Molecular simulation and structure prediction using CHARMM and the MMTSB Tool Set

Free Energy Methods

Charles L. Brooks III

MMTSB/CTBP

2006 Summer Workshop

CHARMM Simulations

The flow of data and information in a CHARMM simulation

Input

- Parameters, topologies
- Sequence, coordinates

Generate

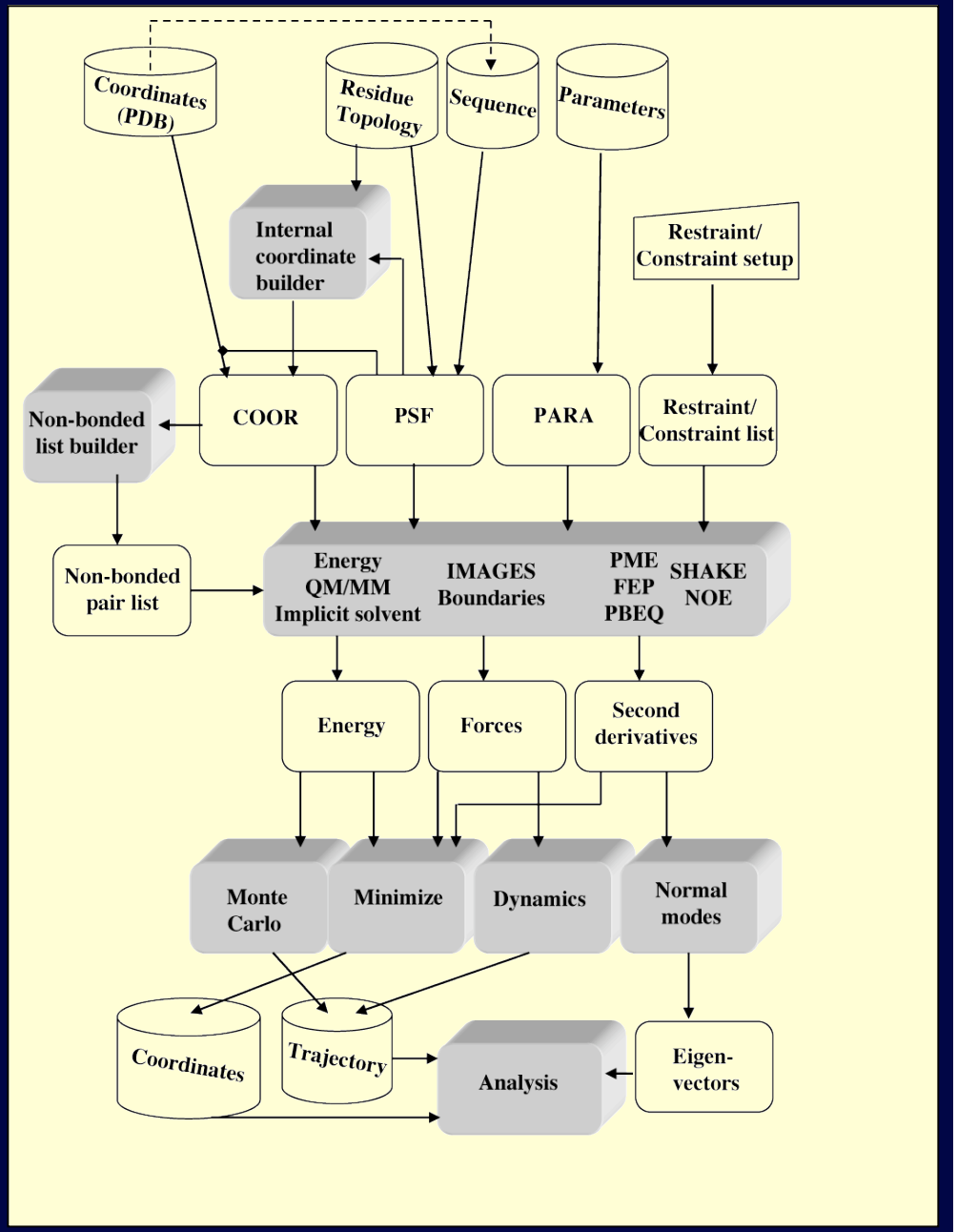
- Protein structure file
- Internal coordinates
- Missing atoms

Input

- Restraints/constraints

Calculate/Analyze

- Manipulate system
  - Energy, dynamics, normal modes, Monte Carlo, minimization
  - Analyze trajectories



The scope and range of methodologies and systems studied with CHARMM

Simulation methodologies

- Classical force field methods
  - Minimization, molecular dynamics, Langevin dynamics, Monte Carlo, normal mode calculations
- Quantum and QM/MM
  - Internal semi-empirical AM1/PM3 and DFFTB methods
  - Interfaces with numerous QM packages
- Sampling methods
  - Umbrella sampling
  - FEP/TI
  - Path and NEB (string of states methods)
  - Multi-canonical and replica approaches

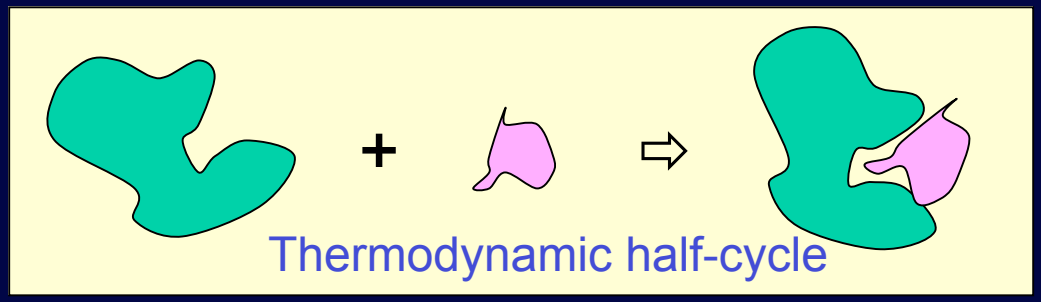
Force fields and systems

- All atom force fields w/ (for proteins and lipids) and w/o atomic polarization
- Extended (polar hydrogen only) force fields
- Many coarse-grained, multi-resolution representations are feasible

