Computational Biochemistry

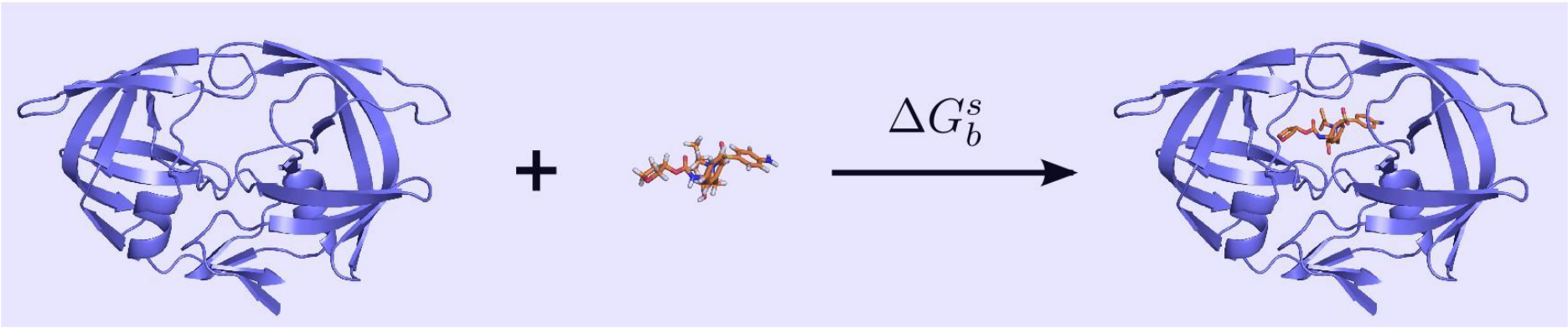
Lecture 6
Interaction Free Energy Prediction

Gibbs free energy

$$\Delta G = \Delta H - T\Delta S$$

- ΔH -- enthalpy
- ΔS -- entropy

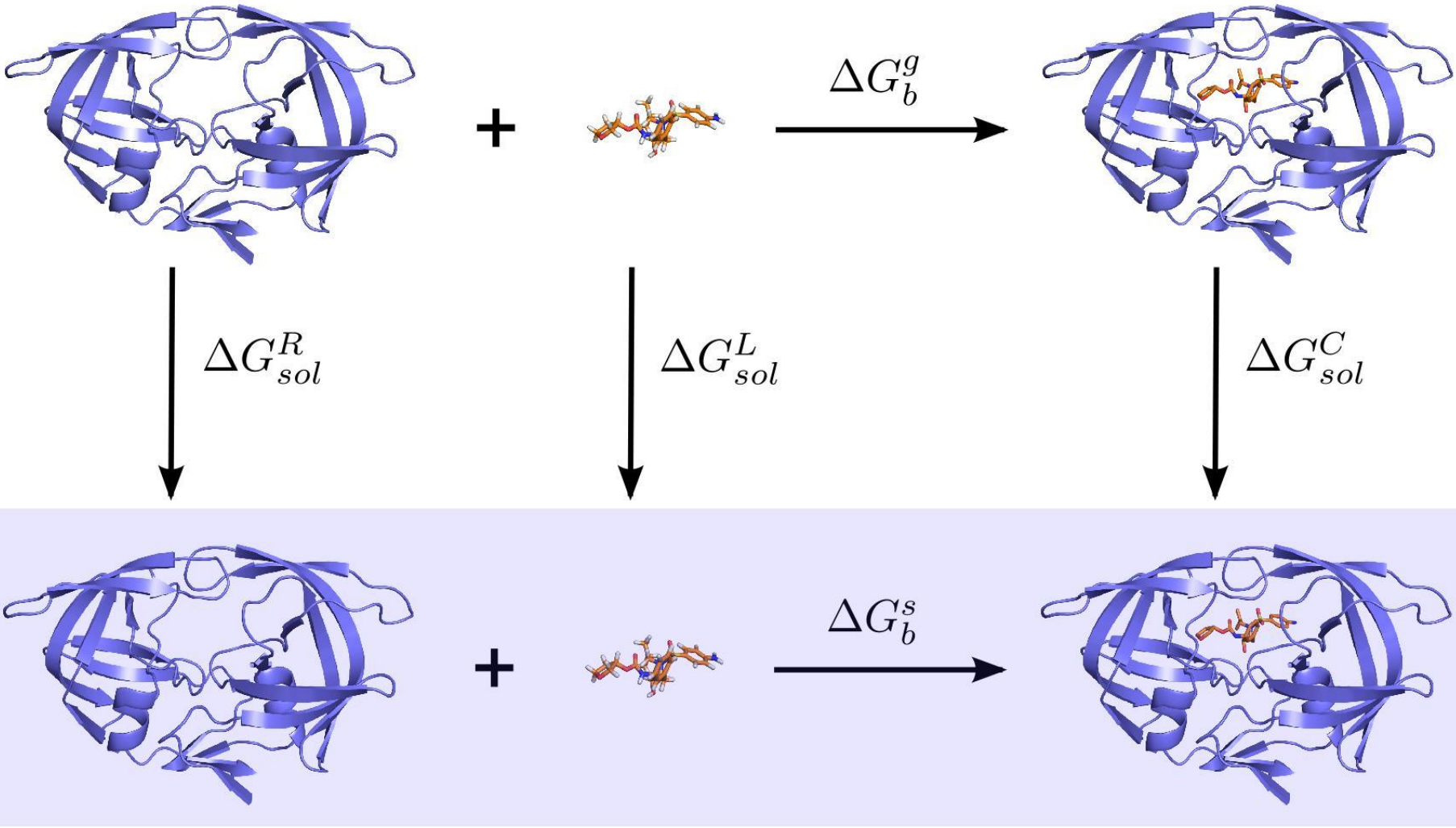
Direct ΔG calculations



$$\Delta G_{bind,solution} = \Delta G_{complex,solution} - (\Delta G_{receptor,solution} + \Delta G_{ligand,solution})$$

