Binding Free Energy Theory and Computation, and The role of conformational transitions.

Emilio Gallicchio

emilio@biomaps.rutgers.edu

http://hpcp.rutgers.edu/~emilio

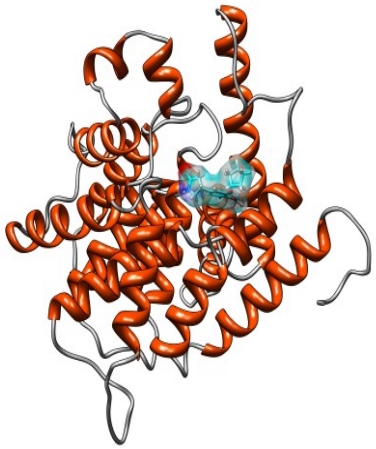
RUTGERS
THE STATE UNIVERSITY OF NEW JERSEY

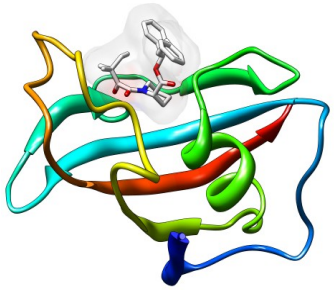
BioMaPS Institute
"Biology at the Interface with the Mathematical and Physical Sciences"

Physics-Based Binding Free Energy Methods

Many Challenges:

- Physiochemical modeling (protonation, etc.)
- Conformational sampling
- Force field accuracy
- Reproducibility





Can we at least agree on a common and well defined model of the thermodynamics of binding?

What do we mean with "Binding"?

Quasi-Chemical Description

$$A_{(sol)} + B_{(sol)} \rightleftarrows AB_{(sol)}$$

$$\mu_{sol,i} = \mu^{\circ}_{sol,i} + k\,T \ln \frac{C_i}{C^{\circ}}$$

$$\Delta G^{\circ}_{AB} = \mu^{\circ}_{AB} - \mu^{\circ}_{A} - \mu^{\circ}_{B} = -k\,T \ln \frac{C_{AB}/C^{\circ}}{(C_A/C^{\circ})(C_B/C^{\circ})} = -k\,T \ln K_{AB}$$

Any formulation of the standart binding free energy must:

- Include a definition of the species "AB"
- Depend on the standard state concentration.

One Possible Definition of "AB": The Indicator Function for the Complex

[Gilson et al., *Biophys. J.* (1997)]

$$K_{AB} = e^{-\beta \Delta G^{\circ}_{AB}} = \frac{e^{-\beta \mu^{\circ}_{AB}}}{e^{-\beta(\mu^{\circ}_{A} + \mu^{\circ}_{B})}} = \frac{C^{\circ}}{8\pi^2} \frac{Z_{AB}}{Z_A Z_B};$$

Z_{AB}, Z_A, Z_B : configurational partition functions in internal coordinates.

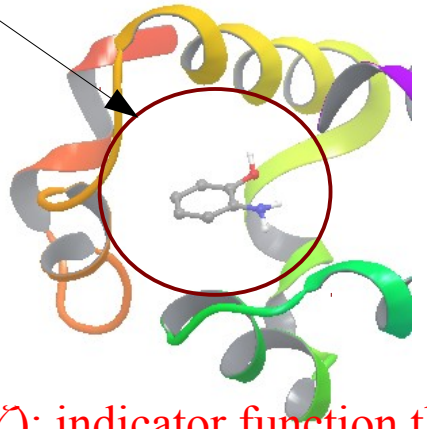
Indicator function: 1 if in pocket, 0 otherwise.

Receptor and ligand internal coordinates.

