Icahn School of Medicine at Mount Sinai

Free-energy perturbation

BSR3101: Computer Aided Drug Design

October 2019

Classification of computational strategies

Is a 3D model of the target available?

No ☞ Ligand-based

- Similarity-based virtual screening
- QSAR modeling
- Pharmacophore search

<u>but also:</u>

- Scaffold hopping
- Identification of novel scaffolds
  (e.g. for IP reasons, or to provide alternate lead series to improve chemistry, ADME, etc.
- Poly-pharmacology
  Target-fishing

Yes ☞ Structure-based

➡Protein prep
➡Ligand QM calculations
➡Binding Site recognition

- Protein-ligand docking
- Structure-based VS
- Molecular Dynamics
- Free-energy methods
- …

Classification of computational strategies

Is a 3D model of the target available?

No ☞ Ligand-based

- Similarity-based virtual screening
- QSAR modeling
- Pharmacophore search

<u>but also:</u>

- Scaffold hopping
- Identification of novel scaffolds
  (e.g. for IP reasons, or to provide alternate lead series to improve chemistry, ADME, etc.
- Poly-pharmacology
  Target-fishing

Yes ☞ Structure-based

➡Protein prep
➡Ligand QM calculations
➡Binding Site recognition

- Protein-ligand docking
- Structure-based VS
- Molecular Dynamics
- Free-energy methods
- …

An example: lead optimization and the meaning of free-energy

Equilibrium binding (K_B)/dissociation (K_D) constant vs Free-energy (ΔG)

$$K_D \text{ (in M)} = \exp\left(-\frac{\Delta G}{0.6 \text{ kcal/mol}} \right)$$

K_D = ligand concentration at which ½ the ligands are bound

For good binding we want: small K_D, large and negative ΔG

ΔG	**K_D**	**Affinity range**
(kcal/mol)	M	
0	1	M
-2	$3.5\ 10^{-2}$	(cM)
-4	$1.3\ 10^{-3}$	mM
-8	$1.5\ 10^{-6}$	µM
-12	$2.0\ 10^{-9}$	nM

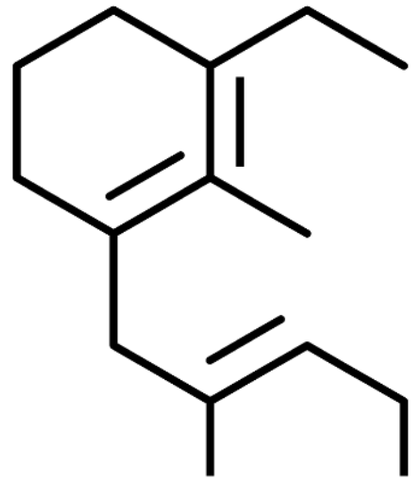
ΔG = energy difference between ligand in complex and ligand in solvent

Comparison between ligands: relative binding free-energy ($\Delta\Delta G$)

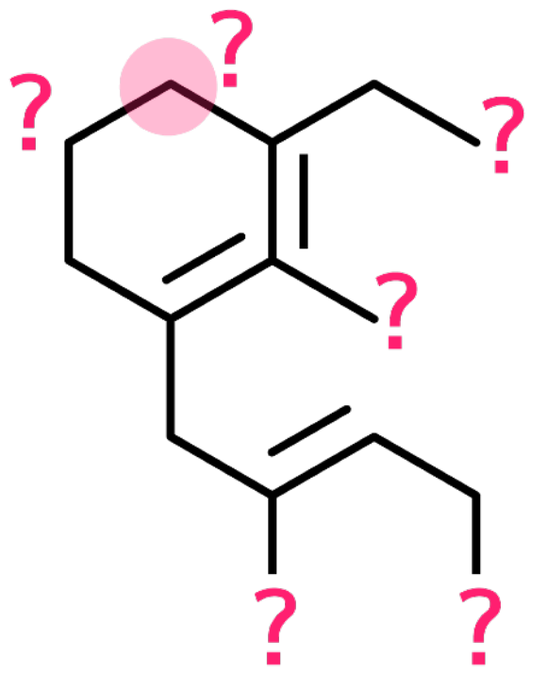
$$\frac{K_D(1)}{K_D(2)} = \exp\left(-\frac{\Delta G_1 - \Delta G_2}{0.6 \text{ kcal/mol}} \right) = \exp\left(-\frac{\Delta\Delta G}{0.6 \text{ kcal/mol}} \right)$$

An example: **lead optimization** and the meaning of free-energy

Given a "lead"
(i.e. an initial molecule that binds to our target)



Identify analogues
(i.e. molecules with similar scaffolds but sightly different)
that bind better (i.e. with higher affinity)



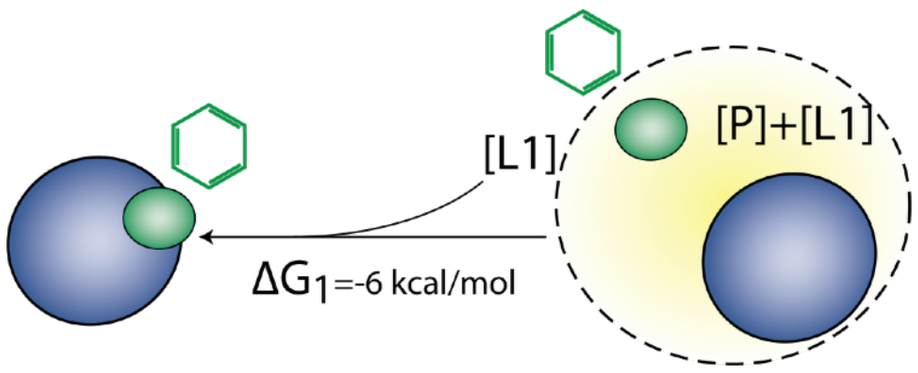
An example: lead optimization and the meaning of free-energy

An example: lead optimization and the meaning of free-energy

benzene

An example: lead optimization and the meaning of free-energy

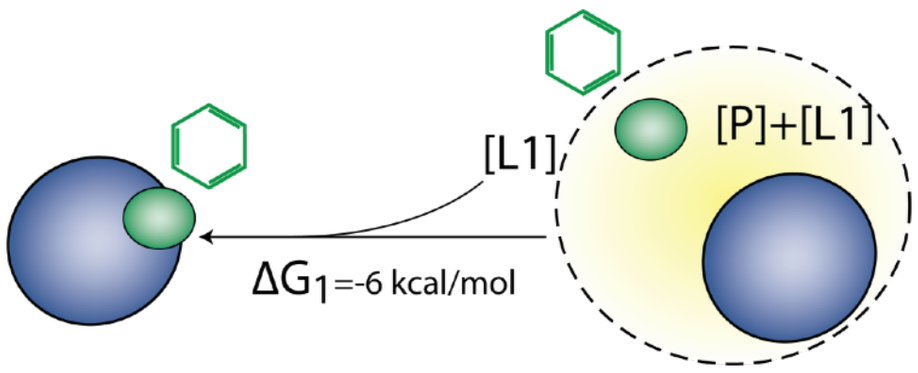
benzene



An example: lead optimization and the meaning of free-energy

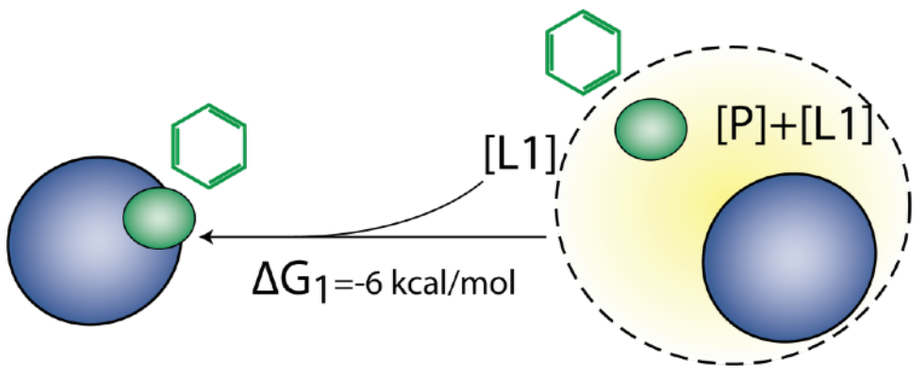
Calculating ΔΔGs is easier than calculating ΔGs

benzene



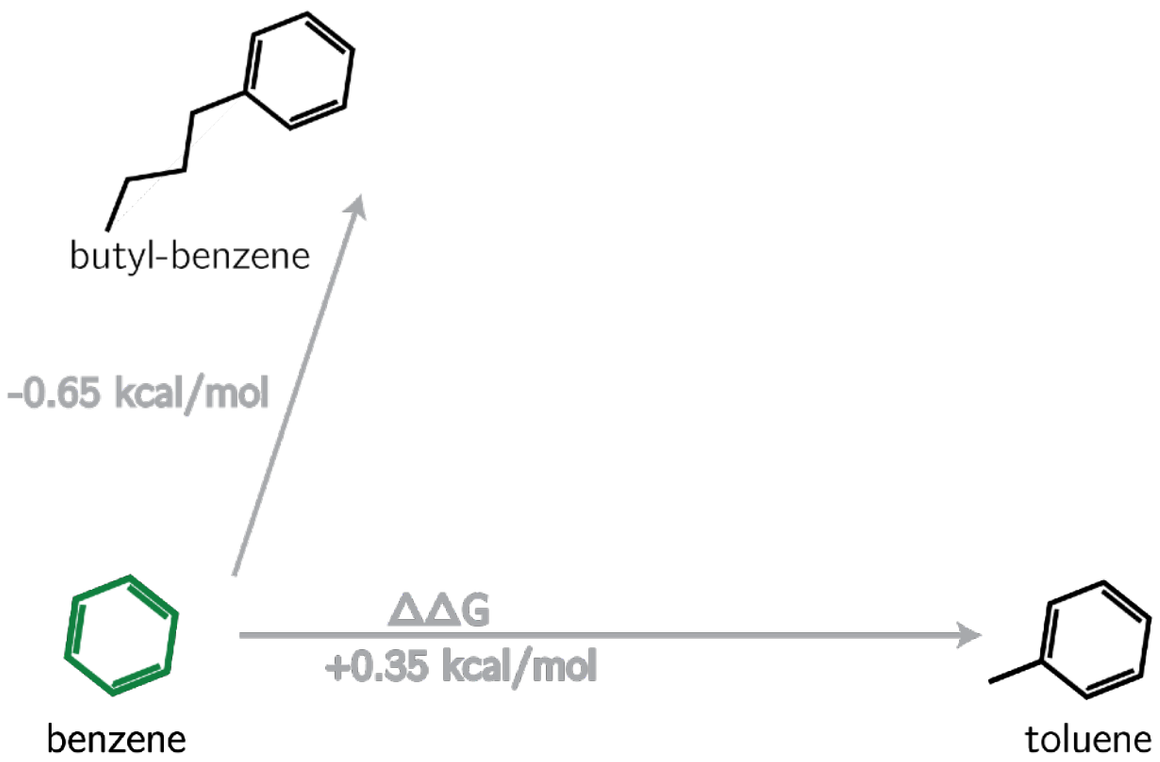
An example: lead optimization and the meaning of free-energy

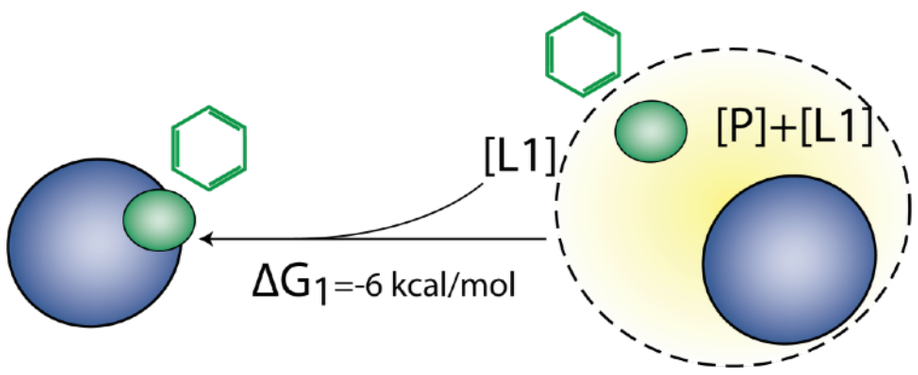
Calculating ΔΔGs is easier than calculating ΔGs



An example: lead optimization and the meaning of free-energy

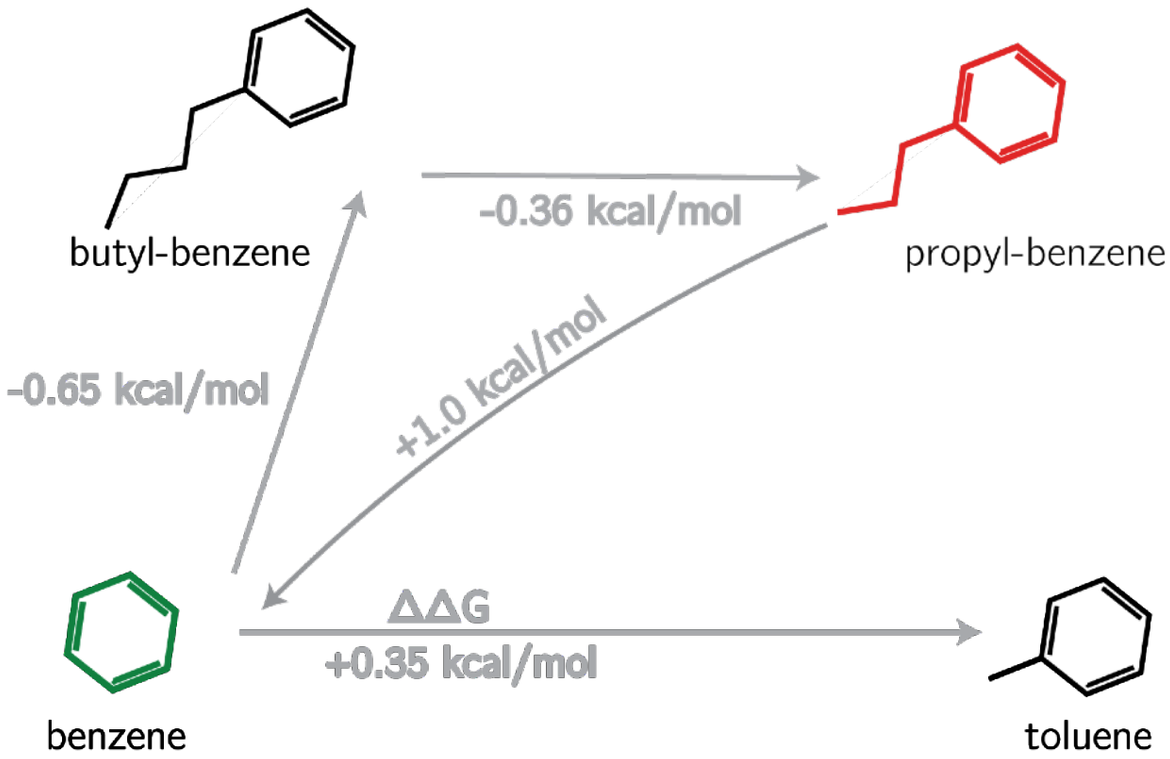
Calculating ΔΔGs is easier than calculating ΔGs

