?

**Molecular dynamics**
- What does MOE's Structure Preparation do?
- What does Protonate3D do?
- What is a periodic box?
- What is a micro-canonical ensemble? What assumptions are available?
- How is a micro-canonical ensemble enforced?
- What are the dimensions of an "energy landscape"?
- What does periodic motion in a MD trajectory say about the local shape of the energy landscape?
- Why do you always energy minimize before running MD?
- Describe a way to cluster a MD trajectory? What are the clusters?
- How can we measure the barrier heights between states?

**Normal mode analysis and principle component analysis**
- Why is it important to model motions in proteins?
- What assumptions are built into NMA?
- What input data is required for NMA?
What does one normal mode consist of? (i.e. What is the data structure?)
How is the first normal mode determined?
Compare and contrast NMA and PCA?
What does it mean that Principle components, or Normal modes, represent a "basis set for $\mathbb{R}^n$ (real space, N-dimensional)?

Docking
What is a "pose"?
What is a "descriptor" in descriptor-based docking?
How many parameters must be searched in energy-based docking?
What does "representation" mean in energy based docking?
What are the descriptors in the UCSF DOCK algorithm?
Be familiar with Receptor/Ligand terminology.
What kind of search is done by ContextShapes? Zdock? PatchDock?
What is the difference between the "bound" state and the "unbound" state?
What does MOE's docking tool do? Is it appropriate for use with a protein ligand?
How can NMA inform us about the ligand binding site or binding pathway for a small ligand, such as seratonin or aspirin?

Protein design
What are the advantages of protein-based drugs over small molecular drugs?
What is the Dead-End Elimination algorithm, in words?
What is a "rotamer"? What does rotamer probability mean? Why is the probability "backbone dependent"?
What sort of theoretical approaches and experimental methods have been used in rational (i.e. non-computational) protein design?
What are the assumptions built into computational protein design using the Dead-End Elimination approach? Can energy be accurately decomposed into pair-wise terms? Why not?

Drug design
What is Lapinki’s rule of 5?
What information do we need to begin to engage the "rational drug design" approach?
Describe the approach we took in class to calculate the energy difference between two drugs and two homolog receptors, using the MOE Dock tool and Rotamer Explorer.
Why do drugs typically fail in Phase three trials?
What does MOE's interaction map tell you?
**Not covered on the final exam**
- Surface area calculation algorithms
- Conjugate gradient
- Newton Raphson and other 2nd deriv-based methods
- Fold-It
- I-sites (except the principle of folding initiation sites)
- Sources of small molecules for drug screening
- ContextShapes, ZDock, PatchDock (except as above)
- DHFR function
- COX-1,2 function