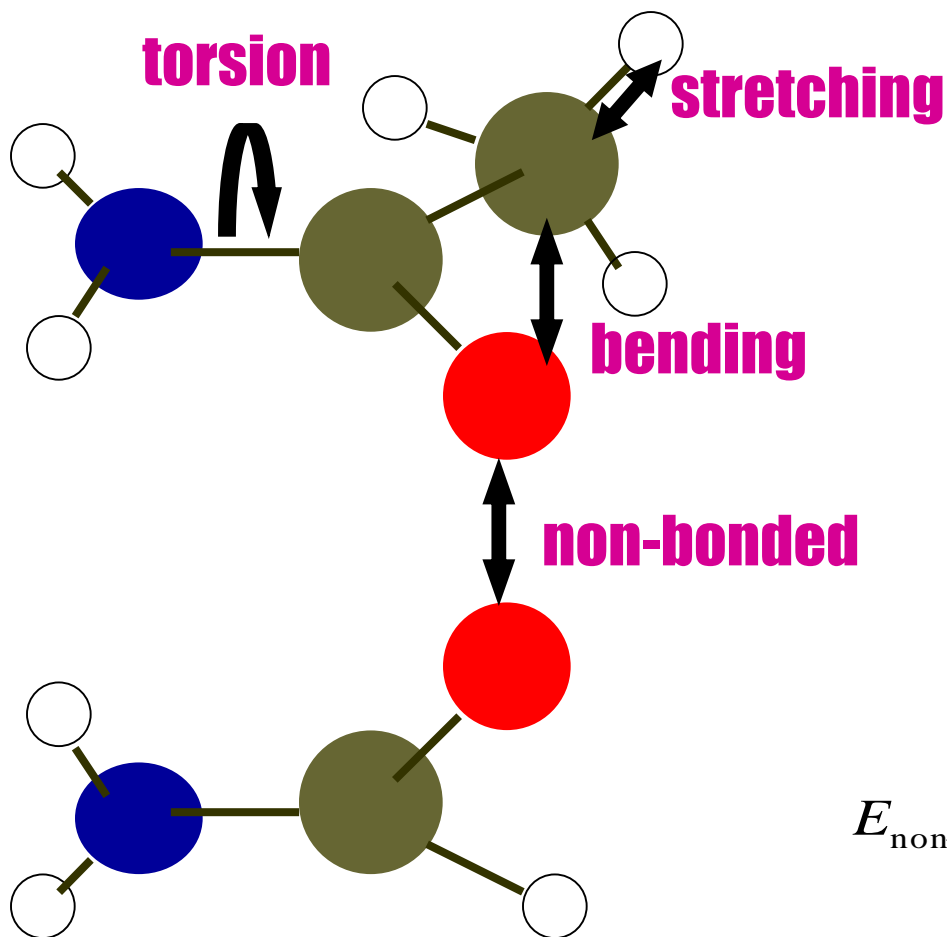# Lecture 6: Force Field Modification

Junmei Wang

Department of Pharmacology, University of Texas

Southwestern Medical Center at Dallas

Junmei.wang@utsouthwestern.edu

"Bare" Molecular Mechanics Atomistic Force Fields:

$$E_{\text{stretch}} = \sum_{\text{bonds}} K_r (r - r_0)^2$$

$$E_{\text{bend}} = \sum_{\text{angles}} K_q (q - q_0)^2$$

$$E_{\text{torsion}} = \sum_{\text{torsions}} \frac{V_n}{2} \left[1 \pm \cos(n - \gamma_n) \right]$$

$$E_{\text{non-bonded}} = \sum_{ij} \left[\frac{q_i q_j}{r_{ij}} + 4_{ij} \left(\frac{\sigma_{ij}^{12}}{r_{ij}^{12}} - \frac{\sigma_{ij}^6}{r_{ij}^6} \right) \right]$$

Why Solvent?

"It cannot be overemphasized that solvation changes the solute electronic structure. Dipole moments in solution are larger than the corresponding dipole moments in the gas phase. Indeed, *any* property that depends on the electronic structure will tend to have a different expectation value in solution than in the gas phase." -**Cramer**
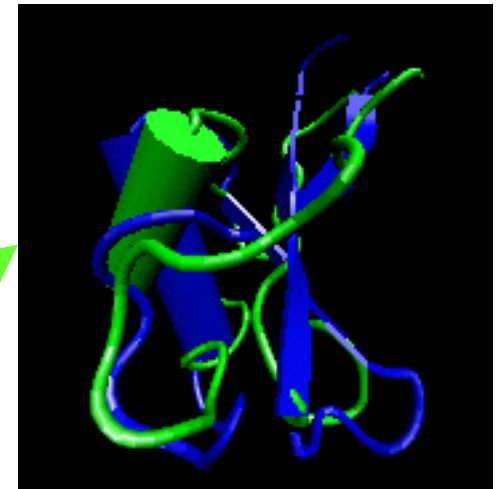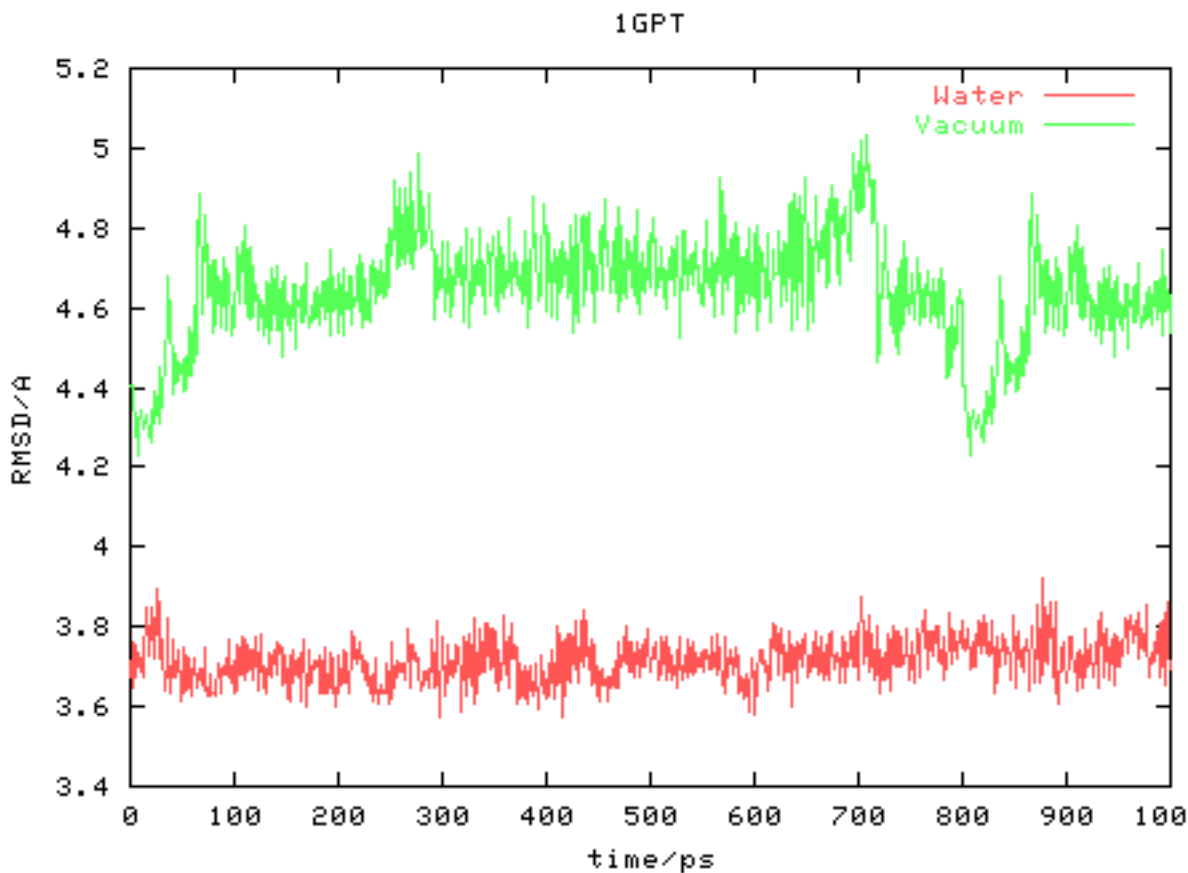
Solvent Effects

- Many reactions take place in solution

- Short-range effects
  - Typically concentrated in the first solvation sphere
  - Examples: H-bonds, preferential orientation near an ion

- Long-range effects
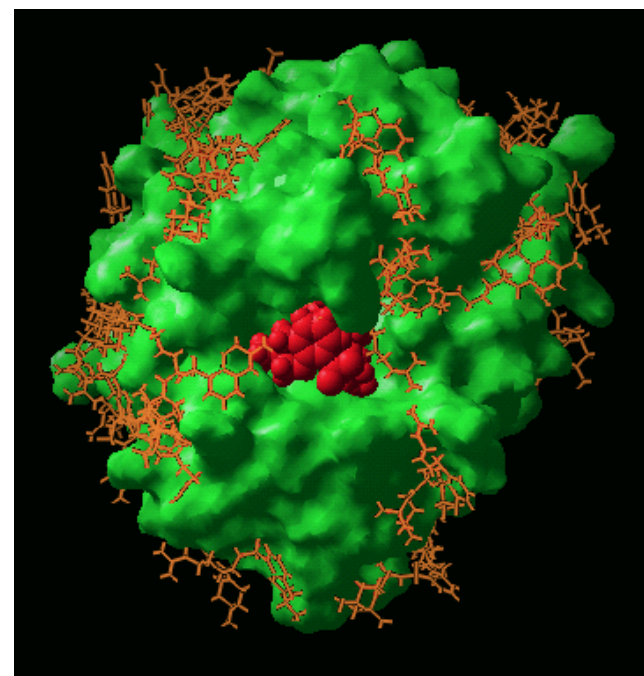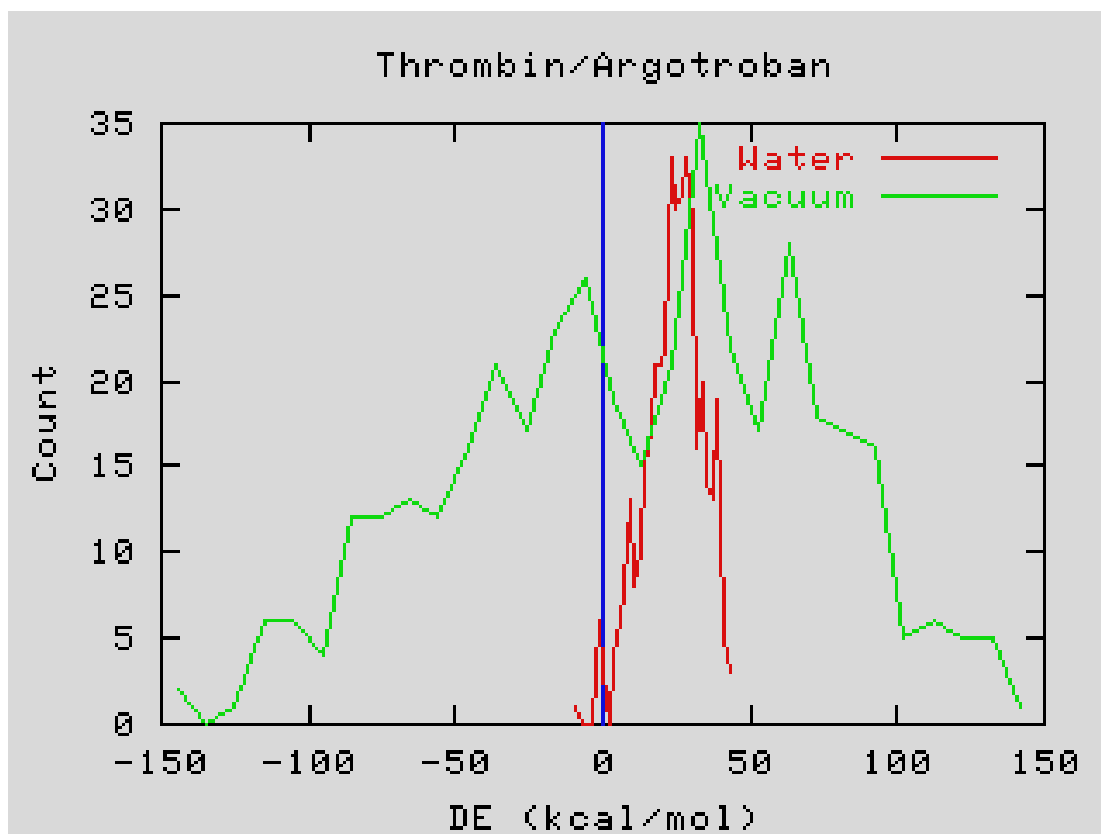 - Polarization (charge screening)
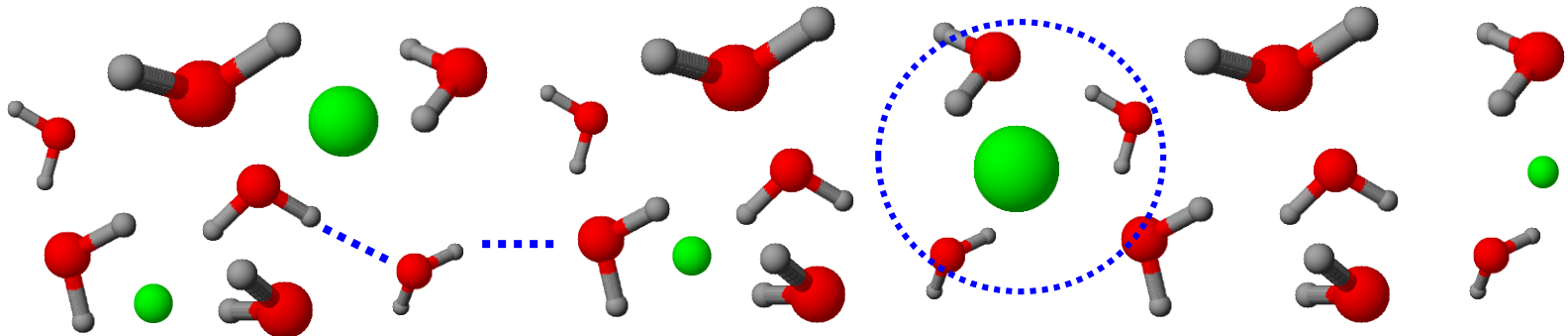
Hydration has a large effect on the conformations of macromolecules

MD simulation, RMSD from native native:
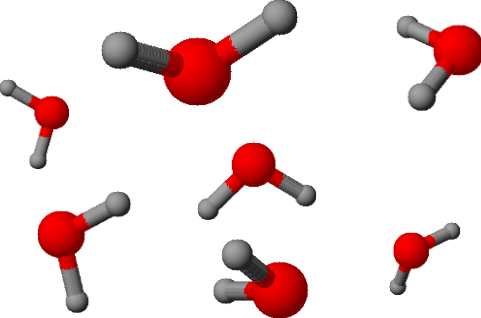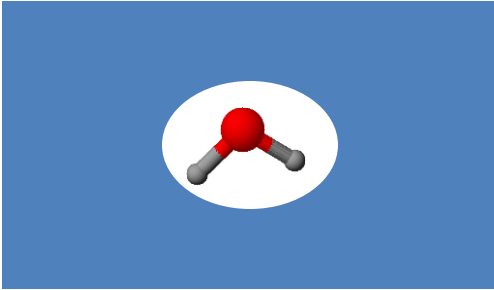
… and on ligand binding

Distribution of complex decoy binding energies:

Two Kinds of Solvatin Models

Models	Explicit solvent models	Continuum solvation models
Features	All solvent molecules are explicitly represented.	Respresent solvent as a continuous medium.
Advantages	Detail information is provided. Generally more accurate.	Simple, inexpensive to calculate
Disadvantages	Expensive for computation	Ignore specific short-range effects. Less accurate.

Explicit Solvation

- Each solvent molecule is represented with a set of atomic interaction centers (just as for the solute).

- Most accurate/detailed.

- Computationally expensive.

- Requires averaging over solvent coordinates.