Free Energy Simulation Methods

Free Energy Methods

- Thermodynamic cycles define free energy methods
  - Consider ligand binding to a receptor



  - The free energy change for this process is:

$$\Delta G(T,P) = G_{P:L(aq)}(T,P) - G_{P(aq)+L(aq)}(T,P)$$

$$= [U_{P:L(aq)} - TS_{P:L(aq)}] - [U_{P(aq)+L(aq)} - TS_{P(aq)+L(aq)}]$$

$$= \Delta U_{P:L(aq)-[P(aq)+L(aq)]} - T\Delta S_{P:L(aq)-[P(aq)+L(aq)]}$$

Free Energy Methods

- Thermodynamic cycles define free energy methods
  - The free energy change for this process is:

$$\Delta G(T,P) = G_{P:L(aq)}(T,P) - G_{P(aq)+L(aq)}(T,P)$$

$$= [U_{P:L(aq)} - TS_{P:L(aq)}] - [U_{P(aq)+L(aq)} - TS_{P(aq)+L(aq)}]$$

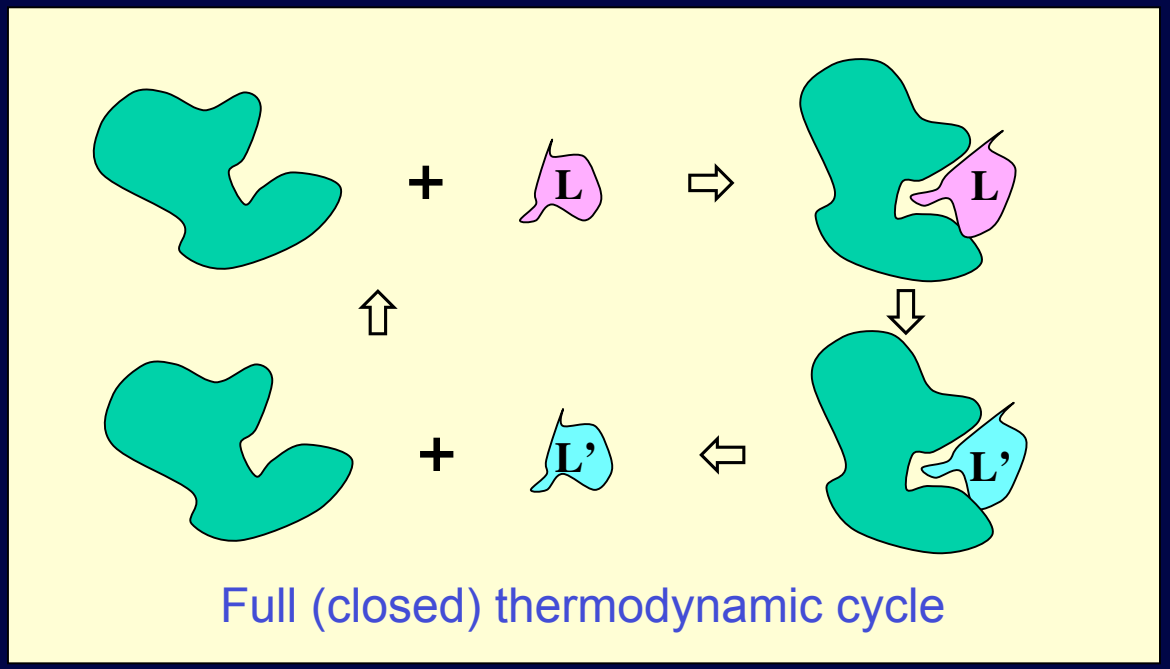
$$= \Delta U_{P:L(aq)-[P(aq)+L(aq)]} - T\Delta S_{P:L(aq)-[P(aq)+L(aq)]}$$

➡ ΔU - provides connections to energy change between ligand and protein, e.g. w/o protein relaxation and ligand conformational changes, this is the INTERACTION ENERGY between protein and ligand!

➡ ΔS - represents change in solvent entropy (h-ϕ effect), and ligand/protein degrees of freedom (+ trans/rot)

Free Energy Methods

- Relative binding affinities define closed thermodynamic cycles
  - Consider two ligands binding to a common receptor



Full (closed) thermodynamic cycle

Free Energy Methods

– The free energy change for this process is:

$$0 = \Delta G_{P:L(aq)-[P(aq)+L(aq)]} + \Delta G_{P:L'(aq)-P:L(aq)} - \Delta G_{P:L'(aq)-[P(aq)+L'(aq)]} - \Delta G_{L(aq)-L'(aq)}$$