The binding free energy for complex formation can be estimated from the difference of the solution-phase free energies of complex ($\Delta G_{complex,solution}$), and those of ligand ($\Delta G_{ligand,solution}$) and receptor ($\Delta Greceptor,solution$)

Direct ΔG calculations



Direct ΔG calculations

$$\Delta G_{bind,solution} = \Delta G_{complex,solution} - (\Delta G_{receptor,solution} + \Delta G_{ligand,solution})$$

$$\Delta G_{complex,solution} = \Delta G_{complex,vacuo} + \Delta G_{complex,solvation}$$

$$\Delta G_{receptor,solution} = \Delta G_{receptor,vacuo} + \Delta G_{receptor,solvation}$$

$$\Delta G_{ligand,solution} = \Delta G_{ligand,vacuo} + \Delta G_{ligand,solvation}$$

$$\Delta G_{bind,solution} = \Delta G_{complex,vacuo} - (\Delta G_{receptor,vacuo} + \Delta G_{ligand,vacuo})$$

$$+ \Delta G_{complex,solvation} - (\Delta G_{receptor,solvation} + \Delta G_{ligand,solvation})$$

$$= \Delta G_{bind,vacuo} + [\Delta G_{complex,solvation} - (\Delta G_{receptor,solvation} + \Delta G_{ligand,solvation})]$$

Direct ΔG calculations

$$\Delta G = \Delta H - T\Delta S$$

$$\Delta G_{bind,vacuo} = \Delta H - T\Delta S$$

The enthalpic contribution (ΔH) to binding is usually approximated by the difference of MM energies between the complex and those for protein and ligand (ΔE_{MM}).

$$\Delta G_{bind,vacuo} = \Delta E_{MM} - T\Delta S$$

Direct ΔG calculations

ΔE_{MM} can be calculated with the difference of the internal energies ($\Delta E_{internal}$) (bond, angle, and dihedral energies), and that of the non-bond interaction energies (ΔE_{ele} and ΔE_{vdW}).

$$\Delta E_{MM} = \Delta E_{internal} + \Delta E_{ele} + \Delta E_{vdW}$$

$$\Delta E_{internal} = \Delta E_{bond} + \Delta E_{angle} + \Delta E_{dihedral}$$

In calculating the binding free energies, a single trajectory approach is often employed, in which snapshots taken from a single MD simulation trajectory are employed for calculation.

$\Delta E_{internal}$ is cancelled since the geometries of the receptor and ligand are the same as those in complex.

ΔE_{MM} depends only on the van der Waals and electrostatic interaction energies.

Entropy

Entropy is an essential component in ΔG and must be considered in order to model many chemical processes, including protein folding, and protein – ligand binding.

Entropy

Configurational Entropy

- rotational
- translational
- vibrational

➢ Calculated by Quasi-harmonic approximation

Conformational Entropy

- Conformational changes

➢ Estimated from an analysis of the covariance matrix

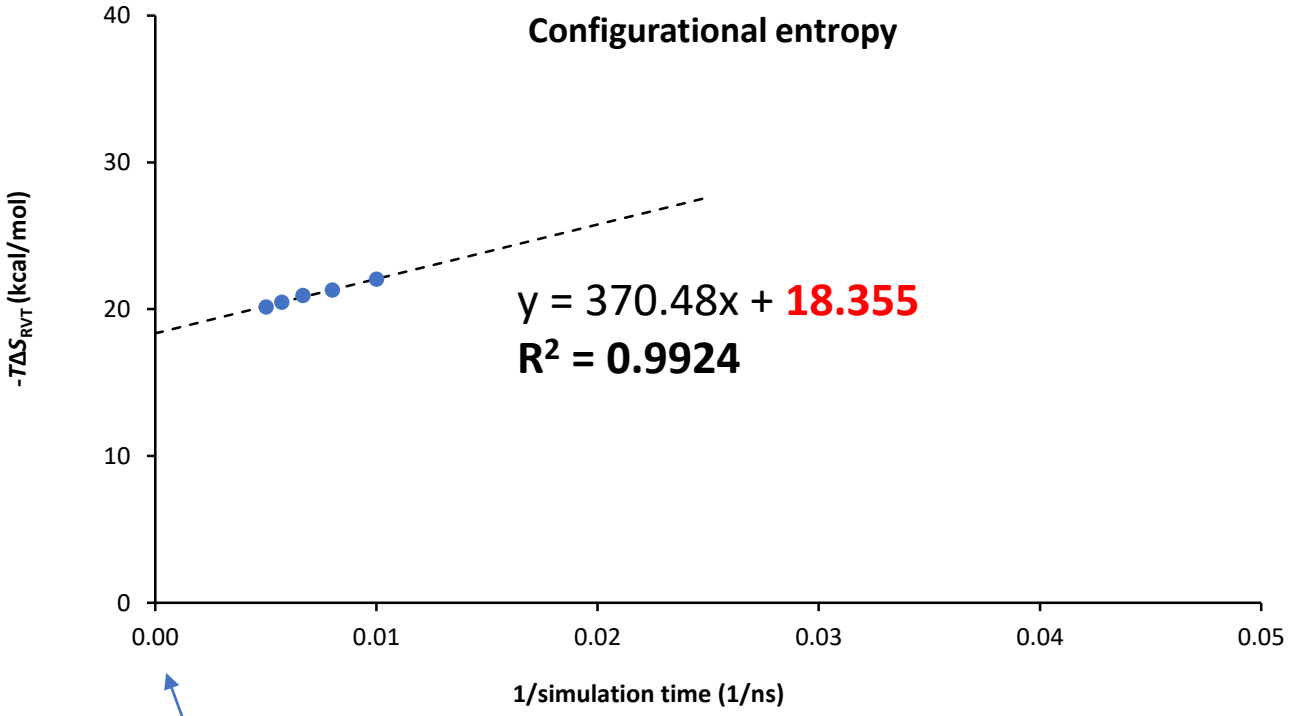
Configurational entropy

The difficulty of calculating configurational entropy, or maybe all contributions to entropy, is sampling.

