

Exam 1 Review

COBB2035 Modern Methods for Structure Based Drug Discovery

Questions about the assignments will be on the exam. Review your submitted code, figures, calculations, and written responses. Questions will be more conceptual or will be trivial to answer if you did the work yourself.

Lecture 0: Introduction to Computational Drug Discovery

- clinical trial phases
- ADMET, pharmacokinetics, pharmacodynamics, QSAR, QSPR, activity cliff: definitions
- SMILES: how to interpret simple molecules (branches, rings)
- Fingerprints: algorithm, Tanimoto
- Lipinski's rule of five
- molecular representations

Lecture 1: Proteins: They are kind of important

- **amino acids - know all 20**
 - recognize structure (do not need to draw structure)
 - associate names, 3 letter abbreviation and 1 letter abbreviation
 - properties (charged, hydrophobic, polar)
 - "special" (why?): glycine, proline, cysteine
 - pKa - for which amino acid(s) does it matter?
- N-terminal/C-terminal (order)
- **molecular interactions**
 - hydrophobic (where are hydrophobic residues more likely to found)
 - hydrogen bonds
 - charge-charge (salt bridges)
 - aromatic
 - cation pi
 - should be able to identify above from structure
- backbone dihedral angles; Ramachandran plots
- primary, secondary, tertiary, quaternary structure; domains (know definition)
- alpha helix, parallel/antiparallel beta sheets (be able to identify)

Lecture 2: Structure Determination: Where do these come from anyway?

- x-ray crystallography - high level understanding of process
 - unit cell, asymmetric unit
 - B-factors, resolution, R-value (interpretation as user)

- Fourier transform
- phase problem, molecular replacement
 - selenium
- NMR - high level understanding of process
 - generates ensemble of models consistent with experimental data
- CryoEM - high level understanding of process
 - resolution
 - ability to resolve multiple states

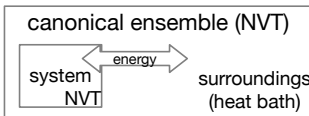
Lecture 3-4: Mechanics

Statistical Mechanics

microstates vs (macro)states

$$S = k_B \log \Omega$$

microcanonical ensemble (NVE)



$$p(x) \propto e^{-\frac{U(x)}{k_B T}} = e^{-\beta U(x)}$$

$$\beta = \frac{1}{k_B T} = \frac{1}{k_B} \left(\frac{\partial S}{\partial E} \right)_{V,N}$$

Partition Function

$$Z = \frac{m}{h} \int \int e^{-\frac{U(x) + \frac{mv^2}{2}}{k_B T}} dx dv$$

$$= \frac{1}{\lambda} \hat{Z} \quad \text{de Broglie wavelength} \quad \lambda = h / \sqrt{2\pi m k_B T}$$

Configurational Partition Function

$$\hat{Z} = \int e^{-\frac{U(x)}{k_B T}} dx$$

$$\langle U \rangle = \hat{Z}^{-1} \int_V U(x) e^{-\frac{U(x)}{k_B T}} dx$$

probability of A state

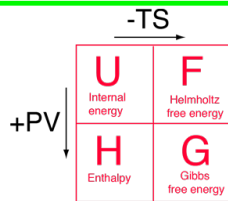
$$p_A = \hat{Z}^{-1} \int_{V_A} e^{-\frac{U(x)}{k_B T}} dx \propto e^{-\frac{F_A}{k_B T}}$$

$$F_A = -k_B T \log Z_A$$

$$\Delta G$$

$$\text{Gibb's Free Energy} \quad \Delta G = \Delta H - T \Delta S$$

$$\text{Helmholtz's Free Energy} \quad \Delta F = \Delta U - T \Delta S$$



$$\Delta G^\circ = RT \log K_d$$

$$K_d = \frac{[A][B]}{[AB]} \quad R = N_A k_B$$

Entropy

$$S = \frac{\langle E \rangle - F}{T}$$

$$e^{-\frac{F}{k_B T}} = e^{\frac{S}{k_B}} e^{-\frac{\langle E \rangle}{k_B T}}$$

$$S = -k_B \sum p_i \log p_i \quad \text{Gibbs}$$

$$H = - \sum_i p_i \log_2 p_i \quad \text{Shannon}$$

Classical Mechanics

velocity $v = \frac{dx}{dt}$ kinetic energy $\text{KE} = \frac{1}{2} m v^2 = \frac{p^2}{2m}$

momentum $p = mv$

$F = ma = \frac{dp}{dt}$ potential energy $U_{\text{pot}}(x) = - \int_{\infty}^x F(r) dr$

$$F = -\nabla_x U$$

physical space
configuration space
phase space

Newtonian
Lagrangian
Hamiltonian

$$\mathcal{H} = \text{KE} + U$$

Equations of Motion

$$\frac{dp}{dt} = - \frac{\partial \mathcal{H}}{\partial x}$$

$$\frac{dx}{dt} = \frac{\partial \mathcal{H}}{\partial p}$$

Verlet Integration

$$\mathbf{x}(t + \Delta t) = 2\mathbf{x}(t) - \mathbf{x}(t - \Delta t) + \frac{F}{m} \Delta t^2$$