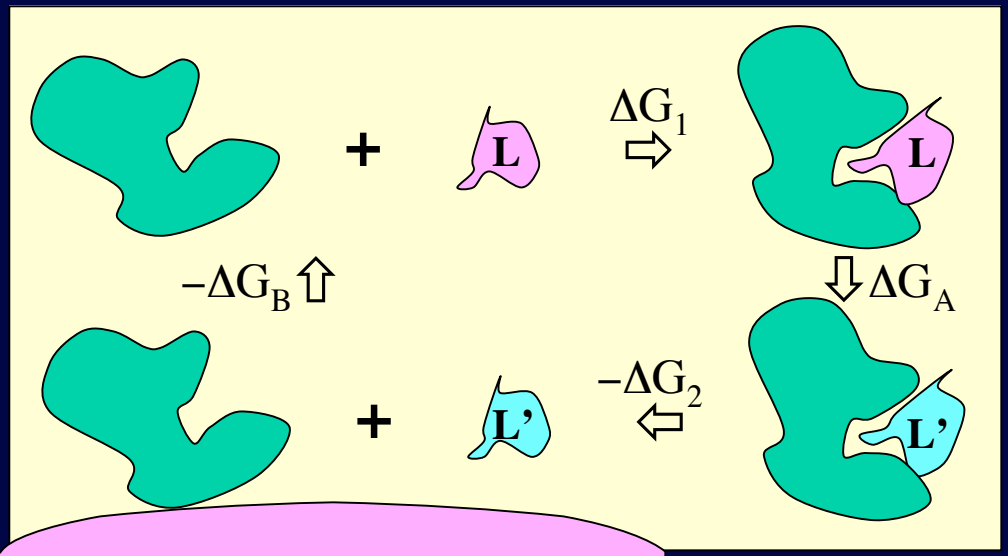
$$0 = \Delta G_{P:L(aq)-[P(aq)+L(aq)]} - \Delta G_{P:L'(aq)-[P(aq)+L'(aq)]} + \Delta G_{P:L'(aq)-P:L(aq)} - \Delta G_{L(aq)-L'(aq)}$$

$$0 = \Delta\Delta G_{\substack{P:L(aq)-[P(aq)+L(aq)] \\ P:L'(aq)-[P(aq)+L'(aq)]}} + \Delta\Delta G_{\substack{P:L'(aq)-P:L(aq) \\ L(aq)-L'(aq)}}$$

$$\Delta\Delta G_{\substack{P:L(aq)-[P(aq)+L(aq)] \\ P:L'(aq)-[P(aq)+L'(aq)]}} = \Delta\Delta G_{\substack{L(aq)-L'(aq) \\ P:L'(aq)-P:L(aq)}}$$

Free Energy Methods

- Free energy perturbation theory uses non-physical legs of thermodynamic cycle



$$\Delta\Delta G_{L,L'} = \Delta G_1 - \Delta G_2 = \Delta G_A - \Delta G_B$$

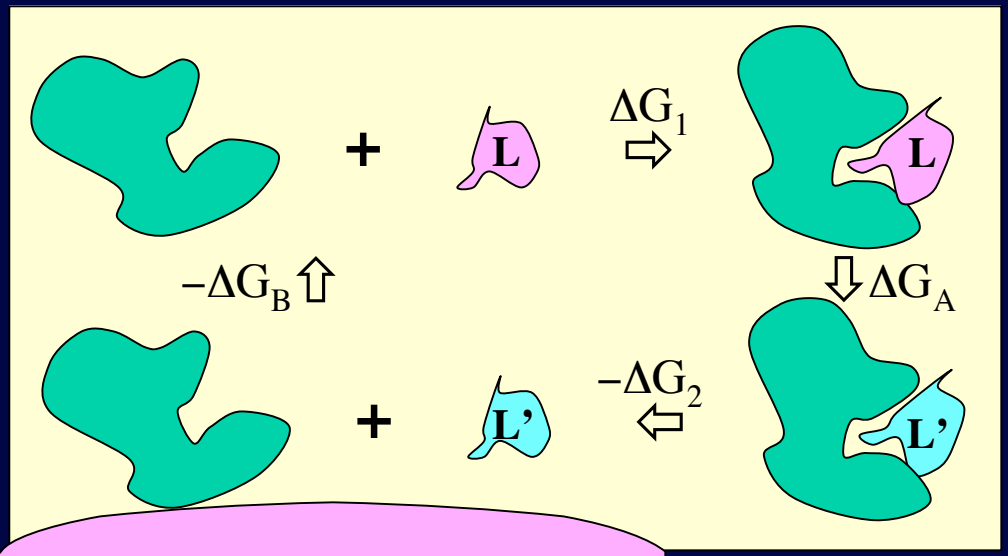
R→P:L, P →P:L'

$$U_{hybrid}(\Gamma, \chi_R, \chi_P, \lambda) = (1-\lambda)\cdot U_R(\chi_R, \Gamma) + \lambda \cdot U_P(\chi_P, \Gamma) + U_{env}(\Gamma)$$

$$\Delta G_{R\to P}(\lambda, \lambda') = -k_B T \cdot \ln\left\langle \exp\left(\frac{-[U_{hybrid}(\lambda') - U_{hybrid}(\lambda)]}{k_B T} \right) \right\rangle_{\lambda}$$

Free Energy Methods

- Thermodynamic integration theory uses non-physical legs of thermodynamic cycle



$$\Delta\Delta G_{L,L'} = \Delta G_1 - \Delta G_2 = \Delta G_A - \Delta G_B$$

R→P:L, P →P:L'

$$U_{hybrid}(\Gamma, \chi_R, \chi_P, \lambda) = (1-\lambda)\cdot U_R(\chi_R, \Gamma) + \lambda \cdot U_P(\chi_P, \Gamma) + U_{env}(\Gamma)$$

$$\Delta G_{R\to P}(\lambda, \lambda') = \int_{\lambda}^{\lambda'} \left\langle \frac{\partial U_{hybrid}}{\partial \eta} \right\rangle_{\eta} d\eta$$

Free Energy Methods - Biased (Umbrella) Sampling

- Consider a system with potential energy function $U_0(r_0,\ldots,r_N)$ and the expression of the Helmholtz free energy for this system

$$Q_0 = \int_{r_0}\cdots\int_{r_N} \exp[-U_0(r_0,\ldots,r_N)/k_BT] = \exp[-A_0/k_BT]$$