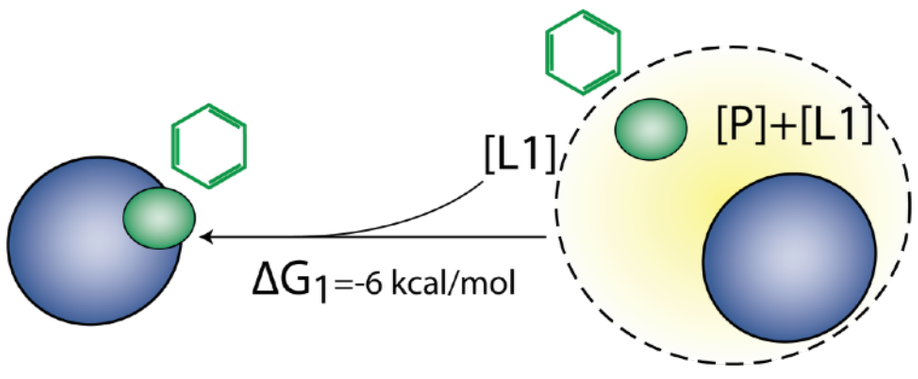




An example: lead optimization and the meaning of free-energy

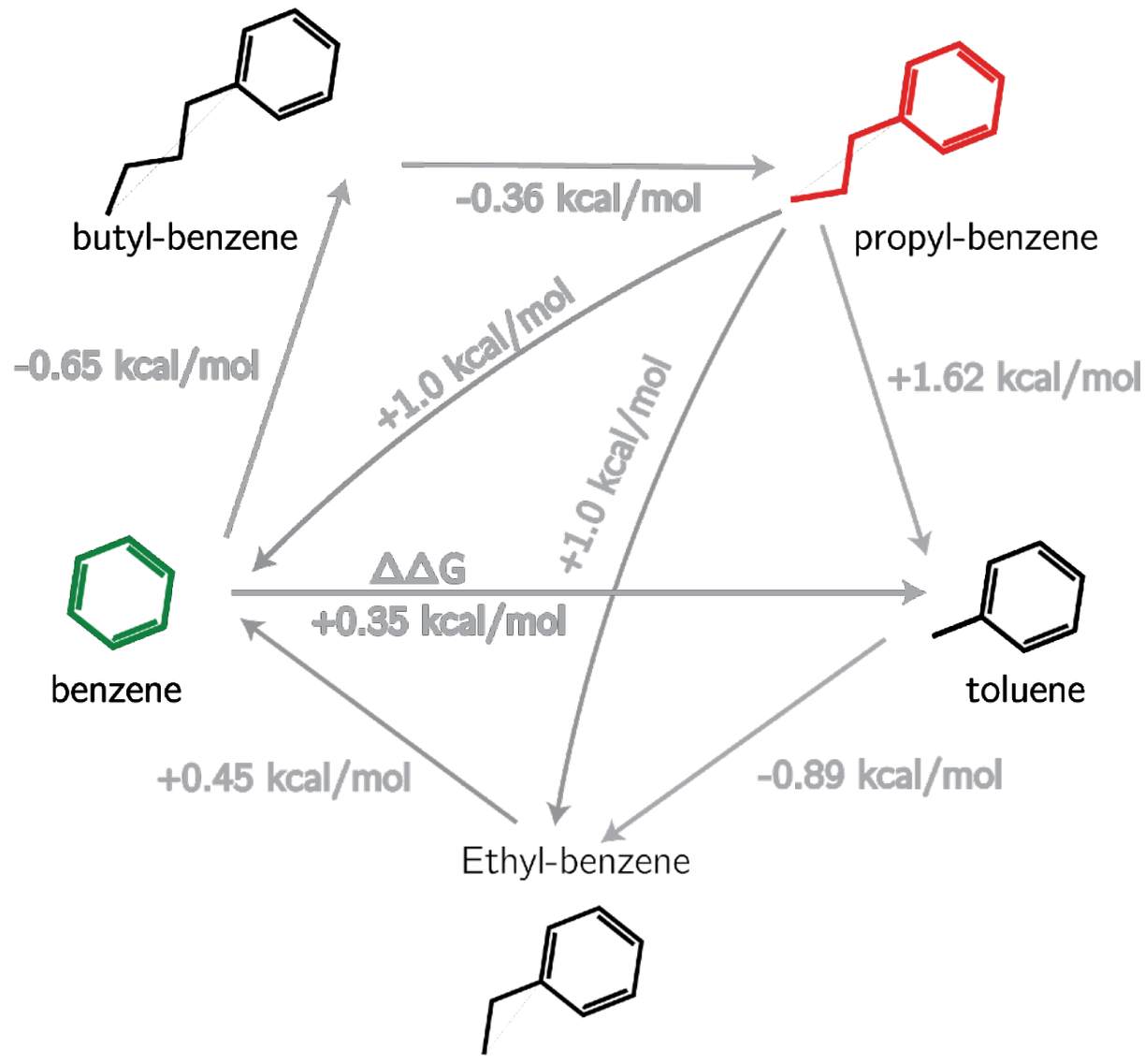
Calculating ΔΔGs is easier than calculating ΔGs

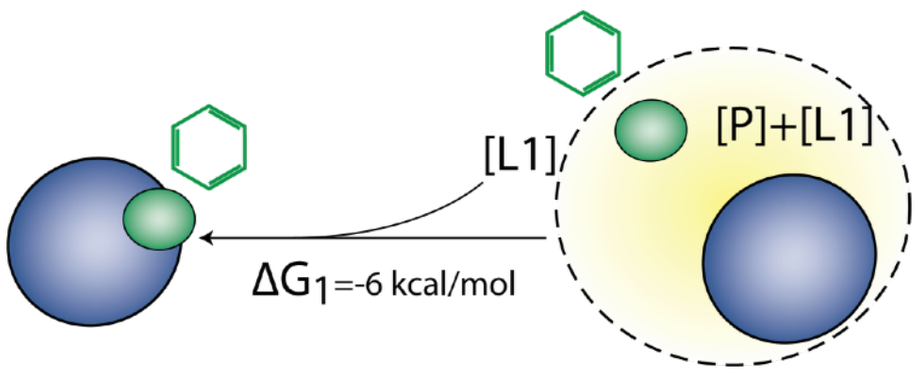




An example: lead optimization and the meaning of free-energy

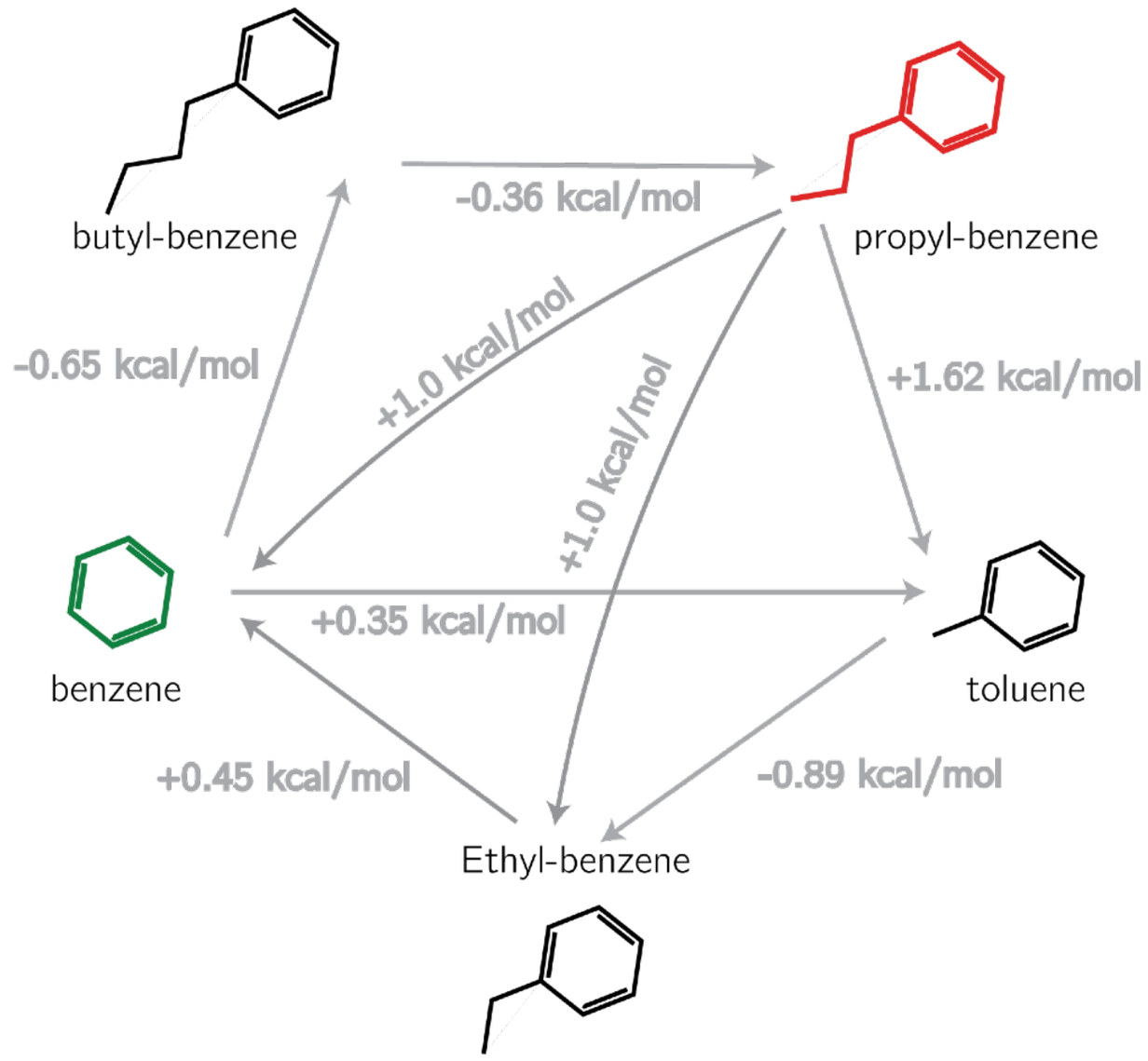
Calculating ΔΔGs is easier
than calculating ΔGs

