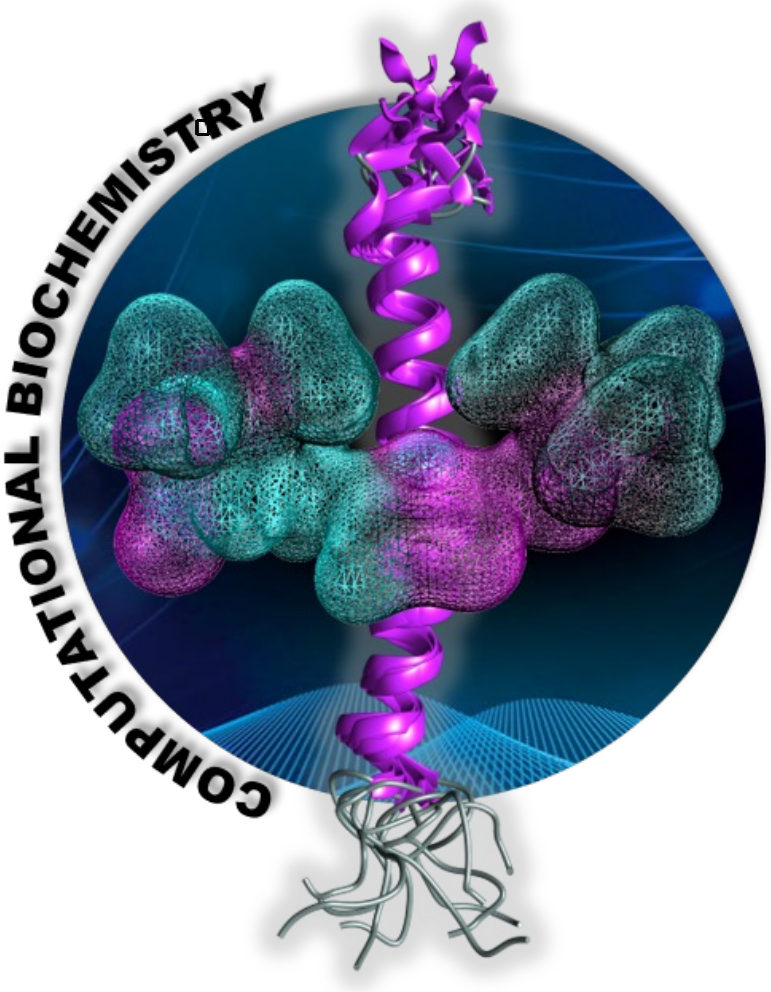


Lesson 8:

Binding Free Energies

Iñaki Tuñón & Vicent Moliner

VNIVERSITAT DE VALÈNCIA

UNIVERSITAT JAUME·I

Outline

Theory (Session 8)

1. Binding Process & Free Energies
2. Calculating Binding Free Energies
3. Alchemical Transformations
    - 3.1. Free Energy Perturbation
    - 3.2. Thermodynamic Integration
    - 3.3. Practical Considerations
4. Approximate Methods: MMPBS & MMGBS

1. Binding Process & Free Energies

The first step in the Michaelis-Menten kinetic scheme is the formation of the enzyme-substrate complex :

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} MC \xrightarrow{k_2} E + P$$

The first step can ve considered as being in equilibrium if $k_2 << k_1, k_{-1}$. In that case the Michaelis constant is the inverse of the equilibrium constant of the binding process:

$$K_M = \frac{1}{K_B} = \frac{k_{-1}}{k_1} = \frac{[E]\cdot[S]}{C^0\cdot[MC]}$$

Where C^0 is the standard concentration, typically chosen to be 1 M. This equilibrium constant can be related to the standard free energy change:

$$\Delta G_B^0 = -RT \cdot lnK_B = RT \cdot lnK_M$$

$$\Delta G_B^0 = \Delta G_{MC}^0 - \Delta G_E^0 - \Delta G_S^0$$

1. Binding Process & Free Energies

One way to block the activity of an enzyme is to add a compound that presents a high affinity for the active site, with a large binding constant :

$$E + I \underset{k_{-1,I}}{\overset{k_{1,I}}{\rightleftharpoons}} EI \qquad K_I = \frac{[E]\cdot[I]}{C^0\cdot[EI]}$$

When designing an enzyme inhibitor one has to look for a large negative value of the binding free energy to maximize the binding constant:

$$\Delta G^0_{B,I} = -RT \cdot \ln(1/K_I) = RT \cdot lnK_I$$

Predicting binding affinities and thus the calculation of binding free energies is one of the holy grails in Drug Design

1. Binding Process & Free Energies

Inhibiting the main protease of SARS-CoV-2

Covalent Inhibitors

$$E + I \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} EI \xrightarrow{k_{cov}} E\text{-}I$$

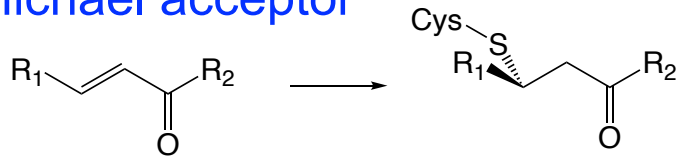
$$\Delta G_B = \Delta G_B(E-I) = \Delta G_B(EI) + \Delta G_{reac}$$

Noncovalent inhibitors

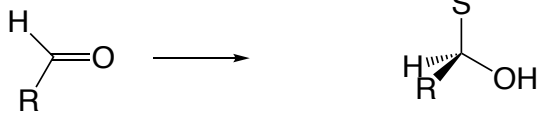
$$E + I \underset{k_{-1,I}}{\overset{k_{1,I}}{\rightleftharpoons}} EI$$

$$\Delta G_B = \Delta G_B(EI)$$

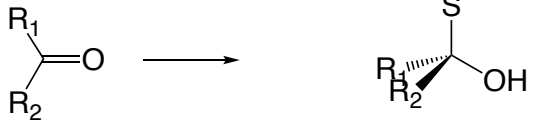
Michael acceptor

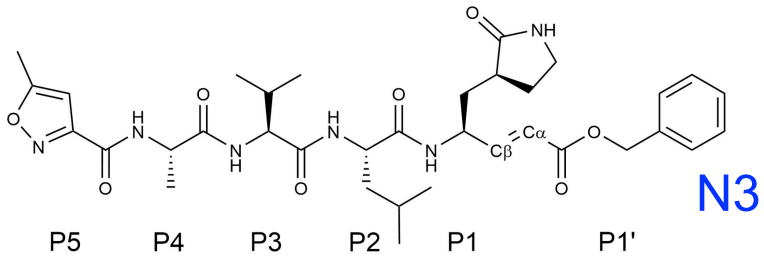


Aldehyde

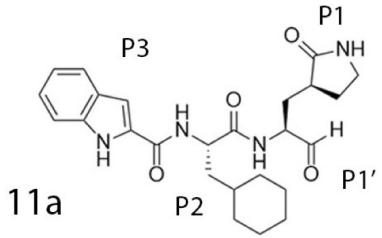


Ketone

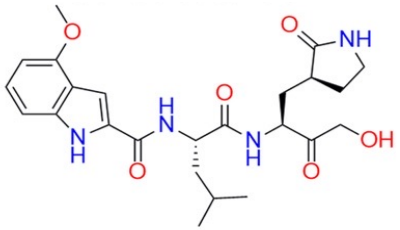




N3



11a



PF-00835231

1. Binding Process & Free Energies

Notes of caution

i) Note that experimental Michaelis and inhibition constants are usually reported with units:

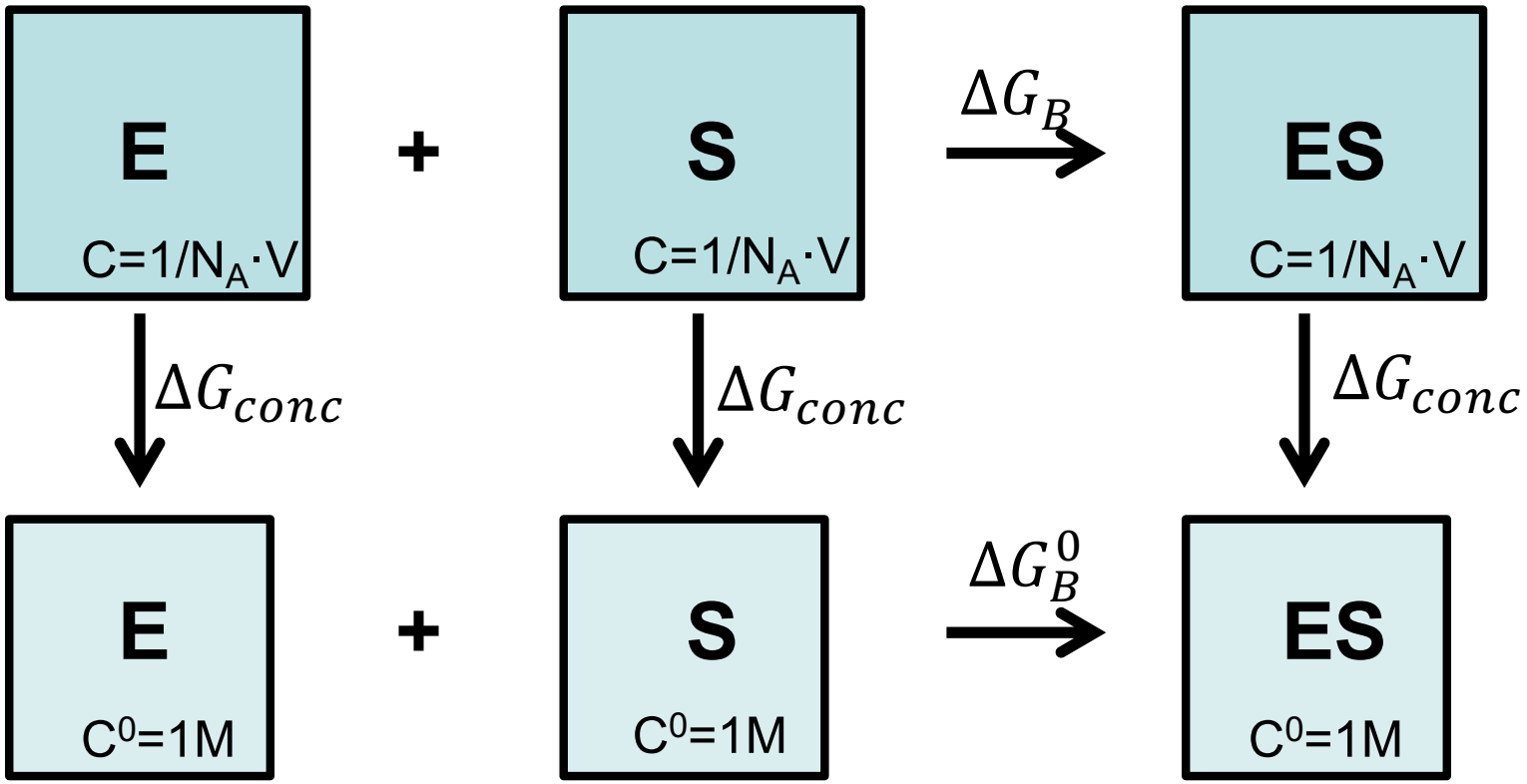
$$K_I = \frac{[E]\cdot[I]}{[EI]} \qquad K_M = \frac{[E]\cdot[S]}{[MC]}$$