- From this expression we can also construct "projections" of the free energy onto interesting "subspaces", e.g., particular coordinates. These projections take the form distributions, and give rise to potentials of mean force:

$$\rho_0[f(r_0,\ldots,r_N)] = \frac{\int_{r_0} dr_0\cdots\int_{r_N} dr_N \exp[-U_0(r_0,\ldots,r_N)/k_BT]\cdot\delta[f(r_0,\ldots,r_N)]}{\int_{r_0} dr_0\cdots\int_{r_N} dr_N \exp[-U_0(r_0,\ldots,r_N)/k_BT]} = \exp[-W_0[f(r_0,\ldots,r_N)]/k_BT]$$

- Suppose we now introduce an additional (biasing) potential $V[f(r_0,\ldots,r_N)]$

Free Energy Methods - Biased (Umbrella) Sampling

- We can now write ρ_0 as

$$\rho_0[f(\gamma,\Gamma)] = \frac{\exp\{\beta\cdot V[f(\gamma,\Gamma)]\}\cdot\int_\gamma d\gamma\int_\Gamma d\Gamma\exp[-\beta\cdot\{U_0(\gamma,\Gamma)+V[f(\gamma,\Gamma)]\}]\cdot\delta[f(\gamma,\Gamma)]}{\int_\gamma d\gamma\int_\Gamma d\Gamma\exp\{\beta\cdot V[f(\gamma,\Gamma)]\}\cdot\exp[-\beta\cdot\{U_0(\gamma,\Gamma)+V[f(\gamma,\Gamma)]\}]}$$

$$\cdot\frac{\int_\gamma d\gamma\int_\Gamma d\Gamma\exp[-\beta\cdot\{U_0(\gamma,\Gamma)+V[f(\gamma,\Gamma)]\}]}{\int_\gamma d\gamma\int_\Gamma d\Gamma\exp[-\beta\cdot\{U_0(\gamma,\Gamma)+V[f(\gamma,\Gamma)]\}]}$$

$$= \frac{\exp\{\beta\cdot V[f(\gamma,\Gamma)]\}\cdot\int_\gamma d\gamma\int_\Gamma d\Gamma\exp[-\beta\cdot\{U_0(\gamma,\Gamma)+V[f(\gamma,\Gamma)]\}]\cdot\delta[f(\gamma,\Gamma)]}{\int_\gamma d\gamma\int_\Gamma d\Gamma\exp[-\beta\cdot\{U_0(\gamma,\Gamma)+V[f(\gamma,\Gamma)]\}]}$$

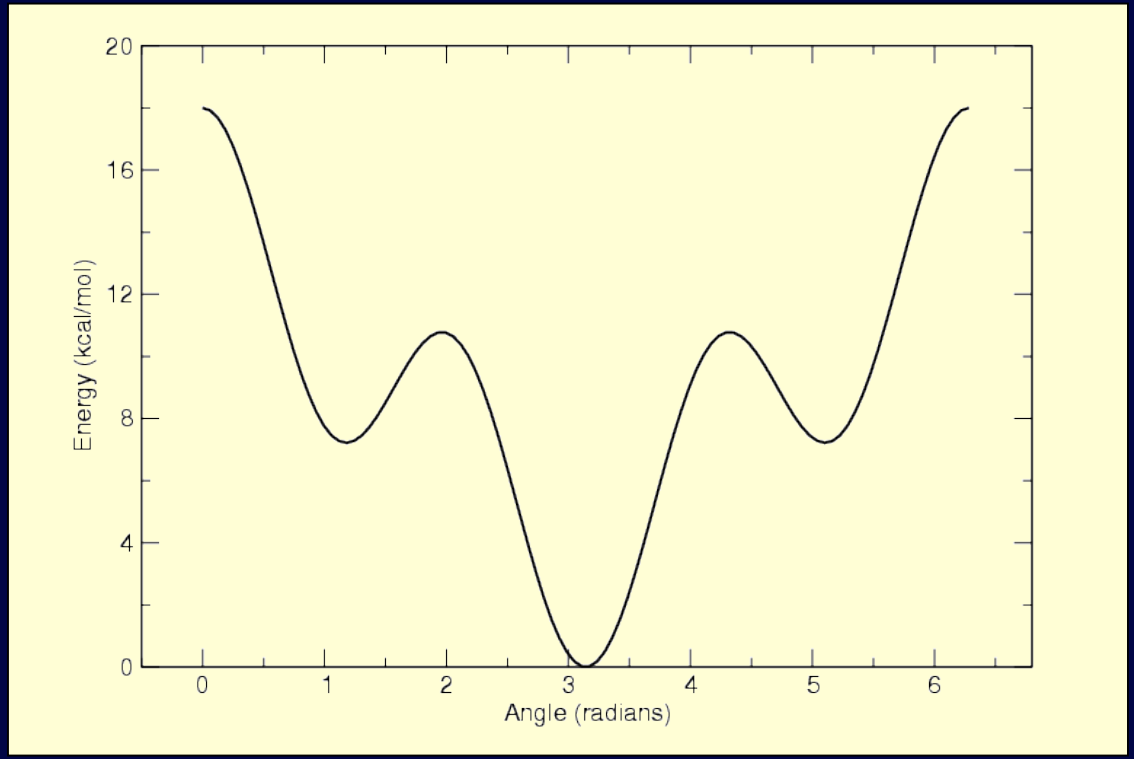
$$\cdot\frac{\int_\gamma d\gamma\int_\Gamma d\Gamma\exp[-\beta\cdot\{U_0(\gamma,\Gamma)+V[f(\gamma,\Gamma)]\}]}{\int_\gamma d\gamma\int_\Gamma d\Gamma\exp\{\beta\cdot V[f(\gamma,\Gamma)]\}\cdot\exp[-\beta\cdot\{U_0(\gamma,\Gamma)+V[f(\gamma,\Gamma)]\}]}$$

$$= \frac{\exp\{\beta\cdot V[f(\gamma,\Gamma)]\}\cdot\rho^*[f(\gamma,\Gamma)]}{<\exp\{\beta\cdot V[f(\gamma,\Gamma)]\}>^*}$$