- Difficult to obtain relative free energies of solute conformations.
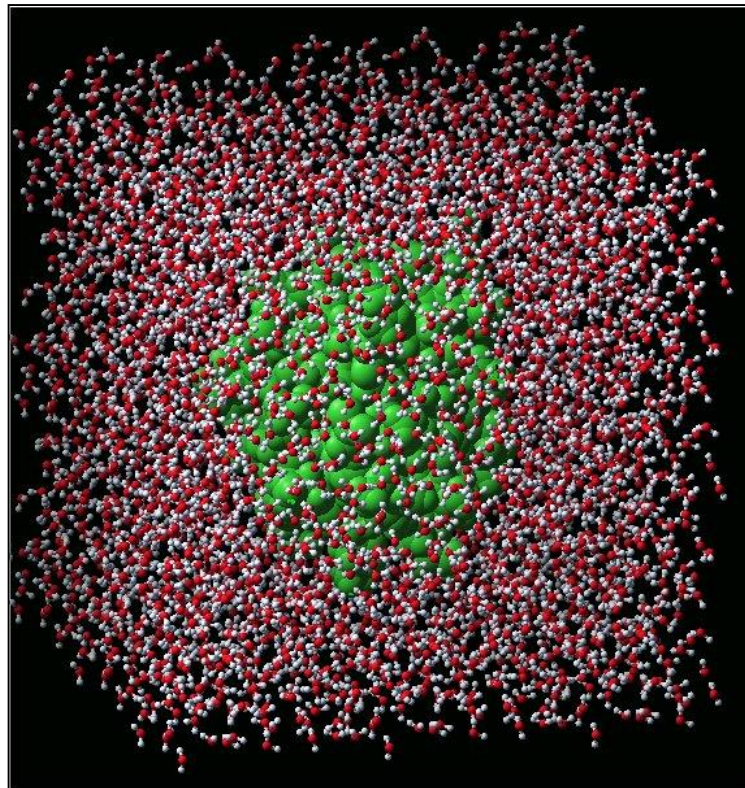
Explicit Water Model Examples

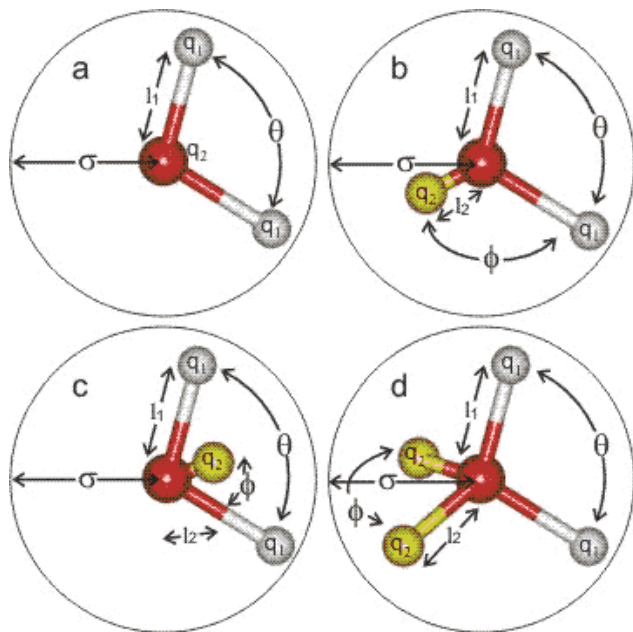|	σ(Å)	ε(kj·mol⁻¹)	l₁(Å)	l₂(Å)	q₁(e)	q₂(e)	θ°	φ°
SPC/E	3.166	0.650	1.00	-	+0.4238	-0.8476	109.47	-
TIP3P	3.1506	0.6364	0.9572	-	+0.4170	-0.8340	104.52	
TIP4P	3.15365	0.6480	0.9572	0.15	+0.5200	-1.0400	104.52	52.26
TIP5P	3.1200	0.6694	0.9572	0.70	+0.2410	-0.2410	104.52	109.47

The SPC/E model adds an average polarization correction to the potential energy function – better density, diffusion constant;

$$E_{pol} = \frac{1}{2} \sum_i \frac{(\mu - \mu^0)^2}{\alpha_i}$$

CHARMM version of the TIP3Pmodel places Lennard-Jones parameters on the hydrogen atoms.

Implicit Solvation

- The solvent is represented by a **continuum** described by macroscopic parameters such as the dielectric constant, density, surface tension, etc.

- Theoretical framework based on solvent PMF.

- Not as accurate, especially for short-range solute-solvent interactions.

- Reduced dimensionality.

- Relative solvation free energies from single point effective potential energy calculations.

$\varepsilon, \rho, ...$

Implicit Solvent Models

continuum solvent
$\varepsilon=80$

$\varepsilon=1-4$
protein
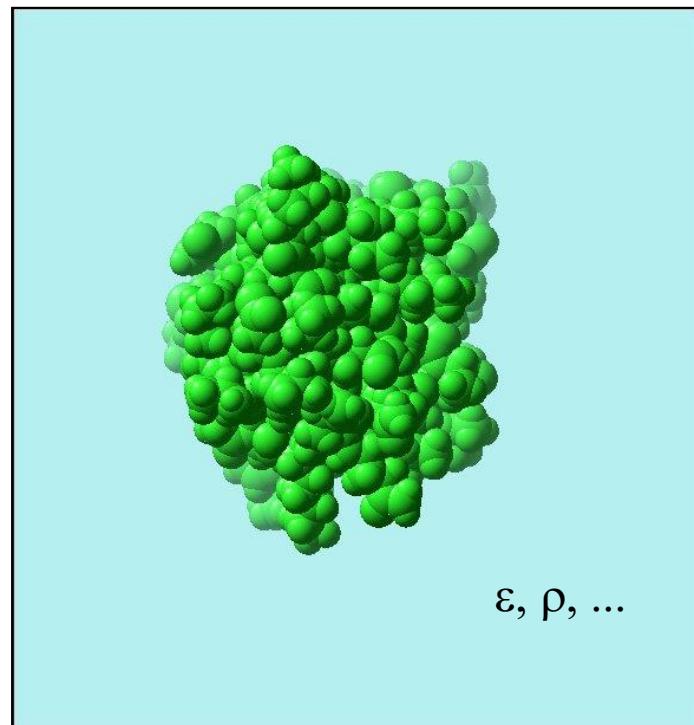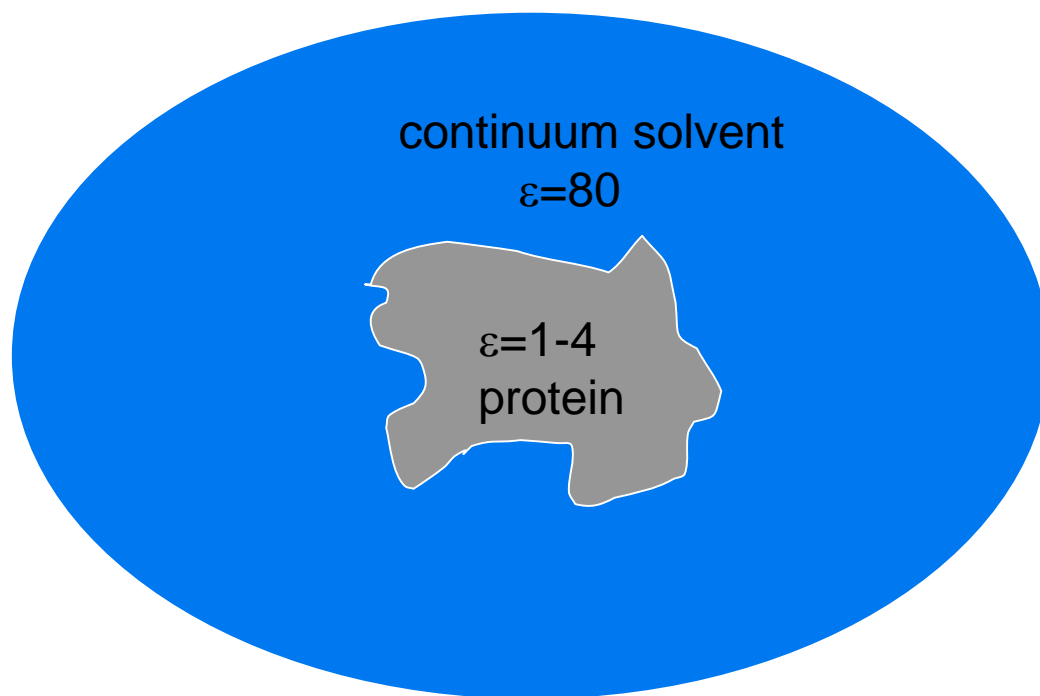
"A continuum model in computational molecular sciences can be defined as a model in which a number of the degrees of freedom of the constituent particles are described in a continuous way, usually by means of a distribution function." -Tomasi, Mennucci, and Cammi

From Explicit to Implicit
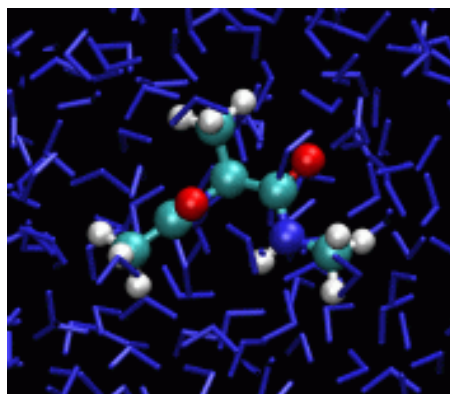
Explicit Solvent
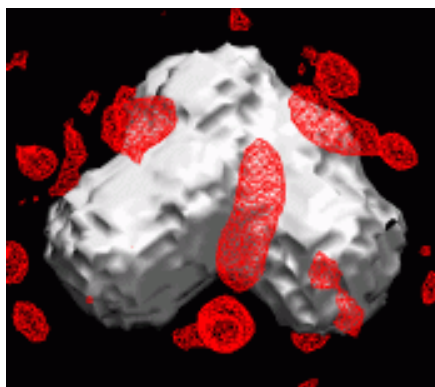
Average Density
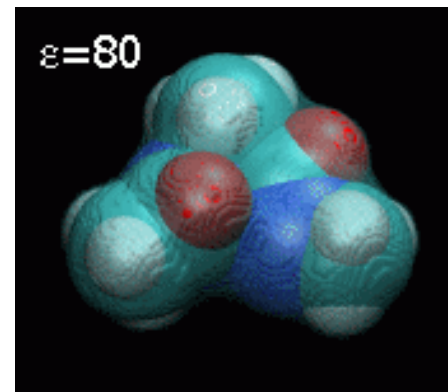
Implicit Solvent

It is possible to construct an **implicit solvent model** by approximating the medium outside the water-excluded volume as a continuum with electrostatic, entropic, and viscous properties that match water.

Molecular Mechanics Atomistic Force Fields With Solvent Effect Taken Into Account

$$V_{potential} = \sum_{bonds} K_r \left(r - r_{eq} \right)^2 + \sum_{angles} K_\theta \left(\theta - \theta_{eq} \right)^2 + \sum_{dihedrals} \frac{V_n}{2} \left[1 + \cos(n\varphi - \gamma) \right] +$$

$$\sum_{\substack{i<j \\ \text{van der Waals}}} \left(\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} \right) + \sum_{\substack{i<j \\ \text{electrostatic}}} \frac{q_i q_j}{R_{ij}} \quad + \boxed{\Delta G_{solv}}$$

Solvation Free Energy Decomposition
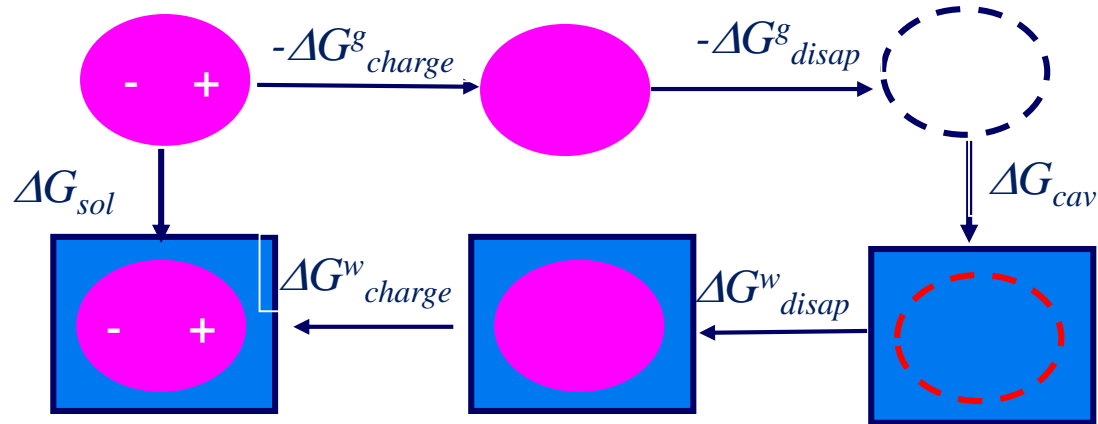
- $\Delta G_{solv} = \Delta G_{polar} + \Delta G_{nonpolar}$

- $\Delta G_{polar} = \Delta G^w{}_{charge} - \Delta G^g{}_{charge}$

- $\Delta G_{nonpolar} = \Delta G^w{}_{disap} - \Delta G^g{}_{disap} + \Delta G_{cav}$

    includes van der Waals interaction energy,

    entropy, reorganization energy

Self-Consistent Reaction Field

- Solvent: A uniform polarizable medium with a dielectric constant ε
- Solute: A molecule in a suitably shaped cavity in the medium

- Solvation free energy:

$$\Delta G_{solv} = \Delta G_{cav} + \Delta G_{disp} + \Delta G_{elec}$$
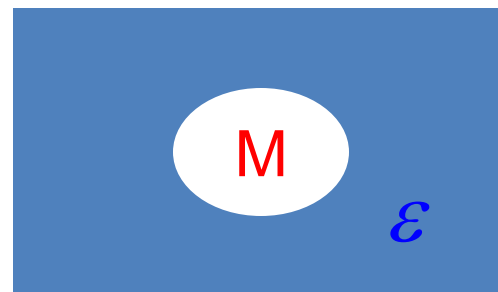
1. Create a cavity in the medium costs energy (destabilization).

2. Dispersion (mainly Van der Waals) interactions between solute and solvent lower the energy (stabilization).

3. Polarization between solute and solvent induces charge redistribution until self-consistent and lowers the energy (stabilization).

Electrostatic Component: ΔG_{polar}

- Poisson-Boltzmann solvers (accurate but numerical and slow).

- Generalized Born models (faster, can be expressed as analytic function).

- Research Trend: improve accuracy and efficiency and coverage

Non-Polar Component: $\Delta G_{nonpolar}$

- Solute surface area models
- Cavity + van der Waals NP models.
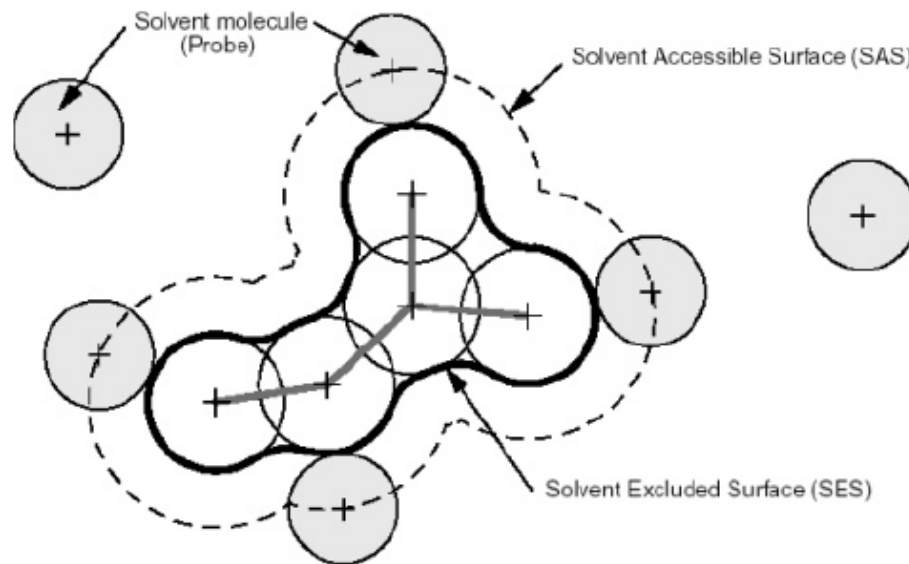
Figure 1. Solvent accessible surface (SAS) traced out by the center of the probe representing a solvent molecule. The solvent excluded surface (SES) is the topological boundary of the union of all possible probes that do not overlap with the molecule.

surface area approaches

Observation: $G_{solvation}$ for the saturated hydrocarbons in water is linearly related to the solvent accessible surface area

exposed solvent accessible surface area (SASA)

$$\Delta G_{residue} = \sum_{atoms,i} \Delta \sigma_i A_i$$

free energy of interaction of a solute with water

atomic solvation parameters based on free energies of transfer

Problems:

- sensitive to σ_i's, parameterization, surface area and change in conformation

- in dynamics you need derivatives of SASA

- what about polarization effects?

Example of an analytical NP model
(the "NP" in AGBNP)

$$\Delta G_{\text{np}} = \Delta G_{\text{cav}} + \Delta G_{\text{vdW}} \qquad \Delta G_{\text{np}} = \sum_i \left[\gamma_i A_i + \alpha_i \omega(B_i) \right]$$

A_i : Surface area of atom i

$\omega(B_i)$: Geometrical predictor based on Born radius

γ_i, α_i : Surface tension and van der Waals adjustable parameters

$$\omega_i \approx \rho_w \int \frac{-4\varepsilon_i \sigma_i^6}{|r - r_i|^6} = -\frac{16\pi\rho_w \varepsilon_i \sigma_i^6}{3C_i^3} \qquad C_i = \left(\frac{3}{4\pi} \int \frac{1}{|r - r_i|^6} \right)^{1/3} \approx B_i$$

Continuum Dielectric Models

Approximate continuum dielectric models

The basic idea is that a dielectric model of hydration should describe these two basic effects:

1. Dielectric polarization around polar groups

- Favorable interaction between exposed charged atoms and the polarized dielectric.
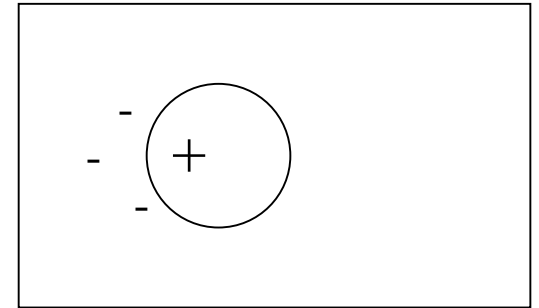
2. Dielectric screening of electrostatic interactions

- The dielectric weakens the interactions between charges
- Distance-dependent dielectric models

$$u_{ij} = \frac{q_i q_j}{\varepsilon(r_{ij}) r_{ij}}$$

$\varepsilon(r)$
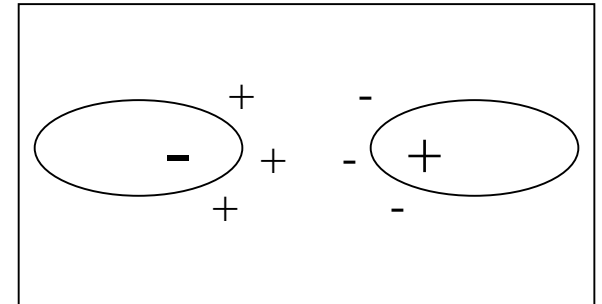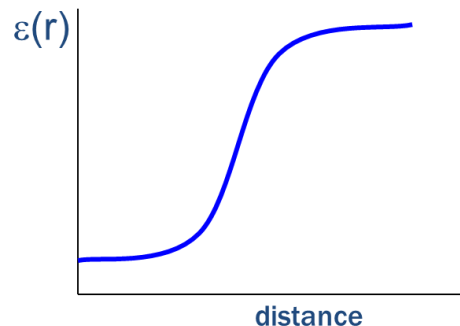
distance

Dielectric Screening

$$U = \sum_{i<j} \frac{q_i q_j}{\varepsilon\, r_{ij}}$$

$\varepsilon(r)$
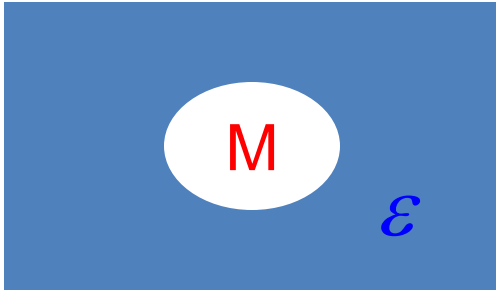
distance

$\varepsilon = 1$ gas phase

$\varepsilon = 80$ liquid phase ?