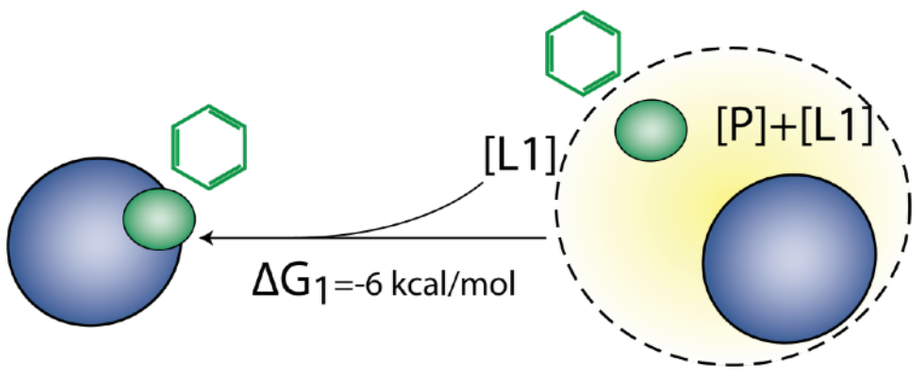


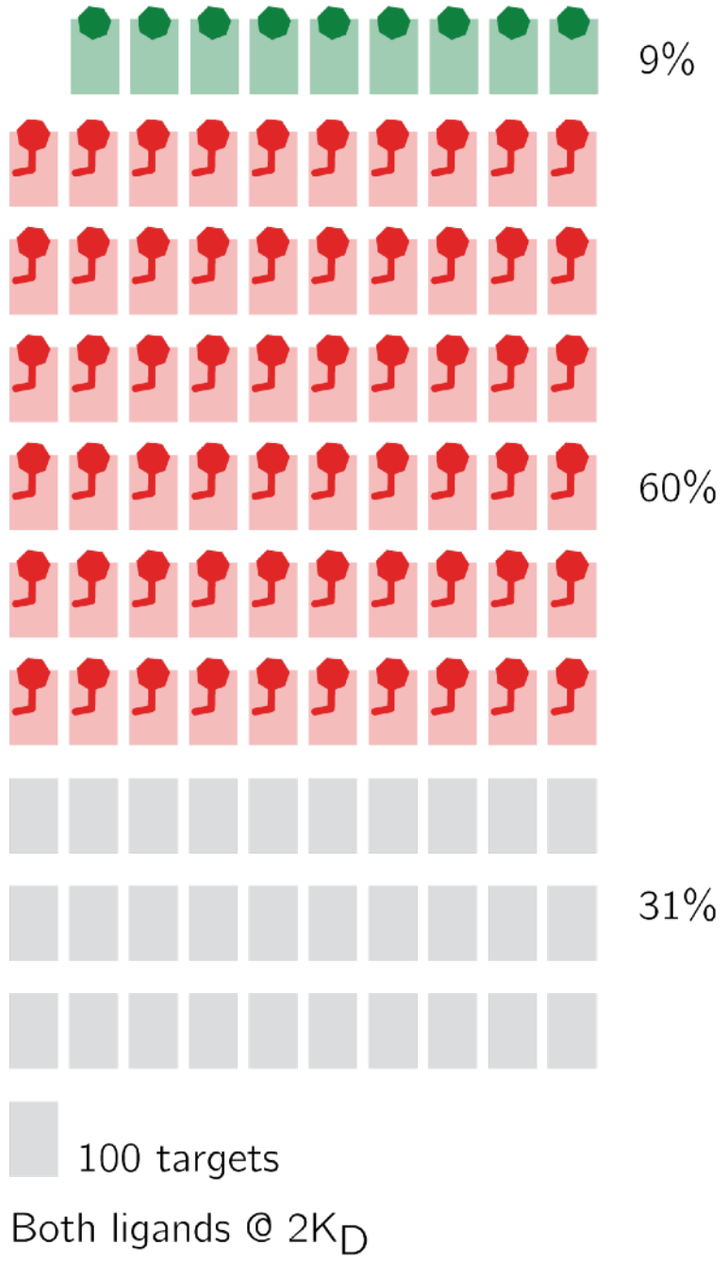


An example: lead optimization and the meaning of free-energy

Calculating ΔΔGs is easier
than calculating ΔGs



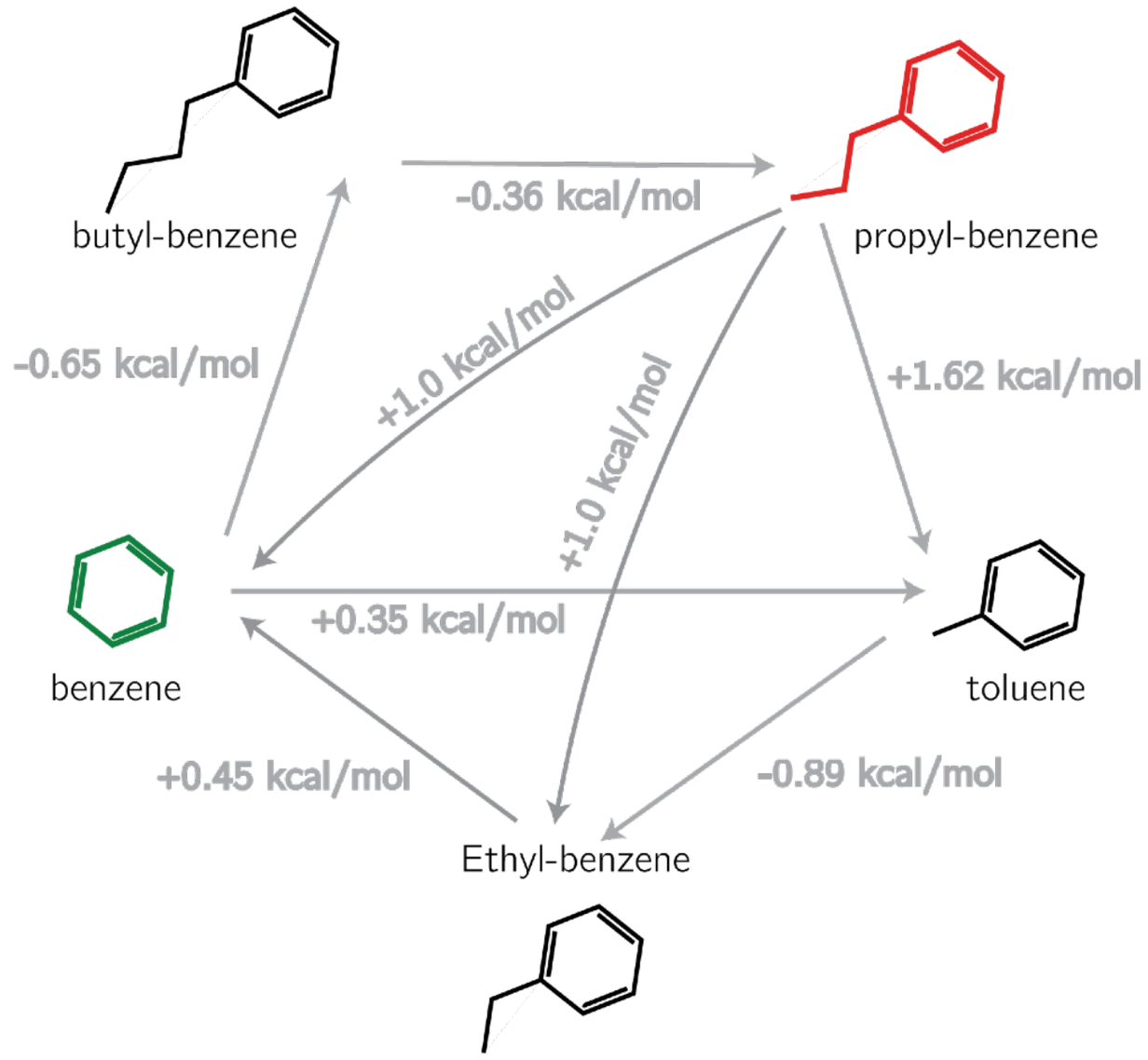


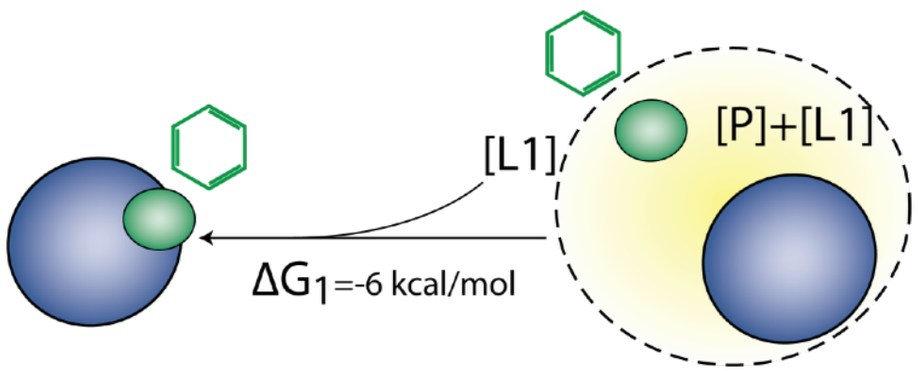


Both ligands @ $2K_D$

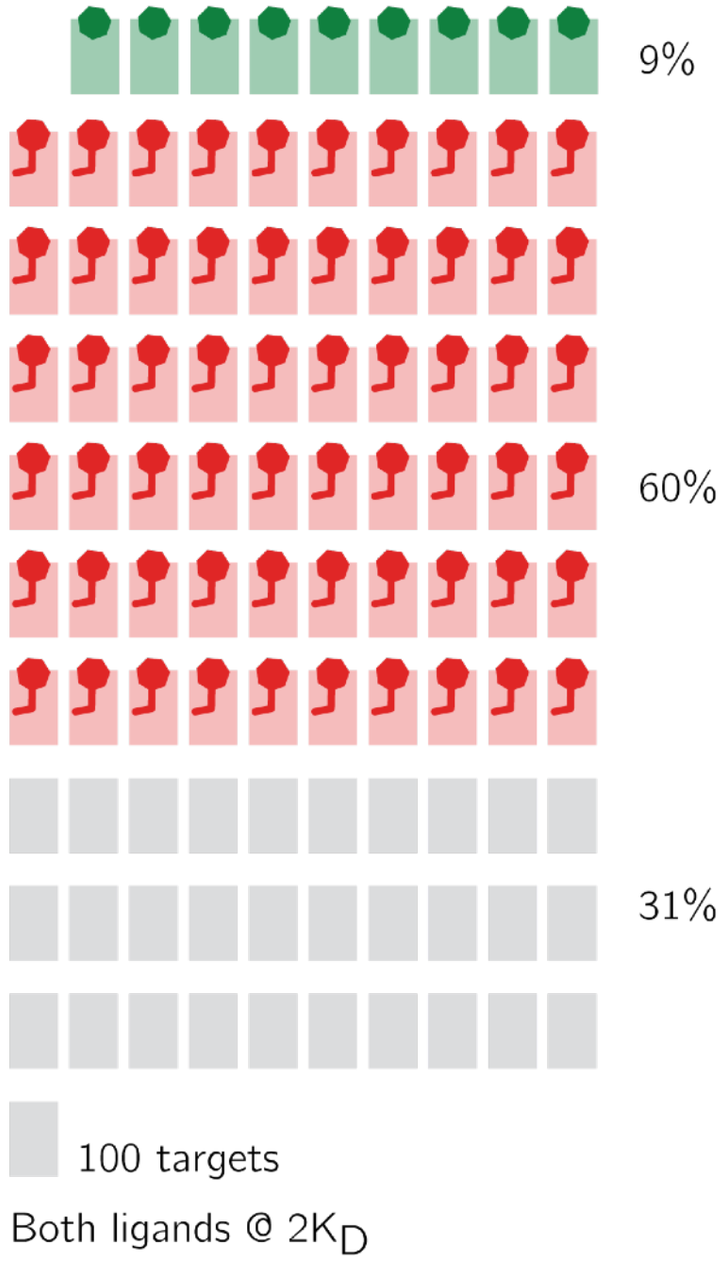
An example: lead optimization and the meaning of free-energy

Calculating ΔΔGs is easier than calculating ΔGs





Even small ΔΔGs can have significant effects on occupancy

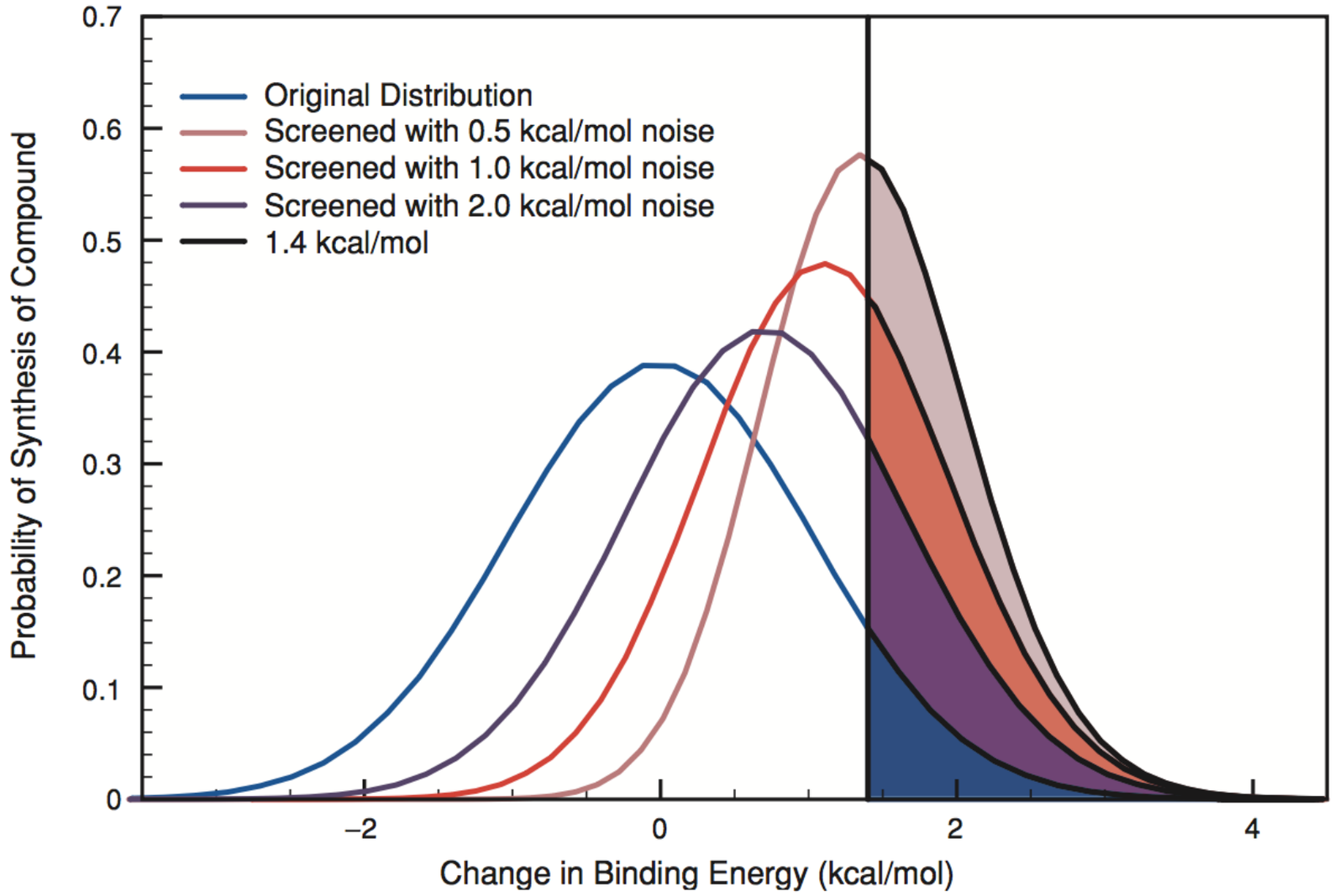


Free-energy methods

Pharma survey (Abbot, ~2005): 80% of affinity improvements are within 1.4 kcal/mol, 8.5% are more than 1.4 and 1% are better than 2.8 kcal/mol losses in binding are distributed similarly.

Suppose we screen N ligands with an imperfect algorithm that has a given error ε.

With an accuracy of ε=0.5 kcal/mol, screening N=10 ligands increases the chances to get a 1.4 kcal/mol improvement to 50% (from 8.5). Even with ε=2 kcal/mol, the probability of is 36% (vs. 8.5%)



Even with moderate error, a reliable method of filtering compounds could significantly improve the efficiency of synthesis in lead optimization.

Free-energy methods

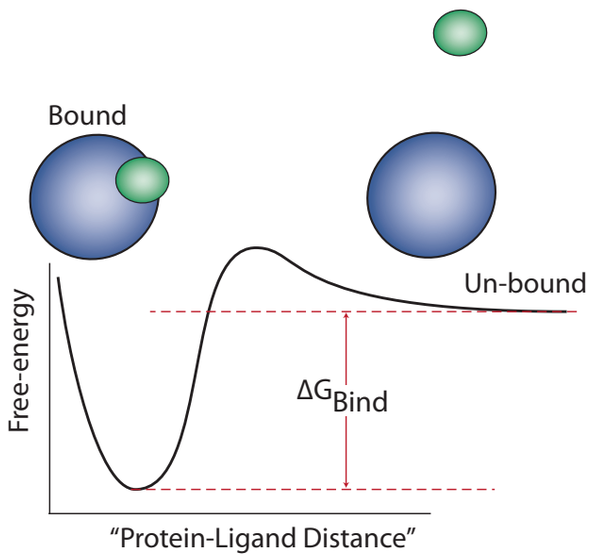
- "Docking" approaches rely upon a variety of approximations, often neglecting statistical mechanical effects such as
  - ➡ conformational entropy
  - ➡ averaging over multiple conformations or binding modes
  - ➡ discrete nature of solvent

  and chemical effects such as
  - ➡ protonation state
  - ➡ tautomer distributions,
  - ➡ their shifts upon binding.
- The neglect of these effects that is likely to be responsible for the gross inaccuracies of current scoring functions when making quantitative estimates of binding interactions.