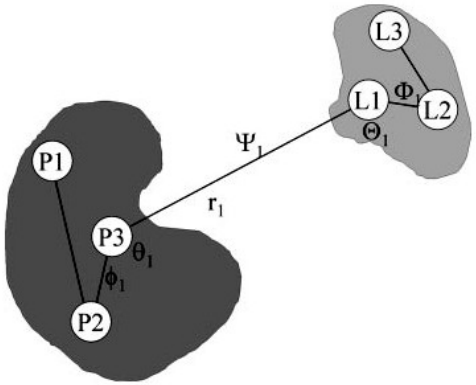
Effective potential (direct + solvent PMF)

$$Z_{AB} = \int d\zeta_B \, I(\zeta_B) \, dx_A \, dx_B \, e^{-\beta V(x_A, x_B, \zeta_B)}$$

External coordinates of ligand relative to receptor



$I(\zeta)$: indicator function that defines the complex

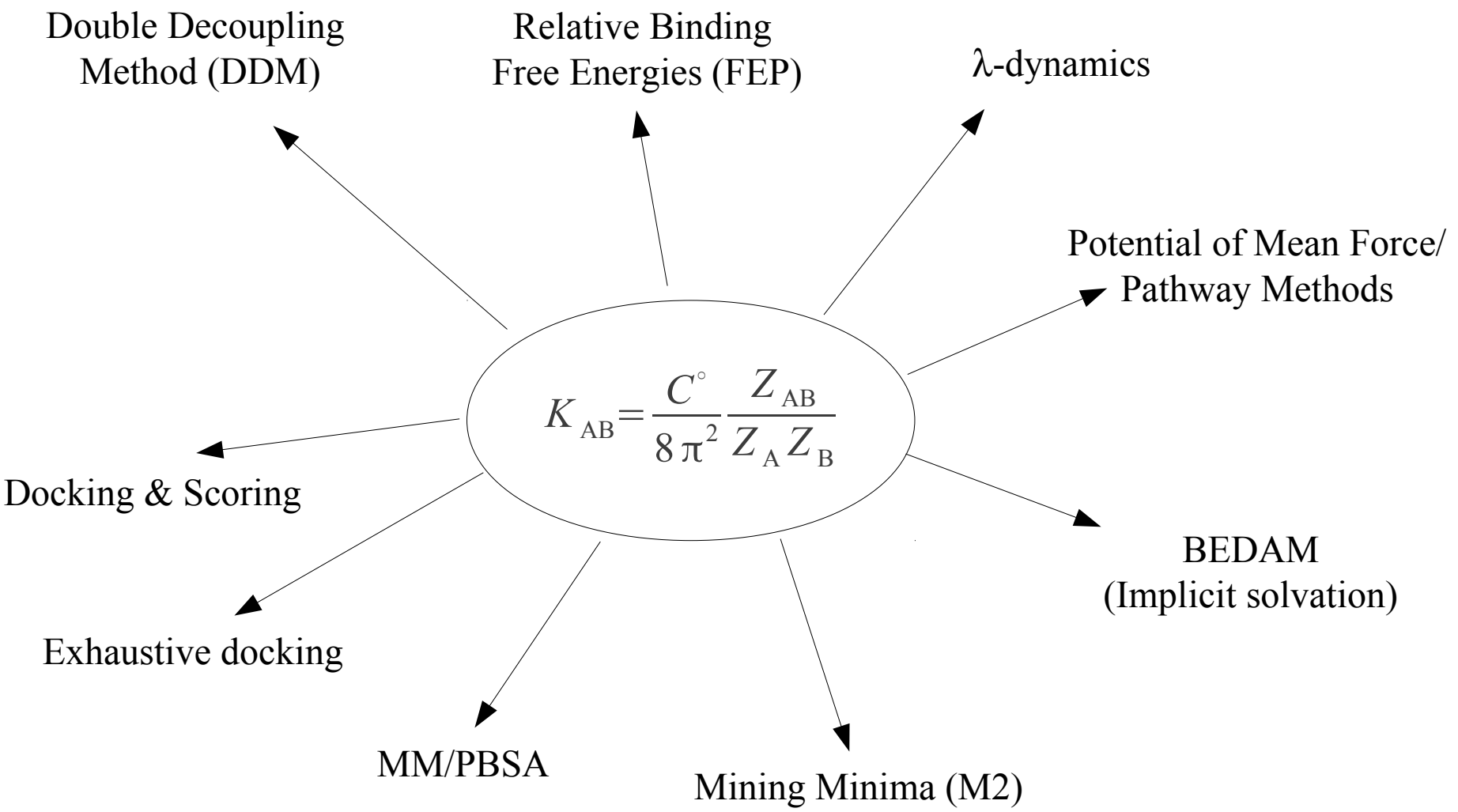


A choice for ζ_B

Boresch, Karplus, et al. JPCB (2003)