This is BULK solvent screening.
At short range, no screening...

Simple Dielectric Screening

Distance dependent dielectric $\quad \varepsilon = r_{ij} \quad or \quad \varepsilon = 4r_{ij}$
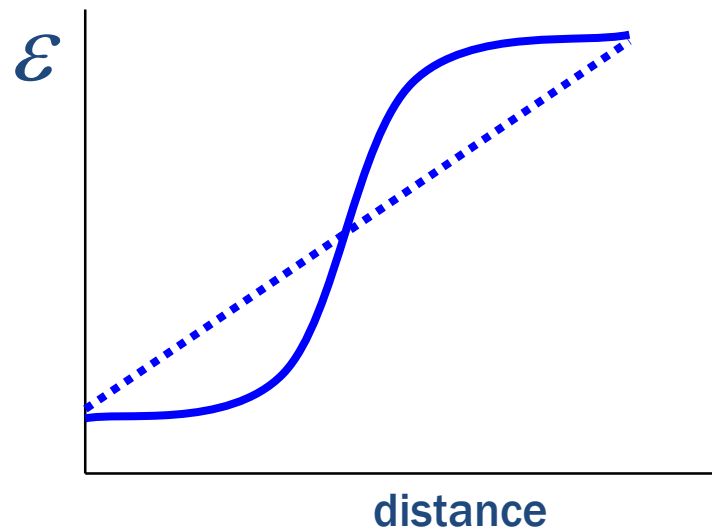
Sigmoidal dielectric function:

$$\varepsilon = D - \frac{D}{2} e^{-r_{ij}S} \left[r_{ij}S^2 + 2r_{ij}S + 2 \right]$$

Poisson-Boltzmann (PB) Model

Polar Solvation

For the general case of a solute of arbitrary shape with several partial charge sites, the electrostatic free energy is given by,

$$\Delta G_{elec} = \int_0^1 d\lambda \, \lambda q \langle \phi \rangle_{v,\lambda q} = -\frac{1}{2} \sum_i q_i \phi_i^{rf}$$

$$\phi_i^{rf} = \phi_i^{aq} - \phi_i^{vac}$$

λ: charging parameter
ϕ: electric field
$\langle\phi\rangle_{v,\lambda q}$: average reaction field

ϕ satisfies the Poisson equation

$$\nabla \cdot [\varepsilon(r)\nabla\phi(r)] = -4\pi\rho(r)$$

analytical solution available for spherical, cylindrical, or planar symmetry

$$\phi(r) = \begin{cases} q/(4\pi\epsilon_0 r) - q(1/\epsilon_0 - 1/\epsilon)/(4\pi a), & r < a \\ q/(4\pi\epsilon r), & r \geq a \end{cases}$$
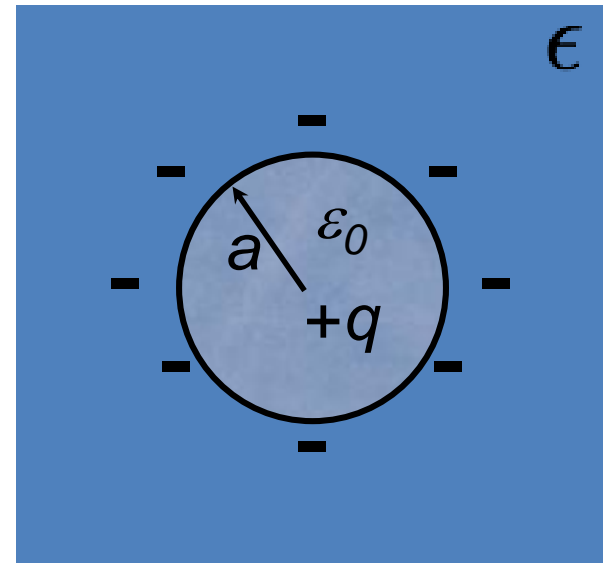
Poisson-Boltzmann Theory

The electrostatic potential related to charge density is given by Poisson's law

$$\nabla \cdot \varepsilon(r) \nabla \phi(r) = -4\pi \rho(r)$$

Mobile ions and the Poisson-Boltzmann equation

$$\rho_m(r) = \sum z_i \rho_{i,bulk} \exp(-z_i \beta \phi(r))$$

Expand at low-salt concentration

$$\nabla \cdot \varepsilon(r) \nabla \phi(r) - \kappa^2 \phi(r) = -4\pi \rho(r)$$

where, $\kappa^2 \sim \beta I$

The Tanford and Kirkwood model for protein

$$\nabla^2 \phi_1(\vec{r}) = -\sum_{i=1}^{N} \frac{q_i}{\varepsilon_1} \delta(\vec{r} - \vec{r}_i)$$

$$\nabla^2 \phi_2(\vec{r}) - \kappa^2 \phi_2(\vec{r}) = 0$$

in: ε_1

out: ε_2
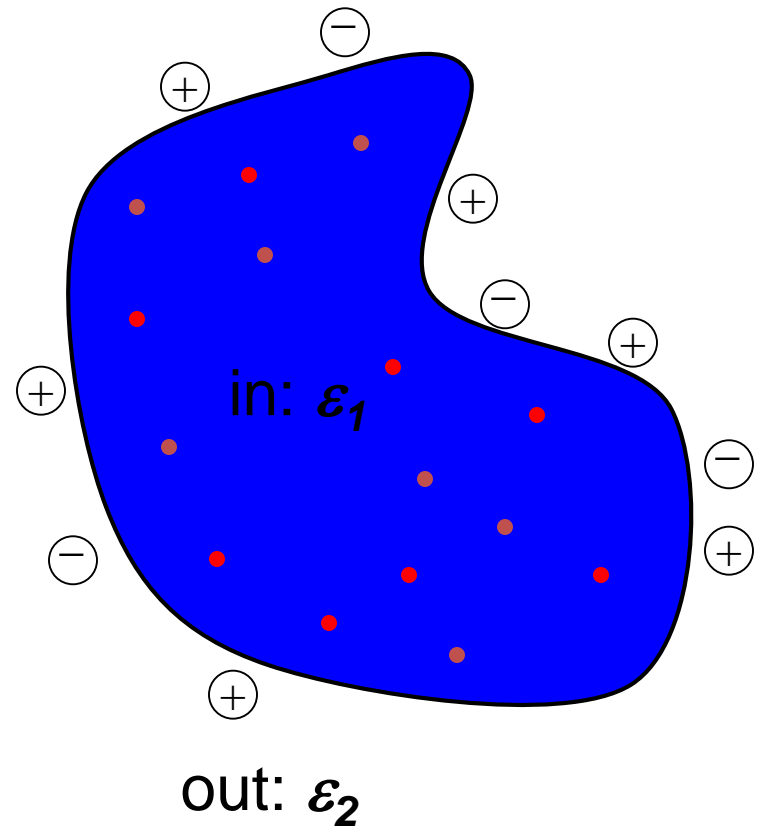
Numerical Solution of PB
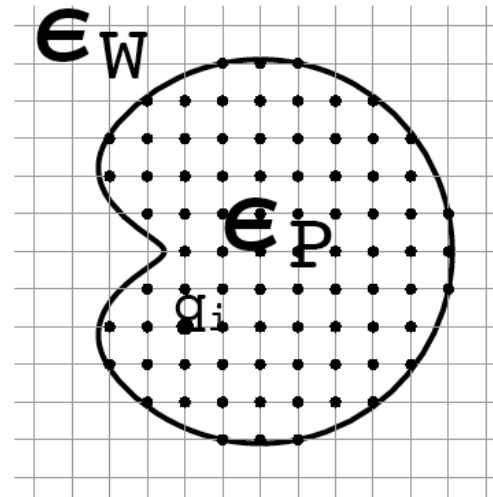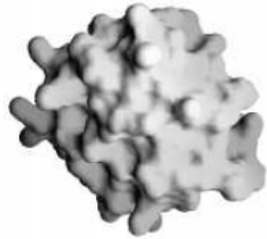
Numerical solution (FD, BE, FEM)

The *finite difference* formulation: spatial derivatives are approximated using neighboring points. A successive overrelaxation method used to get rapid convergence in solving the linear systems obtained from the finite difference discretization;

The *boundary element* method: utilizes analytical solutions obtained in terms of Green's functions and discretization on the domain surface (molecular surface);

The *finite element* method: an adaptive multilevel approach based on tetrahedral elements to create a dense mesh to capture the dielectric discontinuity across the molecular surface.

Numerical Solution of PB

	Advantages	**Disadvantages**
Finite difference and uniform mesh methods	•Fast solvers •Low memory overhead •Cartesian mesh	•Non-adaptive •Poor solution resolution •Previous parallel methods complicated and inefficient
Boundary element methods	•Smaller numerical systems •Easier interaction evaluation	•Less efficient solvers •Only applicable to linear problem
Finite element methods	•Highly adaptive •Relatively fast solvers	•Previous solver and adaptive methods inadequate •Previous parallel methods complicated and inefficient

adapted from Nathan A. Baker's slides, North Dakota State University, 2003

Numerical solutions of the PB equation

- The PB equation is solved on a grid in both surface and volume formulations.

- Finite difference: solves the PB equation on a volume grid (APBS, Delphi, UHBD)

- Finite element: solves integral form of the equation on a volume grid (PBF)

- Boundary element: surface grid.

- PB solvers often available in molecular simulation packages: Amber, CHARMM, IMPACT, etc.

- Main drawback: continuum dielectric models are not suitable for specific short-range solute-solvent interactions, finite size effects, non-linear effects, high ionization states.

- Other limitations are dependence on atomic radii parameters, speed, lack of analytical derivatives, dependence on frame of reference.

Generalized Born (GB) Model

what is the effective solvent polarization?

(solve Poisson equation)

Born: isolated point charge (q) in a spherical cavity of radius r immersed in a dielectric continuum with dielectric constant ε
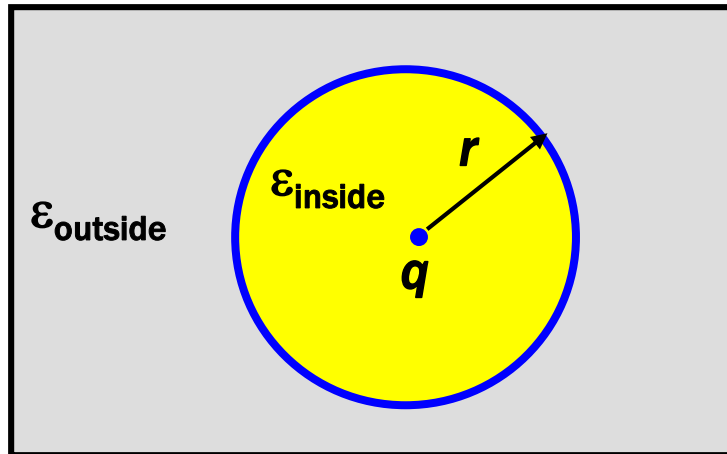
$$\Delta G_{elec} = \int_0^1 d\lambda \, \lambda q \langle \phi \rangle_{v,\lambda q} = -\frac{1}{2} \sum_i q_i \phi_i^{rf}$$

$$\langle \phi \rangle_{v,\lambda q} = \int d\vec{r} \, \frac{1}{|\vec{r}_i - \vec{r}|} \langle \rho_{elec}(\vec{r}) \rangle_\lambda \approx 4\pi \int_0^\infty r^2 dr \frac{1}{R_{ion}} \langle \rho_{elec}(r) \rangle_\lambda$$

$$4\pi \int_0^\infty r^2 dr \langle \rho_{elec}(r) \rangle_\lambda = -\lambda q (1 - \frac{1}{\varepsilon}) \qquad \text{Gauss's theorem}$$

$$\langle \phi \rangle_{v,\lambda q} = -\left(1 - \frac{1}{\varepsilon}\right) \frac{\lambda q}{R_{ion}}$$

$$\Delta G_{elec} = \int_0^1 d\lambda \, \lambda q \langle \phi \rangle_{v,\lambda q} = -\frac{1}{2}\left(1 - \frac{1}{\varepsilon}\right) \frac{q^2}{R_{ion}}$$

$$\boxed{\Delta G_{Born} = -\frac{q^2}{2r}\left(\frac{1}{\varepsilon_{in}} - \frac{1}{\varepsilon_{out}}\right)}$$

Generalized Born Approximation

- Ion of charge q in a spherical cavity of radius a

$$W_{\text{elec}} = -\left(\frac{\varepsilon - 1}{\varepsilon}\right)\sum_{i,j=1}\frac{q_i q_j}{2 f_{\text{ij}}(r_{ij})}$$

$$f_{\text{ij}}(r_{ij}) = \sqrt{r_{ij}^2 + \alpha_{ij}^2 e^{-D_{ij}}} \qquad \alpha_{\text{ij}} = \left(\alpha_i \alpha_j\right)^{0.5} \qquad D_{ij} = \frac{r_{ij}^2}{\left(2\alpha_{ij}\right)^2}$$

- Widely used in biochemistry community
- Allows for partial charges
- Equal solvation energy for positive and negative ions
- Neglects cavitation and dispersion energy
- Born radii, α_i, are not well defined

Overall Features of Generalized Born Models

$$W_{\mathrm{elec}} = -\frac{1}{2}\left(1-\frac{1}{\varepsilon}\right)\sum_{ij}\frac{q_i q_j}{f_{ij}(r_{ij})}$$

The GB model "works" because it describes both dielectric polarization and dielectric screening effects.

Polarization $i=j$ ("self" energy):

$$W_{\mathrm{single}}^{i} = -\frac{1}{2}\left(1-\frac{1}{\varepsilon}\right)\frac{q_i^2}{\alpha_i}$$

Favors the solvent exposure (small α_i) of polar groups (large q).

Dielectric screening $i \neq j$ (pair energies):

$$u_{ij} = \frac{q_i q_j}{\varepsilon(r_{ij})r_{ij}} \qquad \varepsilon_{ij}(r) = S\left(\frac{r}{\sqrt{\alpha_i \alpha_j}}\right) \qquad x \equiv \frac{r}{\sqrt{\alpha_i \alpha_j}}$$

$$S(x) = \left[1-\left(1-\frac{1}{\varepsilon}\right)\frac{1}{\sqrt{1-x^{-2}\exp\left(-x^2/4\right)}}\right]^{-1}$$

$S(x)$
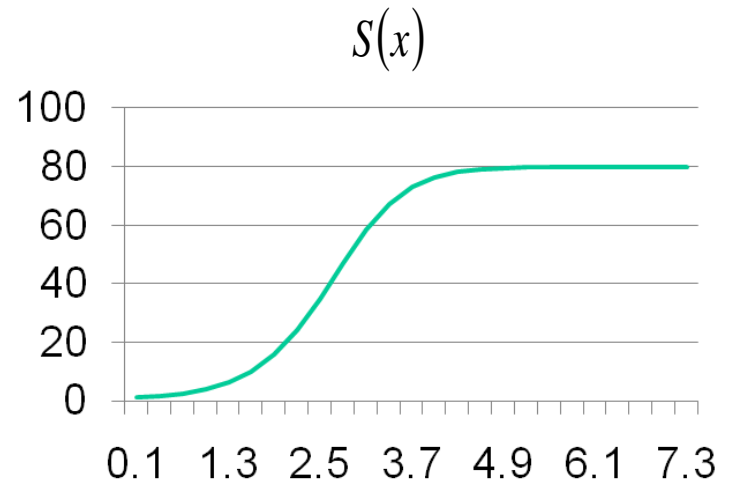
GB implementations

- Most major biomolecular simulation packages (CHARMM, Amber, IMPACT, Gromacs, etc.) include pairwise descreening GB implementations suitable for MD calculations.
- Key ingredients are the atomic radii and the description of the solute volume.
- The atoms overlap problem is generally addressed by empirical scaling coefficients parameterized with respect to higher level calculations – that is the geometric model is parameterized in addition to the energetic model (ACE, GB/SA, GBHCT, GBSW)
- Work on the AGBNP series of models shows that "geometric" parameterization is unnecessary.
- Some implementations (GBMV, SGB) perform numerical integration on a grid (volume or surface) – non-analytic, higher computational cost, difficulties with derivatives, dependence on coordinate frame.
- Some implementations differ in the choice of the GB distance function $f(r)$
- Many of the models include continuum dielectric "correction" terms.
- Recent developments have focused on the "interstitial" volumes problem (GBneck, GBMV, AGBNP2).