In other words, accuracy of entropy calculation depends on whether or not MD simulations could comprehensively access all conformations of the complex.

time (ns)	100	125	150	175	200
$-T\Delta S_{RTV}$	22.0	21.3	20.9	20.5	20.1

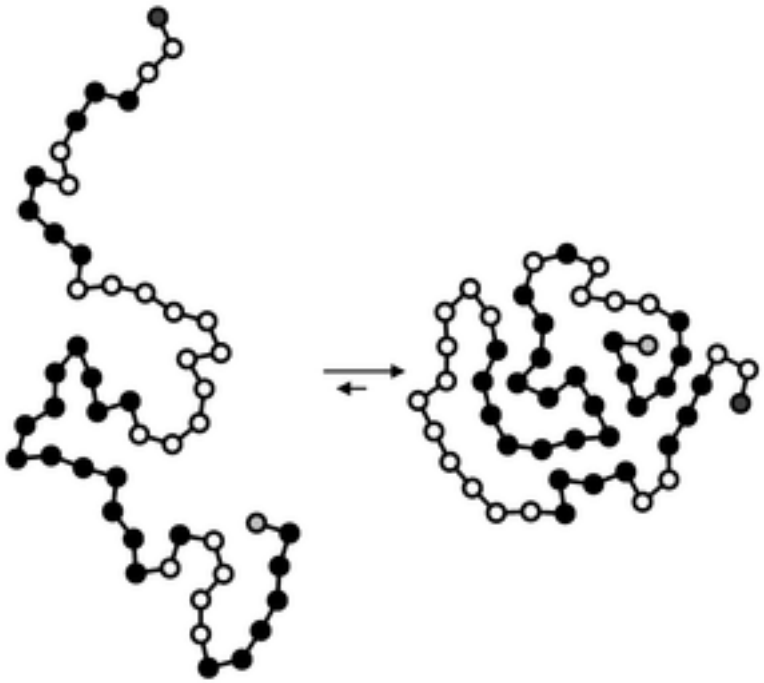
Extrapolating configurational entropy



Infinite simulation time

Conformational entropy

The entropy of heterogeneous random coil or denatured proteins is significantly higher than that of the folded native state tertiary structure

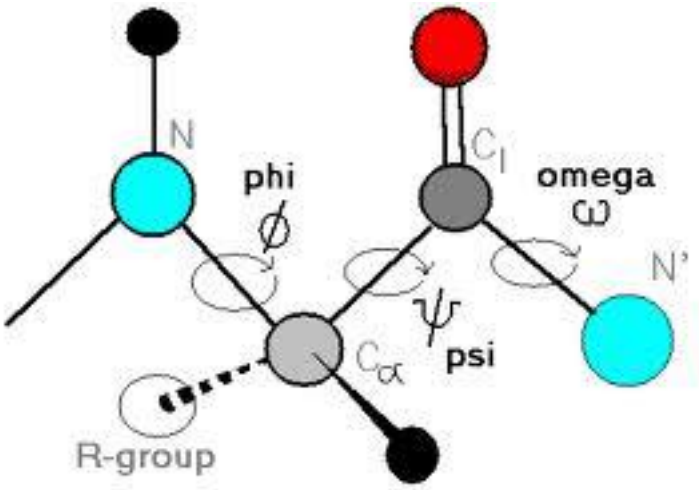


Enthalpy (ΔH) is favorable – due to the formation of hydrogen bonds, salt-bridges, dipolar interactions, van der Waals contacts and other dispersive interactions

Entropy (ΔS) is unfavorable – due to a reduction in the number of degrees of freedom of the molecule – that is, entropy favors disorder

Conformational entropy

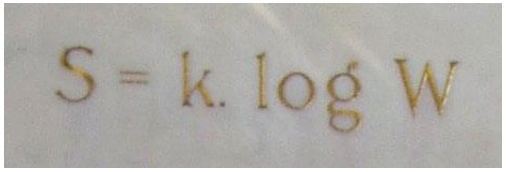
To calculate conformational entropy, the possible conformations may first be discretized into a finite number of states, usually characterized by unique combinations of certain structural parameters, such as rotamers, each of which has been assigned an energy level.



- In proteins, backbone dihedral angles and side chain rotamers are commonly used as conformational descriptors.
- These characteristics are used to define the degrees of freedom available to the molecule.
- Discretize = To convert a continuous space into an equivalent discrete space for the purposes of easier calculation.

Boltzmann's entropy formula

$$S = -R \ln W$$

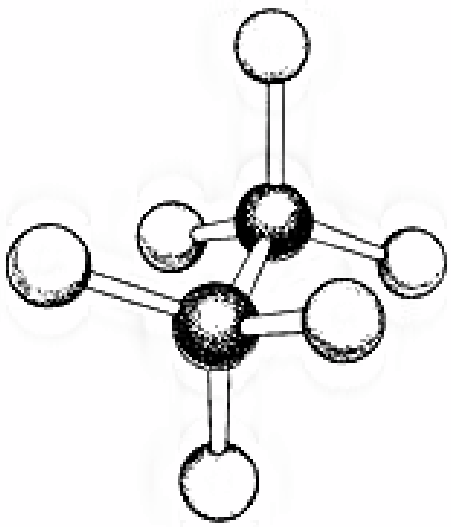


Boltzmann's equation—carved on his gravestone

Where W is the number of different conformations populated in the molecule, R is the gas constant $R = 1.9872036(11)\ cal \cdot K^{-1} \cdot mol^{-1}$

For a single C-C bond (sp3-sp3) there are 3 possible rotamers (gauche+, gauche+, anti-). If we assume that each is equally populated, that is, each bond is 33% g+, 33% g-, and 33% anti. Then $W = 3$.