That are numerically identical to the thermodynamic constant (dimensionless) if given in M units (be careful, sometimes other concentration, units such as nM or μM, are used instead)

1. Binding Process & Free Energies

Notes of caution

ii) Tipically, MD simulations of binding processes are not performed at 1M concentration and then comparison with experimental data (K_B) must be carried out with care:



$$\Delta G_B^0 = -RT \cdot lnK_B$$

$$\Delta G_B^0 = \Delta G_B - \Delta G_{conc}$$

1. Binding Process & Free Energies

The free energy change associated to a change in the concentration from C to C^0 is:

$$\Delta G_{conc} = -RT \cdot ln\frac{V^0}{V} = -RT \cdot ln\frac{C}{C^0} = -RT \cdot ln\frac{1}{N_A \cdot C^0 \cdot V} = RT \cdot \ln(N_A \cdot C^0 \cdot V)$$

So, the relationship between the free energy change calculated using simulation boxes of volumen V and the standard free energy change is:

$$\Delta G_B^0 = \Delta G_B - RT \cdot \ln(N_A \cdot C^0 \cdot V)$$

So, the relationship with the experimentaly determined binding equilibrium constant is:

$$-RT \cdot lnK_B = \Delta G_B - RT \cdot \ln(N_A \cdot C^0 \cdot V)$$

A typical value, at 298 K, for a simulation box of 50 x 50 x 50 $Å^3$ = $1{,}25 \cdot 10^{-22}$ L

$$-RT \cdot \ln(N_A \cdot C^0 \cdot V) = 2.56\ kcal/mol$$

1. Binding Process & Free Energies

Notes of caution

iii) Biological process are tipically carried out under constant temperature and pressure. Thus the change in the Gibbs free energy of the system is the correct predictor for the spontaneity of the process

$$\Delta G_{sys} = \Delta U_{sys} + P\Delta V_{sys} - T\Delta S_{sys} =$$
$$= W_{sys} + Q_{sys} + P\Delta V_{sys} - T\Delta S_{sys} = -P\Delta V_{sys} - Q_{surr} +$$
$$+ P\Delta V_{sys} - T\Delta S_{sys} = -T\Delta S_{surr} - T\Delta S_{sys} = -T\Delta S_{univ}$$

However, in condensed phases the change in the volumen of the system males a negliglible contribution. For example in a Diels-Alder reaction in solution:

$$P \cdot \Delta V \sim 10^{-3} kcal/mol$$

While the reaction free energy is, $\Delta G_R^0 \sim 10\ kcal/mol$

1. Binding Process & Free Energies

Ignoring the PV term means that the Gibbs (G) and Helmholtz (A) free energies are numerically identical:

$$A = U - TS$$
$$G = H - TS = U + PV - TS$$

$$G = A + PV \xrightarrow{\text{condensed phase}} G \approx A$$

This means that we can predict the spontaneity of a process and calculate its equilibrium constant from (N,V,T) simulations that provide the change in the Helmholtz free energy or from (N,P,T) simulations that give us the Gibbs free energy change.

In Statistical Themodynamics each Free energy is related to its corresponding partition function:

$$G = -kT \cdot lnQ(N,P,T) \qquad \mathrm{Q}(N,P,T) = \int \dots \int exp\left(-\frac{H(\boldsymbol{r}^N,\boldsymbol{p}^N) + PV}{kT}\right) d\boldsymbol{r}^N d\boldsymbol{p}^N$$

$$A = -kT \cdot lnQ(N,V,T) \qquad \mathrm{Q}(N,V,T) = \int \dots \int exp\left(-\frac{H(\boldsymbol{r}^N,\boldsymbol{p}^N)}{kT}\right) d\boldsymbol{r}^N d\boldsymbol{p}^N$$

where **r** and **p** refer to position and momenta of the N particles and H is the hamiltonian that defines a given state

2. Calculating Binding Free Energies

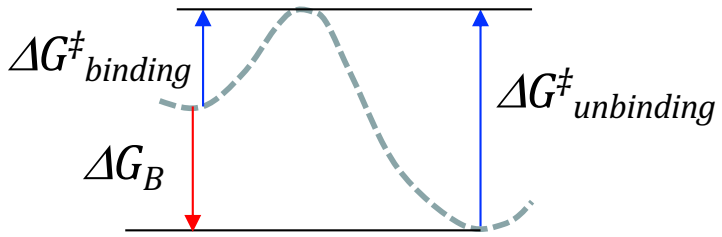
Free energies can be obtained from simulations because they are related to the probability of finding the system in a given state (i):

$$p(i) \sim exp\left(-\frac{G_i}{RT}\right)$$

For example the binding free energy is related to the probability of finding the system in the bound and unbound states:

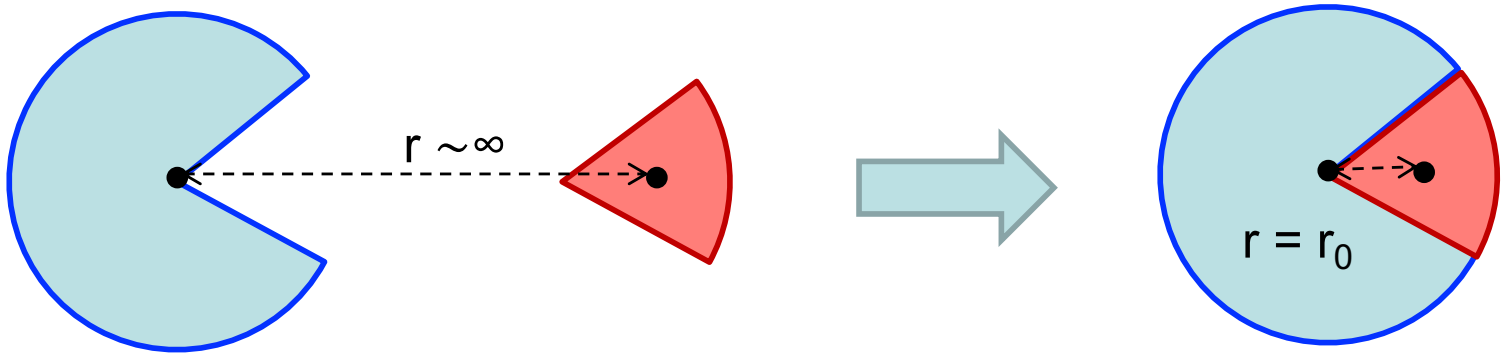
$$\Delta G_B = -RT \cdot \ln\left(\frac{p_{bound}}{p_{unbound}}\right)$$

Direct evaluation of the binding free energy would require to run simulations with converged values of the probabilities, this is where binding/unbinding process take place several times. In most cases, the time required for the binding process and in particular for unbinding clearly prevent the use of free simulations to simple the probabilities of these states.



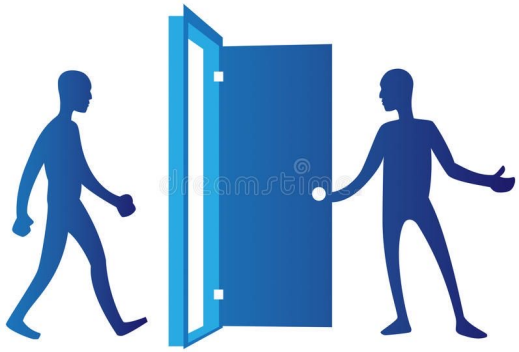
2. Calculating Binding Free Energies

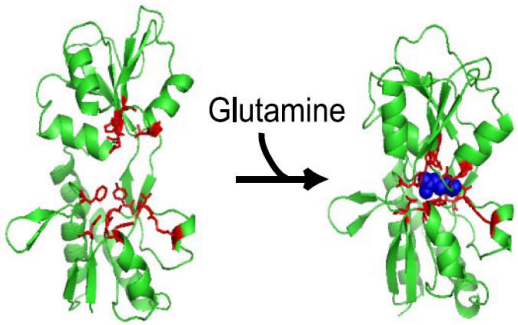
One possibility to obtain binding free energies would be to define a 'reaction coordinate' to drive the system from unbounded to bounded conformations



Receptor (R) Ligand (L) Receptor-Ligand Complex (RL)