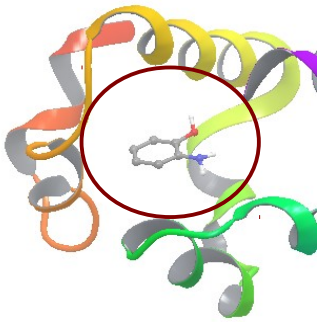
Unification

[Gallicchio et al., *Adv. Prot. Chem* (2012)]



Alchemical Implementation with Implicit solvation (BEDAM)

$$K_{AB} = \frac{C^{\circ}}{8\pi^2} \frac{Z_{AB}}{Z_A Z_B}$$



$$\int d\zeta_B I(\zeta_B) = V_{site}\Omega_{site}$$: effective volume of binding site

$$Z_{AB} = \int d\zeta_B I(\zeta_B) dx_A dx_B \exp[-\beta V(x_A, x_B, \zeta_B)]$$

$$Z_A Z_B = \int dx_A dx_B \exp[-\beta V(x_A)] \exp[-\beta V(x_B)] =$$

$$= \frac{1}{V_{site}\Omega_{site}} \int d\zeta_B I(\zeta_B) dx_A dx_B \exp[-\beta V(x_A)] \exp[-\beta V(x_B)]$$

$$u = u(x_A, x_B, \zeta_B) = V(x_A, x_B, \zeta_B) - [V(x_A) + V(x_B)]$$

Effective binding energy of a given conformation of the complex.

$$K_{AB} = \frac{V_{site}}{V^{\circ}} \frac{\Omega_{site}}{8\pi^2} \langle e^{-\beta u} \rangle_{0, V_{site}}$$