Langevin Motion

$$\frac{dp}{dt} = - \frac{dU}{dx} - \gamma m \frac{dx}{dt} + \mathbf{R}$$

Langevin Integration

$$\mathbf{v}(t + \Delta t/2) = \mathbf{v}(t - \Delta t/2) \alpha + \frac{\mathbf{F}(\mathbf{x}(t))(1 - \alpha) + \sqrt{k_B T(1 - \alpha^2)}/m \mathbf{R}}{\Delta t}$$

$$\mathbf{x}(t + \Delta t) = \mathbf{x}(t) + \mathbf{v}(t + \Delta t/2) \Delta t$$

$$\alpha = \exp(-\gamma \Delta t)$$

ergodic hypothesis

- you should understand all the above relations and how they fit together
- you should have memorized the relations in green boxes and be prepared to apply them
- configuration vs phase space
- Hamiltonian
- microcanonical (NVE), canonical (NVT), NPT
- Boltzmann distribution
- partition function
- how to compute ensemble properties
- equipartition theorem (why do we tend to ignore KE in our calculations?)

- **double square well potentials**

Lecture 5: Mechanics 3

- entropy/enthalpy of water
- entropy/enthalpy compensation
- thermodynamic quantities as derivatives (you will not need to memorize or derive these)
- heat capacity and melting temperature
- intuitive understanding of toy 2D systems

Lecture 6: Mechanics 4

- Jacobian
- Euler's method
- Taylor series
- Verlet (error terms, what does it calculate)
- leapfrog integrators
- Langevin (what are the different terms)
- Overdamped (Brownian)

Lecture 7: Practical Simulation

- ion channels
- topology, trajectory (vocab)
- water models
- periodic systems (why must be neutral?)
- force fields; should be able to recognize (not regurgitate) all terms of Amber
- PME (high-level understanding - what does it compute)

Lecture 8: More Practical Simulation

- should be able to tell from plots of energy/temp/volume what ensemble is simulated
- barostats
- rms, alignment, RMSF
- effect of choice of alignment

Lecture 9: States and Kinetics

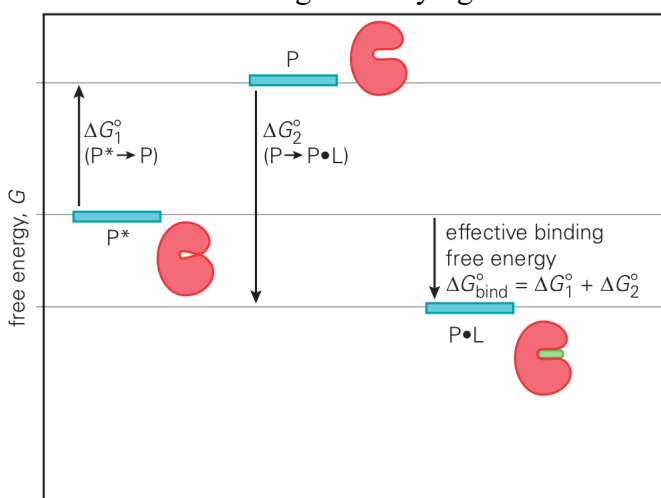
- state definition
- GPCRs
- reaction/progress coordinates
- potential of mean force (mathematical and intuitive visual understanding)
- two state kinetic model
- first passage rate, MFPT, steady state rate
- committor

Lecture 10: Markov State Models

- transition matrix (should be able to apply a small one)
- eigenvectors (esp. with eigenvalue 1)
- k-means clusters
- PCA vs TICA
- implied timescales

Lecture 11: Structural Binding and Alchemical Free Energy

- standard state
- orthosteric vs allosteric
- lock&key, conformational selection, induced fit (1-D cartoons)
- understand what this figure is saying:



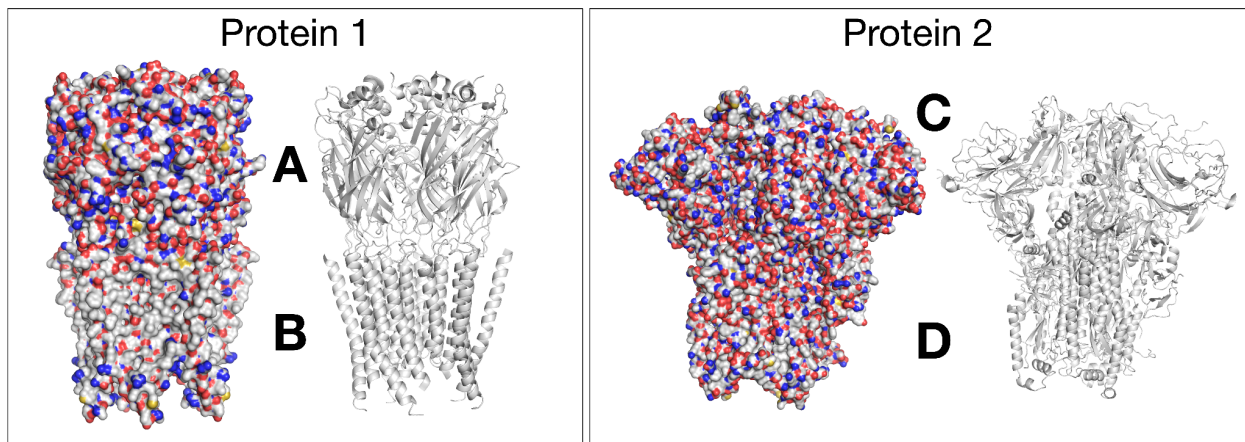
- **components of binding** (decompose ΔG into enthalpic/entropic and receptor/ligand/solvent contributions)
- **thermodynamic cycles**
 - absolute vs relative free energy cycle of ligand binding
 - other cycles (e.g. recitation)
- alchemical transformations, order parameter
- Zwanzig