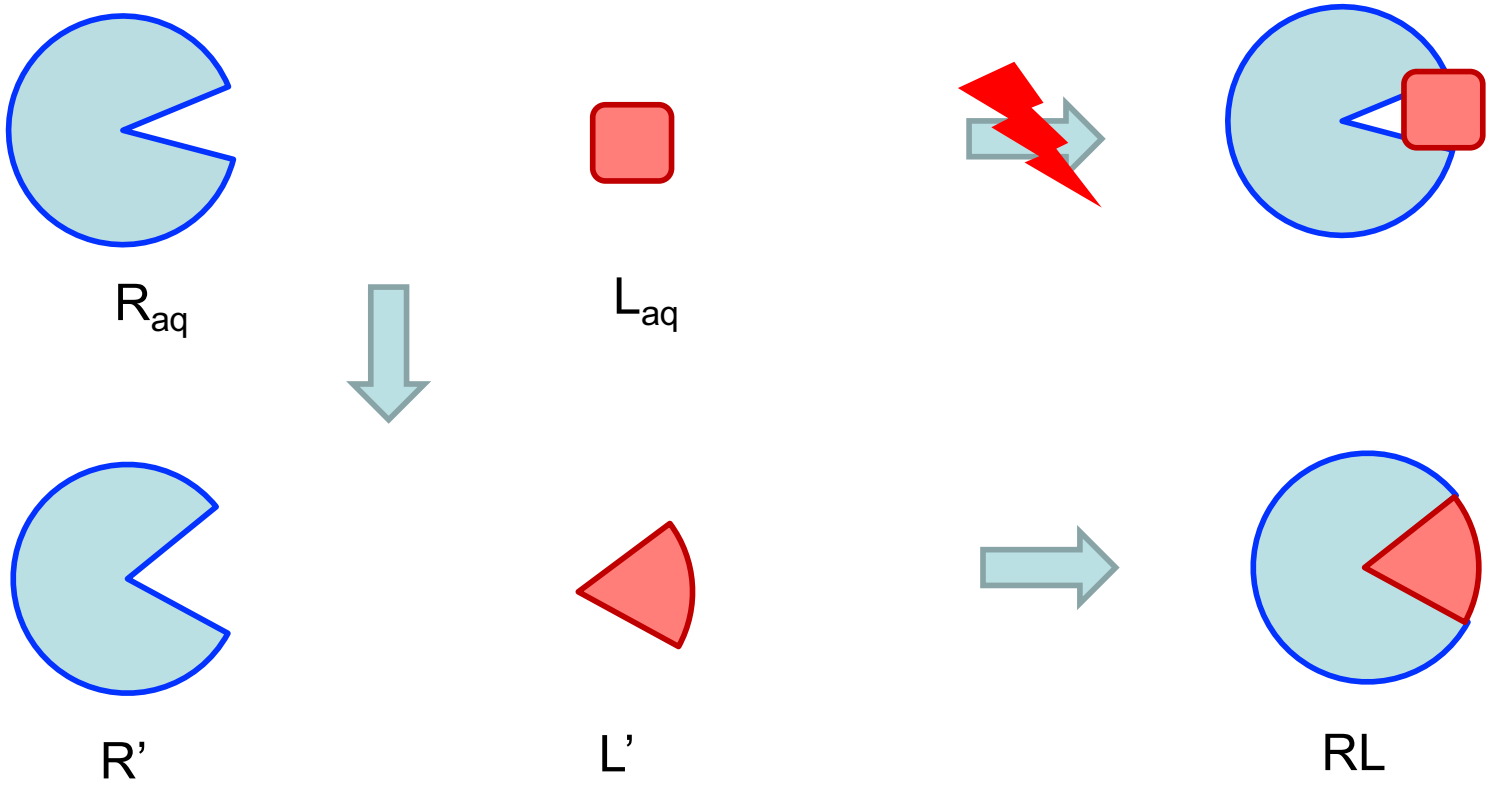
However, we can use the analogy of the room and the door: the distance to a room is a good coordinate to enter that room….provided that the door is opened. Otherwise the relevant coordinate would be that defining the rotational state of the door. In general many changes can happen during receptor binding and thus its can be difficult to define adequate coordinates for the whole process





2. Calculating Binding Free Energies

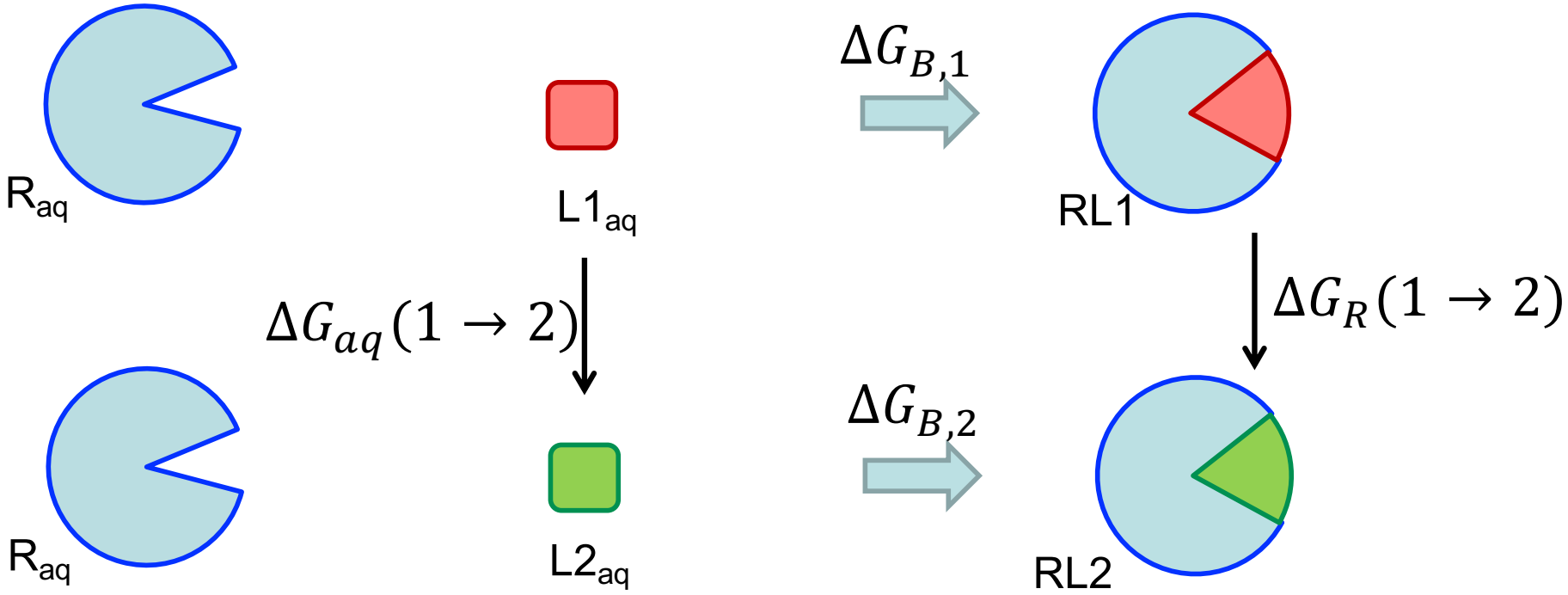
Said in other words, conformational changes of the receptor and/or the ligand can be important during the binding process and if we do not sample them correctly the calculated binding free energies can be completely wrong.



So, we need to define correct thermodynamic cycles and adequate free energy changes to calculate the absolute binding free energy of a ligand

2. Calculating Binding Free Energies

It is easier to calculate differences in binding free energies between similar ligands

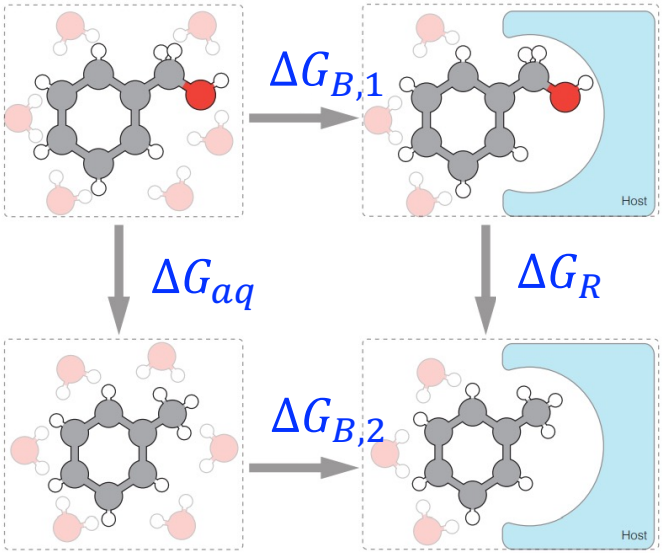


$$\Delta\Delta G_B = \Delta G_{B,2} - \Delta G_{B,1} = \Delta G_R(1 \rightarrow 2) - \Delta G_{aq}(1 \rightarrow 2)$$

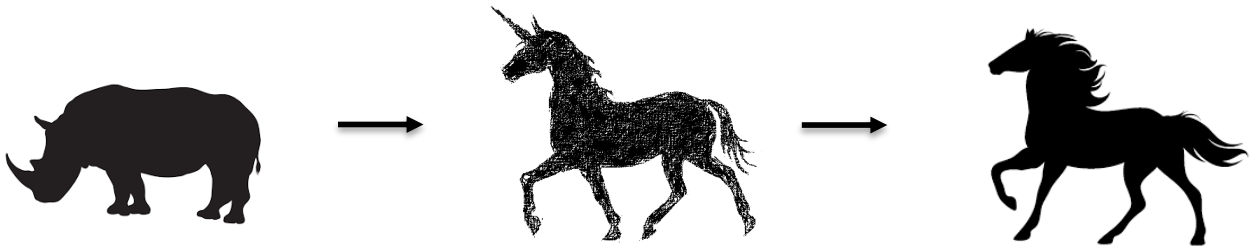
In this case we don't need to evaluate binding or unbinding processes, just the transformation of one ligand into another, both in solution and in the protein environments. This is known as an ***alchemical transformation***

3. Alchemical Transformation

The binding free energy difference between benzyl alcohol and toluene in the same receptor can be computed by calculating free energy cost of mutating benzyl alcohol into toluene within the binding pocket (ΔG_R) and in the solvent (ΔG_{aq}) , saving the need to simulate binding/unbinding processes.

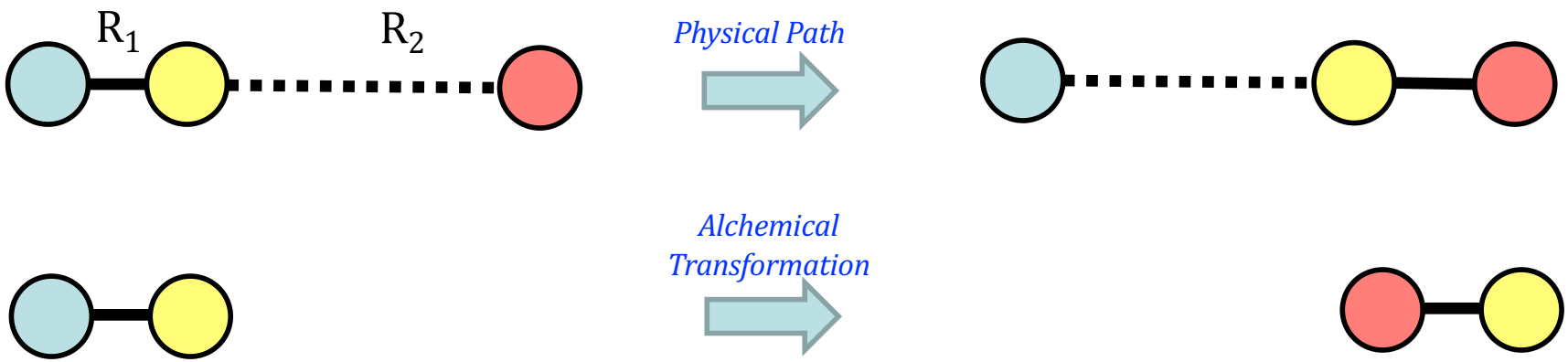


This mutations can be done transforming alchemically one molecule or atom into another, defining ‘hybrid’ atoms or molecules.