Average in uncoupled state within Vsite

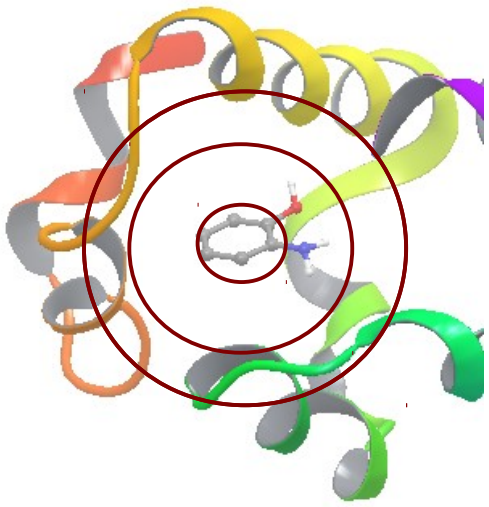
Suitable for numerical computation

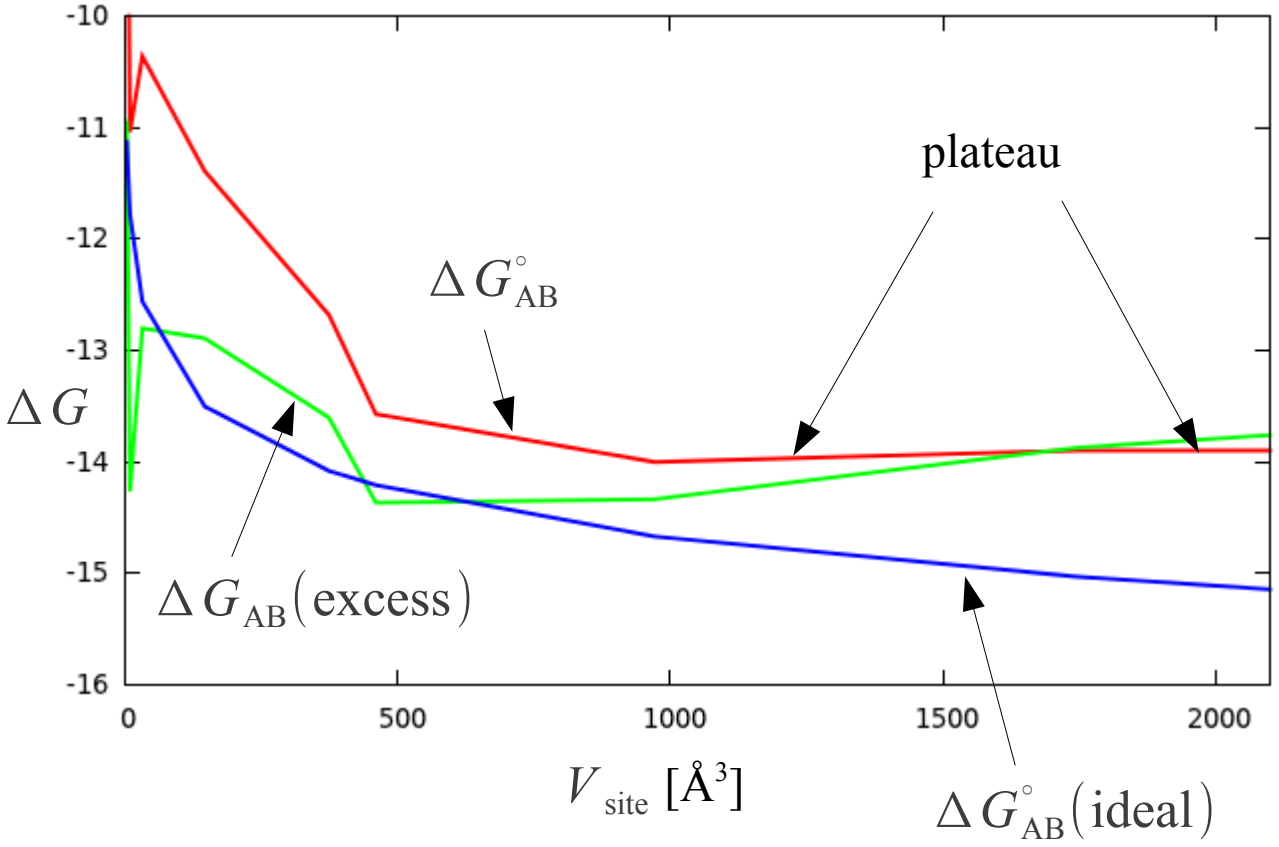
Effect of Varying Vsite

[Gallicchio et al., *JCTC* (2010)]

$$\left(\Omega_{\text{site}} = 8\pi^2\right)$$

$$\Delta G^{\circ}_{\text{AB}} = -kT \ln \frac{V_{\text{site}}}{V^{\circ}} - kT \ln \left\langle e^{-\beta u} \right\rangle_{0, V_{\text{site}}} = \Delta G^{\circ}_{\text{AB}}(\text{ideal}) + \Delta G_{\text{AB}}(\text{excess})$$

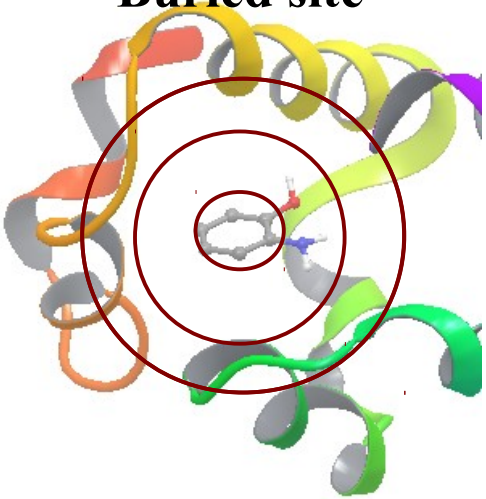




Asymptotic Scaling

$$\Delta G^{\circ}_{\mathrm{AB}} = -kT \ln \frac{V_{\mathrm{site}}}{V^{\circ}} - kT \ln \langle e^{-\beta u} \rangle_{0, V_{\mathrm{site}}} = \Delta G^{\circ}_{\mathrm{AB}}(\mathrm{ideal}) + \Delta G_{\mathrm{AB}}(\mathrm{excess})$$