Other Implicit Solvent Models

PCM – Polarizable Continuum Model

- Shape of cavity determined by shape of solute
  - Overlapping van der Waals spheres (PCM and CPCM) (all atom or united atom)
  - Solvent accessible surface
  - Isodensity surface (IPCM, SCIPCM)
- Electrostatic potential from solute and polarization of solvent must obey Poisson equation

$$\nabla \cdot [\varepsilon(\vec{r}) \nabla \phi(\vec{r})] = 4\pi \rho_M(\vec{r})$$

- Polarization of solvent calculated numerically
  - FE or FD solution of the Poisson equation
  - Apparent surface charge method
  - Generalized Born / surface area

Multipole Expansion Methods

- Aka Kirkwood-Onsager Model (SCRF=Dipole)
- Solute with dipole, μ, in a spherical cavity of radius a.

$$G_{\mathrm{P}} = -\frac{1}{2}\left[\frac{2(\varepsilon\text{-}1)}{(2\varepsilon\text{+}1)}\right]\frac{\mu^2}{a^3}$$

- Easily generalized for multipole expansions

$$G_{\mathrm{P}} = -\frac{1}{2}\sum_{l=0}^{L}\sum_{m=-l}^{l}\sum_{l'=0}^{L}\sum_{m'=-l'}^{l'} M_l^m f_{ll'}^{mm'} M_{l'}^{m'}$$

- Multipole expansions are slow to converge

Multipole Expansion Methods

- QM requires a new potential term in **F**

$$\mathbf{V} = -\mathbf{r} \cdot \mathbf{R} \qquad \mathbf{R} = \frac{2(\varepsilon - 1)}{(2\varepsilon + 1)a^3} \boldsymbol{\mu}$$

- Allows solute to respond to the reaction potential resulting from polarization of the solute

- MPE easily rolled into the SCF/CPHF equations

- Very sensitive to the cavity radius *a*

- Determine *a* from the molecular volume [Volume and iop(6/44=4)]

Apparent Surface Charge (ASC) methods

- The polarization of the solute's charge distribution, ρ_M, must obey Poisson equation

$$-\nabla[\varepsilon(\mathbf{r})\nabla V(\mathbf{r})] = 4\pi\rho_M(\mathbf{r}) \qquad V(\mathbf{r}) = V_M(\mathbf{r}) + V_R(\mathbf{r})$$

- On the cavity surface, Γ, two jump conditions exist

$$[V] = V_{in} - V_{out} = 0 \text{ on } \Gamma$$

$$[\partial V] = \left(\frac{\partial V}{\partial n}\right)_{in} - \varepsilon\left(\frac{\partial V}{\partial n}\right)_{out} = 0 \text{ on } \Gamma$$

- From the second jump condition, the apparent surface charge, $\sigma(\mathbf{s})$, can be defined

$$V_\sigma(\mathbf{r}) = \int_\Gamma \frac{\sigma(\mathbf{s})}{|\mathbf{r} - \mathbf{s}|} d^2\mathbf{s} = V_R(\mathbf{r})$$

Boundary Element Method

- BEM used to solve ASC equation

- Γ approximated by tesserae small enough to consider $\sigma(\mathbf{s})$ almost constant within each tessera

- A set of point charges, q_k, are defined based on the local value of $\sigma(\mathbf{s})$ in a tessera of area A_k

$$V_\sigma(\mathbf{r}) \; ; \; \sum_k \frac{\sigma(\mathbf{s}_k)A_k}{|\mathbf{r} - \mathbf{s}_k|} = \sum_k \frac{q_k}{|\mathbf{r} - \mathbf{s}_k|}$$

- Adaptable for linearized Poisson-Boltzmann applications: nonzero ionic strength solvents

- FMM speed up BEM calculations

ASC Methods: PCM

- The Polarizable Continuum Model (PCM) is the oldest ASC method.

- The PCM surface charge is

$$\sigma(\mathbf{s}) = \frac{\varepsilon - 1}{4\pi\varepsilon} \frac{\partial}{\partial \mathbf{n}} \left(V_{\mathrm{M}} + V_{\sigma} \right)_{\mathrm{in}}$$

- Three major formulations
  - DPCM (SCRF=PCM)
  - IPCM (SCRF=IPCM)
  - SCIPCM prone to stability issues (SCRF=SCIPCM)
  - CPCM = COSMO with k=0.5 (SCRF=CPCM)
  - IEFPCM = IVCPCM = SS(V)PE recommended method (SCRF=IEFPCM)

Binding Free Energy Calculations

Free Energy of Ligand Binding
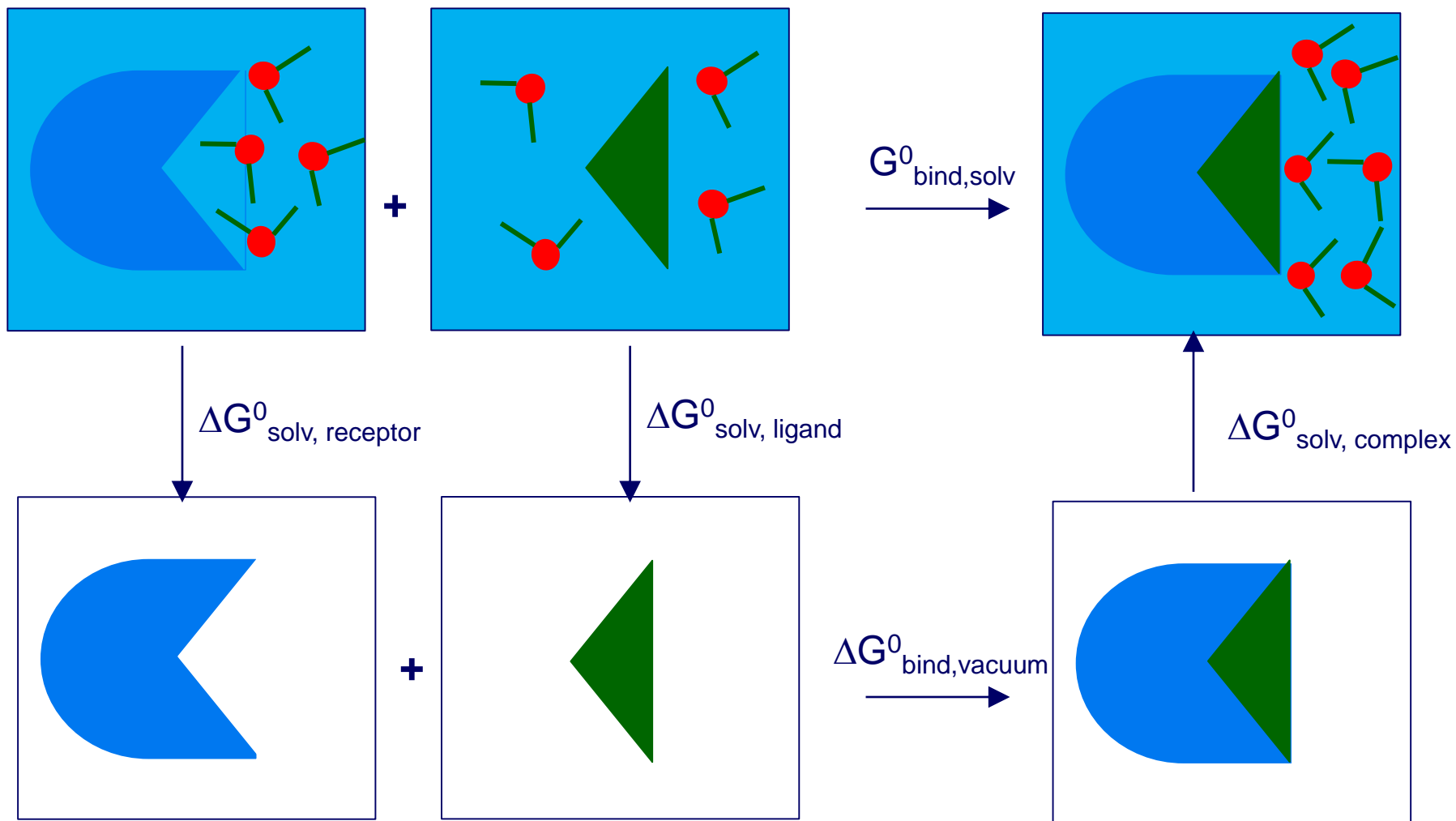
$$\Delta G^0_{bind,solv} = \Delta G^0_{bind,vacuum} + \Delta G^0_{solv,complex} - \left(\Delta G^0_{solv,ligand} + \Delta G^0_{solv,receptor}\right)$$

Free Energy of Ligand Binding (Cont'd)

$$\Delta G^0_{vacuum} = \Delta E^0_{MM} - T\Delta S^0_{NMA} \quad (1)$$

$$\Delta G^0_{solv} = G^0_{electrostatic,\varepsilon=78.3} - G^0_{electrostatic,\varepsilon=1} + \Delta G^0_{nonpolar} \quad (2)$$

$$G^{GB}_{electrostatics} = \sum_{i=1}^{N} \sum_{j=i+1} \frac{q_i q_j}{\varepsilon r_{ij}} - \frac{1}{2}\left(1 - \frac{1}{\varepsilon}\right)\sum_{i=1} \frac{q_i^2}{\alpha_i} \quad (3)$$

MM-PBSA or MM-GBSA

$$\nabla.\varepsilon(r)\nabla\phi(r) = -4\pi\rho(r) \quad (4)$$

$$\Delta G^{SA}_{nonpolar} = \gamma SAS + b \quad (5)$$

Wang, J.; Hou, T.; et al. *Current Computer-Aided Drug Design*, 2, 95-103, 2006.
Hou, T.; Wang, J.; et al. *J. Chem. Info. Model.,* 51:69-82, 2011
Hou, T.; Wang, J.; et al. *J. Comput. Chem.*, 32, 866-877, 2011
Xu, L.; Sun, H.; Li, Y.; Wang, J.; Hou, T. *J. Phys. Chem. B,* 117, 8408-8421, 2013

A Blind Test: Prediction of the Complex Structure and Binding Free Energy for HIV-1 RT/Efavirenz
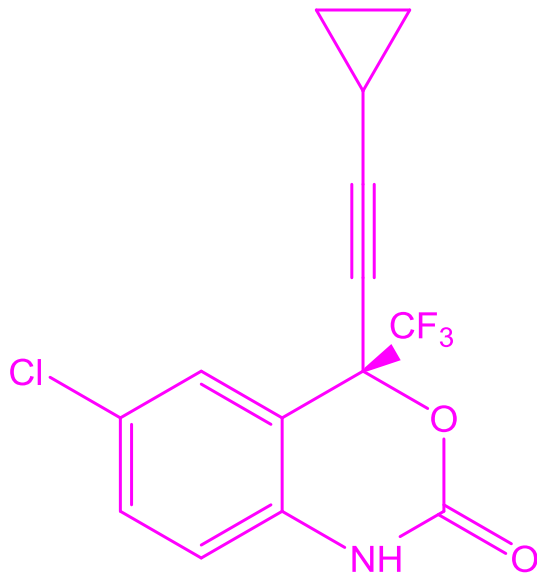
Efavirenz (Sustiva™)

Crystal Structure of HIV-1 RT/8CI-TIBO (1UWB, Resolution 3.2 Å)

Computational Strategy for Modeling Protein Complexes

Dock small molecule to protein

↓

Run MD simulation for each docking pose

↓

Calculate binding affinities (MM-PBSA)

↓

Select most favorable binding mode
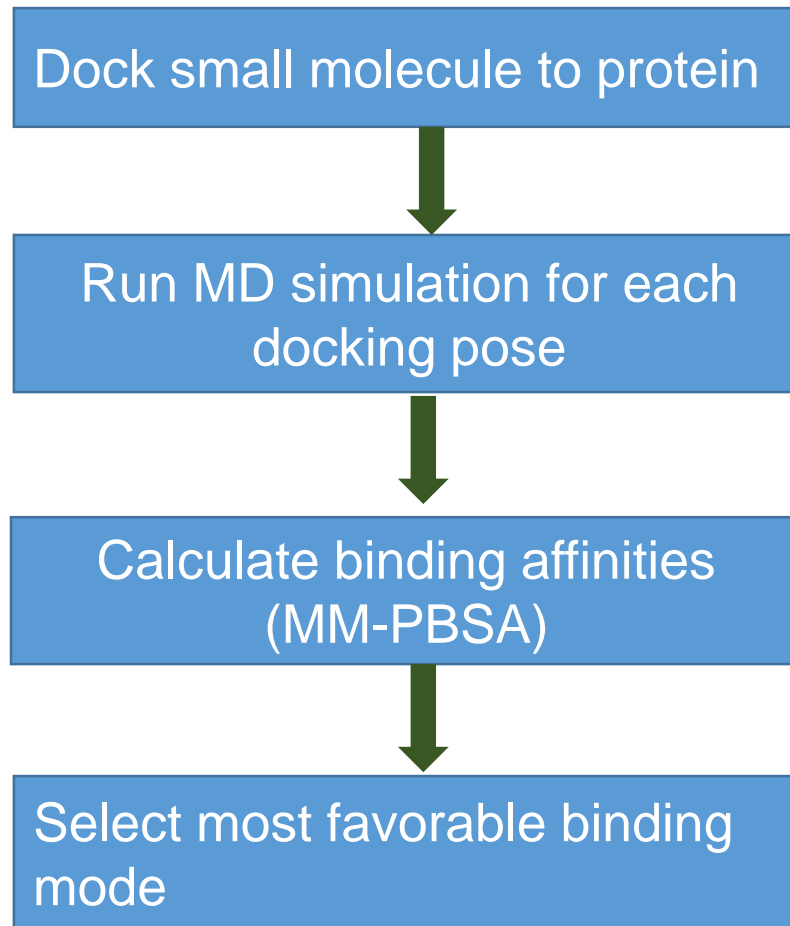
Relative Positions and Orientations of Efavirenz to 8Cl-TIBO Suggested by Docking

1 (-13.2)   2 (3.0)   3 (-6.4)

4 (-6.2)   5 (-6.3)

Experimental
Expt. –11.6 kcal/mol

Magenta: 8-Cl TIBO
Cyan: Efavirenz
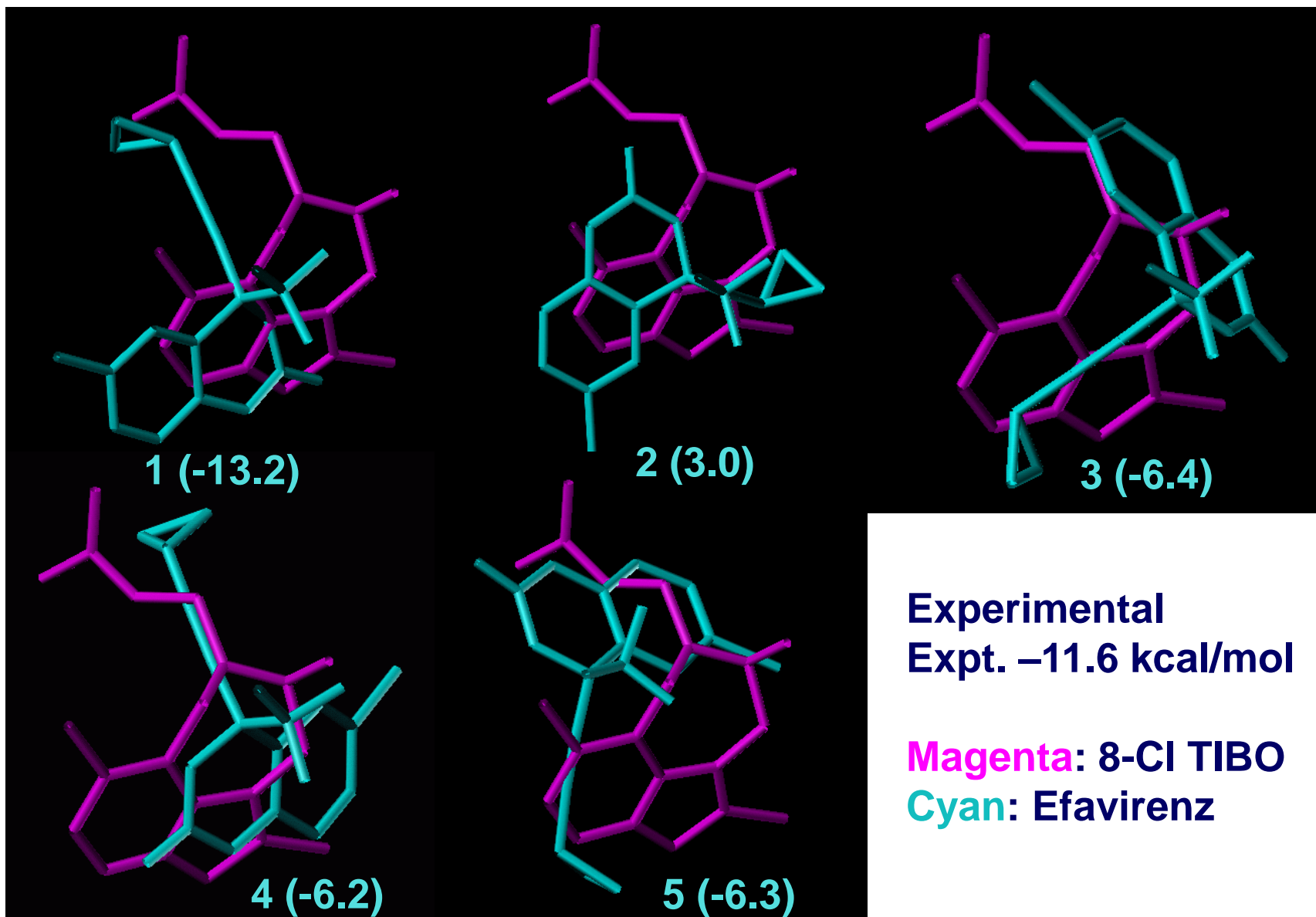
How Well does MD Reproduce the Crystal Structure?
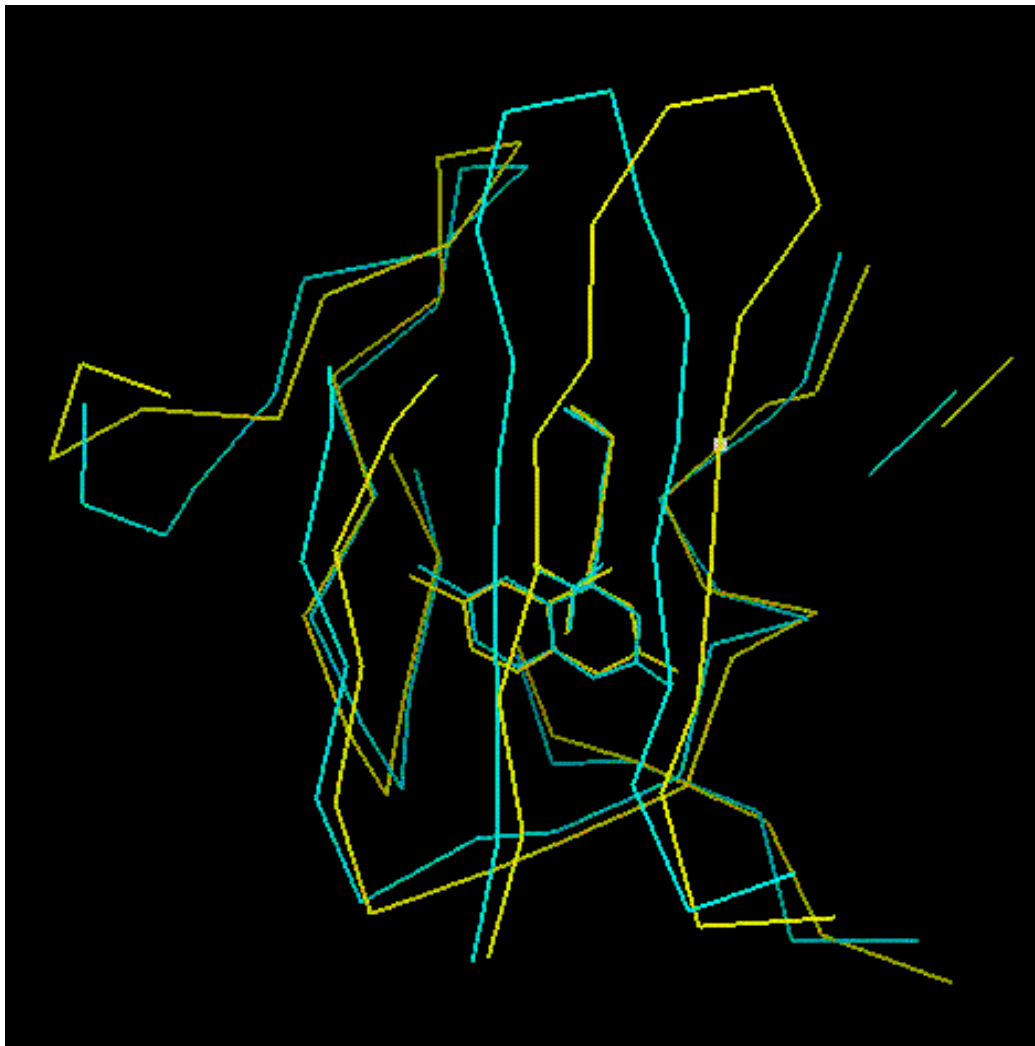
Alignment of the model structure (yellow) and the crystal structure (cyan) for HIV-1 RT/efavirenz.