Free-energy methods

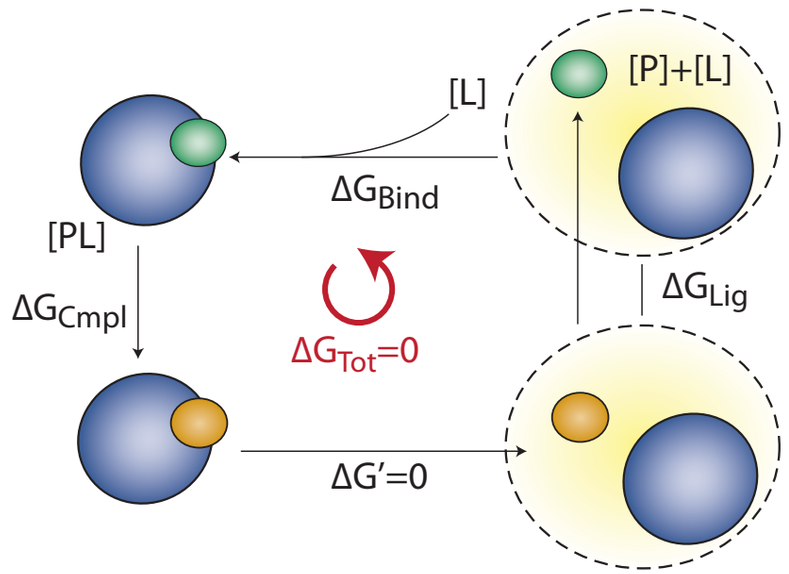
Potential of Mean Force (PMF) Methods

The binding/unbinding processes is simulated directly, biasing the simulation with external forces so that the overall work needed to remove the ligand from the binding pocket can be calculated



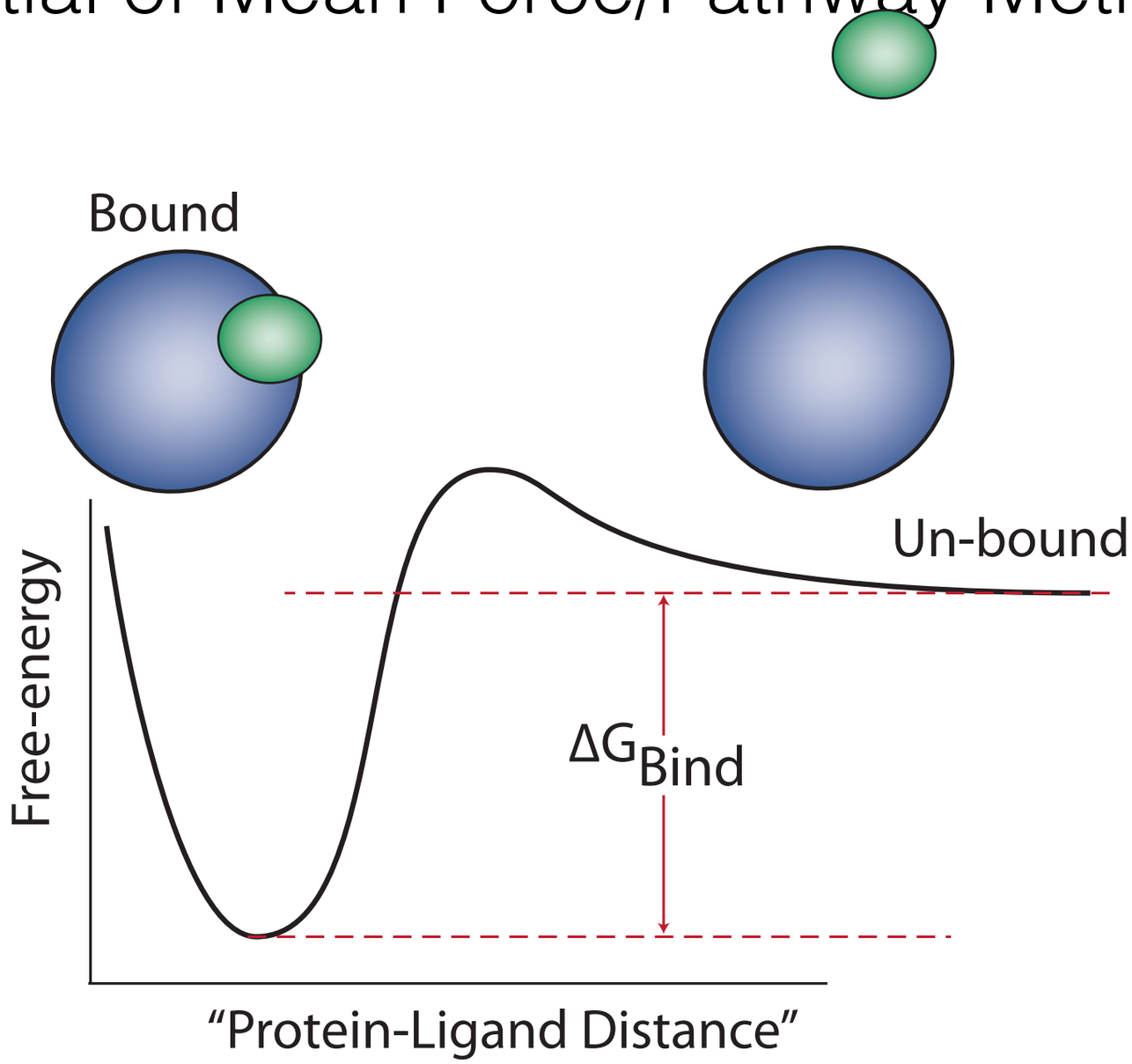
Alchemical Methods

Instead of simulating the binding/unbinding processes directly, the ligand is alchemically transmuted into either another chemical species or a noninteracting "dummy" molecule through intermediate, possibly nonphysical stages.



Because free energy is a state function, the choice of intermediates is in principle arbitrary, but in practice, can have great impact on the efficiency of the calculation.

Potential of Mean Force/Pathway Methods



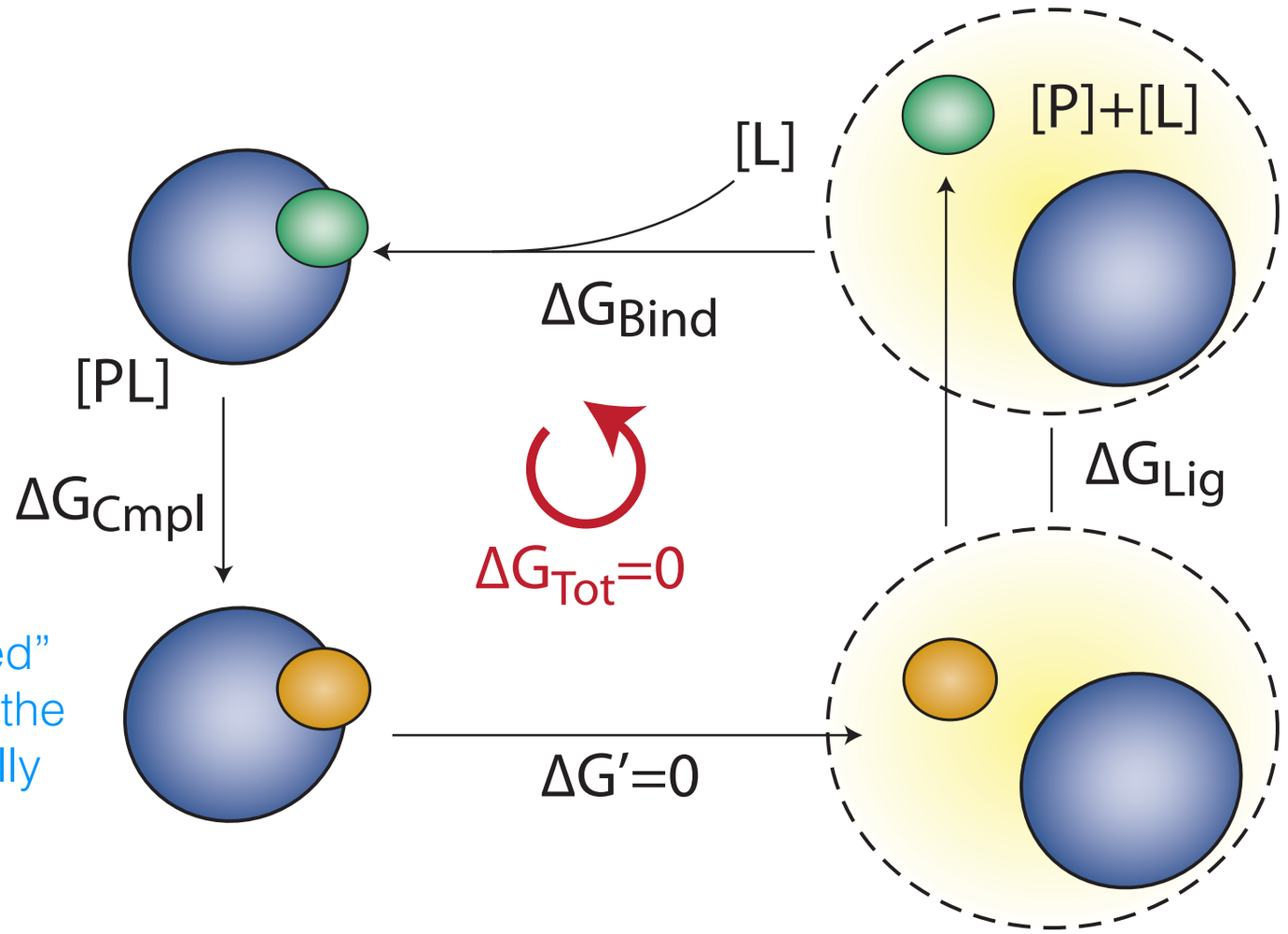
$$\mathrm{prob}(R) \propto \exp\left(\mathcal{H}/k_B T\right)$$

Can be obtained by

- unbiased MD
- enhanced MD (metadynamics, aMD, …)
- umbrella sampling
- …

Free-energy methods

Example: Double-decoupling method for absolute binding free-energy



The ligand is "uncoupled" from the protein, i.e. all the interactions are artificially switched off.

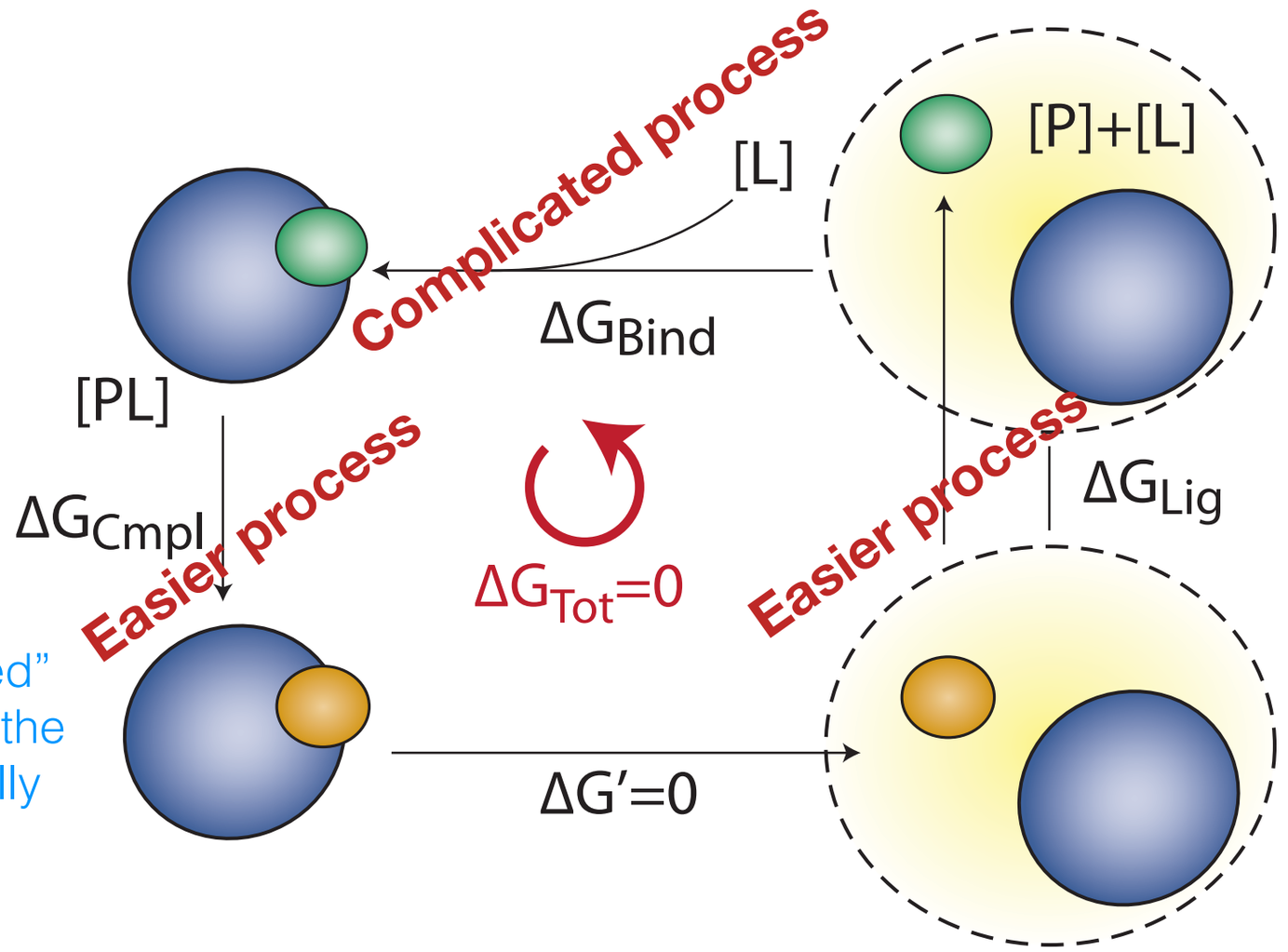
$$\Delta G_{Tot} = \Delta G_{Bind} + \Delta G_{Cmpl} - \Delta G' - \Delta G_{Lig} = 0$$

$$\boxed{\Delta G_{Bind} = \Delta G_{Lig} - \Delta G_{Cmpl}}$$

Instead of directly calculating the ΔG_{Bind} term, we calculate the easier terms ΔG_{Cmpl} and ΔG_{Lig}

Free-energy methods

Example: Double-decoupling method for absolute binding free-energy



The ligand is "uncoupled" from the protein, i.e. all the interactions are artificially switched off.