3. Alchemical Transformation

In general, in the context of free energy calculations, ***alchemical transformations*** is a term to refer to unphysical paths to go from one initial state to a final state. This term is used to distinguish these methods from those that are carried out along physical transformations, as it is done for example when using a reaction coordinate to drive a system using umbrella sampling or other sampling methods.



Two of the most popular methods to obtain free energy differences using alchemical transformations are:

- Free Energy Perturbations
- Thermodynamic Integration

3.1. Free Energy Perturbation

For convenience we will work with Helmolthz free energies and the NVT ensemble, although the expressions are equally valid if we substitute:

- $\Delta A \rightarrow \Delta G$ and
- $H(\mathbf{r}^N,\mathbf{p}^N) \rightarrow H(\mathbf{r}^N,\mathbf{p}^N) + PV$

Let's express the free energy difference between two states (I and II) from their respective partition functions:

$$\Delta A = A(II) - A(I) = -kT \ln \frac{Q_{II}}{Q_I} =$$

$$= -kT \ln \frac{\int \cdots \int exp\left(-\frac{H_{II}(\mathbf{r}^N, \mathbf{p}^N)}{kT}\right) d\mathbf{r}^N d\mathbf{p}^N}{\int \cdots \int exp\left(-\frac{H_I(\mathbf{r}^N, \mathbf{p}^N)}{kT}\right) d\mathbf{r}^N d\mathbf{p}^N}$$

where H_I and H_{II} are the hamiltonians that describe each of the states (for example benzyl alcohol (I) and toluene (II) in the receptor)

3.1. Free Energy Perturbation

We can introduce *[exp(H_I/kT)·exp(-H_I/kT)]* in the numerator (- it's just unity!)

$$\Delta A = -kT \ln \frac{\int \cdots \int \exp\left(-\frac{H_{II}}{kT}\right) \exp\left(-\frac{H_I}{kT}\right) \exp\left(\frac{H_I}{kT}\right) d\mathbf{r}^N d\mathbf{p}^N}{\int \cdots \int \exp\left(-\frac{H_I}{kT}\right) d\mathbf{r}^N d\mathbf{p}^N} =$$

$$= -kT \ln \int \cdots \int \exp\left(-\frac{H_{II}}{kT}\right) \exp\left(\frac{H_I}{kT}\right) \frac{\exp\left(-\frac{H_I}{kT}\right)}{\int \cdots \int \exp\left(-\frac{H_I}{kT}\right) d\mathbf{r}^N d\mathbf{p}^N} d\mathbf{r}^N d\mathbf{p}^N =$$

$$= -kT \ln \int \cdots \int \exp\left(-\frac{(H_{II} - H_I)}{kT}\right) \rho_I d\mathbf{r}^N d\mathbf{p}^N =$$

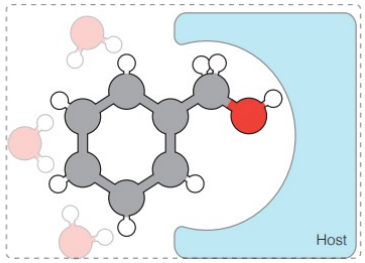
$$= -kT \ln \langle \exp[-(H_{II} - H_I)/kT] \rangle_I$$

This is the probability density of state I

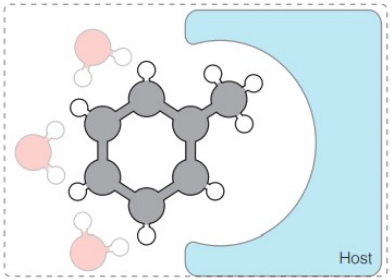
Average of the energy difference exponential evaluated at state I

This expression suggests that the free energy difference between two states can be obtained running a simulation of state I and evaluating the energy difference between states I and II

3.1. Free Energy Perturbation



state I



state II

- How to do it:

$I \rightarrow II$

$$\Delta A(II - I) = -kT \ln\langle \exp[-(H_{II} - H_I)/kT] \rangle_I$$

$II \rightarrow I$

$$\Delta A(I - II) = -kT \ln\langle \exp[-(H_I - H_{II})/kT] \rangle_{II}$$

i) Simulate state I (or state II)
ii) Select a number of configurations of state I (or state II)
iii) Evaluate the energy of state II (state I) in these configurations
iv) Average the exponential of the energy difference / kT
v) Obtain Helmholtz free energy difference as the logarithm multiplied by kT

- Problems: *If states I and II are too different (the phase spaces of both do not overlap) then the energy difference will be too large and the configurations will not significantly contribute to the average*

⇒ However, we can imagine intermediate states between *I* & *II*

3.1. Free Energy Perturbation

- How to do it:

$$\Delta A = A_{II} - A_I = (A_{II} - A_{N-1}) + (A_{N-1} - A_{N-1}) + \ldots + (A_2 - A_I) =$$

$$= -kT\sum_{i=1}^{N-1} ln\langle exp[-(H_{i+1} - H_i)/kT]\rangle_i = kT\sum_{i=1}^{N-1} ln\langle exp[-(H_i - H_{i+1})/kT]\rangle_{i+1}$$

We can create as many intermediate states as needed using a coupling parameter between states I and II

$$H(\lambda) = (1 - \lambda)H_I + \lambda H_{II}$$ ⇨ *Linear coupling (the most common!)*

when λ=0 we are in the state I, when λ=1 we are in state II
λ can take any value between 0 and 1

3.1. Free Energy Perturbation

- A toy example:

We want to evaluate the free energy difference associated to the process of charging an atom in aqueous solution

$$B^{0}_{(aq)} \Longleftrightarrow B^{-1}_{(aq)}$$

state I *state II*

They do not directly overlap because charging requires substantial reordering of water molecules

$$H_I = \frac{p^2}{2m} + \sum_S \left[\frac{A_S}{r_S^{12}} - \frac{B_S}{r_S^{6}} \right] + H_{sol}$$

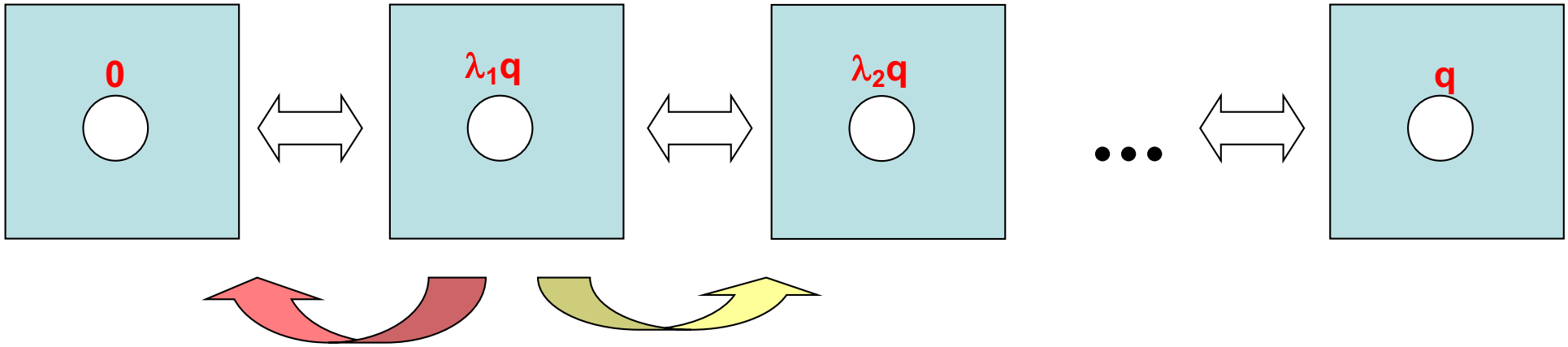
$$H_{II} = \frac{p^2}{2m} + \sum_S \left[\frac{A_S}{r_S^{12}} - \frac{B_S}{r_S^{6}} \right] + \sum_S \frac{qq_s}{r_S} + H_{sol}$$

$$H(\lambda) = (1-\lambda)H_I + \lambda H_{II} = H_I + \sum_S \frac{\lambda q q_s}{r_S}$$

Note that this is an unphysical transformations. We can not charge the atom with a fractional charge!

3.1. Free Energy Perturbation

$$H(\lambda) = \frac{p^2}{2m} + \sum_s \left[\frac{A_s}{r_s^{12}} - \frac{B_s}{r_s^6} \right] + H_{solvent} + \sum_s \frac{\lambda q q_s}{r_s}$$



$$\Delta A = -kT \sum_{\lambda=0}^{\lambda=1} ln\left\langle exp\left[-\left(H(\lambda_{i+1}) - H(\lambda_i)\right)/kT\right]\right\rangle_{\lambda_i} = kT \sum_{\lambda=0}^{\lambda=1} ln\left\langle exp\left[-\left(H(\lambda_{i-1}) - H(\lambda_i)\right)/kT\right]\right\rangle_{\lambda}$$

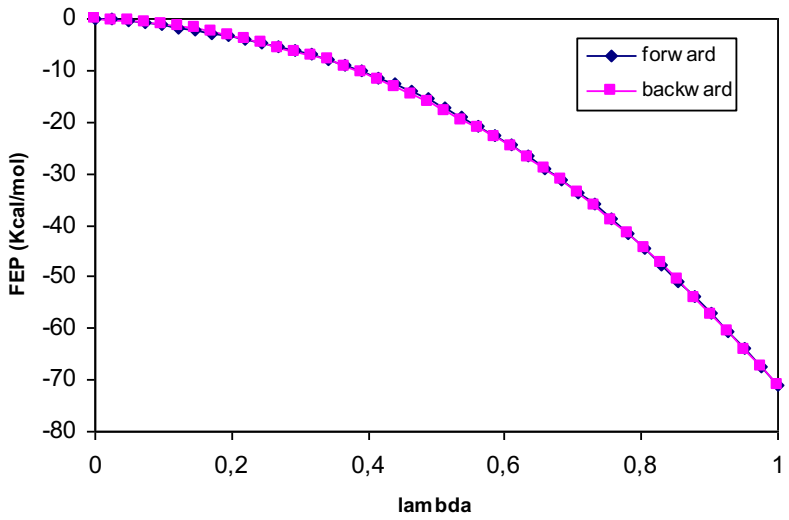
i) Run simulation with $H(\lambda_i)$ hamiltonian
ii) Select configurations and try solute with $H(\lambda_{i+1})$ (and λ_{i-1})
iii) Evaluate averages
iv) Run simulation with next λ