- By computing the distribution over the biased sampling, ρ^*, one can construct the exact distribution by correcting the result

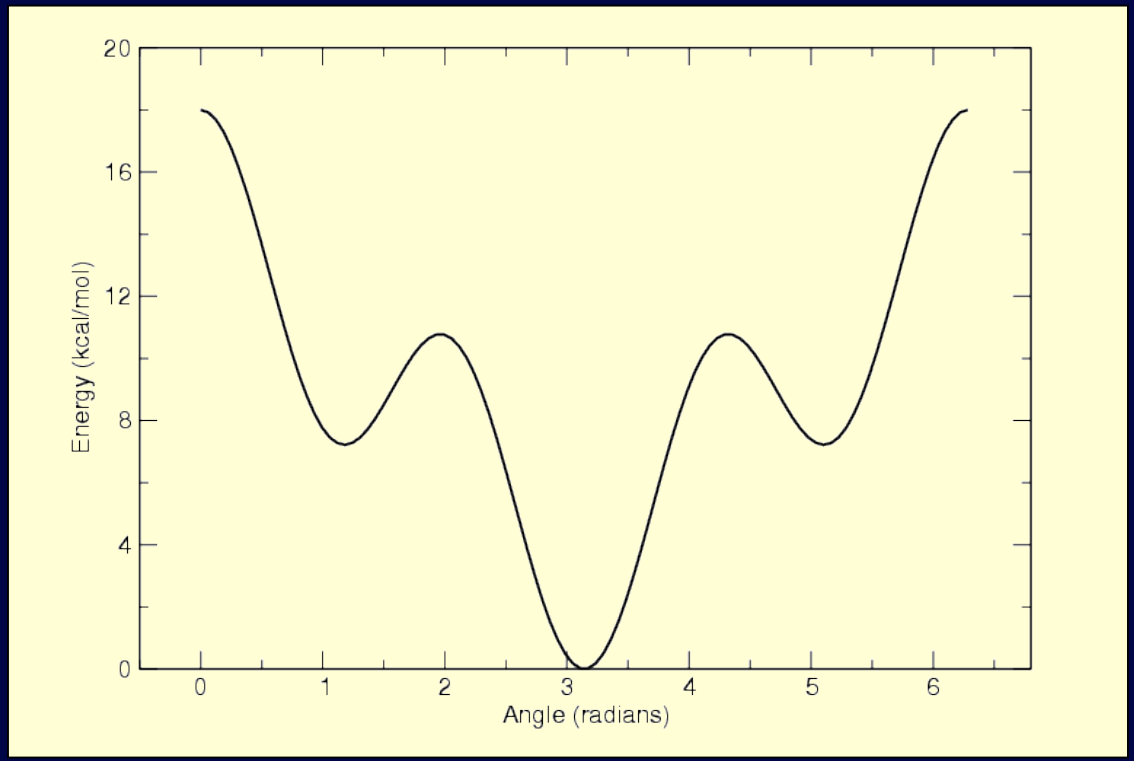
Umbrella sampling and free energy simulations

- Consider sampling an energy surface as shown below
  - How can we efficiently sample all of the space?



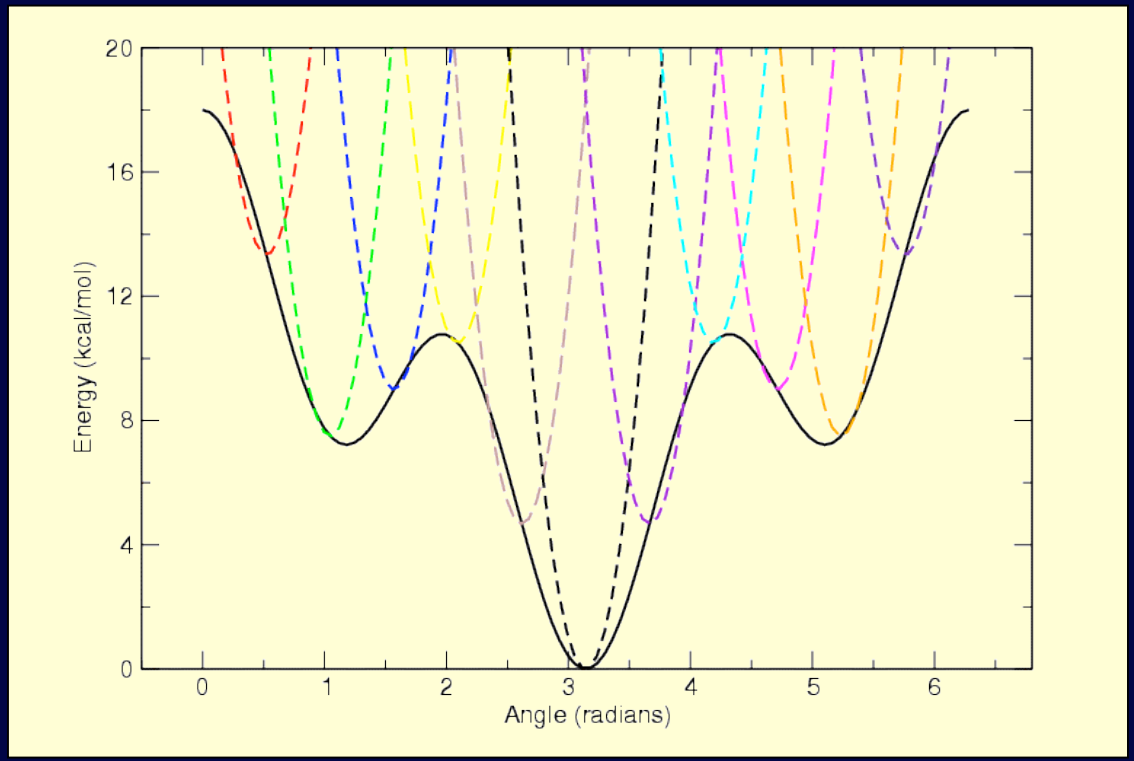
Umbrella sampling and free energy simulations