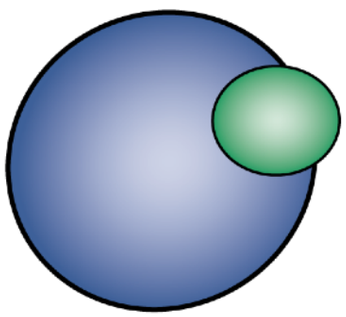
$$\Delta G_{Tot}=\Delta G_{Bind}+\Delta G_{Cmpl}-\Delta G'-\Delta G_{Lig}=0$$

$$\boxed{\Delta G_{Bind}=\Delta G_{Lig}-\Delta G_{Cmpl}}$$

Instead of directly calculating the ΔG_{Bind} term, we calculate the easier terms ΔG_{Cmpl} and ΔG_{Lig}

Free-energy **Perturbation** methods

The energy is written as a sum of a "unperturbed" term and a perturbation

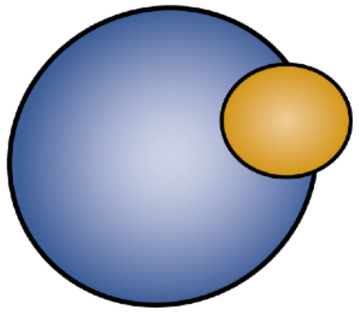


$$\mathcal{H}_1(R) = \mathcal{H}_0(R) + \Delta\mathcal{H}(R)$$

Non-interacting parts

Ligand-protein interactions

ΔG



$\mathcal{H}_0(R)$

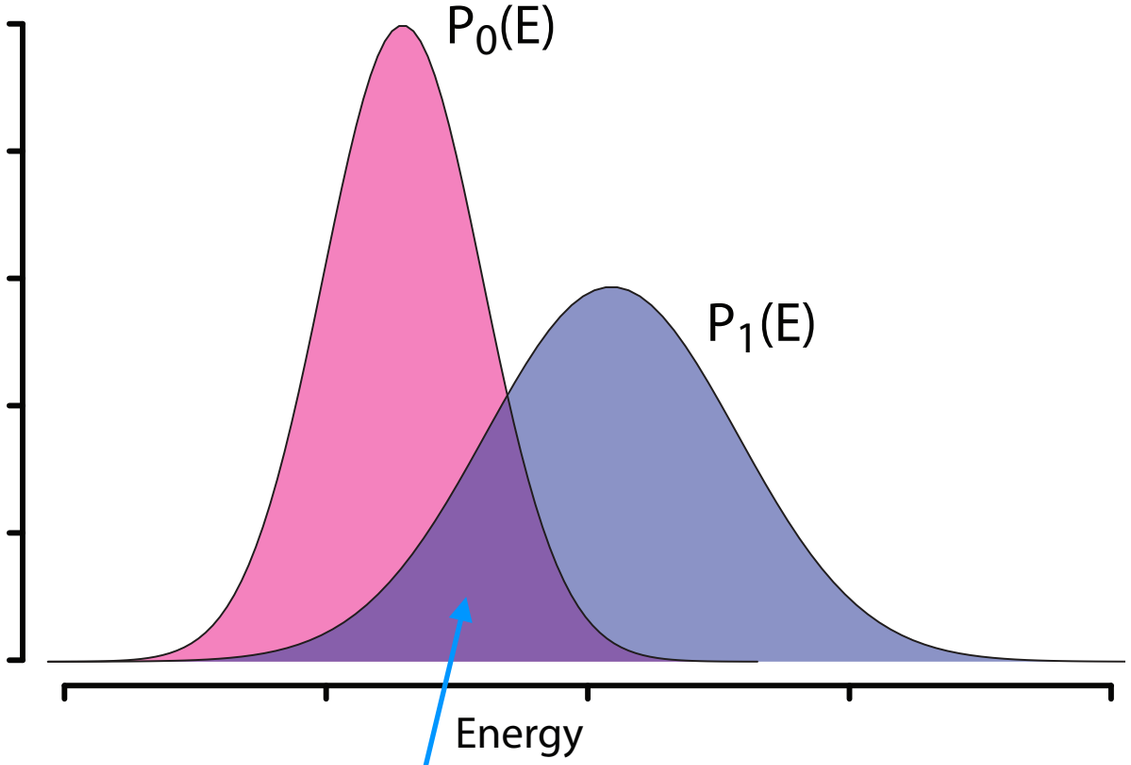
Run MD to sample **the unperturbed system**, and calculate the **mean** of the exponential of the perturbation

$$\Delta G = -k_B T \ln \langle \exp\left(\Delta\mathcal{H}/k_B T\right) \rangle_0$$

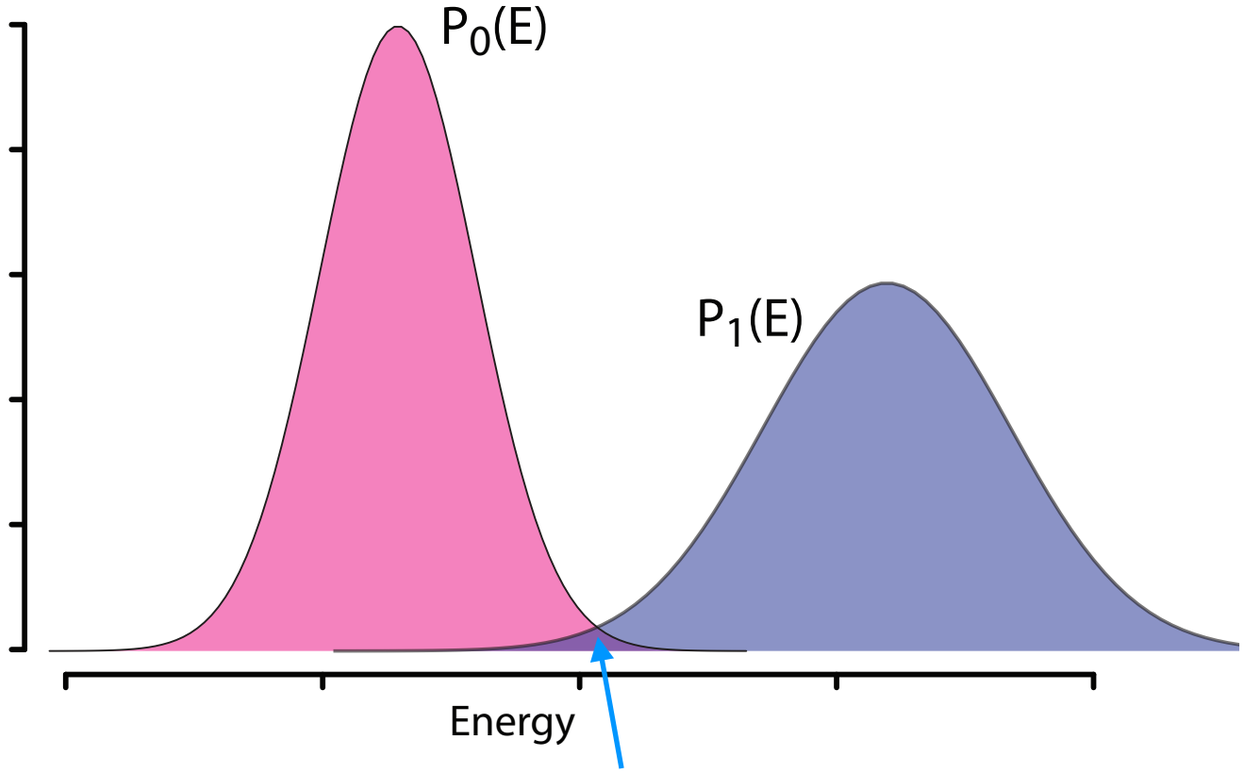
Free-energy **Perturbation** methods

$$\Delta G = -k_B T \ln \langle \exp\left(\Delta \mathcal{H} / k_B T\right) \rangle_0$$

Calculate the **mean** of the exponential of the perturbation **on the conformations sampled by the unperturbed system**

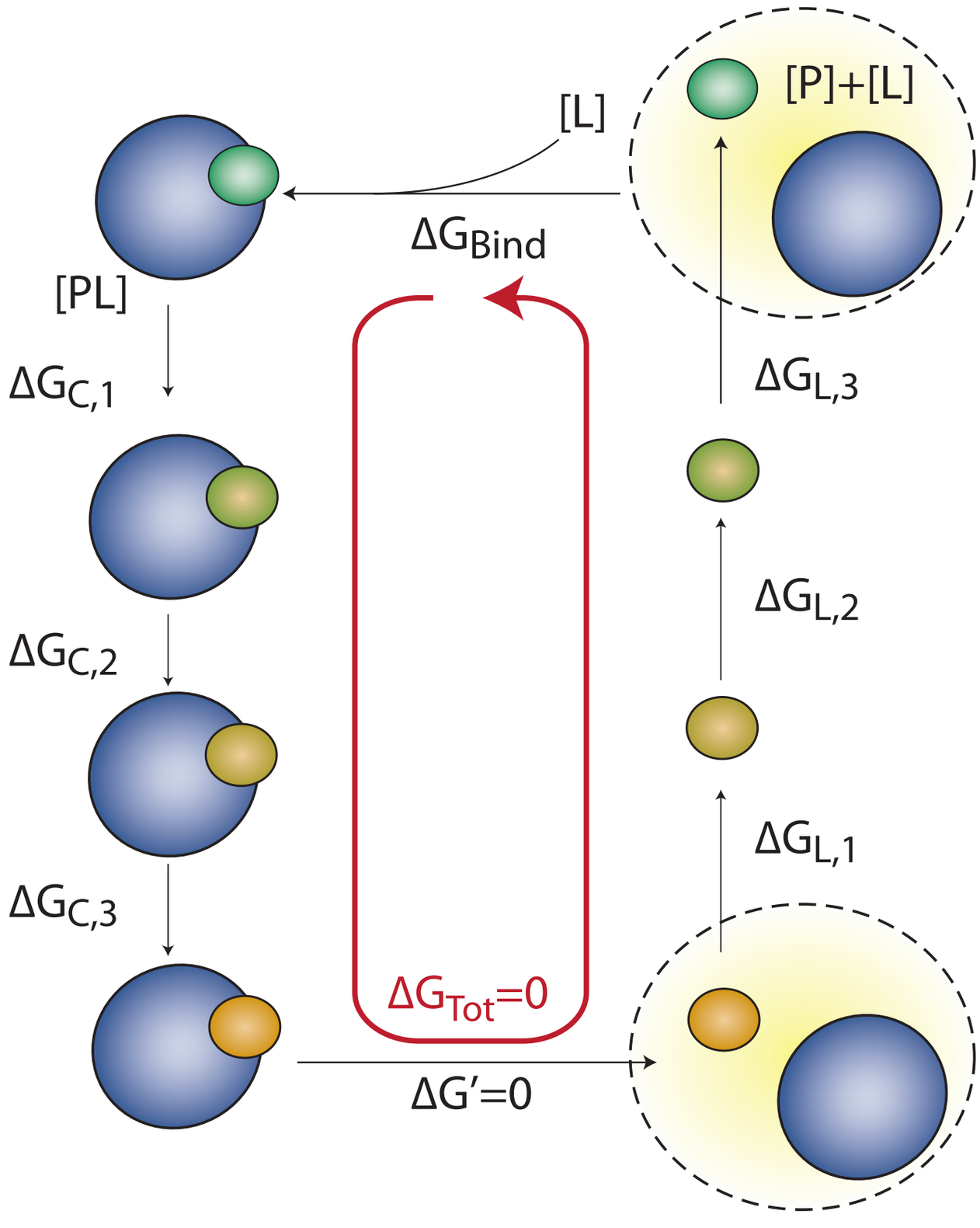


Good overlap allows to calculate the required mean accurately

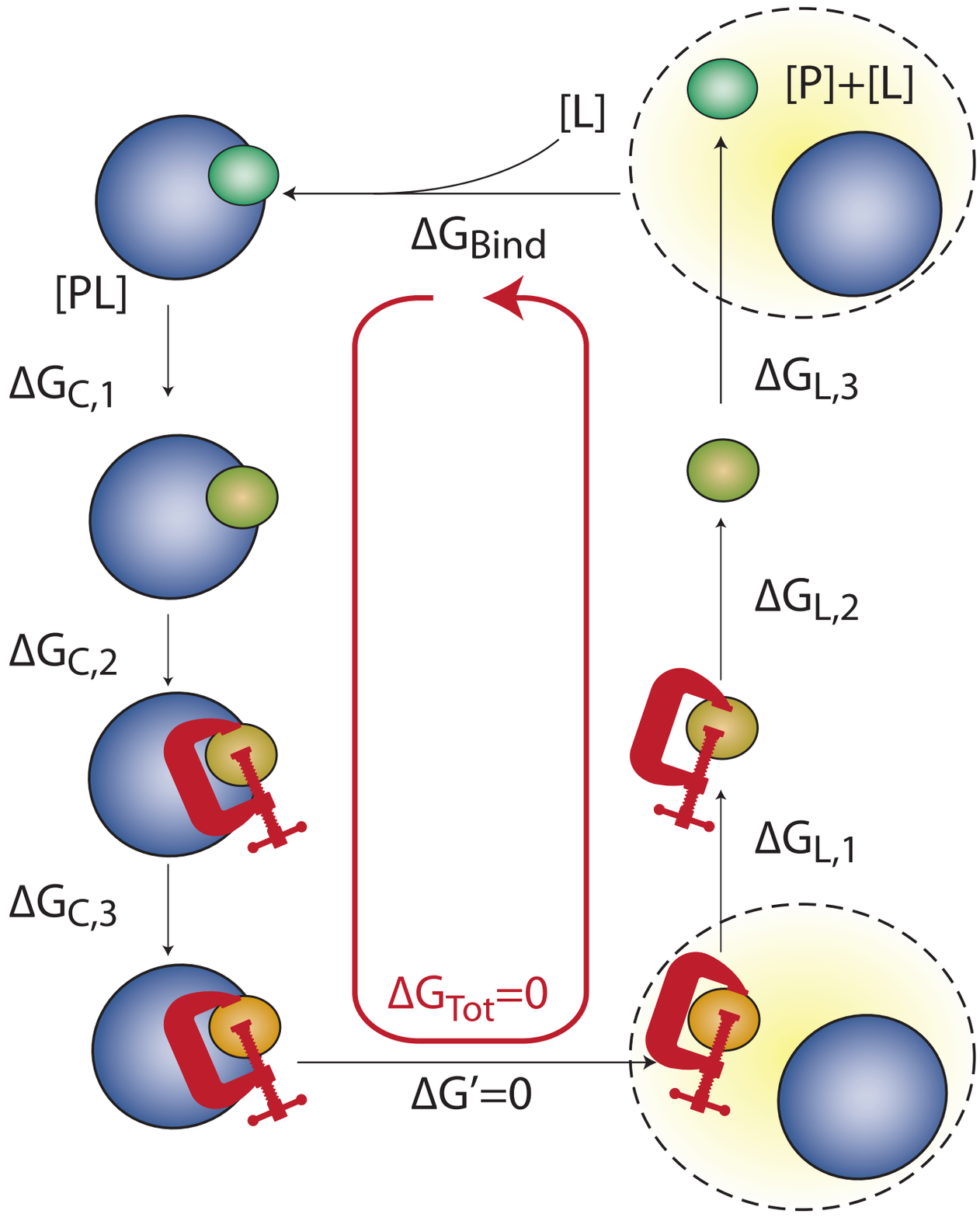


If the two states are too different, the poor overlap make the calculation of the mean very inaccurate