So, $S = -R \ln 3 = -2.2\ cal \cdot K^{-1} \cdot mol^{-1}$ per rotatable bond



How much energy is this at 300K?

0.66 kcal/mol – can you derive this?

But, what if the rotamers are not populated equally?

Conformational Entropy as a Function of State Populations

The conformational entropy associated with a particular conformation is then dependent on the probability associated with the occupancy of that state.

Conformational entropies can be defined by assuming a Boltzmann distribution of populations for all possible rotameric states[1]:

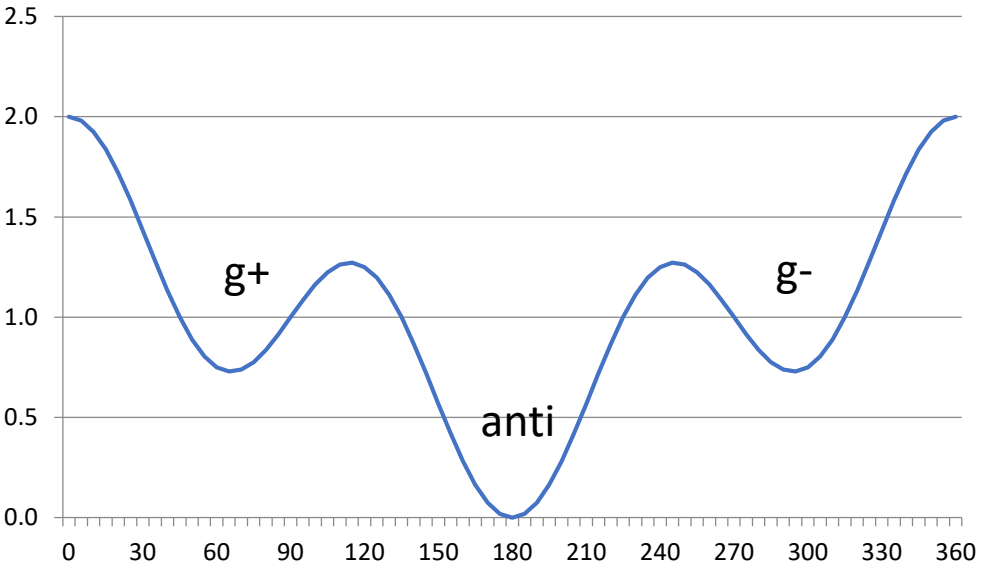
$$S = -R\sum_{i} p_i \ln(p_i)$$

where R is the gas constant and p_i is the probability of a residue being in rotamer i.

1. Pickett SD, Sternberg MJ. (1993). Empirical scale of side-chain conformational entropy in protein folding. *J Mol Biol* **231**(3):825-39.

Deriving Probabilites/Populations from Energies

But how do we derive the probabilities (or populations) that a particular state will be occupied? Boltzmann to the rescue!



$$\frac{N_i}{N} = p_i = \frac{e^{-E_i/(k_B T)}}{\sum_j e^{-E_j/(k_B T)}}$$

Boltzmann distribution

E_{g+} = 0.75 kcal/mol
E_{anti} = 0.00 kcal/mol
E_{g-} = 0.75 kcal/mol

Probabilites/Populations

For the three rotamers: E_{g+} = 0.75 kcal/mol, E_{anti} = 0.0 kcal/mol, E_{g-} = 0.75 kcal/mol

For rotamer 1 (E_{g+}): $e^{-E_i/(k_BT)} = e^{-0.75/0.59} = e^{-1.25} = 0.28$

For rotamer 3 (E_{g-}): $e^{-E_i/(k_BT)} = e^{-0.75/0.59} = e^{-1.25} = 0.28$

For rotamer 2 (E_{anti}): $e^{-E_i/(k_BT)} = e^{-0.0/0.59} = e^{0} = 1.00$

$$\frac{N_i}{N} = \frac{e^{-E_i/(k_BT)}}{\sum_j e^{-E_j/(k_BT)}}$$

And the sum: $\sum_j e^{-E_j/(k_BT)} = 0.28 + 1.0 + 0.28 = 1.56$

Now the populations (or probabilities, p_i) can be computed easily for each rotamer as:

$$\frac{N_{g+}}{N} = p_{g+} = \frac{N_{g-}}{N} = p_{g-} = \frac{0.28}{1.56} = 0.18$$

And p_{anti} = 0.64, can you derive this?