3.1. Free Energy Perturbation

$B^{0}_{(aq)} \Longleftrightarrow B^{-1}_{(aq)}$

state λ=0 *state λ=1*

- One solute with TIP3P water molecules
- Simulation Box ~18 Å with PBC and cutoff of 8 Å
- 40 simulation *windows* employed (λ values).
- 5ps +10 ps of NVT MD simulation at each window (time step 1 fs)



Forward = -71.0 kcal/mol

Backward = -71.2 kcal/mol

Average = -71.1±0.1 kcal/mol

3.2. Thermodynamic Integration

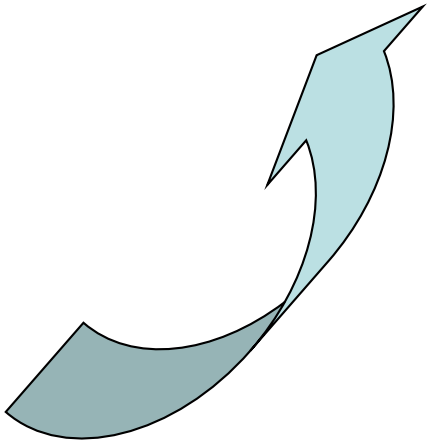
If Free Energy is a continuous function of the coupling parameter λ

$$\Delta A = \int \frac{\partial \Delta A(\lambda)}{\partial \lambda} d\lambda = \int \frac{\partial \left[-\, kT\, ln\, Q(\lambda) \right]}{\partial \lambda} d\lambda = \int - \frac{kT}{Q(\lambda)} \frac{\partial Q(\lambda)}{\partial \lambda} d\lambda$$

Using the chain rule we can get the derivative of Q(λ)

$$\frac{\partial Q(\lambda)}{\partial \lambda} = \frac{1}{N!\, h^{3N}} \int \cdots \int \frac{\partial}{\partial \lambda} exp\left(- \frac{H}{kT} \right) d\boldsymbol{r}^{N} d\boldsymbol{p}^{N} =$$

$$= \frac{1}{N!\, h^{3N}} \int \cdots \int \left(- \frac{1}{kT} \right) \frac{\partial H}{\partial \lambda} exp\left(- \frac{H}{kT} \right) d\boldsymbol{r}^{N} d\boldsymbol{p}^{N}$$



3.2. Thermodynamic Integration

$$\Delta A = \frac{1}{N!\,h^{3N}} \int \left[\frac{1}{Q(\lambda)} \int \cdots \int \frac{\partial H}{\partial \lambda} \exp\left(-\frac{H}{kT}\right) d\mathbf{r}^N d\mathbf{p}^N \right] d\lambda =$$

$$= \int \left[\int \cdots \int \frac{\partial H}{\partial \lambda} \frac{\exp\left(-\frac{H}{kT}\right)}{\int \cdots \int \exp\left(-\frac{H}{kT}\right) d\mathbf{r}^N d\mathbf{p}^N} d\mathbf{r}^N d\mathbf{p}^N \right] d\lambda$$

$$= \int \left[\int \cdots \int \frac{\partial H}{\partial \lambda} \rho_{NVT} d\mathbf{r}^N d\mathbf{p}^N \right] d\lambda = \int \left\langle \frac{\partial H}{\partial \lambda} \right\rangle_\lambda d\lambda$$

Probability density obtained with H(λ)

Ensemble average

- How to do it:

$$\Delta A = \int_0^1 \left\langle \frac{\partial H(\lambda)}{\partial \lambda} \right\rangle_\lambda d\lambda$$

If using a linear coupling $H(\lambda) = (1 - \lambda)H_I + \lambda H_{II}$

The derivative is just $\frac{\partial H(\lambda)}{\partial \lambda} = H_{II} - H_I$

i) Run simulations for different λ values (0→1)

ii) For each λ value evaluate the average of the H(λ) derivative

iii) Evaluate the integral (with numerical methods such as Simpson's rule or gaussian cuadrature)

3.2. Thermodynamic Integration

- A toy model

$$B^{0}_{(aq)} \Longleftrightarrow B^{-1}_{(aq)}$$

state λ=0 *state λ=1*

$$H_I = \frac{p^2}{2m} + \sum_s \left[\frac{A_s}{r_s^{12}} - \frac{B_s}{r_s^{6}} \right] + H_{sol}$$

$$H_{II} = \frac{p^2}{2m} + \sum_s \left[\frac{A_s}{r_s^{12}} - \frac{B_s}{r_s^{6}} \right] + \sum_s \frac{qq_s}{r_s} + H_{sol}$$

$$H(\lambda) = (1-\lambda)H_I + \lambda H_{II} = H_I + \sum_s \frac{\lambda q q_s}{r_s}$$

$$\frac{\partial H(\lambda)}{\partial \lambda} = \sum_s \frac{qq_s}{r_s}$$

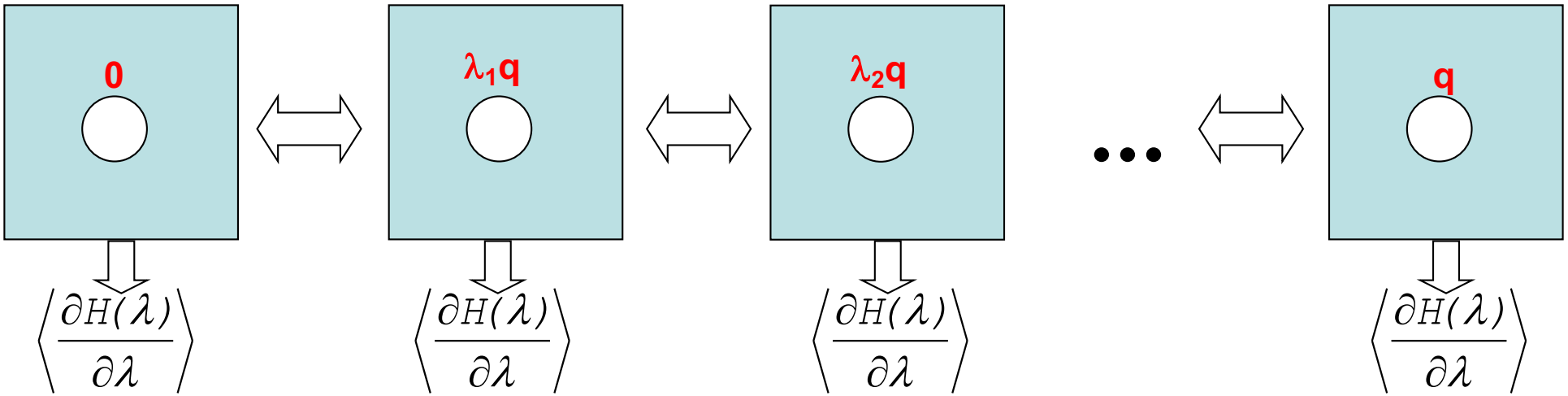
$$\Delta A = \int_0^1 \left\langle \frac{\partial H(\lambda)}{\partial \lambda} \right\rangle_\lambda d\lambda = \int_0^1 \left\langle \sum_s \frac{qq_s}{r_s} \right\rangle_\lambda d\lambda$$

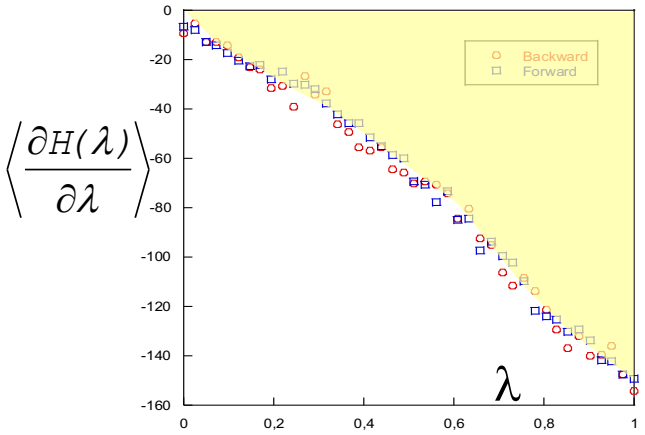
3.2. Thermodynamic Integration

- A toy model

$B^{0}_{(aq)} \Leftrightarrow B^{-1}_{(aq)}$

state λ=0 *state λ=1*





Forward = -71.7 kcal/mol

Backward = -70.6 kcal/mol

Average = -71.1±0.5 kcal/mol

(FEP -71.1 kcal/mol)

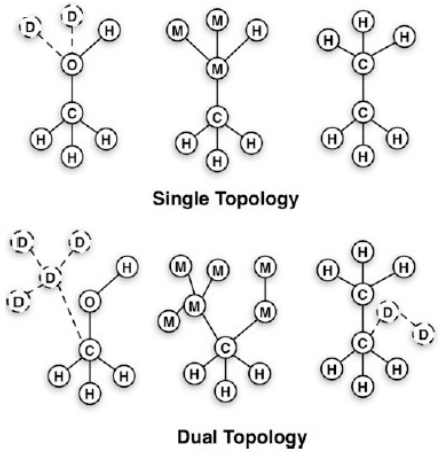
3.3. Practical Considerations

How can we build the end states?

Constructing alchemical pathways between two states can be done using two different approaches:

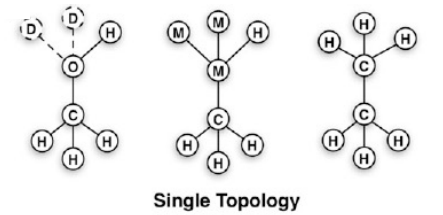
Single Topology: A single molecule is defined containing all the atoms that must be present in the initial and final states. Some of these atoms maybe dummy atoms at the beginning or at the end of the transformation, this is they will have zero partial charges and zero van der Waals parameters.

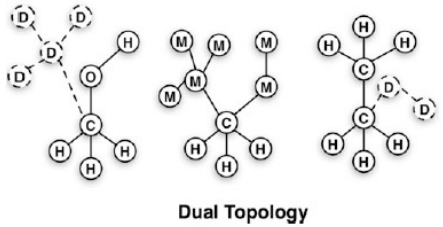
Dual Topology: Two molecules are defined, each one corresponding to the initial and final states. These two molecules may share some common atoms while other atom coexist but without interacting among them.