Free-energy **Perturbation** methods



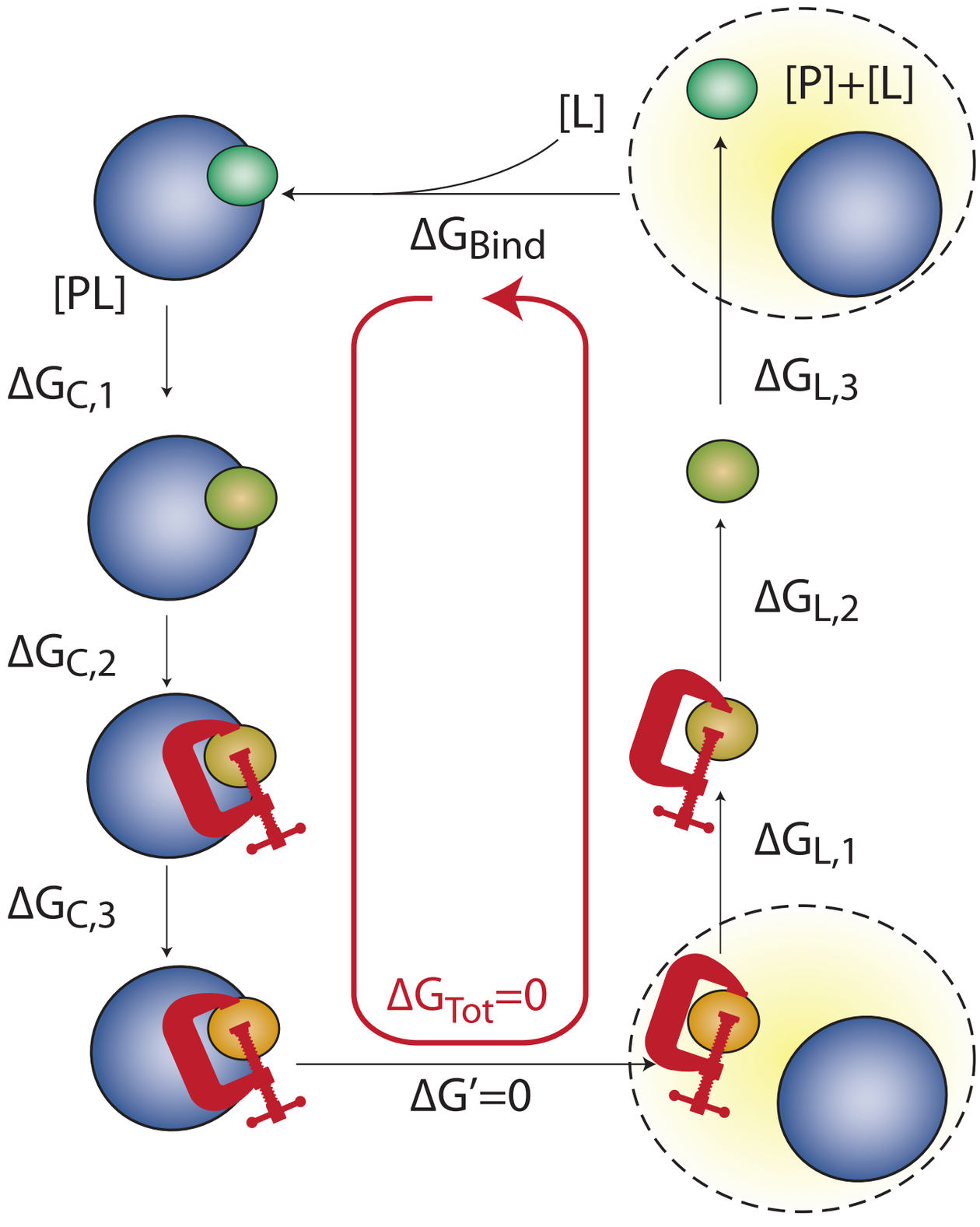
Free-energy **Perturbation** methods



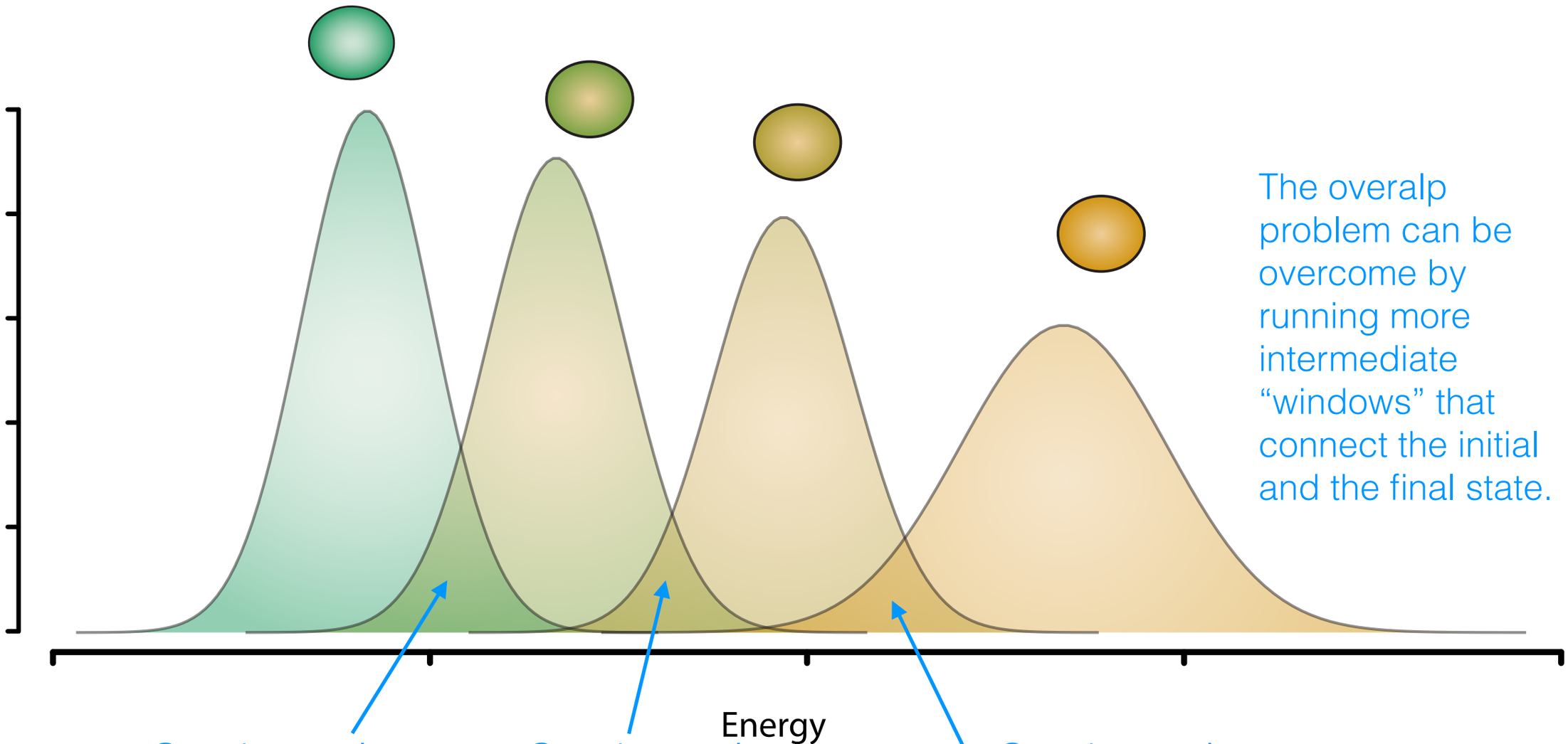
- Constraints

Free-energy **Perturbation** methods

- Constraints
- Decoupling schemes for electrostatics and VdW terms
- Soft-core potentials



Free-energy **Perturbation** methods



The overalp problem can be overcome by running more intermediate “windows” that connect the initial and the final state.

Good overalp allows to estimate $\Delta G_{C,1}$

Good overalp allows to estimate $\Delta G_{C,2}$

Good overalp allows to estimate $\Delta G_{C,3}$

$$\Delta G_C = \Delta G_{C,3} + \Delta G_{C,2} + \Delta G_{C,3}$$

Free-energy methods

Alchemical free energy methods can be used to compute either absolute binding affinities (for an individual ligand to a receptor) or relative binding affinities (a difference between two or more related ligands).

Absolute Free-energy

$$\Delta G_{L1} = G_{L1}(\text{bound}) - G_{L1}(\text{unb.})$$

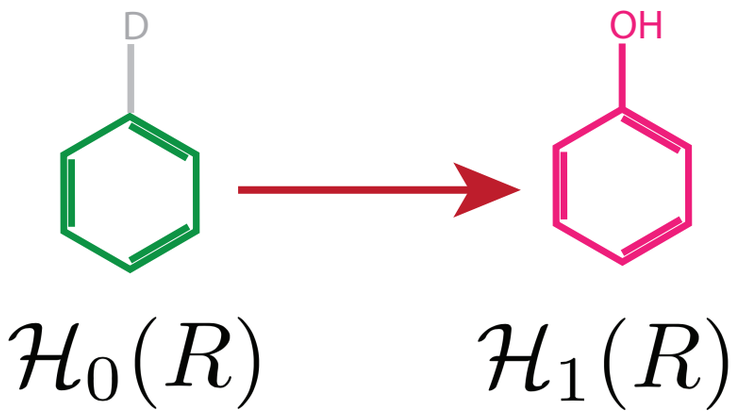
- **Directly comparable to experiments**
- absolute free energy calculations are more demanding, and require thorough sampling of the bound and unbound states of the molecule

Relative Free-energy

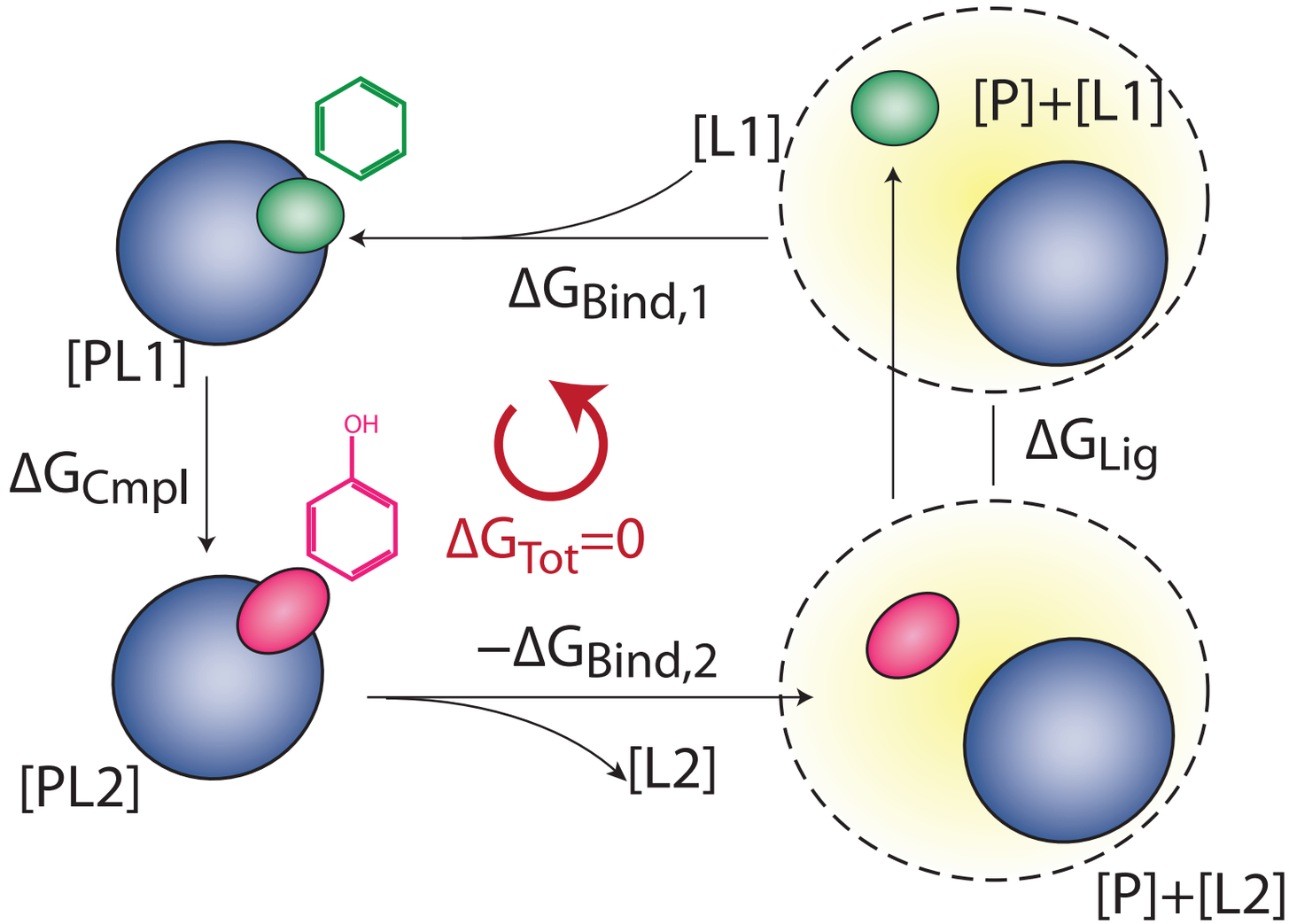
$$\Delta\Delta G = \Delta G_{L1} - \Delta G_{L2}$$

- **Common in lead optimization efforts**, where optimization through small, sequential chemical modification is of primary interest
- accurate relative free energies could determine whether modifications have increased affinity and selectivity.
- relative free energy calculations can be more efficient, requiring fewer alchemical states to bridge the phase space between two related molecules.