- Consider sampling an energy surface as shown below
  - How can we efficiently sample all of the space?
- Add umbrella potentials to guide sampling



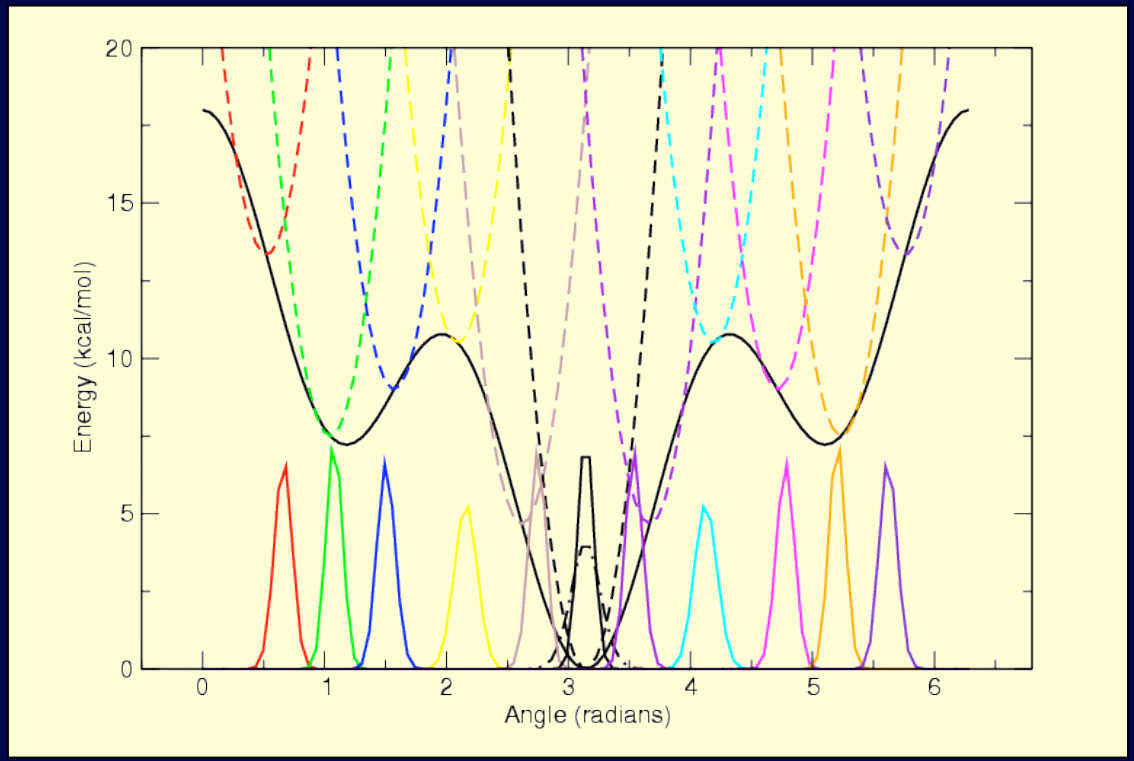
Umbrella sampling and free energy simulations

- Consider sampling an energy surface as shown below
  - How can we efficiently sample all of the space?
- Add umbrella potentials to guide sampling



Umbrella sampling and free energy simulations

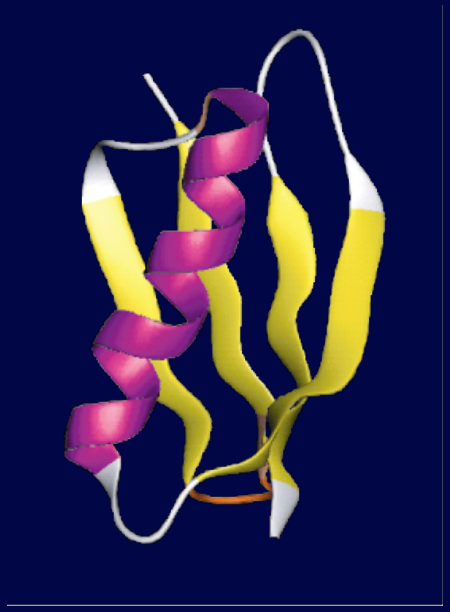
- Consider sampling an energy surface as shown below
  - How can we efficiently sample all of the space?
- Add umbrella potentials to guide sampling

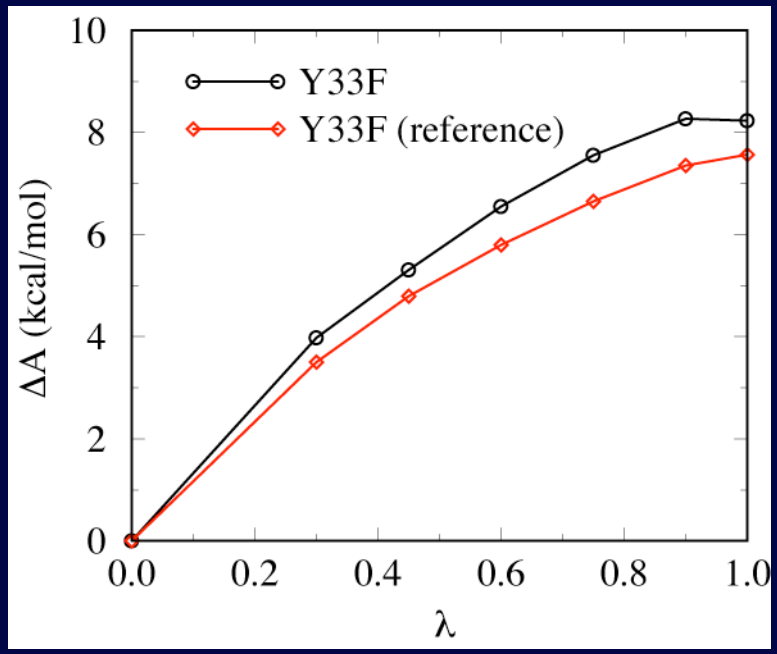


- Use free energy connections to unbias

Free Energy Methods

- Applications can include relative binding affinities as well as changes in protein (or nucleic acid) stability from sequence mutations
  - Consider question of relative stability of mutant GB1 domain after Tyr33→Phe mutation