Buried site



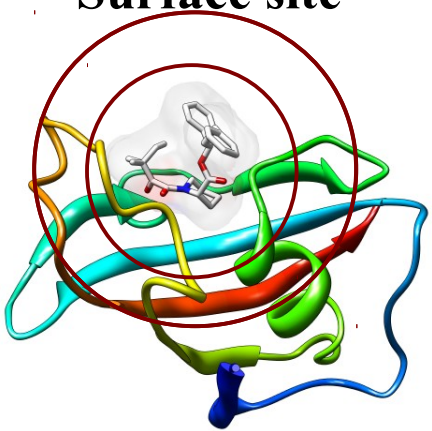
$e^{-\beta u} \simeq 0$ inside receptor

$\langle e^{-\beta u} \rangle_{0, V_{\mathrm{site}}} \sim \dfrac{1}{V_{\mathrm{site}}}$ for sufficiently large V_{site}

Cancels $\ln V_{\mathrm{site}}$ term

$\Delta G^{\circ}_{\mathrm{AB}} \rightarrow$ constant

Surface site



$e^{-\beta u} \simeq 1$ outside receptor

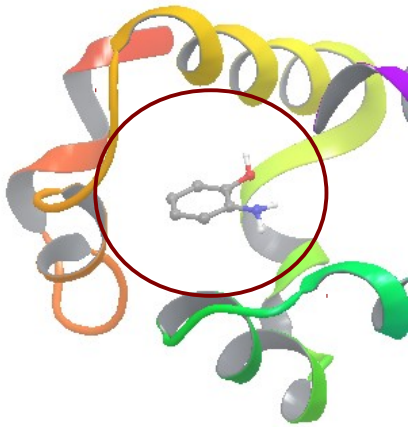
$\Delta G^{\circ}_{\mathrm{AB}}(\mathrm{excess}) \rightarrow 0$ for $V_{\mathrm{site}} \rightarrow \infty$

$$\Delta G^{\circ}_{\mathrm{AB}} \rightarrow \ln \frac{1}{V_{\mathrm{site}}} \rightarrow -\infty \quad \text{as } V_{\mathrm{site}} \rightarrow \infty$$

Basic Facts about Vsite

$$\Delta G^{\circ}_{AB} = -kT \ln \frac{V_{site}}{V^{\circ}} - kT \ln \langle e^{-\beta u} \rangle_{0, V_{site}} = \Delta G^{\circ}_{AB}(\text{ideal}) + \Delta G_{AB}(\text{excess})$$

- Ligand con not “escape” from the binding site.
- It is not meaningful to report the binding free energy without defining the binding site region
- The simulation box is probably not a good definition
- The region spanned in the simulation is probably not a good definition
- Ideal term should be always combined with the excess term
- Vsite is not a auxiliary restraint: it's part of the system definition, not a computational device.
- Any binding free energy value can be reproduced with the “right” Vsite



"What Vsite should I Choose?"