Relative Free-energy methods



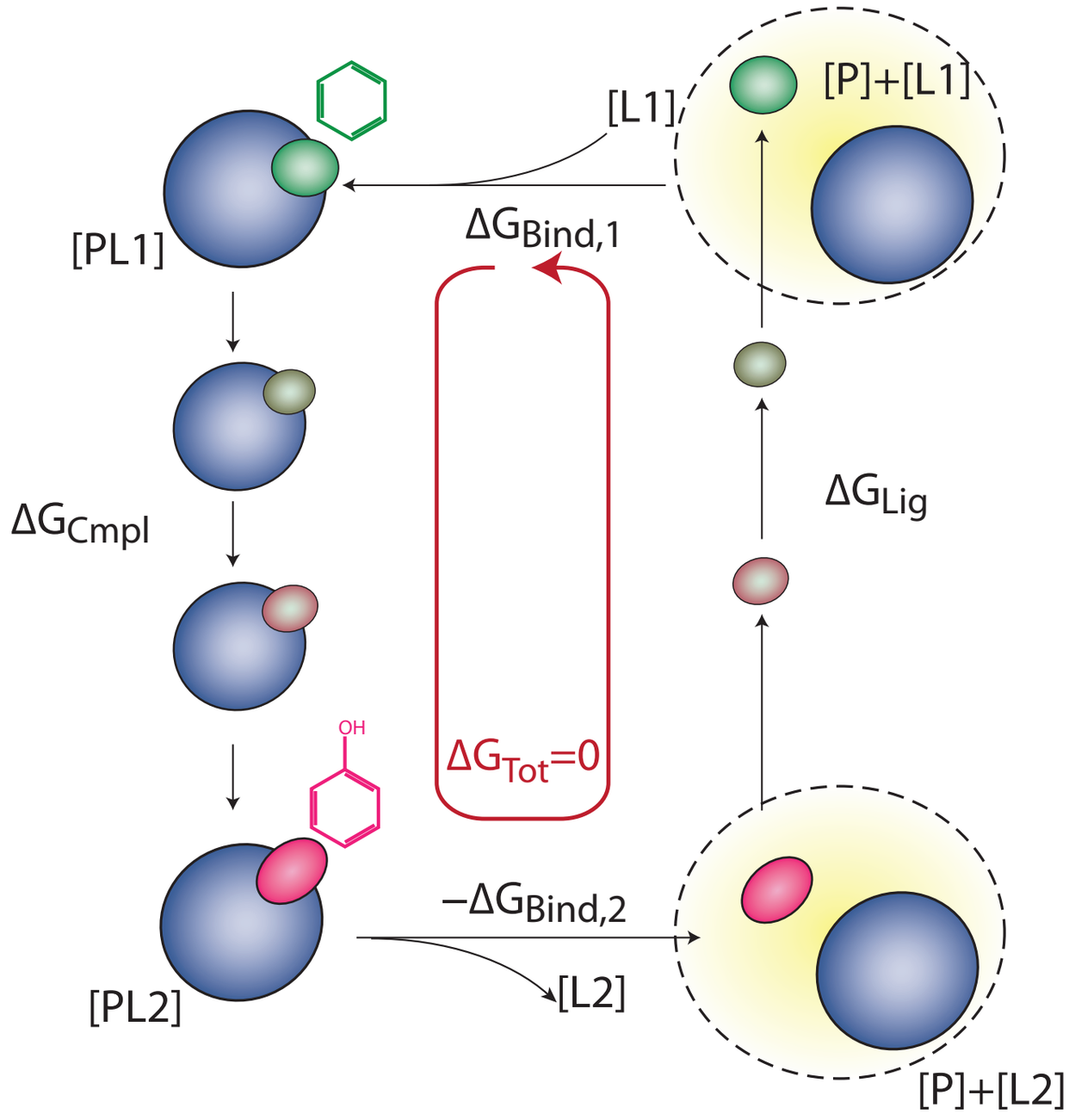
The same approach can be used to directly calculate the difference in binding of two ligands



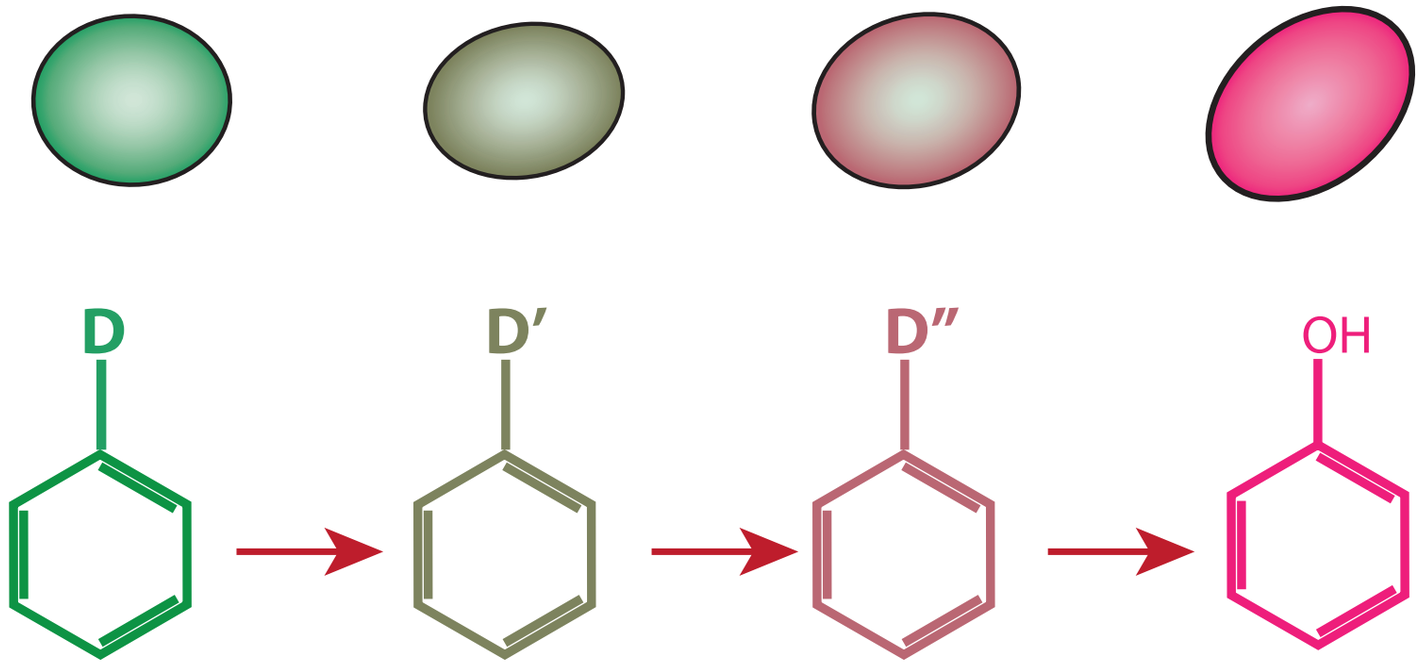
$$\Delta G_{Tot}=\Delta G_{Bind,1}+\Delta G_{Cmpl}-\Delta G_{Bind,2}-\Delta G_{Lig}=0$$

$$\boxed{\Delta G_{Bind,1}-\Delta G_{Bind,2}=\Delta G_{Lig}-\Delta G_{Cmpl}}$$

Relative Free-energy methods



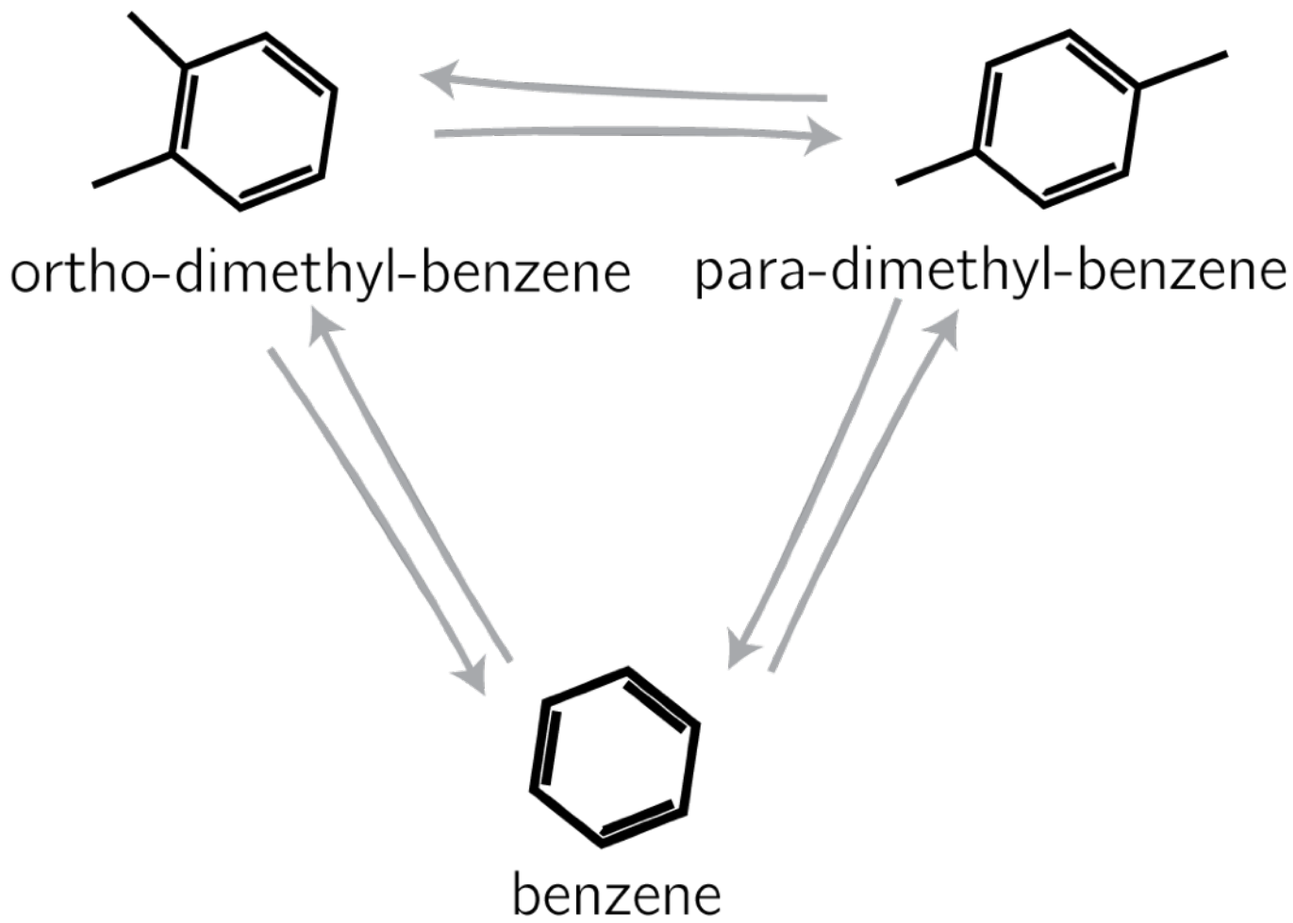
Relative Free-energy methods



$$\mathcal{H}_0 \rightarrow \mathcal{H}_{\lambda=0.33} \rightarrow \mathcal{H}_{\lambda=0.66} \rightarrow \mathcal{H}_1$$

In this case, the intermediate states are non-physical molecules that represent hybrid interactions that interpolate between the two extremes.

Tutorial



1. Setup gromacs input for calculation
2. Analysis of output

A few technical details

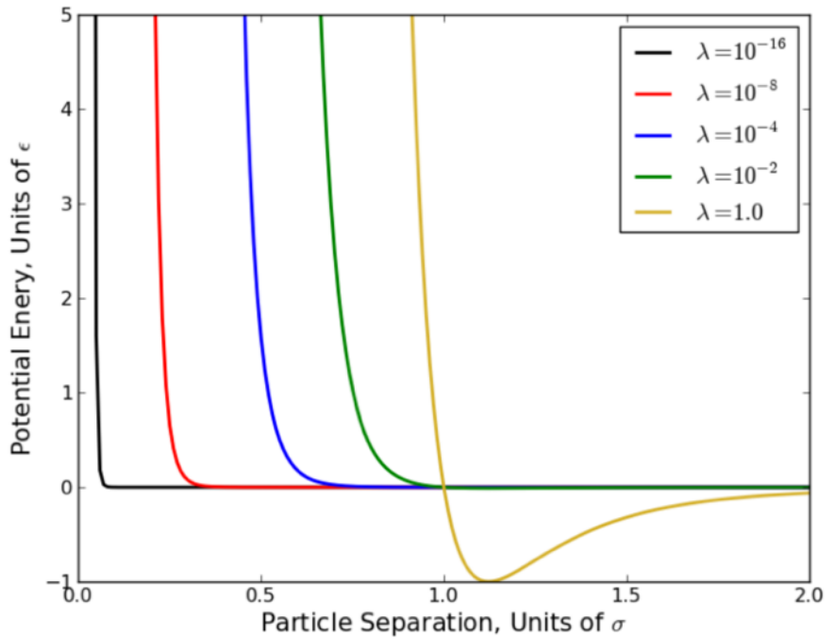
1. Alchemical paths
2. Soft-cores
3. Ways to calculate DG
4. Convergence and problem diagnostics
5. Series of ligands (error estimation)

A few technical details

Alchemical paths and soft-cores

Overlap is not constant as lambda changes, therefore a linear lambda distribution can be sub-optimal.