Alchemical transformation of methanol into ethane using single and dual topology approaches

3.3. Practical Considerations





One advantage of using dual topology is that the two molecules are free to sample their configurational space, which can help to converge the simulations

Instead, dual topology requires the decoupling of more atoms from their environment (annihilating their charges and van der Waals parameters), which could require the use of more intermediate states.

The choice usually depend on the simulation code used.

3.3. Practical Considerations

How can we build intermediate states?

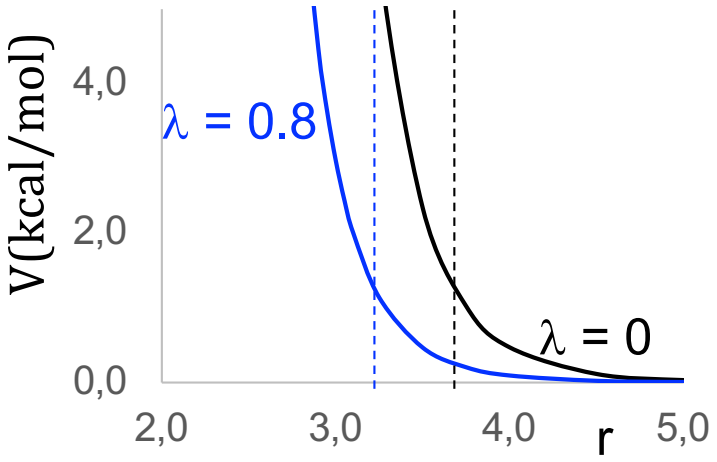
Intermediate states are usually obtained using a linear pathway between the initial and the final states:

$$H(\lambda) = (1-\lambda)\cdot H_I + \lambda \cdot H_{II}$$

Note that λ values do not need to be equidistant.
However, the linear choice can be problematic for small or large values of λ (λ close to 0 or close to 1) when atoms are being annihilated or created, due to the steep dependence of the repulsive Lennard Jones term with the distance.

Imagine you are trying to make disappear an atom with Lennar Jones parameters: σ = 3.9 Å and ε = 0,15 kcal·mol^{-1}. The repulsion term will vary as:

$$V(\lambda) = (1-\lambda)\cdot 4\varepsilon\left(\frac{\sigma}{r}\right)^{12}$$



When the transformation has been carried out up to 80%, the exclusion volume of the atom (defined from the distance at which the repulsion energy is > 2·RT) is still 70% of the initial value

If one runs a trajectory at λ=1 (without the atom) then some configurations will have a distance below 3,0 Å. Then when one tries to evaluate the energy at for example λ=0,8 a too large energy will be obtained, leading to convergence problems (also known as the 'end catastrophe')

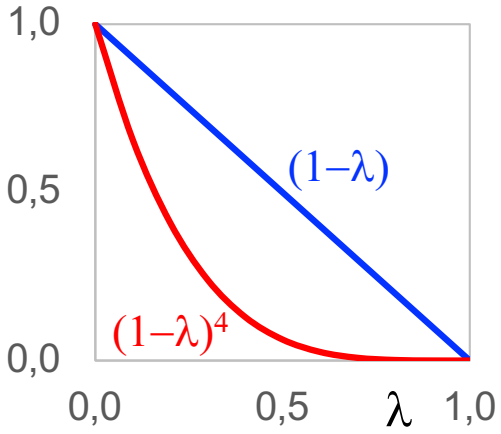
3.3. Practical Considerations

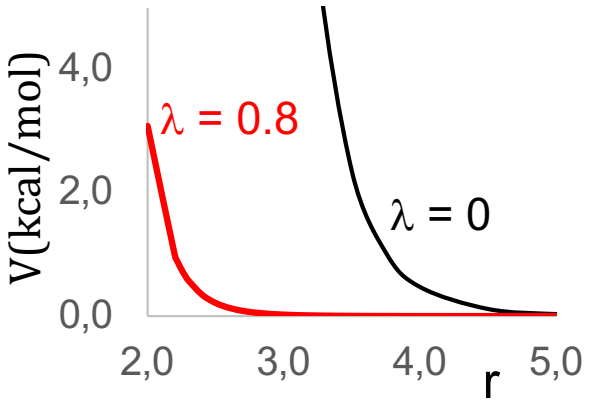
One possibility is to use a nonlinear transformation between the initial and final states:

$$H(\lambda) = (1-\lambda)^k \cdot H_I + [1-(1-\lambda)^k] \cdot H_{II}$$

where $k \geq 4$.

In that case the mixing between the initial and final states is less steep at the end points of the process and then a larger damping of the repulsive potential is obtained than when using a linear transformation.





Still, for any nonzero value of λ, whatever the power is used, there will be particles with a small impenetrable core that may cause problems during molecular dynamics simulations

3.3. Practical Considerations

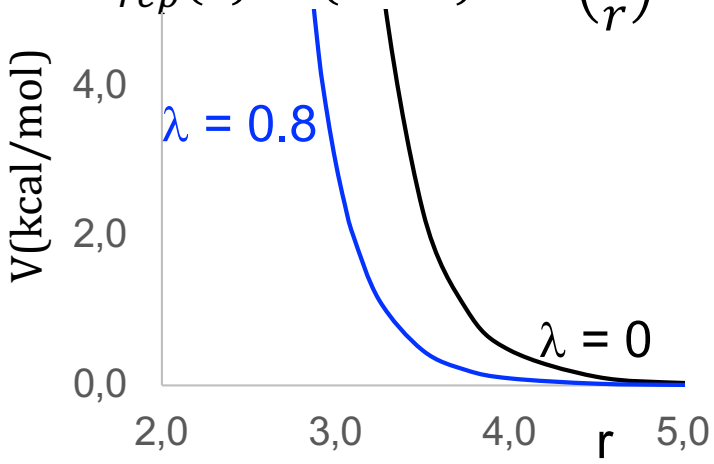
Another option is to substitute the original Lennard Jones potential of the atoms appearing/disappearing by the so-called 'soft-core' potential:

$$V_{LJ}(r,\lambda) = 4\varepsilon(1-\lambda)\left(\frac{1}{(\alpha\cdot\lambda+(r/\sigma)^6)^2}-\frac{1}{(\alpha\cdot\lambda+(r/\sigma)^6)}\right)$$

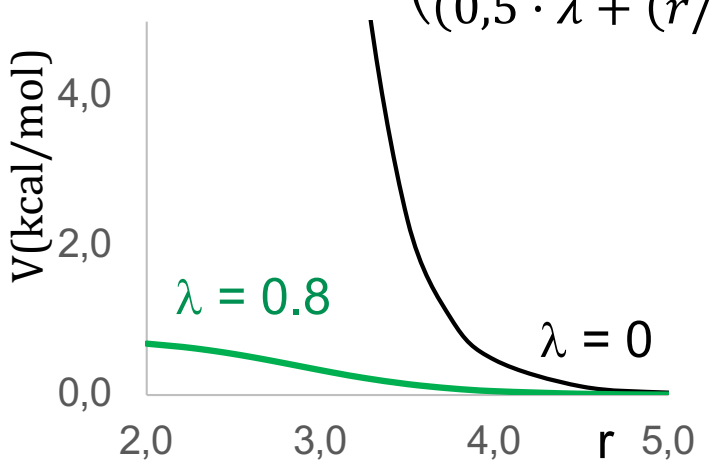
where, usually, α = 0,5

If we compare now the repulsive part of the original Lennard-Jones potential and the soft-core potential, for an atom with σ = 3.9 Å and ε = 0,15 kcal·mol^{-1}, we can observe the effect when we have already performed a large fraction of the process comparing λ=0 with λ =0,8

$$V_{rep}(\lambda) = (1-\lambda)\cdot 4\varepsilon\left(\frac{\sigma}{r}\right)^{12}$$



$$V_{rep}(r,\lambda) = 4\varepsilon(1-\lambda)\left(\frac{1}{(0{,}5\cdot\lambda+(r/\sigma)^6)^2}\right)$$

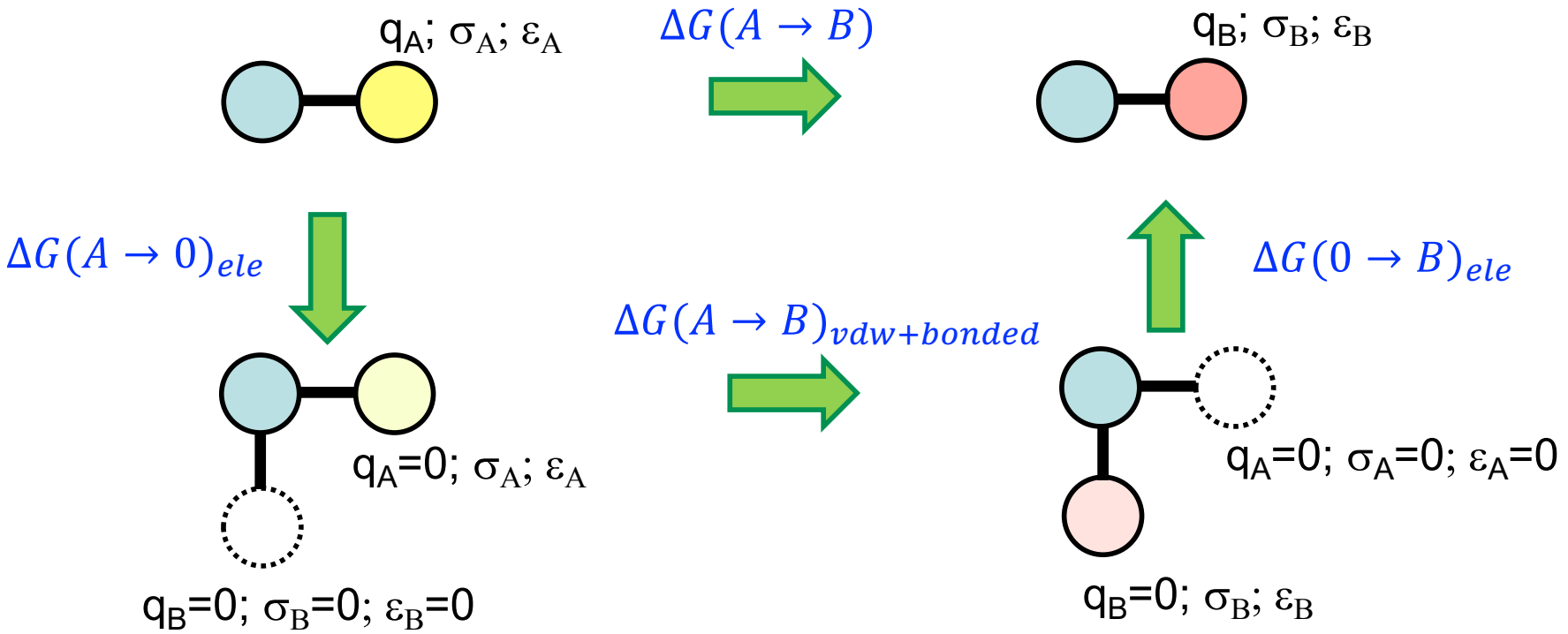


With the soft-core version of the potential, at λ=0,8 a large part of the disappearance process has been already done.