[Gallicchio et al., *Adv. Prot. Chem* (2012)]

Answer #1: "Whatever you like"

- Binding region definition should include all highly populated conformations of the complex
- So err in excess; after all the binding free energy decreases slowly with increasing Vsite

Answer #2: "Ask our collaborator"

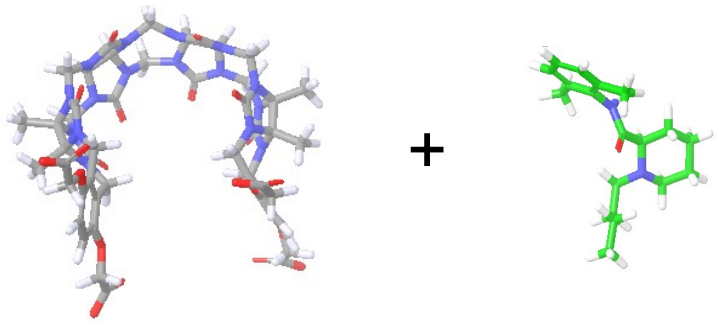
- So at least we can compare apples with apples

Answer #3: "Look at the experimental paper: do your best"

- Binding region definition should include all "important" conformations of the complex corresponding to the experimental reporter signal
- EC_{50}/IC_{50}, NMR, ITC, SPR?
- [Mihailescu & Gilson. Biophys J. (2004); R. D. Groot, JCP (1992)]

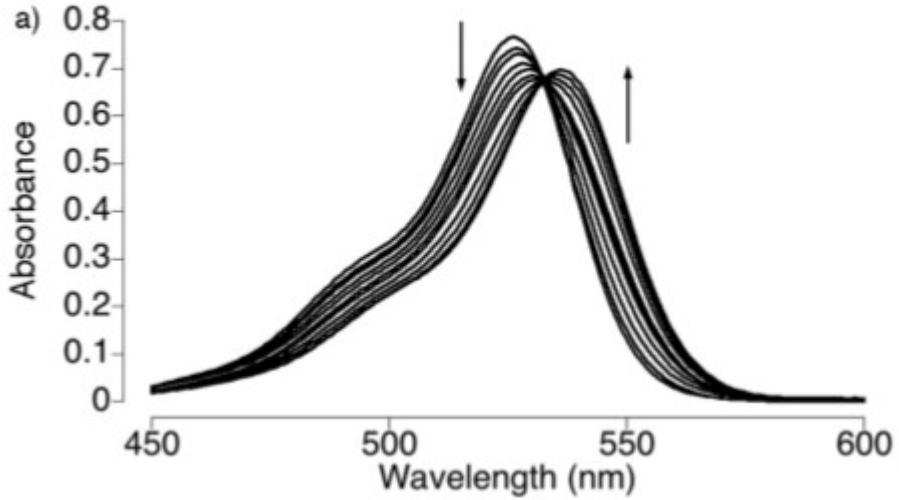
Best Case Example

[Ma, Zavalij, Isaacs. *JOC* (2010); Gallicchio, Levy. JCAM (2012)]

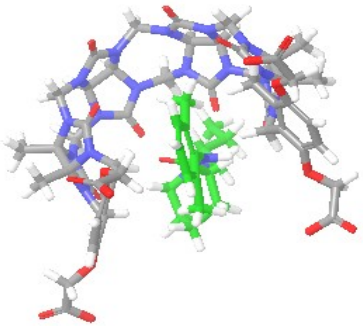


+

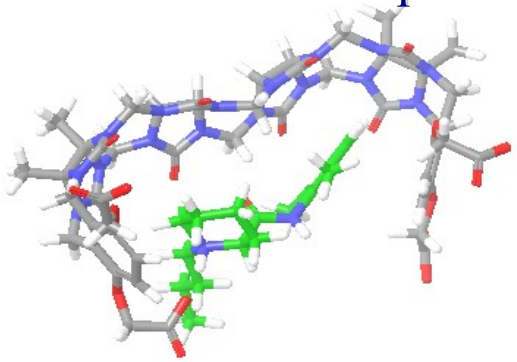
UV/Vis Titration



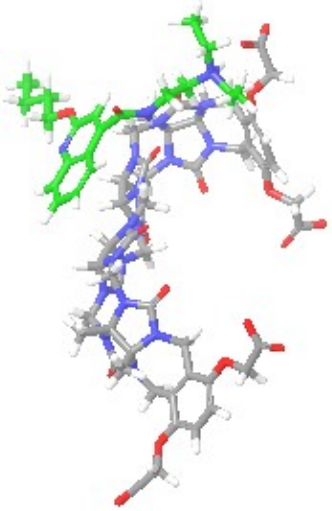
This is a complex



Is this is a complex?