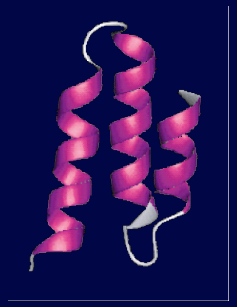


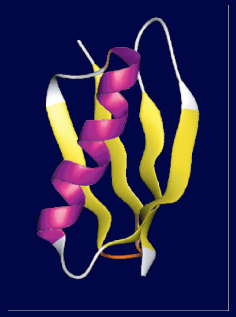


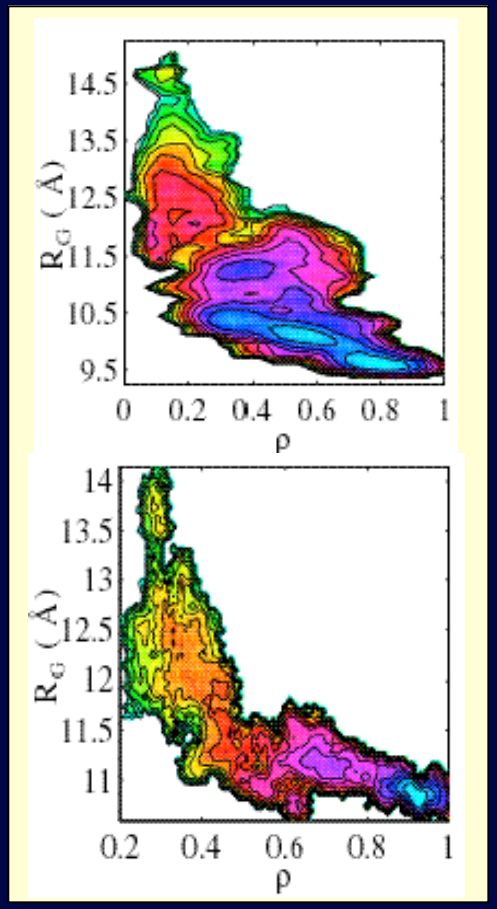
Removing OH from benzyl ring destabilizes protein by ~0.7 kcal/mol

Free Energy Methods - Conformations

- These same free energy ideas can be used to probe free energy changes for conformational shifts
- Protein folding - an illustration







$\rho \rightarrow$ fraction of native contacts (interactions)

$R_G \rightarrow$ radius of gyration

"diagonal" free energy landscape occurs for topologies with mainly local structure (helices)

"L-shaped" landscapes suggest requirement of collapse preceding native contact formation

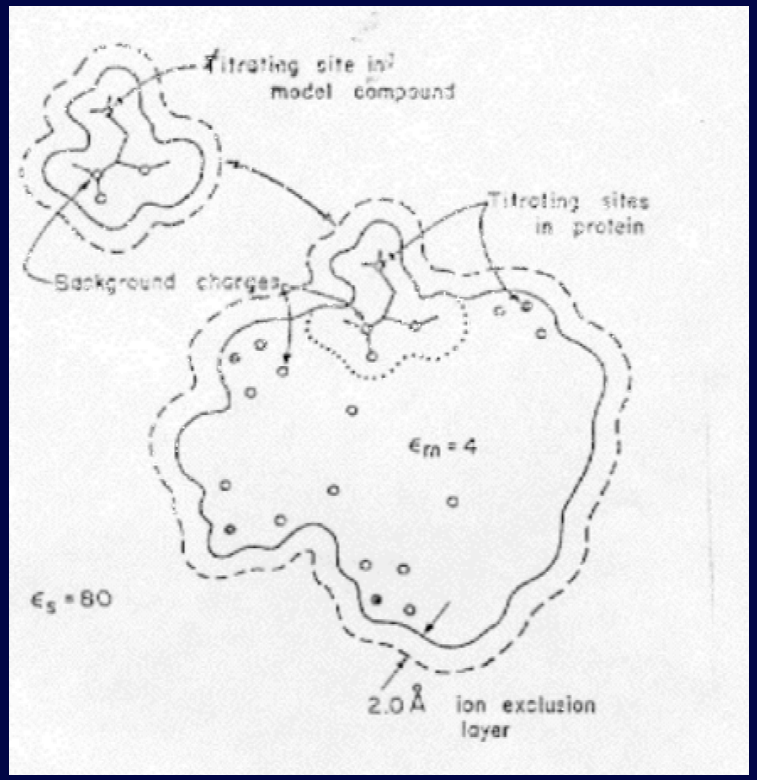
Free Energy Methods - Approximations

- Probing pH effects in proteins, pK_a calculations

$$A_pH \rightarrow A_p + H_s$$
$$\uparrow\uparrow \qquad\qquad \downarrow\downarrow$$
$$A_sH \rightarrow A_s + H_s$$

$$\Delta\Delta G = 2.303 \cdot k_B T(pK_a(ref) - pH) + \Delta G_{(A_pH \rightarrow A_p + H_s)} - \Delta G_{(A_sH \rightarrow A_s + H_s)}$$