The RMSD of the 54 C_α is 1.1 Å.

Wang et al. *J. Am. Chem. Soc.*, 123, 5221-5230 (cited 380 times).

Active And Inactive Conformations of Human Orexin GPCR

RMSD: 3.19 Å

1. Biology background

Respond to orexin neuropeptides in the central nervous system to regulate sleep and other behavioural functions in humans.

Suvorexant (SUV), is a drug to treat insomnia

1. Experimental Structures

Orexin GPCR (4S0V): resolved at 2.5 Å, but is an inactive conformation.

Neurotensin receptor NTS1 (4GRV): resolved at 2.80 Å, is an active conformation (in complex with neurotensin)

```
4S0V/1-282   1 - - - - - - - - PKEYEWWLIAGYIIVFVVALIGNVLVCVAVWKNHHMRTVTNYFIVNLSLADVLVTITCLPATLVV-DIT-ETWFFGQSLCKVIPYLQ85
4GRV/1-293   1 NSDLDVNTDIYSKVLVTAIYLALFVGTVGNSVTLFTLARK-SLQSTVHYHLGSLALSDLLILLLAMPVELYNFIWVHHPWAFGDAGCRGYYFLR94

4S0V/1-282  86 TVSVSVSVLTLSCIALDRWYAICHP- - - - - - - -STAKRARNSIVIIWIVSCIIMIPQAIVMECSTVF- - -KTTLFTVCDERWGGEIYPKMYHICFF170
4GRV/1-293  95 DACTYATALNVASLSVARYLAICHPFKAKTLMSRS-RTKKFISAIWLASALLAIPMLFTMGLQNRSADGTHPGGLVCTPI-MDTATVKVVIQVNT187

4S0V/1-282 171 LVTYMAPLCLMVLAYLQIFRKLWCRQKQIRARRKTARMLMVVLLVFAICYLPISILNVLKRVFGM-FAH--DRETVYAWFTFSHWLVYANSAANP262
4GRV/1-293 188 FMSFLFPMLVISILNTVIANKLTVMVV--QALRHGVLVARAVVIAFVVCWLPYHVRRLMFCYISDEQWTTFLFDFYHYFYMLTNALAYASSAINP280

4S0V/1-282 263 IIYNFLSGKF-REEFKAAFSC                                                                          282
4GRV/1-293 281 ILYNLV-SANFRQV- - - - - - -                                                                          293
```
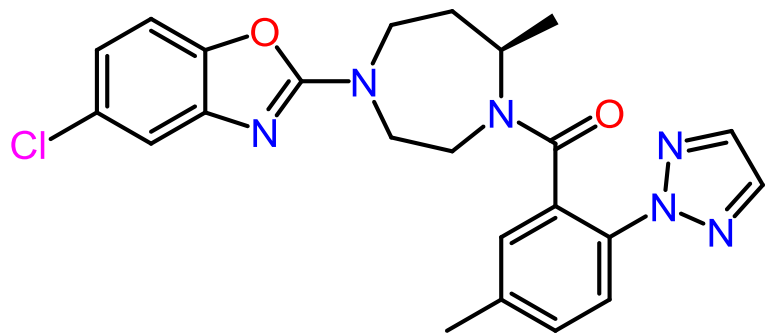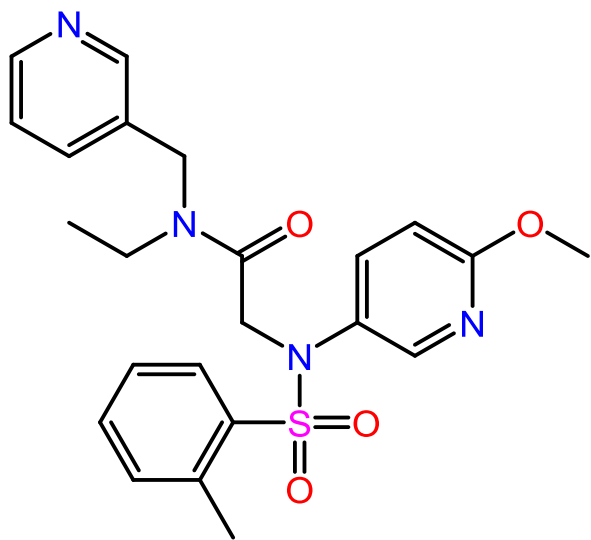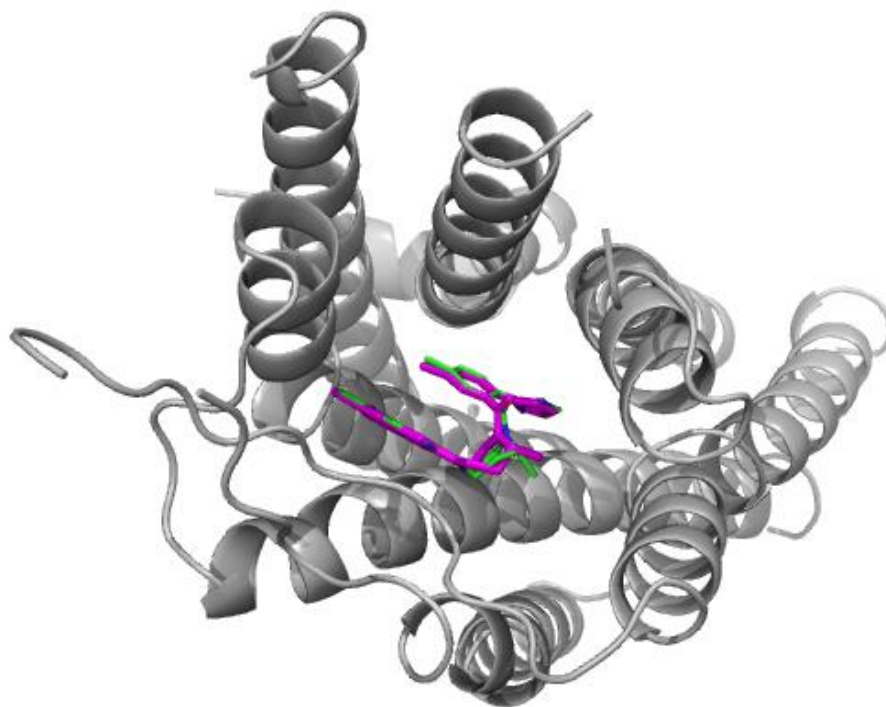
Sequence Identity: 20.1 %

Glide Successful Docks Suvorexant Into Binding Pocket

Suvorexant (SUV)
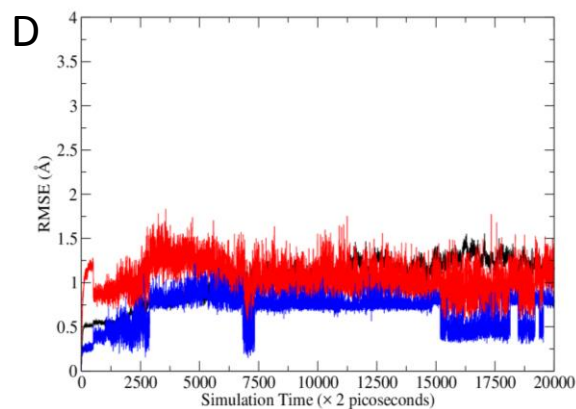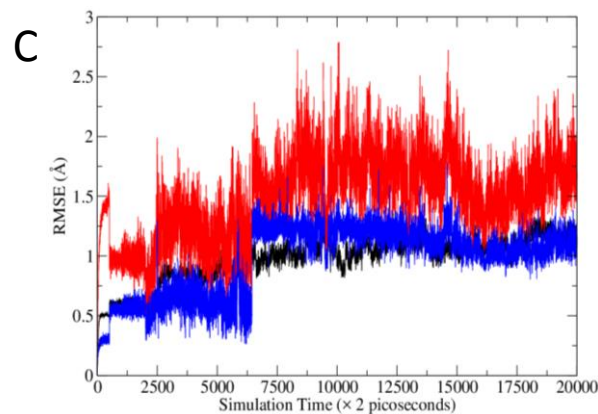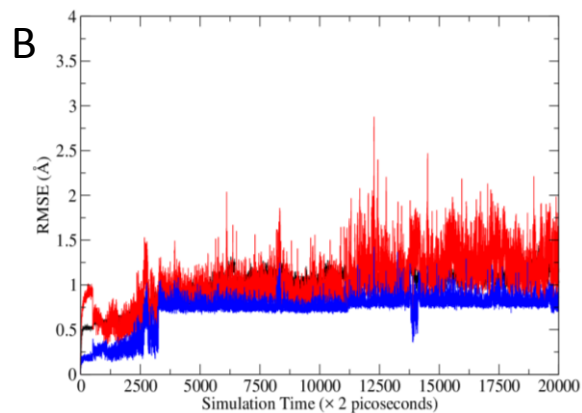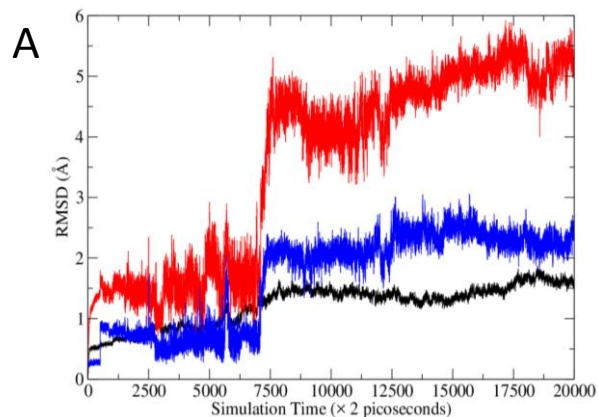**A drug used to treat insomnia**

LIG, a selective OR2 inhibitor

Glide Score: -8.19
RMSD   :   0.42 Å

RMSD plots of MD simulations

Black: receptor
**Red**: ligand without fitting
Blue: ligand with fitting

(A) hOX$_1$R/LIG   (B) hOX$_1$R/Suvorexant
(C) hOX$_2$R/LIG   (D) hOX$_2$R/Suvorexant

MM-PB/SA Analysis

Ligand	Energy Terms	hOX1R	hOX2R
Suvorexant	ΔE_{vDW}	-57.6 ± 0.1	-58.2 ± 0.1
	ΔE_{EEL}	-8.4 ± 0.1	-4.6 ± 0.1
	ΔG_{PB}	37.3 ± 0.2	33.6 ± 0.0
	ΔG_{SA}	-4.2 ± 0.0	-4.1 ± 0.0
	ΔG_{PBSA}	33.1 ± 0.2	29.5 ± 0.0
	$\Delta E_{EEL}+\Delta G_{PB}$	28.9 ± 0.2	29.0 ± 0.1
	$T\Delta S$	-23.6 ± 0.0	-24.2 ± 0.0
	ΔG_{bind}	**-9.3 ± 0.2**	**-9.1 ± 0.1**
LIG	ΔE_{vDW}	-53.5 ± 0.2	-55.2 ± 0.1
	ΔE_{EEL}	-22.8 ± 0.2	-21.9 ± 0.3
	ΔG_{PB}	53.8 ± 0.1	52.5 ± 0.4
	ΔG_{SA}	-4.3 ± 0.0	-4.5 ± 0.0
	ΔG_{PBSA}	49.4 ± 0.1	48.0 ± 0.4
	$\Delta E_{EEL}+\Delta G_{PB}$	31.0 ± 0.2	30.6 ± 0.1
	$T\Delta S$	-23.3 ± 0.0	-22.9 ± 0.0
	ΔG_{bind}	**-3.6 ± 0.1**	**-6.1 ± 0.1**

Hot Spots Identified by MM-GB/SA Analysis
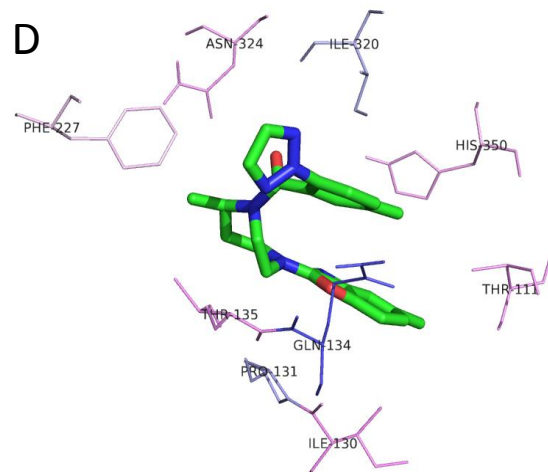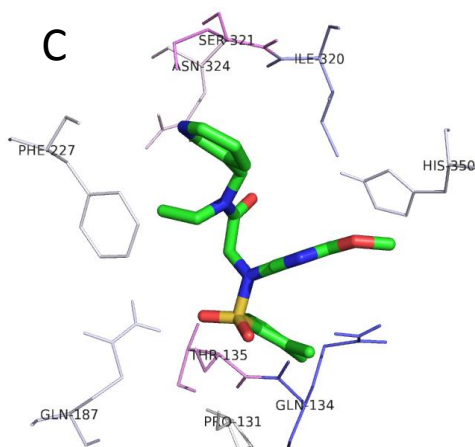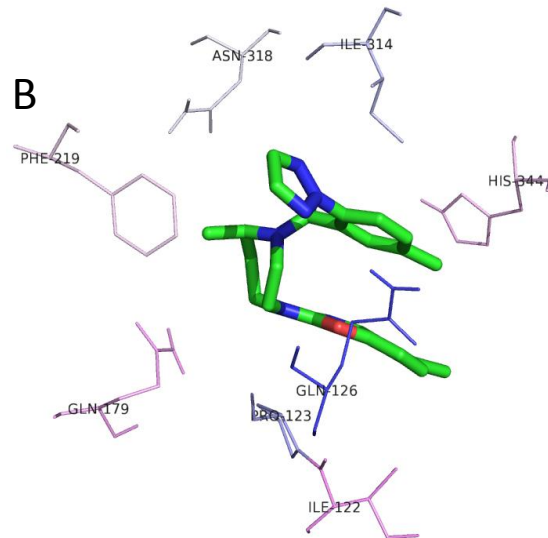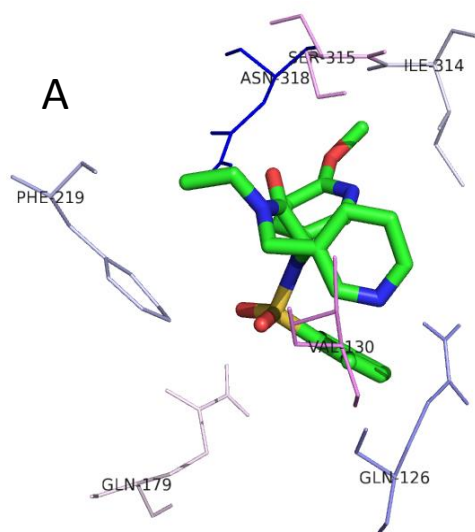
A: OR1/LIG
B: OR1/SUV
C: OR2/LIG
D: OR2/SUV

Traditional Free Energy Methods

1. Free energy perturbation (FEP)

2. Thermodynamic integration (TI)

3. Potential of mean force (PMF)

Free energy perturbation (FEP) and

Thermodynamic integration (TI)

Free energy perturbation (FEP)

Free energy differences can be calculated relatively easily and several methods have been developed for this purpose. The starting point for most approaches is Zwanzig's perturbation formula for the free energy difference between two states A and B:

$$\Delta G(A \to B) = G_B - G_A = -kT \ln \left\langle \exp^{-(H_B - H_A)/kT} \right\rangle_A$$

$$\Delta G(B \to A) = G_A - G_B = -kT \ln \left\langle \exp^{-(H_A - H_B)/kT} \right\rangle_B$$

$$\Delta G(A \to B) = -\Delta G(B \to A)$$

The equality should hold if there is sufficient sampling.