How about this one?



Efficient conformational sampling techniques for alchemical binding free energy calculations

- Quick convergence with minimal prior structural knowledge
- General applicability and high degree of automation

Convergence and Conformational Variability

Convergence is not a problem when sampling is constrained within one binding mode/conformational state.

However:

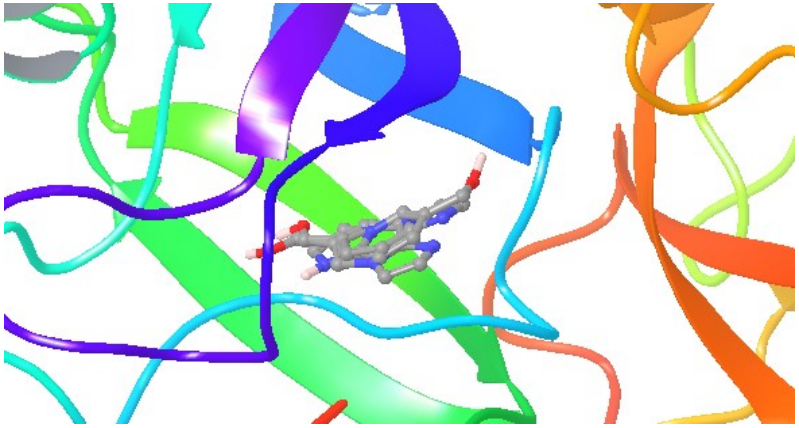
- Multiple binding modes can contribute to binding
- Physiological binding modes can be uncertain
- Relative FE's: possible changes in binding mode upon transformation

To achieve reliable and unsupervised protocols we should avoid restraints as much as possible

Example: The SAMPL3 Challenges' Systems

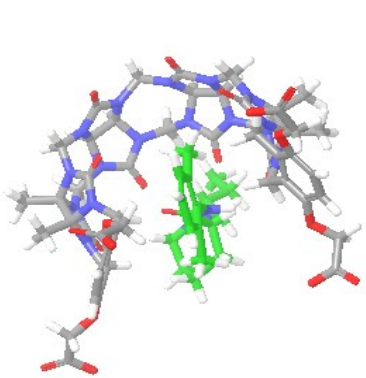
[Gallicchio, Levy. J. Comp. Aid. Mol. Design (2012)]

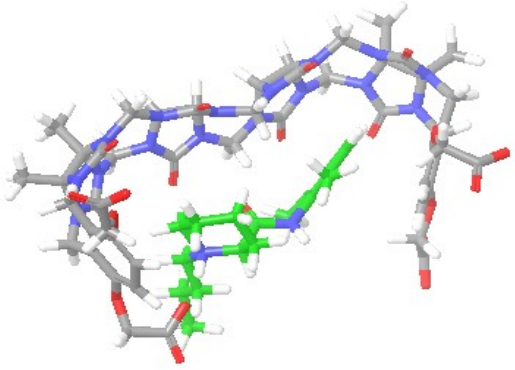
34 molecular fragments binding to trypsin



Location of binding site?
Multiple potential orientations

7 Host-Guests: extensive conformational variability

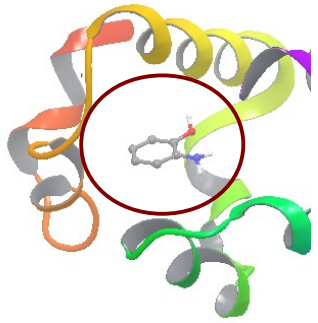




Very extensive conformational variability

The Binding Energy Distribution Analysis Method (BEDAM)

Implicit Solvation (OPLS/AGBNP2)



Hybrid potential: $V_\lambda(x) = V_0(x) + \lambda u(x)$

Effective Potential at uncoupled state (both ligand and receptor interact only with solvent continuum)

Binding energy = perturbation (effective potential energy change for displacing ligand from solution to receptor site)

λ=0: uncoupled state

λ=1: coupled state

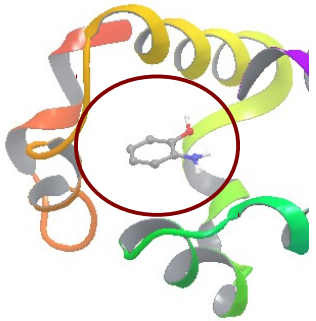
Direct transfer from implicit solvent environment to the complex.