3.3. Practical Considerations

How can we build intermediate states?

The transformation can be carried out in a single step or splitting the electrostatic and the van der Waals transformations. This can be an advantage if when repulsion disappears two charged atoms becomes too close resulting in too large energies



Nowadays, considering the simulation times available, especially using GPUs, and with the use of soft-core potentials there are no clear evidences supporting one scheme or the other

3.3. Practical Considerations

How reproducible are Alchemical Transformations?

- Same software in simple cases (e.g. relative hydration free energy calculations between structurally similar small organic molecules) - highly precise estimates with standard deviation under 0.01 kcal/mol.

- Different simulation packages can introduce additional issues depending on integrators, thermostats, barostats, treatment of long-range electrostatics, etc. – variability up to 1 kcal/mol.

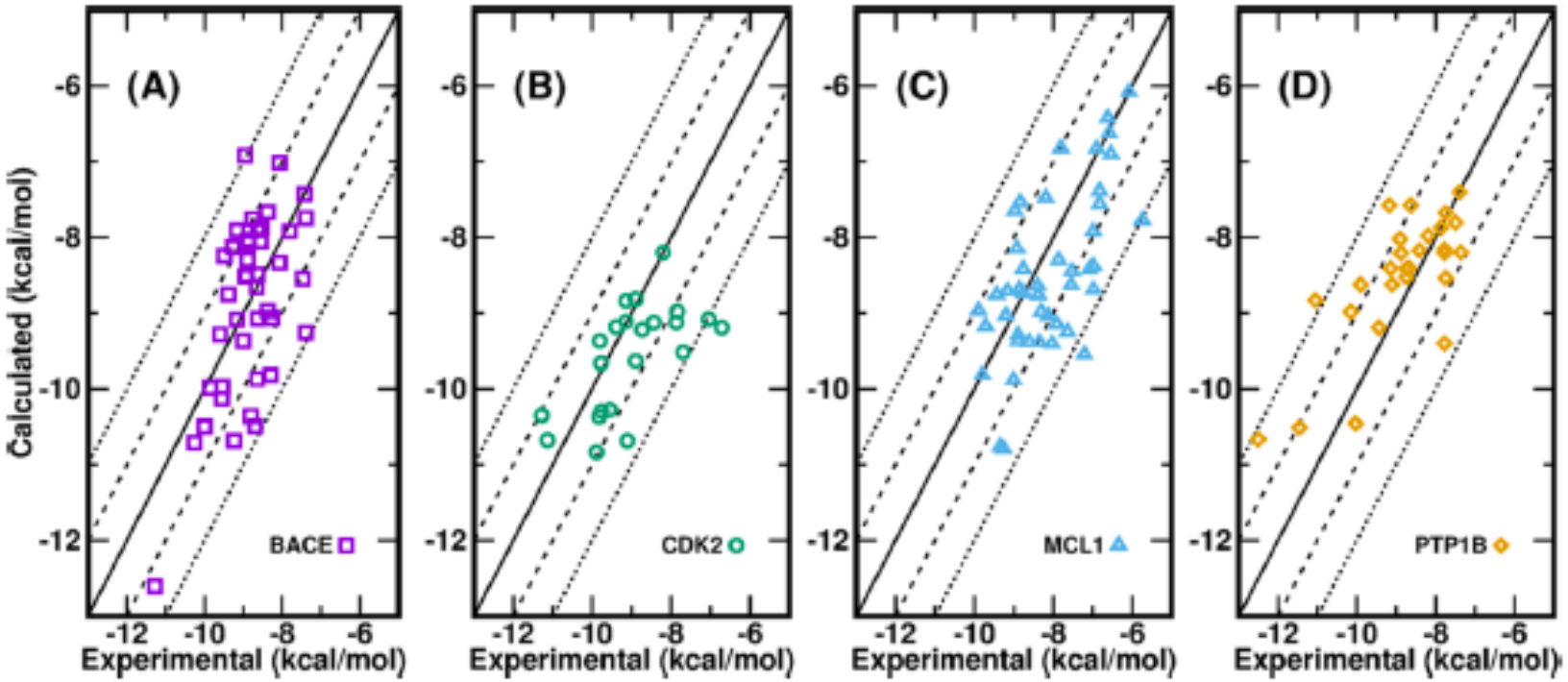
How accurate are Alchemical Tranformations?

- Root mean squared deviation relative to the experiment uses to be 1-2 kcal/mol
- The deviation depend on several factor such as the force field, sampling method, sampling time, number of intermediate steps, ….
- Additional factors that may increase the error: slow protein or ligand rearrangements, uncertainties in ligand binding mode, charged ligands can make these calculations far less reliable (because larger changes are involved).

3.3. Practical Considerations

Comparison between experimental and calculated binding free energies of 134 protein-ligand complexes with 4 different proteins.

Simulations were done with Amber using Thermodynamic Integration to evaluate the free energy change. Each value is the results of 5 replicas (in solution and in the receptor)



ACS Omega 5, 4611-4619, 2020

4. MM/PBSA & MM/GBSA Methods

The MM/PBSA method (Molecular Mechanics Poisson-Boltzmann Surface Area) was developed by Kollman et al. in the late 90s [J. Am. Chem. Soc. 1998, 120, 9401-9409] and has enjoyed a growing popularity since then. In this method the free energies of binding is estimated from the free energy differences of reactants and products calculated separately:

$$\Delta G_B = G_{RL} - G_R - G_L$$

Each of the free energies appearing on the previous equation is calculated as the sum of several terms, that includes the internal energy of each state, its entropy an the solvation free energy:

$$G_i = \langle E_{MM,i} \rangle - T \cdot S_i + \Delta G_{sol,i}$$

where the brackets above emphasize the fact that the internal energy is an average on that state. This term is further decomposed into:

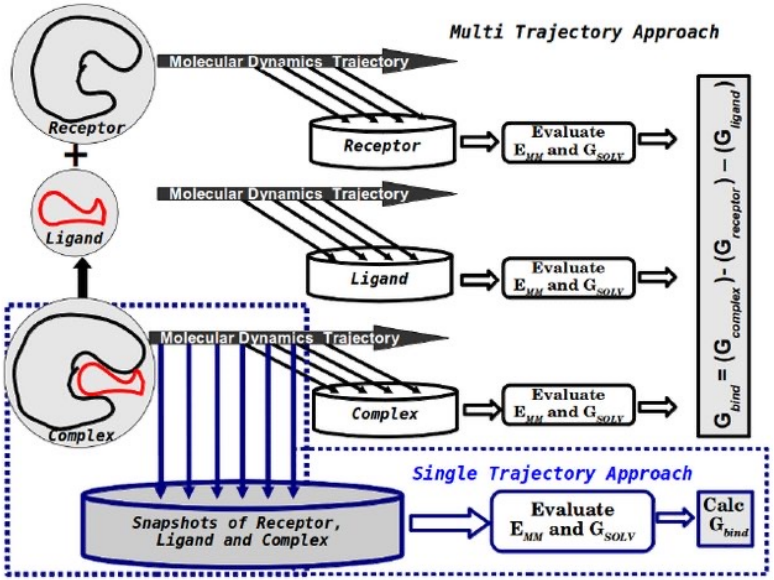
$$E_{MM} = E_{cov} + E_{ele} + E_{vdw}$$

The first term accounts for the bond + angle + torsion energy terms of the force field, the second one for the nonbonded electrostatic interactions and the last one for the van der Waals energy.

4. MM/PBSA & MM/GBSA Methods

$$\Delta G_B = G_{RL} - G_R - G_L$$

Strictly speaking, each of the free energies appearing in this equations should be calculated from independent simulation of the complex (RL), the unbound receptor (R) and the ligand alone (L), all of them in solution. However, the common practice in MM/PBSA and MM/GBSA methods is to run a **single trajectory** of the complex and then create the ensemble average of the free receptor and the ligand alone by simply removing the other partner for the RL simulation and solvating the remaining atoms with a continuum model.



A single trajectory of the complex is run in explicit solvent. Snapshots are selected and then the coordinates of the atoms of the ligand, the receptor and the complex (no solvent molecules) are extracted in separated files and used to evaluate the terms needed for the evaluation of the Binding free energy

4. MM/PBSA & MM/GBSA Methods

$$\Delta G_B = G_{RL} - G_R - G_L$$

One of the advantages of the single trajectory approach is that the covalent contribution to the internal energy exactly cancels out. The terms are large and fluctuate considerably, making convergence very difficult:

$$\Delta E_{cov} = \langle E_{cov,RL} \rangle_{RL} - \langle E_{cov,R} \rangle_{RL} \langle E_{cov,L} \rangle_{RL} = 0$$

On the other hand, the single trajectory approach ignores the change in the geometry of the receptor and ligand upon binding, which can be an important factor for the affinity. However, the standard deviation has been found to be up to five times larger if individual simulations for the three states are run instead of a single trajectory for the complex.

So, the binding free energy can be obtained within this approach as:

$$\Delta G_B = \Delta E_{ele} + \Delta E_{vdw} - T \cdot \Delta S + \Delta \Delta G_{sol}$$

The first two terms are obtained from the evaluations of the MM energy in the single trajectory approach:

$$\Delta E_{ele} + \Delta E_{vdw} = \langle E_{MM,RL} \rangle_{RL} - \langle E_{MM,R} \rangle_{RL} \langle E_{MM,L} \rangle_{RL}$$

4. MM/PBSA & MM/GBSA Methods

The solvation term

For the evaluation of the solvation term, the explicit solvent is removed and the system (RL, T or L) is placed in a continuum solvent.