The LJ hard-core r^{-12}
has a divergent derivative; tven for very small lambda, the repulsive core influences strongly the sampling, leading to inefficient convergence.



Some guidelines

Bonded terms can be modified/turned off linearly. Changes in parameters can be done linearly.

Maximize similarity between states by removing/decoupling as few atoms as possible. Do not open and close rings.

Deleting or adding atoms should always be done with a soft core potential.

All charge on atoms must be turned off prior to atomic repulsion (otherwise you can get an infinite attractive potential and crash your simulation.)

Charge should be maintained across all λ (simply having charged molecules is fine, but the net of the system should remain constant. If you must change the net charge, there are complicated ways to do so.)

A few technical details

Algorithms to estimate ΔG

$$\Delta G = -k_B T \ln \langle \exp\left(\Delta \mathcal{H} / k_B T\right) \rangle_0$$

1. Exponential average (Zwanzig's equation)
2. Bennet Acceptance ratio
3. Multistate BAR

…

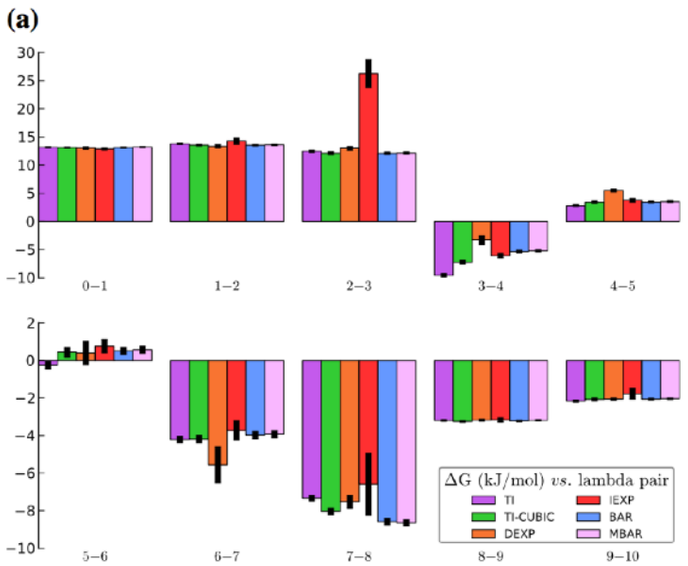
$$\Delta G = \int_0^1 \langle \frac{\partial \mathcal{H}}{\partial \lambda} \rangle_\lambda d\lambda$$

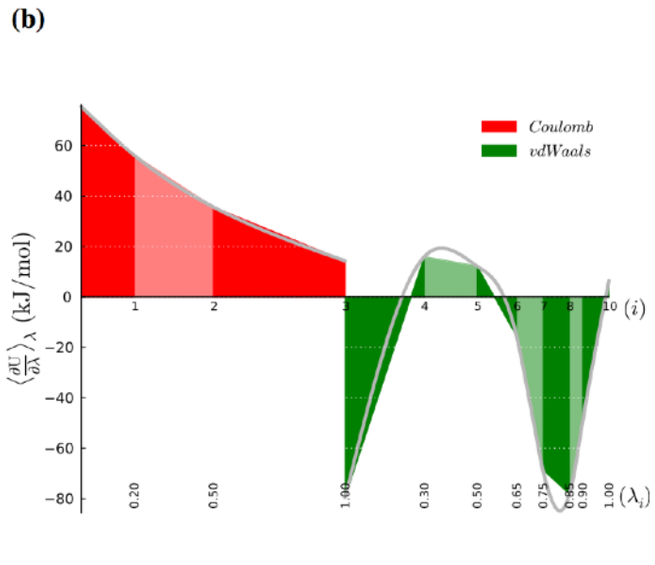
4. Thermodynamic integration (accuracy in TI is not a function of window overlap but of the curvature of the dU/dl function.)

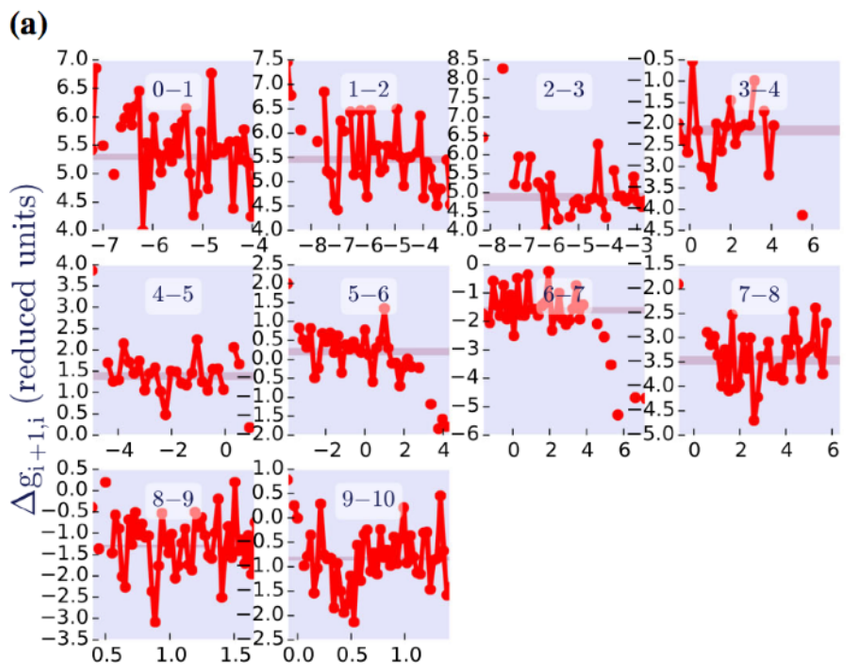
Cross-comparison of different methods can highlight problems.

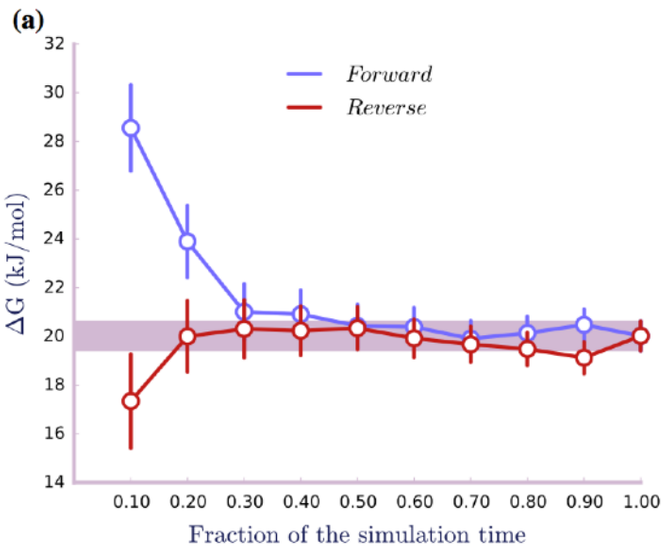
A few technical details

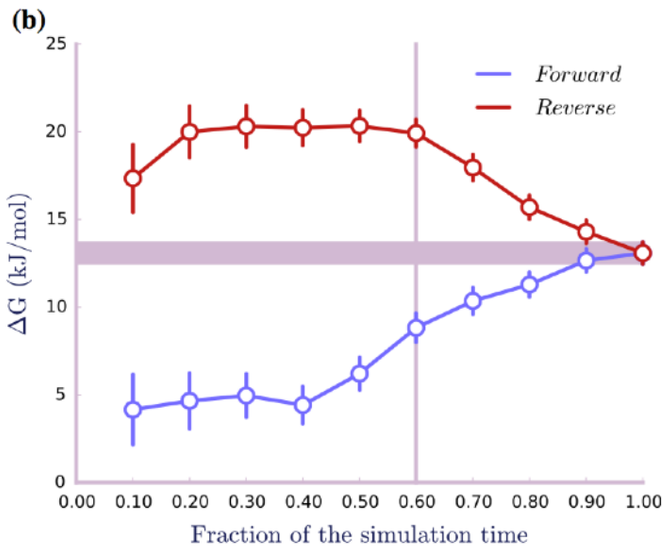
Convergence and problem diagnostics

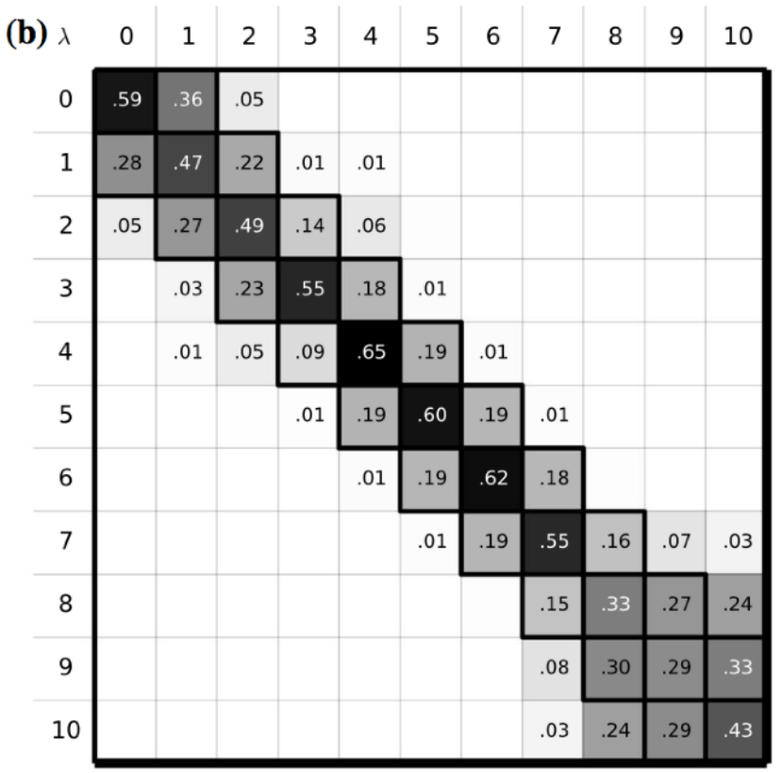




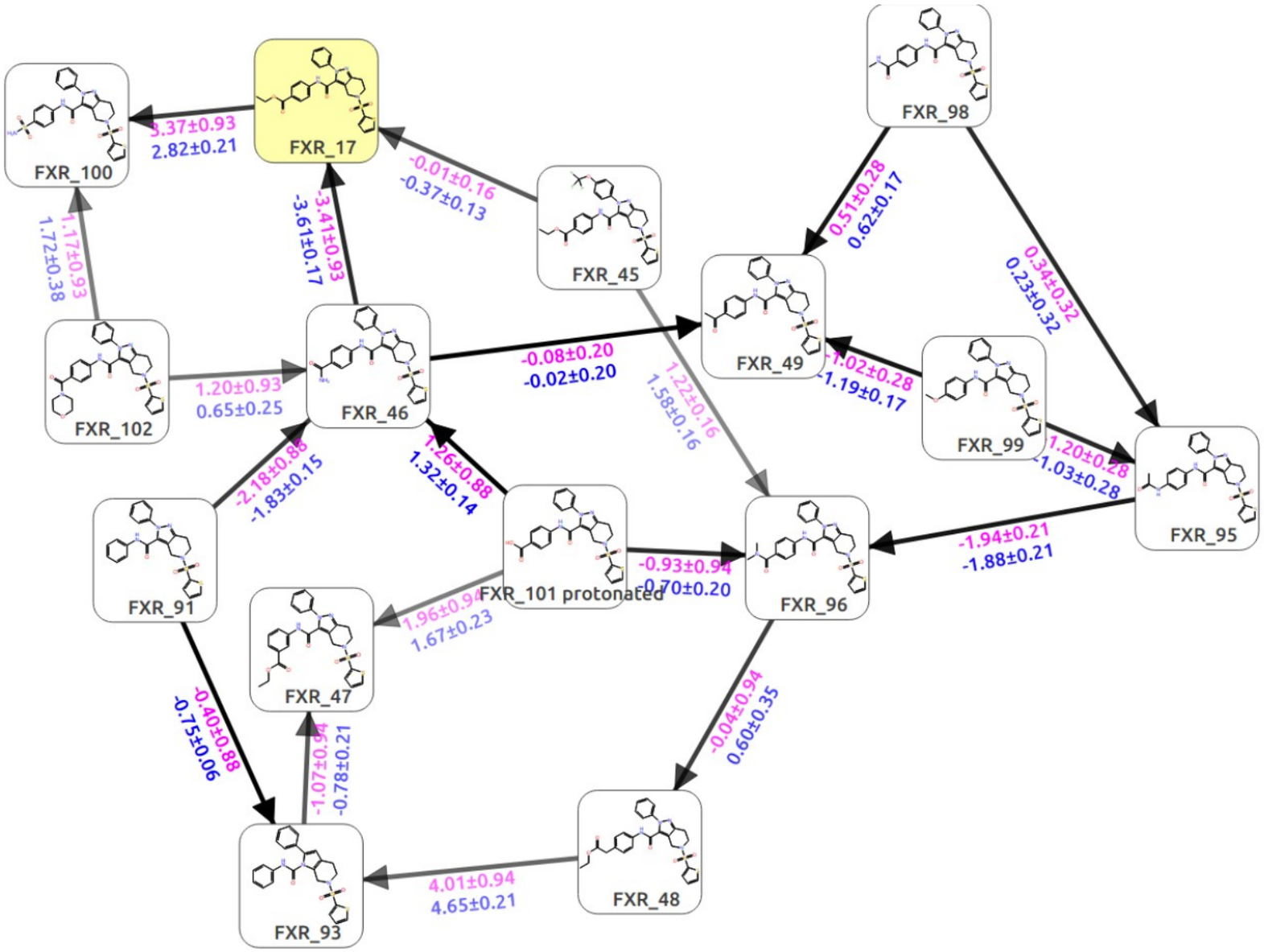








A few technical details



- Raw errors: predicted from bootstrapping and analytical error of free energy (e.g. BAR)
- Cycle closure errors: are calculated though the hysteresis of each cycle