Entropies from Boltzmann Probabilites

$$S = -R\sum_i p_i \ln(p_i)$$

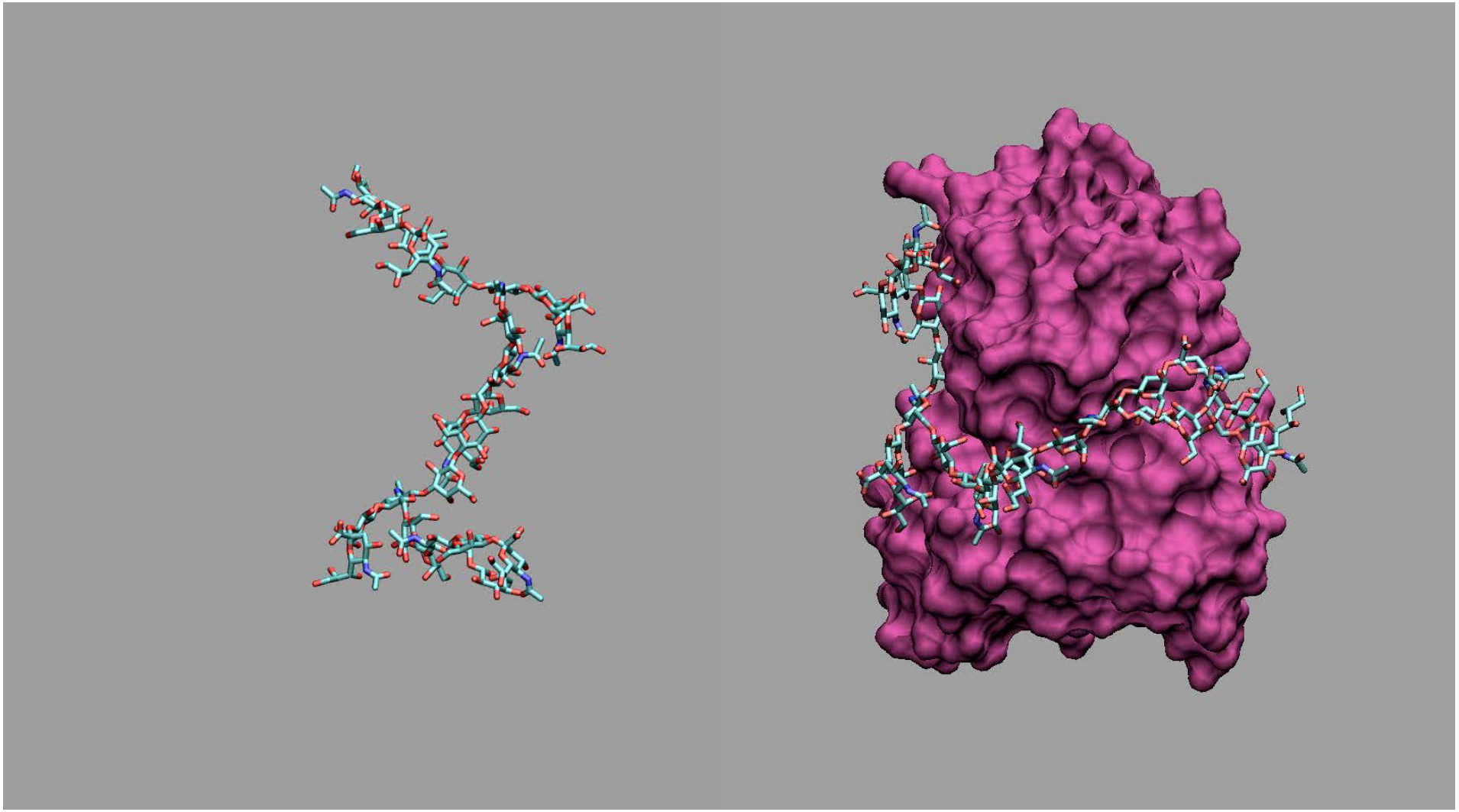
where R is the gas constant (0.001987 kcal/mol•K) and p_i is the probability of a residue being in rotamer *i*.

Rotamer	Relative Energy (kcal/mol)	Probability of being Populated	$p_i \ln(p_i)$	Entropy $-Rp_i \ln(p_i)$ kcal/mol/K	Entropic Energy Contribution at 300K
gauche+	0.75	0.18	-0.309	0.00061	0.18
gauche-	0.75	0.18	-0.309	0.00061	0.18
anti-	0.0	0.64	-0.286	0.00057	0.17
Total	----	1.00	-0.904	0.00179	0.54

Conclusion? A single rotatable bond has about 0.5 kcal/mol of entropic energy

Thus, if a single bond becomes rigid upon binding to a receptor, it will cost about 0.5 kcal/mol

MD Simulations of Free and Bound GBSIII CPS



Calculated Binding Energy for GBSIII CPS

Energy Component	Interaction Energies
$< \Delta E_{Electrostatic} >$	-167.5 ± 20
$< \Delta E_{VDW} >$	-126.9 ± 9
$< \Delta E_{MM} >$	--
$< \Delta G_{Solvation} >$	**211.9 ± 17**
$< \Delta G_{GB+MM} >$	-82.5 ± 11
$< -T\Delta S >$	77.6 ± 18
$< \Delta G_{Binding} >$	-4.9 ± 11

Enthalpy – entropy compensation due to ligand flexibility.

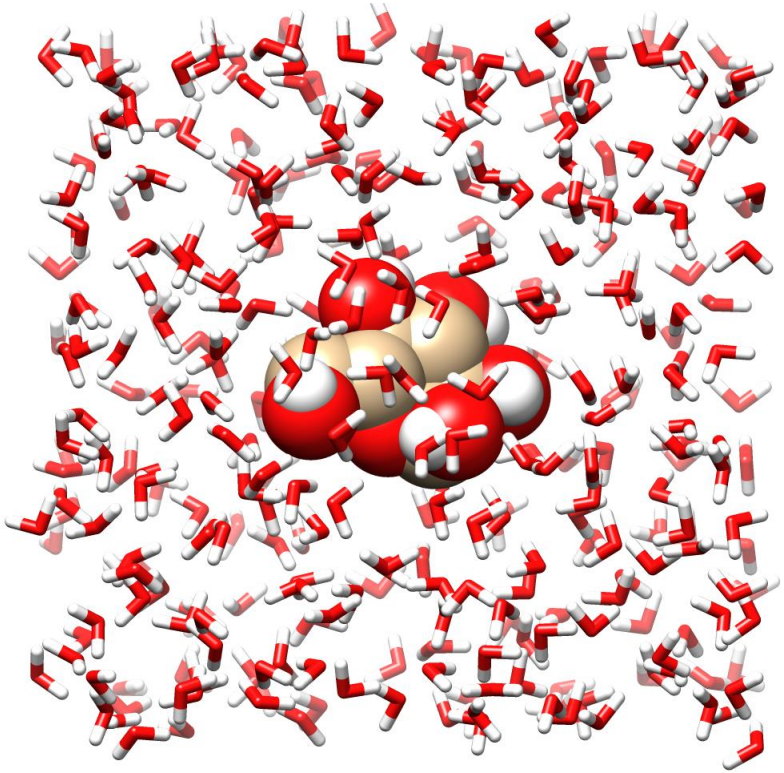
Kadirvelraj et al. (2006) *Proc. Natl. Acad. Sci. USA*