However, if the two states are not similar enough, this is difficult to achieve and there will be a large hysteresis effect (i.e. the forward and backward results will be very different).

FEP with alchemical transformation

To obtain accurate results with the perturbation formula, the energy difference between the states should be < 2 kT, which is not satisfied for most biomolecular processes. To deal with this problem, one introduces a hybrid Hamiltonian

$$H(\lambda) = (1 - \lambda)H_A + \lambda H_B$$

and performs the transformation from A to B gradually by changing the parameter λ from 0 to 1 in small steps. That is, one divides [0,1] into n subintervals with $\{\lambda_i, i = 0, n\}$, and for each λ_i value, calculates the free energy difference from the ensemble average

$$\Delta G(\lambda_i \to \lambda_{i+1}) = -kT \ln\left\langle \exp[-(H(\lambda_{i+1}) - H(\lambda_i))]/kT \right\rangle_{\lambda_i}$$

FEP with alchemical transformation - continued

The total free energy change is then obtained by summing the contributions from each subinterval

$$\Delta G(0 \rightarrow 1) = \sum_{i=0}^{n-1} \Delta G(\lambda_i \rightarrow \lambda_{i+1})$$

The number of subintervals is chosen such that the free energy change at each step is < 2 kT, otherwise the method may lose its validity.   Points to be aware of:

Most codes use equal subintervals for λ_i. But the changes in ΔG_i are usually highly non-linear. One should try to choose λ_i such that ΔG_i remains around 1-2 kT for all values.

The simulation times (equilibration + production) have to be chosen carefully. It is not possible to extend them in case of non-convergence (have to start over).

Thermodynamic integration (TI)

Another way to obtain the free energy difference is to integrate the derivative of the hybrid Hamiltonian H(λ):

$$\frac{dG}{d\lambda} = \frac{\int \frac{\partial H}{\partial \lambda} e^{-H/kT} dpdq}{\int e^{-H/kT} dpdq} = \left\langle \frac{\partial H}{\partial \lambda} \right\rangle_{\lambda}$$

$$\Delta G = \int_{0}^{1} \left\langle \frac{\partial H(\lambda)}{\partial \lambda} \right\rangle_{\lambda} d\lambda$$

This integral is evaluated most efficiently using a Gaussian quadrature.
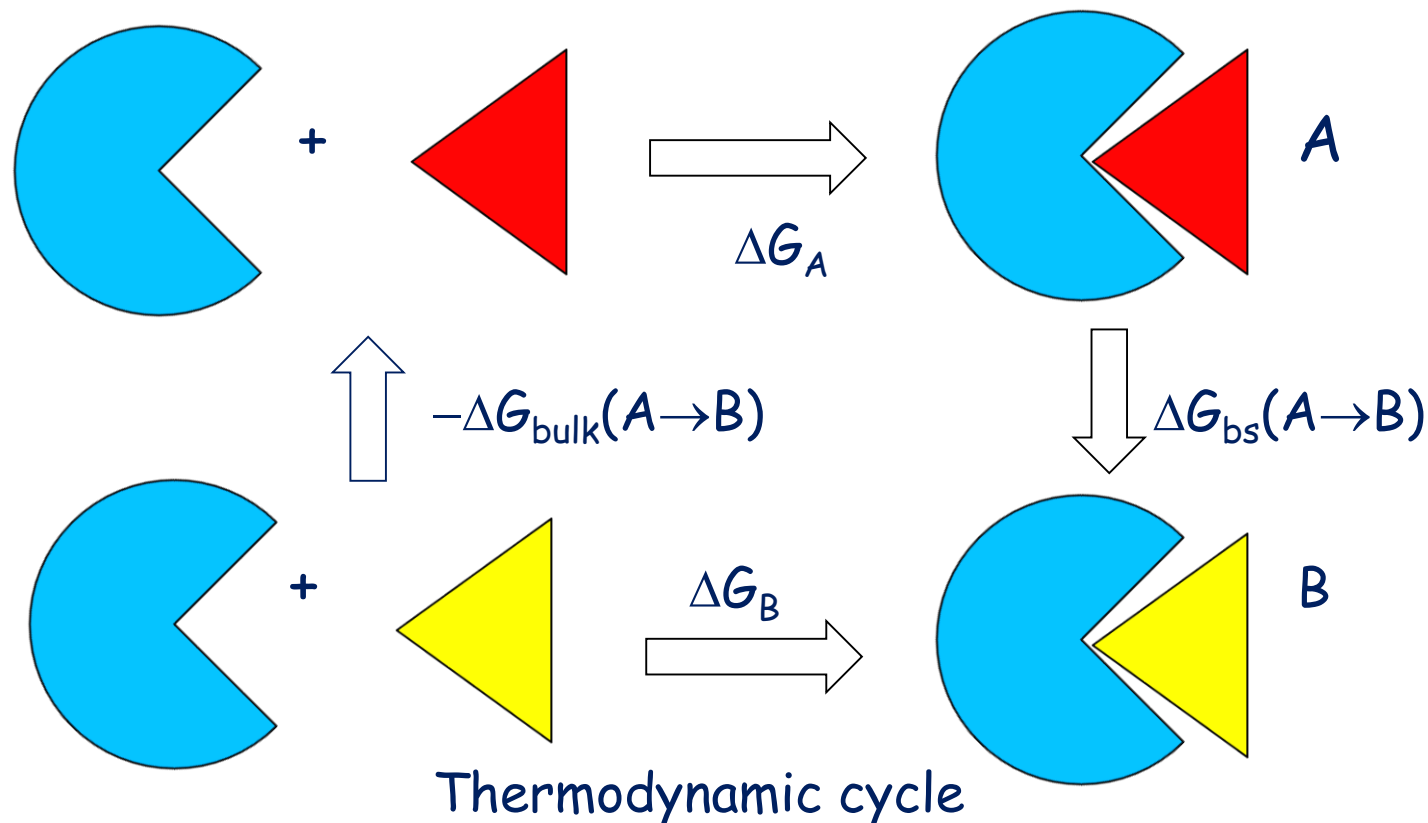In typical calculations for ions, 7-point quadrature is sufficient.
(But check that 9-point quadrature gives the same result for others)
The advantage of TI over FEP is that the production run can be extended as long as necessary and the convergence of the free energy can be monitored (when the cumulative ΔG flattens, it has converged).

Example: Free energy change in mutation of a ligand

A very common question is how a mutation in a ligand (or protein) changes the free energy of the protein-ligand complex.

$$\Delta G_B - \Delta G_A = \Delta G_{bs}(A \rightarrow B) - \Delta G_{bulk}(A \rightarrow B)$$

ΔG_A

A

$-\Delta G_{bulk}(A \rightarrow B)$

$\Delta G_{bs}(A \rightarrow B)$

ΔG_B

B

Thermodynamic cycle

Applications

1. Ion selectivity of potassium channels

$$\Delta\Delta G_{sel}(K^+ \rightarrow Na^+) = \Delta G_b(Na^+) - \Delta G_b(K^+)$$

$$= \Delta G_{bs}(K^+ \rightarrow Na^+) - \Delta G_{bulk}(K^+ \rightarrow Na^+)$$

2. Selectivity of amino acid transporters (e.g. glutamate transporter)

$$\Delta\Delta G_{sel}(Asp \rightarrow Glu) = \Delta G_b(Glu) - \Delta G_b(Asp)$$

$$= \Delta G_{bs}(Asp \rightarrow Glu) - \Delta G_{bulk}(Asp \rightarrow Glu)$$

3. Free energy change when a side chain is mutated in a bound ligand. Similar calculation as above. Important in developing drug leads from peptides.
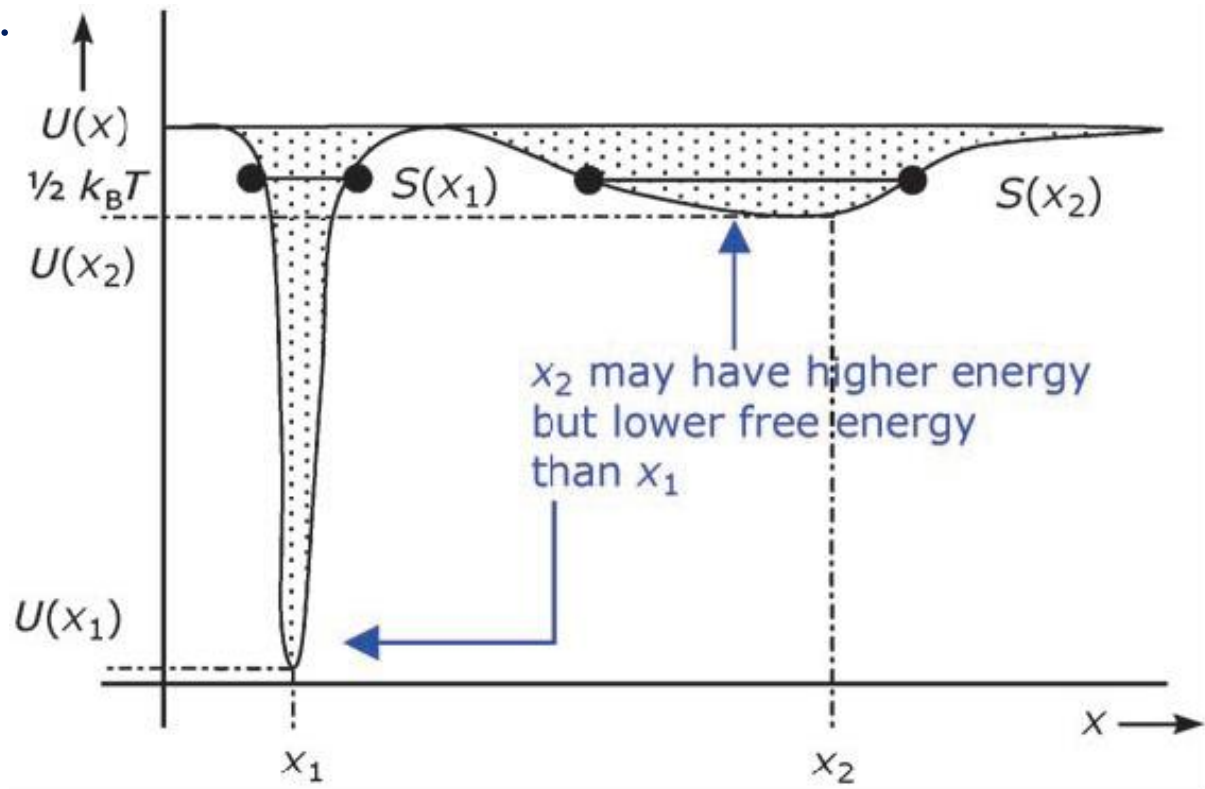
Sampling Issue

At zero temperature, the potential function U is sufficient to characterize the system completely. At room temperature, the fundamental quantity is the free energy, F = U − TS, which creates the sampling problem. Example: F= − 24, U= − 41, and TS= − 17 (kJ/mol) for liquid water at STP.

Statistical weight:

$$P(x) \sim e^{-U(x)/kT}$$

But if $S_2 \gg S_1$

we may have

$F_2 < F_1$

$U(x)$

½ k_BT

$U(x_2)$

$S(x_1)$

$S(x_2)$

x_2 may have higher energy but lower free energy than x_1

$U(x_1)$

x_1

x_2

x

Points to consider for FEP

In FEP, one has to decide on the number of windows and the

equilibration time in advance. The windows are created serially, so

if the equilibration time is inadequate, it has to be repeated using

longer equilibration time and the initial data are wasted.

A second potential problem in FEP calculations is the requirement

that ΔG_i remains around 1-2 kT for all windows. Because the

change in the free energy is nonlinear, it is very difficult to guess

the number of windows one should use. For the same reason,

using fixed intervals is not optimal. Exponentially spaced

intervals would reduce the required number of windows by half.

Example: Na⁺ binding energy in glutamate transporter
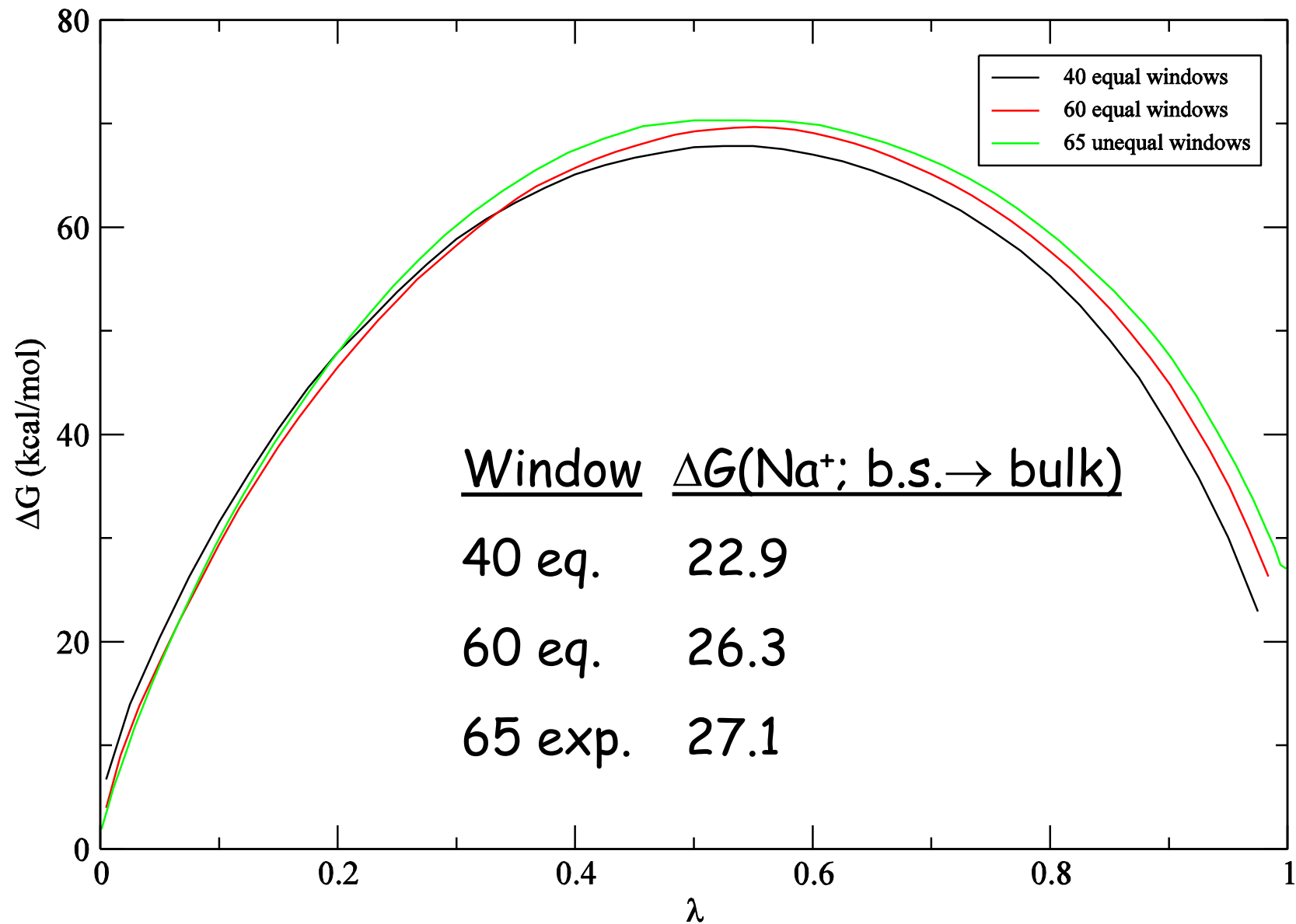
Free energy change ΔG at each step of FEP calculation
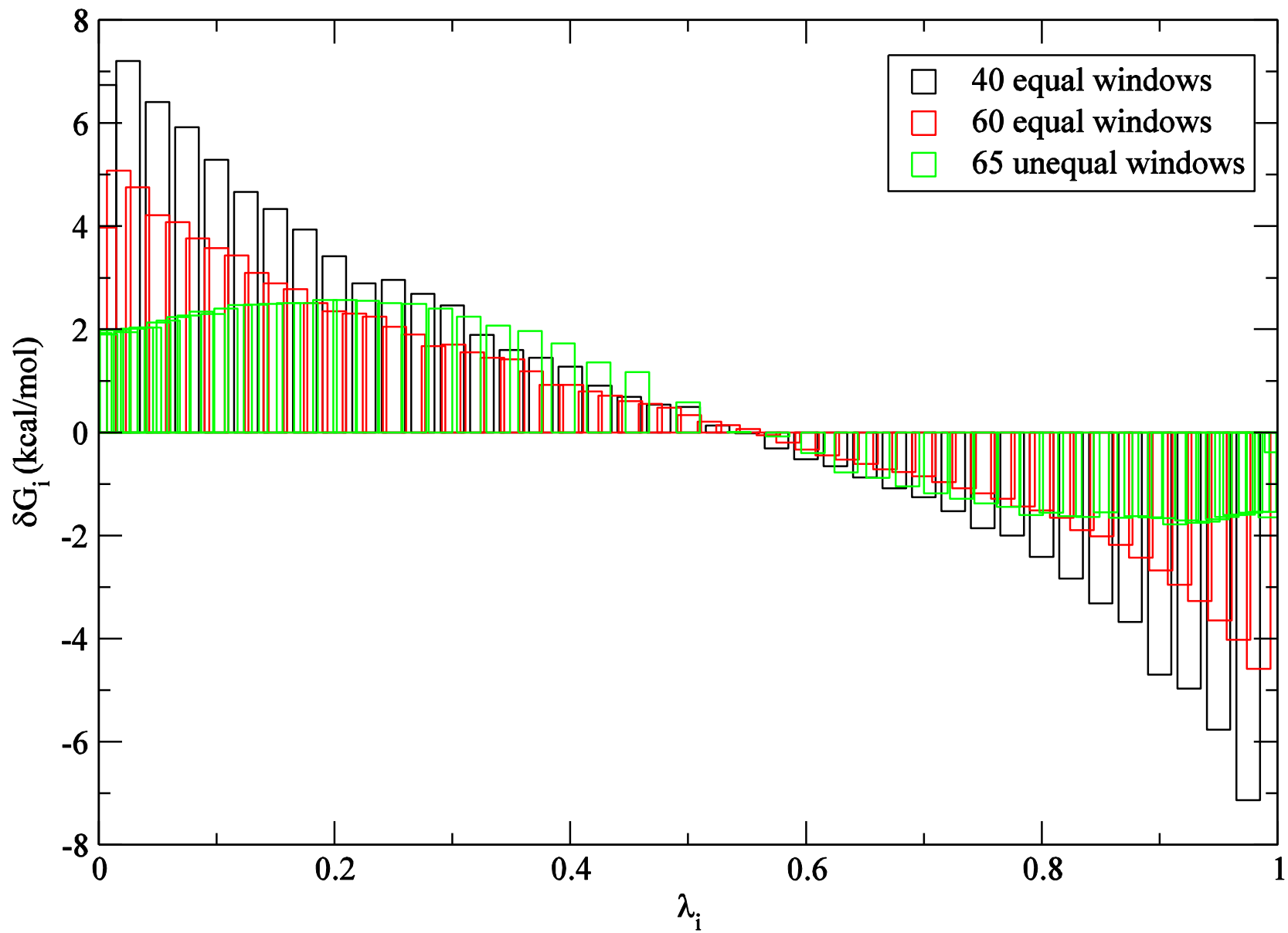
Points to consider for TI