- When $\Delta\Delta G = 0$, $pH = pK_a$

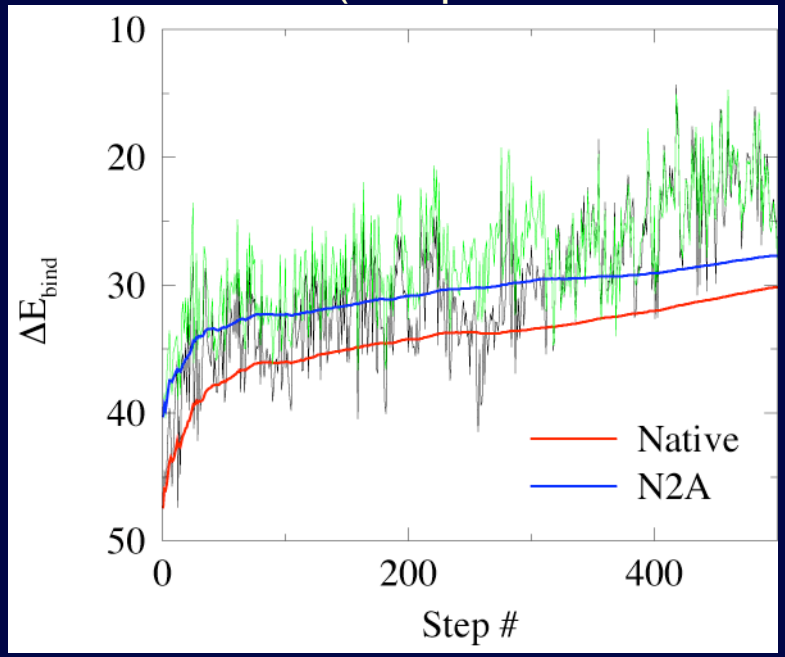


- All sites interact, in general. However, in the single site approximation one can relatively rapidly estimate pK_a values w/ GB or PB.

Free Energy Methods - Approximations

- Can one use simulations of endpoints to estimate free energy differences?
- Example - association free energy for protein-protein interface (GCN4 Leucine zipper dimer association)
  - Model - MD of GCN4 dimer in solution
  - GB analysis of trajectory for dimer and each monomer to estimate association free energy (MM/GB, MM/GBSA, MM/PBSA)
  - Approximate mutational differences for Asn16→Ala (computational "alanine scanning")

Removal of Asn16 from dimer interface destabilizes dimer by 2.5 kcal/mol



CHARMM Free Energy Methods

- Chemical perturbations
  - TSM (Thermodynamic simulation methods)
  - BLOCK (general interaction energy blocking method)
  - PERT (single topology perturbation code)
- Conformational perturbations
  - Many flavors of umbrella potentials
  - ADUMB (adaptive umbrella)
  - Multi-canonical
  - TPS
  - Etc.

CHARMM Free Energy Methods

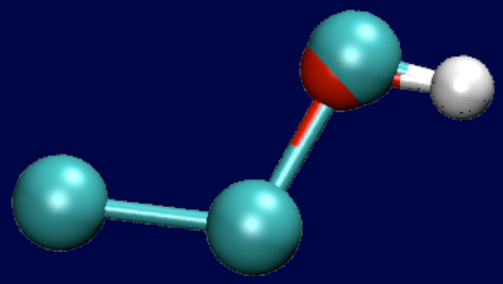
- Defining dual topologies - an example
  - Ethanol to propane

```
RESI ETP 0.000
GROU
atom C1 CH3E 0.    ! environment atom
atom C2 CH2E 0.265 ! COLO atom the charge is the reactant charge
atom O1 OH1 -0.7   ! reactant atom
atom H1 H    0.435 C3 ! reactant atom note the non-bonded exclusion with
GROU
atom C3 CH3E 0.    ! product atom

BOND C1 C2       !environment term
BOND C2 O1 O1 H1 !reactant terms
BOND C2 C3       !product term

! the angles MUST be specified
! note the absence of O1 C2 C3 between reactant and product atoms
ANGLe  C1 C2 C3  !product term
ANGLe  C1 C2 O1  C2 O1 H1  !reactant terms

! this will be a V(R) term.
DIHED C1 C2 O1 H1
```

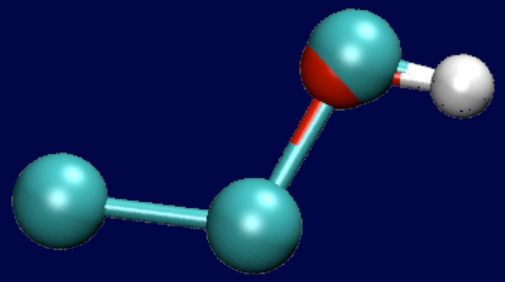


CHARMM Free Energy Methods

- Defining dual topologies - an example
  - Ethanol to propane

```
! don't really need it but what the heck.
DONO H1 O1
ACCE O1

IC C1 C2 O1 H1  1.54 111. 180. 109.5 0.96
IC C2 O1 H1 BLNK 0. 0. 0. 0. 0.
IC C1 C2 C3 BLNK 0. 0. 0. 0. 0.
IC C2 C3 BLNK BLNK 0. 0. 0. 0. 0.
PATCH FIRST NONE LAST NONE
```



CHARMM Free Energy Methods

- Setting up TSM - an example
  - Ethanol to propane

```
tsm

! Assign reactant list:
REAC sele atom etp 1 O1 .or. atom etp 1 H1 end

! Assign product list:
PROD sele atom etp 1 C3 end

! Set lambda - we will use TI or TP.
! The lambda dependence of the Hamiltonian will be linear.
! This is the default and the POWEr 1 command is actually unecessary.

LAMBda .125 POWEr 1
! The common methyl group is a colo atom.  Since the charge in the
! rtf was for the reactant the RCHArge command is actually unecessary.
COLO ETp 1 C2  PCHArge 0. RCHArge 0.265
!
! This is a thermalization run - so no save statement.
! Just terminate the FES setup with an END statement.
END
```