How can we calculate energies from an MD simulation?

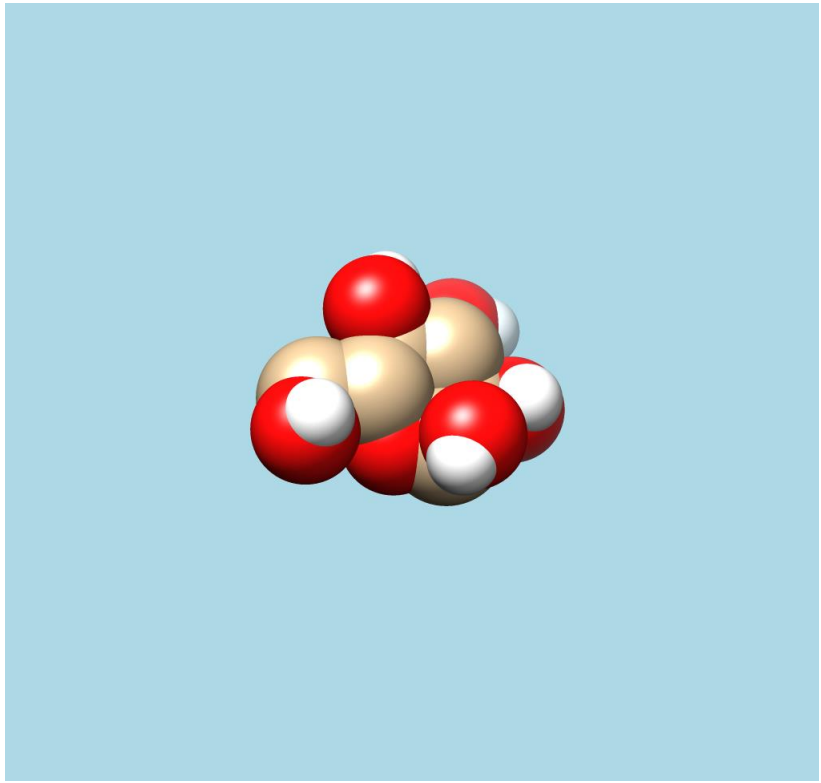
This is very easy in principle.

1) Collect multiple snapshots from the trajectory
2) Compute their energies using the force field
3) Average the energies. Voila!

Explicit *versus* implicit solvation

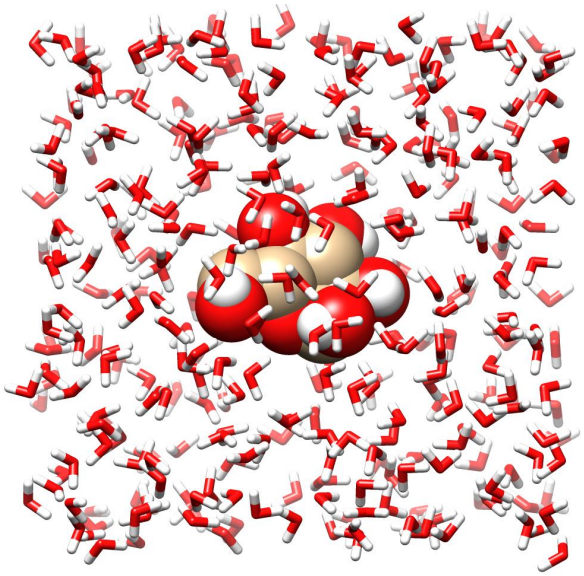


Every solvent (water) molecule is included explicitly, and all interactions between molecules are computed from the force field



Solvent is approximated as a uniform field around the solute (dielectric constant), and its influence on the solute is calculated from an implicit solvent model (generalized Born or Poisson-Boltzmann)

Explicit *versus* implicit solvation

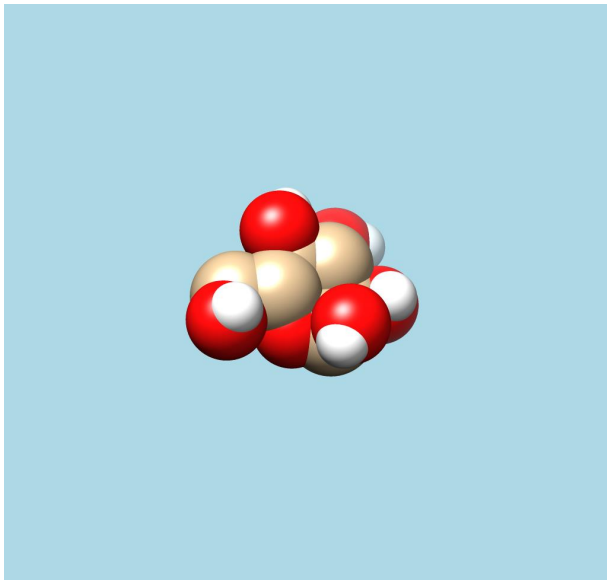


Pro:

- Good treatment of bound water
- Balanced treatment of direct polar interactions (hydrogen bonds)
- Better treatment of ions
- Better treatment of solute structure and dynamics

Con:

- Many water models to choose from
- Difficult to include when estimating solvent-solute interaction energies



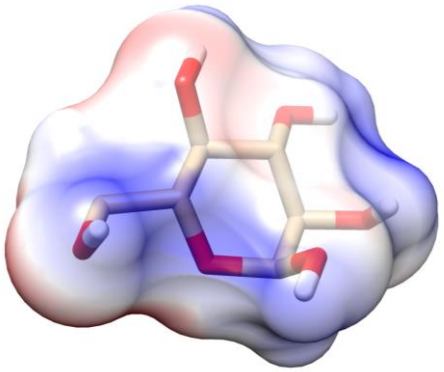
Pro:

- No need to know the individual water positions
- Useful in estimating solvent-solute interaction energies

Con:

- Unable to model “bound” waters
- Inaccurate estimates of the strength of direct polar interactions (hydrogen bonds)
- Poor treatment of ions
- Many implicit models to choose from

Implicit Solvation: Cavity Formation and van der Waals Energy



$$\Delta G_{Solvation} = \boxed{\Delta G_{Cavity}} + \boxed{\Delta G_{vdW}} + \Delta G_{Ele}$$

The van der Waals interaction with water is approximately proportional to *the surface area of the molecule in contact with the water*.

Similarly, the energy to create a cavity in the water depends on the size of the molecule, and can be approximated by surface area of the molecule in contact with the water.

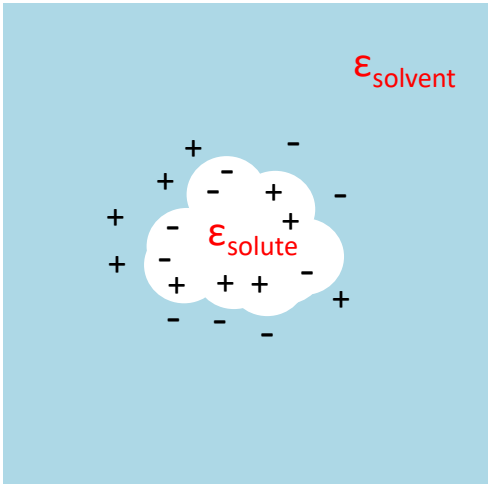
Both the cavity formation energy and van der Waals interaction energy employ similar approximations:

$$\Delta G_{Cavity} = S_{tension} \sum_{i=1}^{Natoms} SASA_i \qquad \Delta G_{vdW} = \sum_{i=1}^{Natoms} \sigma_i SASA_i$$

Where $SASA_i$ is the solvent accessible surface area of atom *i*, $S_{tension}$ is the surface tension of the solvent, and σ_i is an empirical *solvation parameter* for atom *i*. The $SASA_i$ and σ_i values depend on the atom type (N-amide, N-amine, C-sp^3, C-sp^2, C-carbonyl, O-alcohol, O-ether, O-carbonyl, etc).

Implicit solvation: electrostatic energy

$$\Delta G_{Solvation} = \Delta G_{Cavity} + \Delta G_{vdW} + \Delta G_{Ele}$$



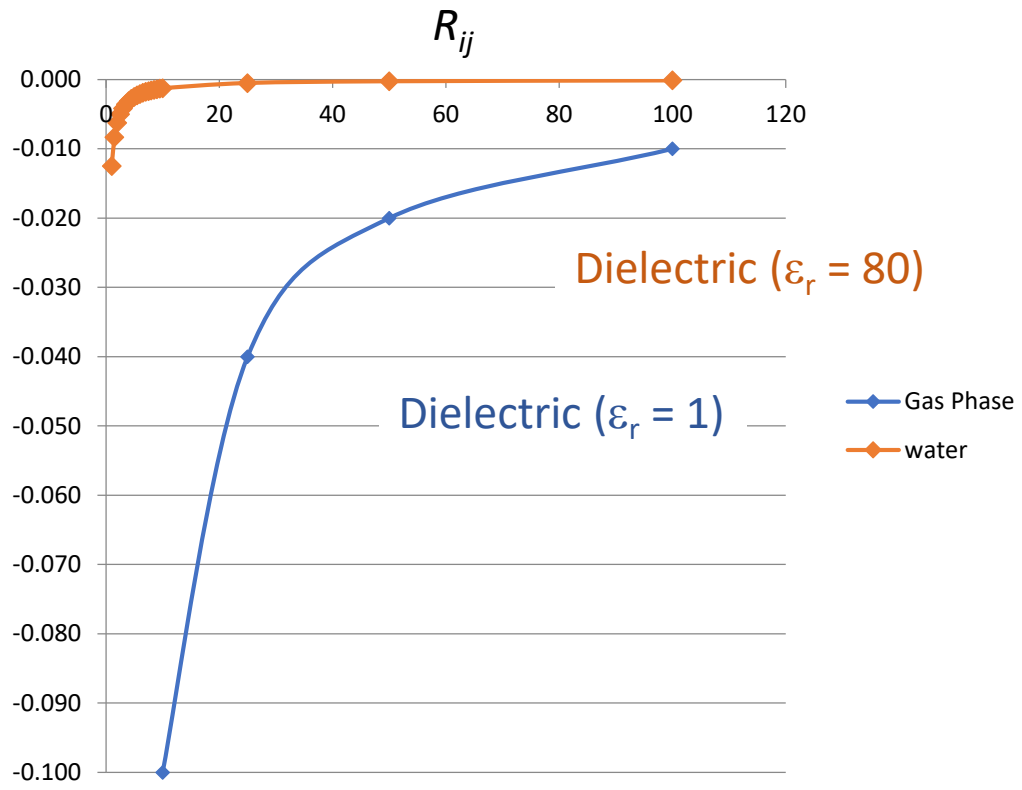
The electrostatic interaction with solvent is related to the polarity of the atoms and the dielectric constants of the solvent and the solute.