In TI , one only need to specify the number of windows in advance. The data can be divided into equilibration and production parts later. Moreover, one can continue accumulating data if there is a problem with convergence, thus there is no wastage of data.

Convergence can be monitored by plotting the running average of the free energy. Flattening out of the curve is usually taken as a sign for convergence.

Because small number of windows are used in TI, equilibration may prove difficult in some systems. An initial FEP calculation with large number of windows can resolve this problem (choose the TI windows from the nearest FEP window).

Example: Na⁺ and Asp binding energies in glut. transporter

TI calculation of the binding free energy of Na+ ion to the binding site 1 in Gltph. Integration is done using Gaussian quadrature with 7 points. Thick lines show the running averages, which flatten out as the data accumulate. Thin lines show averages over 50 ps blocks of data.
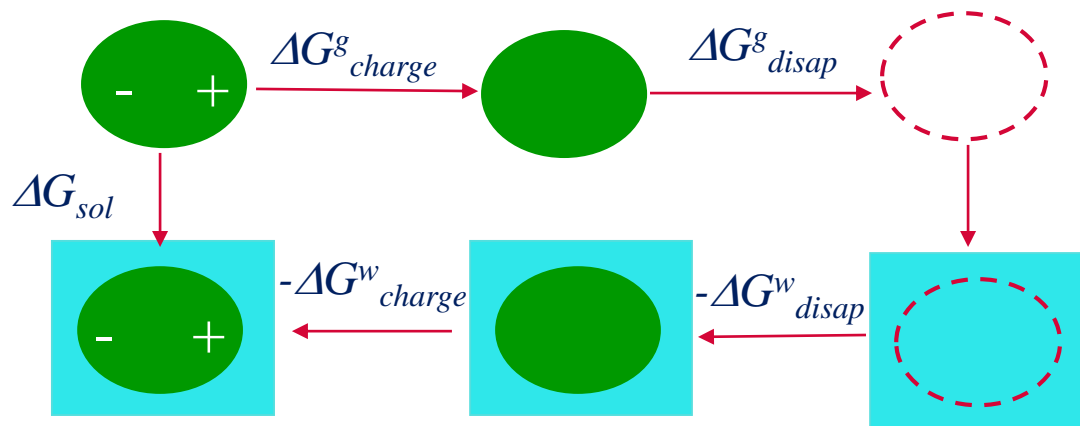

open Gltph ion1
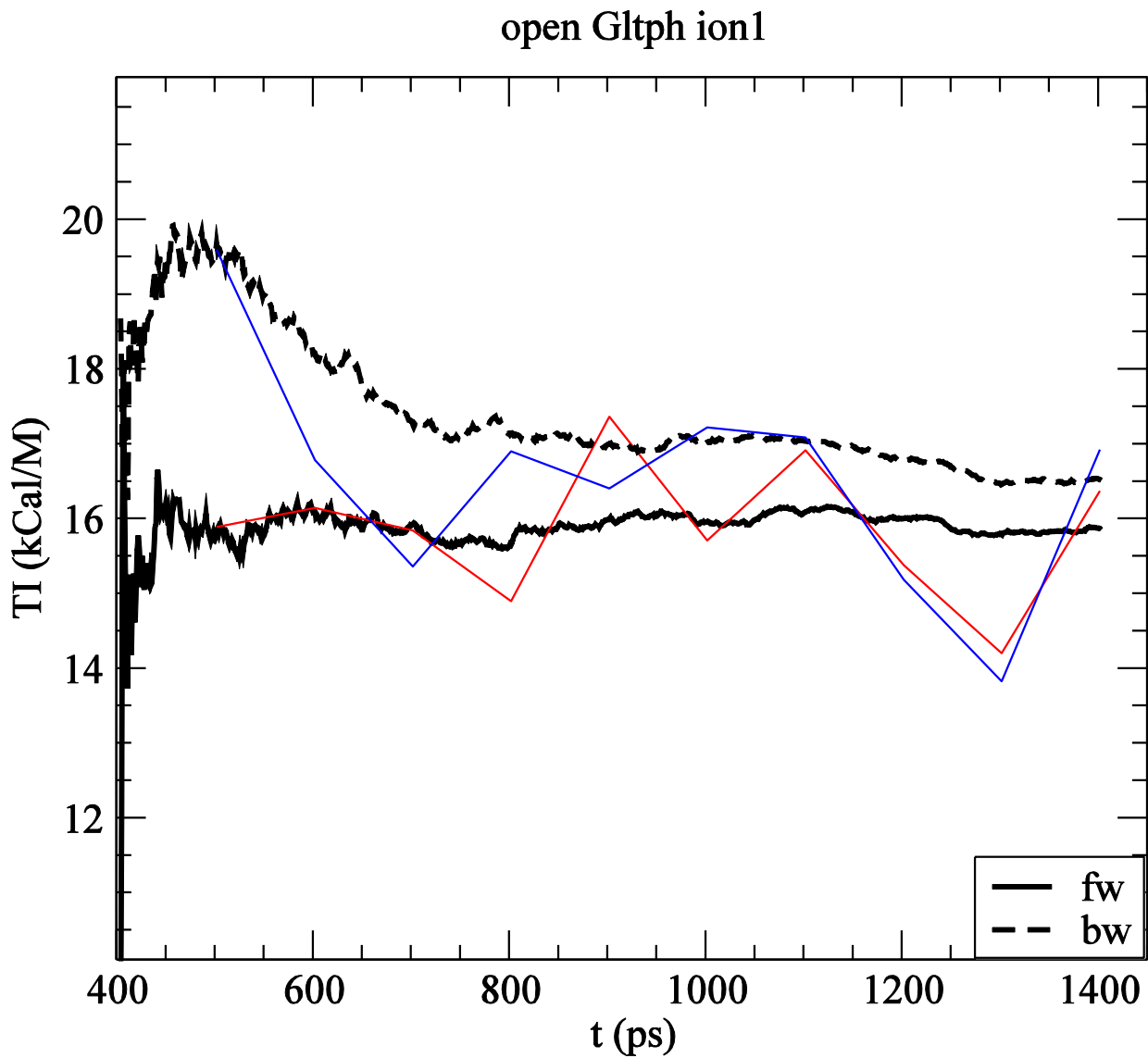
Asp binding energy in glutamate transporter

TI calculation of the
binding free energy of
Asp to the binding site
in Gltph.

Asp is substituted with
5 water molecules.

First 400 ps data
account for equilibration
and the 1 ns of data are
used in the production.

open Gltph asp
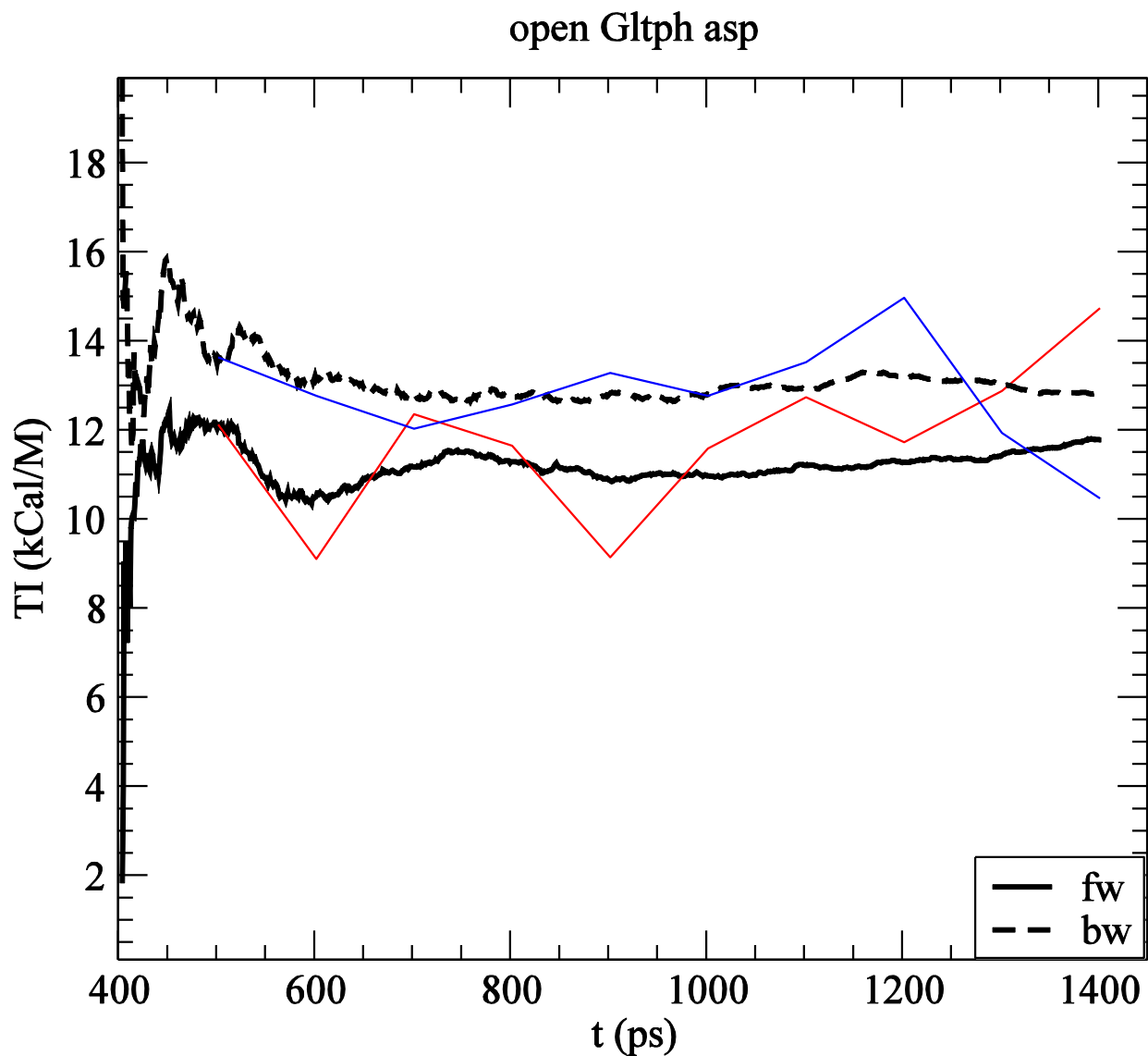
Computational Details – Solvation Free Energy Calculations

$$\Delta G = \int_0^1 < \partial V / \partial \lambda >_\lambda d\lambda$$

$$\Delta G \approx \sum_{i=1}^n w_i < \partial V / \partial \lambda >_{\lambda_i}$$

$$V_\lambda = f(\lambda)V_0 + [1 - f(\lambda)]V_1$$

$$V_\lambda = \lambda V_0 + (1 - \lambda)V_1$$

➢12-window Gaussian integration
1ns each window

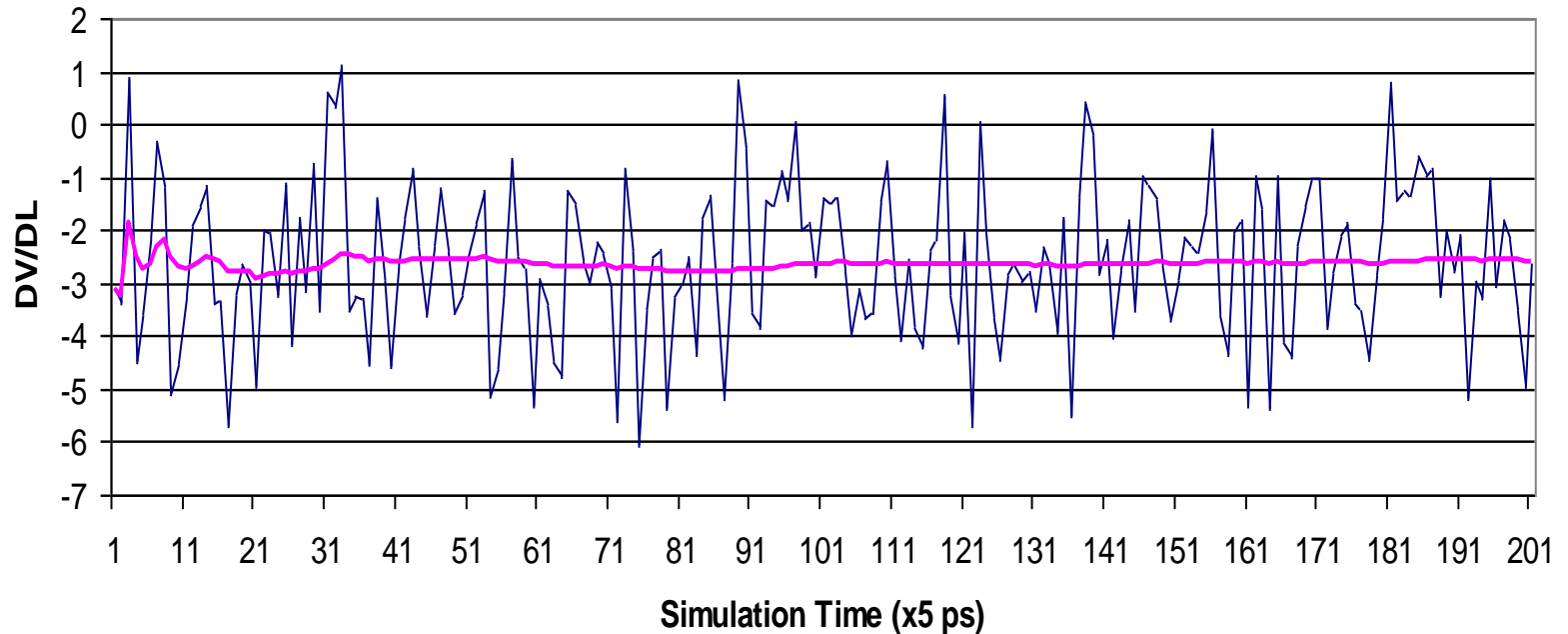# DV/DL ~ Simulation Time Plot clambda = 0.31608 for Methanol