The solvation free energy term is then calculated as the sum of two contributions that account for the electrostatic (polar) solute-solvent interactions and the nonpolar terms (cavitation and solute-solvent van der Waals interactions):

$$\Delta G_{sol} = \Delta G_{npol} + \Delta G_{pol}$$

Cavity formation

Dispersion-repulsion interactions (van der Waals)

Mutual polarization of solute and solvent

$$\Delta G_{sol} = \underbrace{\Delta G_{cav} + \Delta G_{vdW}}_{\Delta G_{npol}} + \Delta G_{pol}$$

4. MM/PBSA & MM/GBSA Methods

The polar term is obtained solving the Poisson-Boltzmann equation for a distribution of point charges (the MM charges of the ligand, the receptor or the complex) embedded in a continuum

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

where ϕ is the electrostatic potential, ε the dielectric constant and ρ the charge distribution. The equation can be completed adding free ions in the solvent according to the Boltzmann distribution:

$$\rho_{ion}(r) = \sum_i C_i^0 \, z_i e \cdot exp\left(-\frac{z_i e\phi(r)}{RT}\right)$$

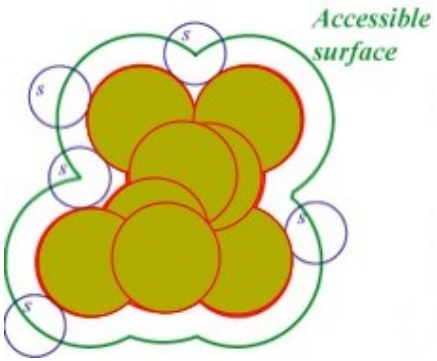
where C^0 is the concentration and $z_i \cdot e$ the charge of the ion.

The resulting equation can be linearized and solved using a finite difference method (PB methods) or a Generalized Born approximation (GB methods) in which the solute is described as a set of spheres centred on the charges and with an internal dielectric constant that differs from the external continuum solvent.

$$\Delta G_{pol} = \frac{1}{2}\int \rho(r) \cdot \phi(r) \cdot dV$$

4. MM/PBSA & MM/GBSA Methods

The nonpolar term arises from the formation of the solute's cavity and the van der Waals interactions (repulsion and dispersion) between the solute and the solvent. In the MM/PBSA and MM/GBSA methods these terms are simply represented by a linear relation to the solvent accessible surface area of the solute (SASA)



$$\Delta G_{npol} = \gamma \cdot SASA + b$$

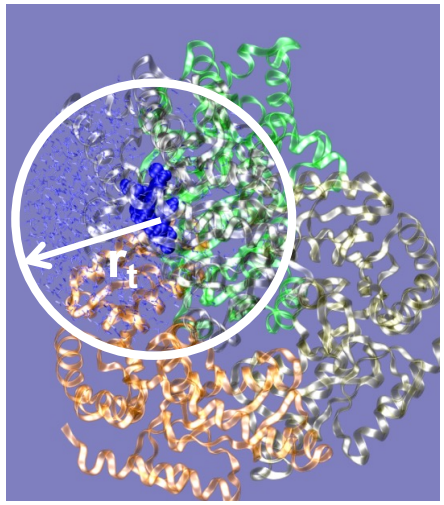
The surface tension and the independent term are usually set to be constant for all solutes. For example, in Amber the values are:

$\gamma = 5.42 \cdot 10^{-3}$ kcal·mol^{-1}·Å^{-2}
b=0.92 kcal·mol^{-1}

4. MM/PBSA & MM/GBSA Methods

The entropic term

In the original formulation the entropic term is obtained after minimization and normal mode analysis. In order to have a system-size treatable with this approach a truncation radius (r_t) is used, deleting all the water molecules and receptor residues found at a distance > r_t from any ligand atom.



$$\Delta S = S_{RL} - S_R - S_L$$

$$S_i = \mathrm{R} \cdot \sum_j \left[\frac{u_j}{\exp(u_j) - 1} - ln\left(1 - \exp(-u_j)\right) \right]$$

$$u_j = \frac{h\upsilon_j}{k_B T} \; ; \mathrm{j} = 1..3\mathrm{N} - 6$$

The entropy calculation uses to be the most costly part in MM/PBSA calculations. In addition, it is problematic since:

i) The normal mode approximation fails for low frequency modes
ii) The truncation introduces some arbitrarity and possible conformational changes during minimization
iii) Its evaluation usually gives the largest statistical uncertainty of all the terms

Very often this term is simply omitted, which can be a good approximation to evaluate relative binding free energies but not for absolute affinities.

4. MM/PBSA & MM/GBSA Methods

Results for seven biotin analogues to avidin (in $kJ \cdot mol^{-1}$) and standard deviations for two of them (1 & 7). From: Expert Opinion on Drug Discovery, 10, 449-461, 2015.
Ligands 1-3 have charge -1, ligands 4-7 are neutral

Term	Btn1	Btn2	Btn3	Btn4	Btn5	Btn6	Btn7	SD1	SD7
E_{el}	−1224	−1295	−1287	−174	−83	−50	−109	46	21
E_{vdW}	−148	−149	−132	−200	−128	−128	−49	16	11
G_{pol}	1224	1321	1259	266	146	123	124	30	13
G_{np}	−17	−17	−17	−21	−16	−16	−11	0	0
TS	−81	−96	−70	−82	−67	−66	−28	46	57
$E_{el} + G_{pol}$	0	26	−27	92	63	72	15	22	20
ΔG_{bind}	−187	−145	−222	−114	−49	−34	−53	47	62

i) E_{el} and G_{pol} are the larger terms for charged species but they cancel each other
ii) T·S term has the largest standard deviation
iii) G_{npol} is negative (cavitation favours association) and quite constant
iv) E_{vdw} dominates

4. MM/PBSA & MM/GBSA Methods

Advantages:

i) It is a fast method
ii) It is useful to rationalize differences in binding free energies
iii) It can provide reasonable trends within families of ligands

Disadvantages:

i) Structural changes upon binding are ignored
ii) Convergence problems: several independent estimations can result in large statistical uncertainties
iii) No systematic way to improve it (QM description for the ligand, better solvation free energies, polarizable force fields, ….)
iv) It is not accurate enough for predictive drug design

End

That's all…..

…well, that's never true.

Further Reading:

Evolution of Alchemical Free Energy Methods in Drug Discovery
L. F. Song & K. M. Merz Jr.
J. Chem. Inf. Model. 60, 5308-5318, 2020

Relative Binding Free Energy Calculations in Drug Discovery: Recent Advances and Practical Considerations
Z. Cournia et al.
J. Chem. Inf. Model. 57, 2911-2937, 2017

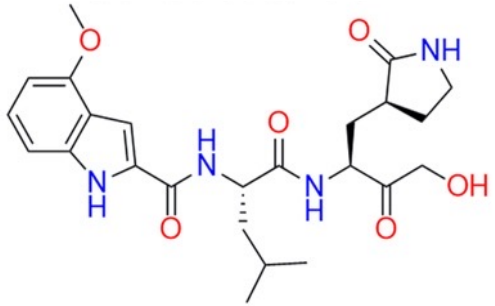
An Introduction to Best Practices in Free Energy Calculations
M. R. Shirts & D. L. Mobely
Biomolecular Simulations: Methods and Protocols. Springer, NY 2013

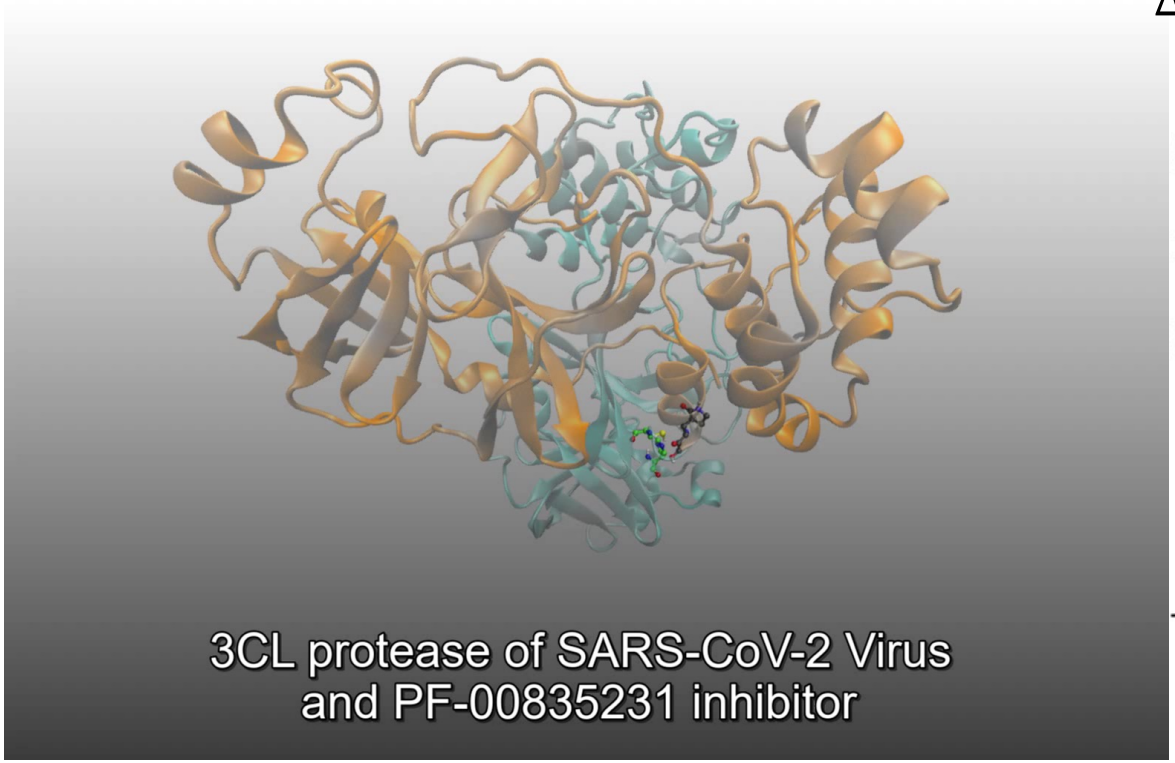
The MM/PBSA and MM/GBSA methods ro estimate ligand-binding affinities
S. Genheden & U. Ryde
Expert Opinion on Drug Discovery 10, 449-461, 2015

Fast, Accurate, and Reliable Protocols for Routine Calculations of Protein−Ligand Binding Affinities in Drug Design Projects Using AMBER GPU-TI with ff14SB/GAFFS.
X. He et al.
ACS Omega 5, 4611-4619, 2020

1. Binding Process & Free Energies

PF-00835231 ketone-based inhibitor





$$\Delta G_B = \Delta G_B(E-I) = \Delta G_B(EI) + \Delta G_{reac}$$