Lecture 12: FEP and MCMC

- BAR intuition
- TI intuition
- Hamiltonian replica exchange
- **Metropolis Monte Carlo**
 - properties of MCMC and contrast with molecular dynamics
 - acceptance probability, proposal probability, move set
 - detailed balance

Sample Questions

1. Name two amino acids that can form a salt bridge with one another.
2. Which mutation is **least** likely to significantly change the folding free energy of a protein?
 - () R→E
 - () V→W
 - () A→Q
 - () V→I
3. Assuming sufficient simulation time, which of the following processes can **not** be observed in a molecular dynamics simulation?
 - () The disassociation of two proteins
 - () The folding of a protein
 - () The construction of RNA by RNA polymerase
 - () The conformational sampling of residue side chains or short loop regions
4. Consider the two proteins shown below rendered both with surface and cartoon styles. Both proteins have a predominantly beta-sheet domain and a predominantly alpha helix domain. There is only one transmembrane domain. Which one is it?



5. Consider a mutation of a solvent accessible serine (WT) on the surface of a protein to an lysine (mut) and the decomposition of the free energy change into enthalpic and entropic terms and solvent-solvent, solvent-protein, and protein-protein terms. Assume the protein does not undergo a significant conformational change.

5a. What is the most likely relationship between the entropy of the protein in the two systems?

$$\begin{aligned} S_{\text{protein}}^{WT} &< S_{\text{protein}}^{\text{mut}} \\ S_{\text{protein}}^{WT} &= S_{\text{protein}}^{\text{mut}} \\ S_{\text{protein}}^{WT} &> S_{\text{protein}}^{\text{mut}} \end{aligned}$$

5b. What is the most likely relationship between the entropy of the water in the two systems?

$$\begin{aligned} S_{\text{water}}^{WT} &< S_{\text{water}}^{\text{mut}} \\ S_{\text{water}}^{WT} &= S_{\text{water}}^{\text{mut}} \\ S_{\text{water}}^{WT} &> S_{\text{water}}^{\text{mut}} \end{aligned}$$

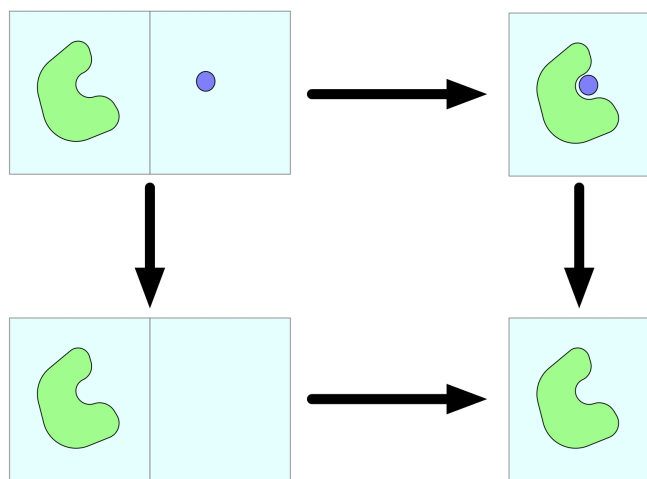
6. Circle the terms of the AMBER force field below that are scaled during an alchemical free energy calculation.

$$\begin{aligned} U(r^N) = & \sum_{\text{bonds}} k_b(l - l_0)^2 + \sum_{\text{angles}} k_a(\theta - \theta_0)^2 + \sum_{\text{torsions}} \frac{1}{2} V_n [1 + \cos(n\omega - \gamma)] \\ & + \sum_{j=1}^{N-1} \sum_{i=j+1}^N \left\{ \epsilon_{i,j} \left[\left(\frac{r_{0ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{r_{0ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right\} \end{aligned}$$

7. **True or False.** Kinetic trap states will affect the mean first passage time but not the steady state rate.

8. **True or False.** The probability of a microstate x is $e^{-\frac{U(x)}{k_B T}}$

9. Circle the arrows in the thermodynamic cycle below that would be simulated in an alchemical free energy calculation.



10. Which distribution is most disordered (highest entropy)?