Charging in Gas Phase
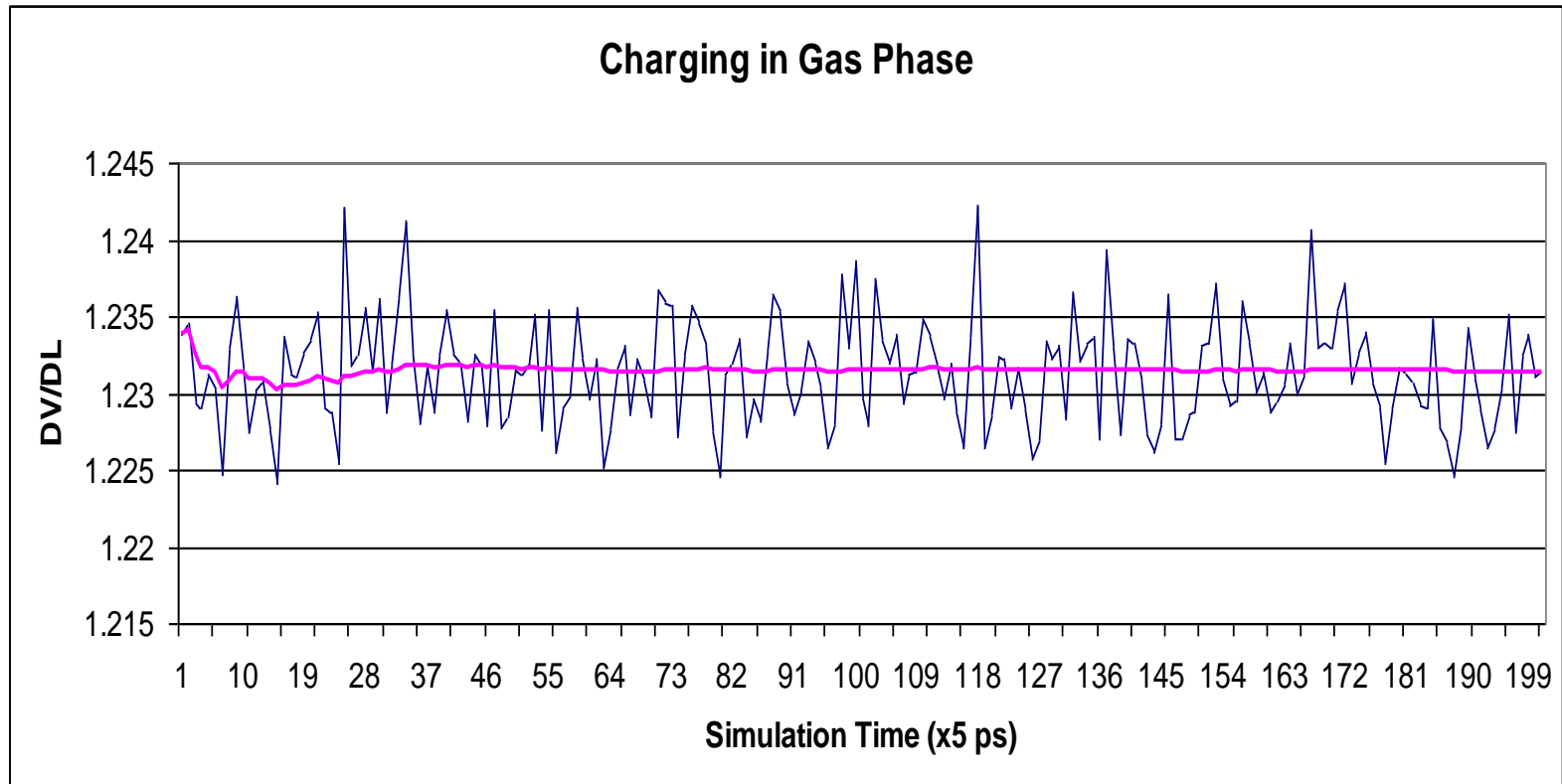
DV/DL ~ Simulation Time Plot clambda = 0.31608 for Toluene

Charging in Water
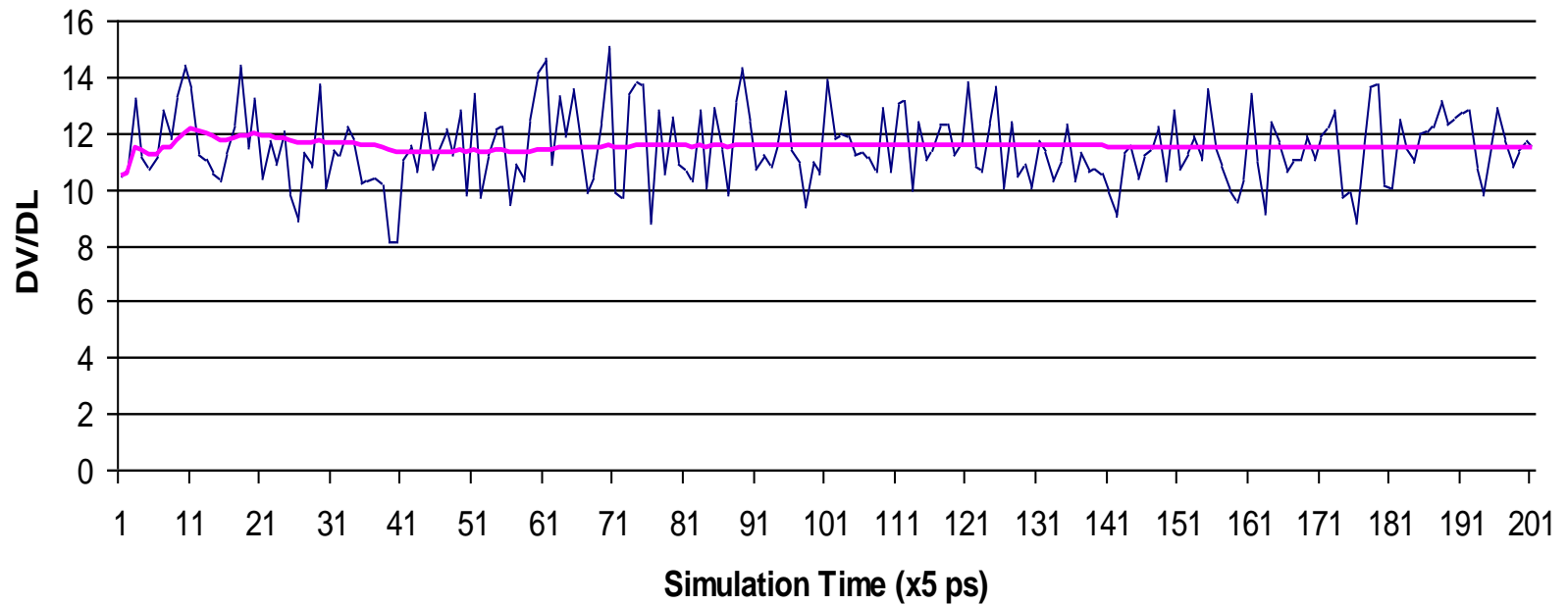
DV/DL ~ Simulation Time Plot clambda = 0.31608 for Toluene

Disappearing in Water

Performance of Solvation Free Energy Calculations

	Compound	Expt.	GAFF	Applequist/ GAFF	Model B1	Model C1
1	methanol	-5.07	-3.97	**-5.91**	-7.87	-6.61
2	benzaldehyde	-4.02	-3.19	**-4.74**	-8.3	-7.51
3	acetic acid	-6.7	-7.92	**-4.56**	-	-
4	methyl amine	-4.6	-4.75	**-4.67**	-5.76	-5.35
5	dimethyl amine	-4.29	-2.41	-3.08	-5.44	**-4.44**
6	trimethyl amine	-3.23	0.27	-0.16	**-2.22**	-1.99
7	acetamide	-9.72	-9.15	-11.38	**-10.27**	-10.55
8	ammonium	-81.53	-68.91	-70.94	**-73.46**	-73.16
9	N-guanidinium	-66.07	-59.61	**-66.57**	-	-
10	acetate ion	-80.65	**-94.81**	-99.54	-107.26	-104.1

Potential of mean force (PMF)

Potentials of Mean Force

- May wish to examine the Free Energy as a function of some inter- or intramolecular coordinate. (ie. Distance, torsion angle etc.)

- The free energy along the chosen coordinate is known as the Potential of Mean Force (PMF).

- Calculated for physically achievable processes so the point of highest energy corresponds to a TS.

- Simplest type of PMF is the free energy change as the separation (r) between two particles is varied.

- PME can be calculated from the radial distribution function ($g(r)$) using:

$$A(r) = -k_B T \ln g(r) + \text{constant}$$

  - $g(r)$ is the probability of finding an atom at a distance r from another atom.

Potentials of Mean Force

- Problem: The logarithmic relationship between the PMF and $g(r)$ means a relatively small change in the free energy (small multiple of k_BT may correspond to $g(r)$ changing by an order of magnitude.

  – MC and MD methods do not adequately sample regions where the radical distribution function differs drastically from the most likely value.

- Solution: Umbrella Sampling.

  – The coordinates of interest are allowed to vary over their range of values throughout the simulation. (Subject to a potential modified using a forcing function.)

Umbrella Sampling

- The Potential Function can be written as a perturbation:

$$\mathcal{V}'(\mathbf{r}^N) = \mathcal{V}(\mathbf{r}^N) + W(\mathbf{r}^N)$$

  – Where $W(\mathbf{r}^N)$ is a weighting function which often takes a quadratic form:

$$W(\mathbf{r}^N) = k_W (\mathbf{r}^N - \mathbf{r}_0^N)^2$$

  – Result: For configurations far from the equilibrium state, $\mathbf{r}_0^N$, the weighting function will be large so the simulation will be biased along some relevant reaction coordinate.

  – The Boltzmann averages can be extracted from the non-Boltzmann distribution using:

$$\langle A \rangle = \frac{\langle A(\mathbf{r}^N) \exp\left[+ W(\mathbf{r}^N) / k_B T \right] \rangle_W}{\langle \exp\left[+ W(\mathbf{r}^N) / k_B T \right] \rangle_W}$$

  - Subscript W indicates that the average is based on the probability $P_W(\mathbf{r}^N)$, determined from the modified energy function $\mathcal{V}'(\mathbf{r}^N)$.

Points to consider in umbrella sampling

Two main parameters in umbrella sampling are the force constant, k

and the distance between windows, d. In bulk, the position of the

ligand will have a Gaussian distribution given by

$$P(z) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(z-z_0)^2/2\sigma^2}, \qquad \langle z \rangle = z_0, \quad \sigma = \sqrt{k_B T / k}$$

The overlap between two Gaussian distributions separated by d

$$\% \, overlap = 1 - erf(d/\sqrt{8}\sigma)$$

The parameters should be chosen such that 10% > % overlap > 5%

If the overlap is too small, PMF will have discontinuities

If it is too large, simulations are not very efficient.

Steered MD (SMD)

Steered MD (SMD) simulations and Jarzynski's equation

Steered MD is a more recent method where a harmonic force is applied to an atom on a peptide and the reference point of this force is pulled with a constant velocity. It has been used to study unfolding of proteins and binding of ligands. The discovery of Jarzynski's equation in 1997 enabled determination of PMF from SMD, which has boosted its applications.

Jarzynski's equation:

$$e^{-\Delta F / kT} = \left\langle e^{-W / kT} \right\rangle$$

Work done by
the harmonic force

$$W = \int_i^f \mathbf{F}.\mathbf{ds}, \quad \mathbf{F} = k[\mathbf{r} - (\mathbf{r}_0 + \mathbf{v}t)]$$

This method seems to work well in simple systems and when ΔG is large but beware of its applications in complex systems!

# Example: PMF for binding of charybdotoxin to K⁺ channel

From the previous examples, we have seen that ions equilibrate quite fast (~100 ps) and < 1 ns production run is sufficient for PMF.
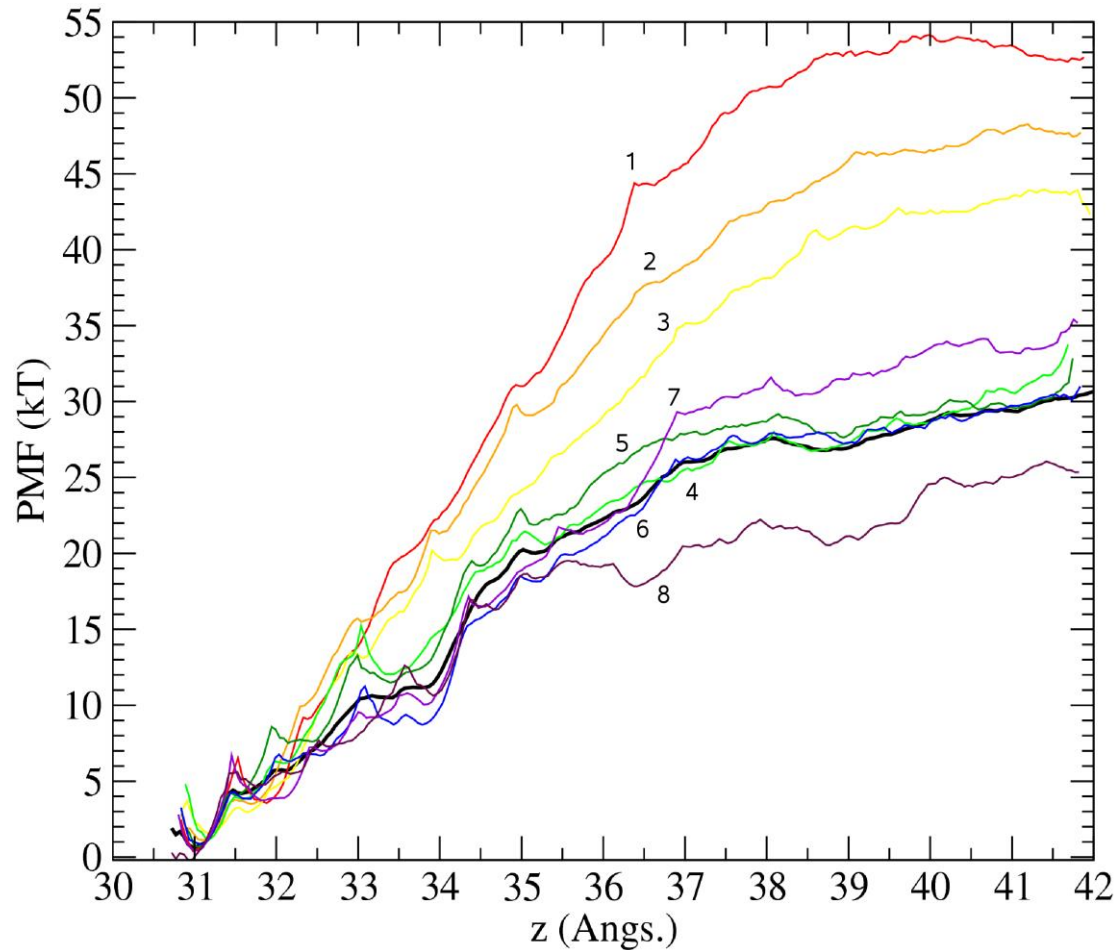
For complex ligands, the situation is obviously more complicated.
For one thing, the ligand may be distorted, which will lead to erroneous results.
One also requires much longer equilibration of the system (typically > 1 ns), and longer production runs (> 1 ns).
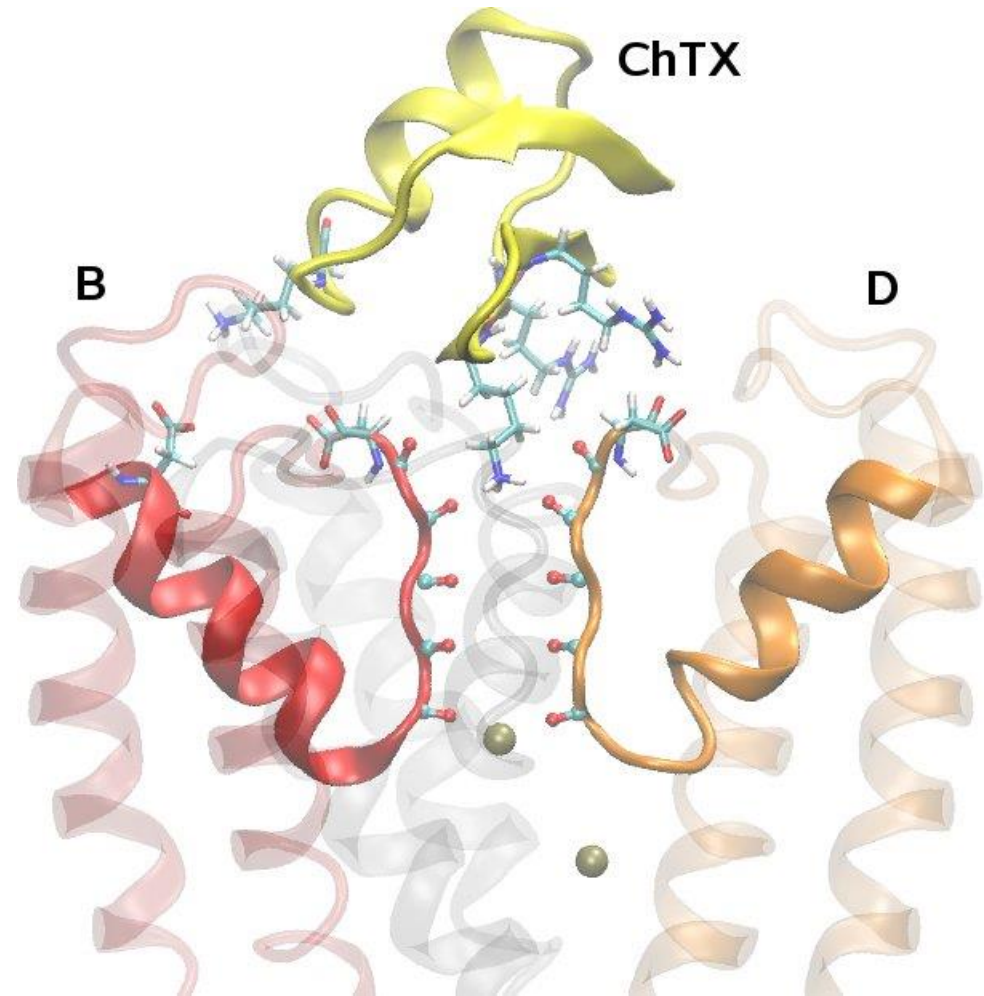
Convergence of the toxin PMF

Force constant: k=20 kcal/mol/A² Umbrella windows: 0.5 A

Each color represents 400 ps of sampling. The first 1.2 ns is dropped for equilibration and PMF is obtained from the last 2 ns (black line)

Lab Section

Run MD Simulations

1. **Run MD simulations**

    sander

    pmemd

    pmemd.MPI

    pmemd.cuda

2. **Replay MD Trajectories With VMD**

Analyze MD Snapshots

1. **Programs**

   ptraj, cpptraj

2. **Input file for ptraj and cpptraj**

MM-PB/GBSA

1. **Programs**

   mmpbsa.py

2. **Input files**

3. **Output files**

Delphi

Use Pymol to plot electrostatic potential