We need to know:

1) The partial charges on the atoms (placed at the nuclei).
2) The distance of each charge from the solvent.
3) The dielectric constants (ε) for the solute and solvent.

Effect of dielectric constant

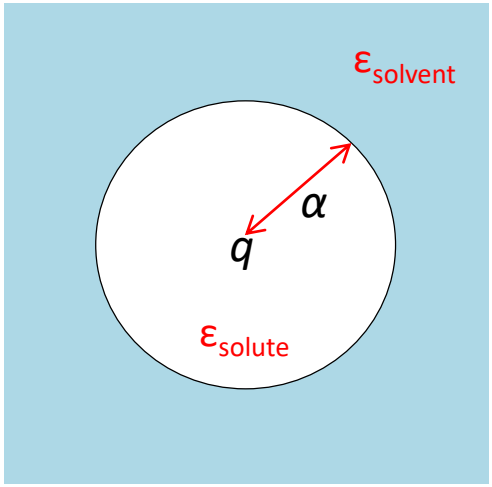
$$V_{Electrostatic} = \frac{q_i q_j}{\varepsilon_r R_{ij}}$$



Employing a large dielectric constant effectively eliminates electrostatic interactions.

When is this reasonable? Only when the charges are far apart

Implicit Solvation: Electrostatic Energy (Born approximation)



$$\Delta G_{Solvation} = \Delta G_{Cavity} + \Delta G_{vdW} + \Delta G_{Ele}$$

For a single atom in implicit solvent, the Born approximation says:

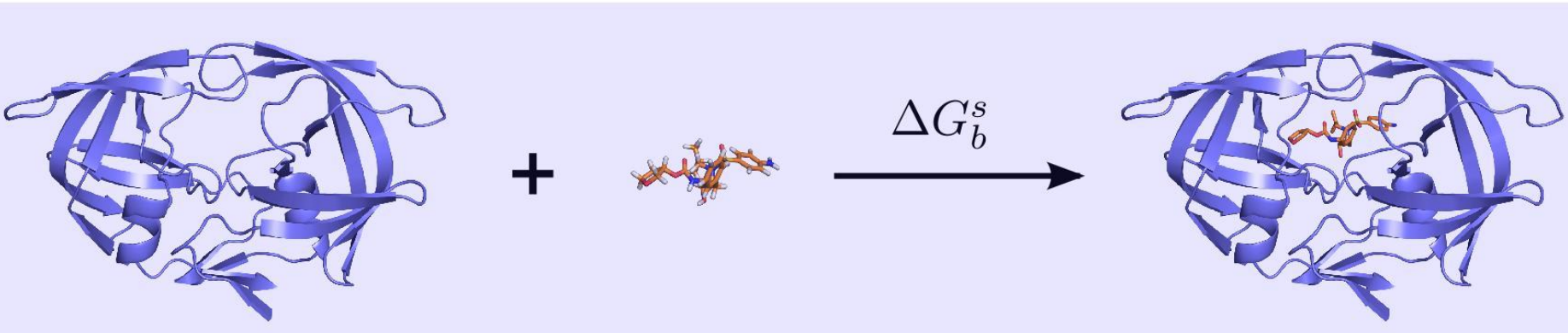
$$\Delta G_{Ele} = -\frac{1}{2}\left(\frac{1}{\varepsilon_{solute}} - \frac{1}{\varepsilon_{solvent}}\right)\frac{q^2}{\alpha}$$

where q is the partial charge on the atom (placed at the nucleus), α is the distance of the charge from the solvent, ε is the dielectric constant (for the solute and solvent).

The Born equation can be generalized (*generalized Born equation*) for a molecule of any shape. The distance (α) is known as the ***effective Born radius***. If the atom is solvent exposed, the Born radius will be close to the van der Waals radius. However, for a buried atom, the Born radius can become quite large. In any case the Born radius is not the same as the van der Waals radius.

Different parameterizations of the Born method vary principally in how they determine the effective Born radii.

ΔG calculations from MD simulations



$$\Delta G_{Binding} = \Delta G_{Complex} - \Delta G_{Ligand} - \Delta G_{Receptor}$$

From MD – use average energies for each component, indicated by angled brackets:

$$\langle \Delta G_{Binding} \rangle = \langle \Delta G_{Complex} \rangle - \langle \Delta G_{Ligand} \rangle - \langle \Delta G_{Receptor} \rangle$$

MM/GBSA Calculations

Molecular Mechanics/Generalized Born Surface Area (MM/GBSA)

$$\left\langle \Delta G_{Binding} \right\rangle = \left\langle \Delta G_{Complex} \right\rangle - \left\langle \Delta G_{Ligand} \right\rangle - \left\langle \Delta G_{\mathrm{Re}ceptor} \right\rangle \quad (1)$$

$$\Delta G = \Delta H - T\Delta S \quad (2)$$

$$\left\langle \Delta G \right\rangle \approx \left\langle \Delta V_{ForceField} \right\rangle + \left\langle \Delta G_{Desolvation} \right\rangle - \left\langle T\Delta S \right\rangle \quad (3)$$

So for each component (Complex, Receptor and Ligand)

1) compute the average energy from the force field
2) estimate the desolvation free energy (generalized Born)
3) estimate the entropy.

Determine the binding energy using equation 1.