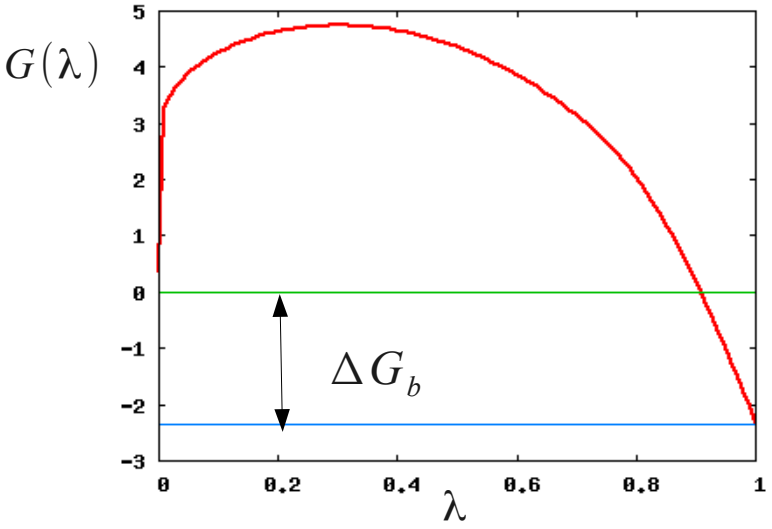
(one simulation leg rather than two as with explicit solvation)

Gallicchio, Lapelosa, Levy, JCTC (2010); Gallicchio & Levy, Curr. Op. Struct. Biol. (2011); Gallicchio & Levy, Adv. Prot. Chem. (2011); Lapelosa, Gallicchio, Levy, JCTC (2012). Gallicchio, Levy J. Comp. Aid. Mol. Design (2012); Gallicchio, Mol. Biosc (2012).

The Binding Energy Distribution Analysis Method (BEDAM)

1. Run simulations at several intermediate λ states (stratification)
2. λ-hopping Hamiltonian Parallel Replica Exchange
3. Collect binding energies (*u*'s)
4. MBAR analysis to get *G*(λ)



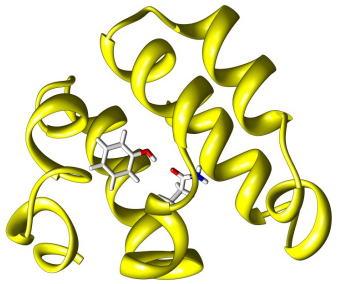


$$\Delta G_b^\circ = -kT\ln\left(C^\circ V_{\text{site}}\right) - kT\left\langle e^{-\beta u}\right\rangle_{0,V_{\text{site}}}$$

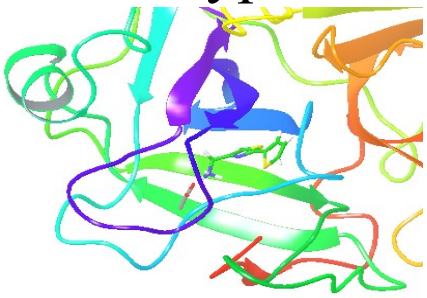
Gallicchio, Lapelosa, Levy, JCTC (2010); Gallicchio & Levy, Curr. Op. Struct. Biol. (2011); Gallicchio & Levy, Adv. Prot. Chem. (2011); Lapelosa, Gallicchio, Levy, JCTC (2012). Gallicchio, Levy J. Comp. Aid. Mol. Design (2012); Gallicchio, Mol. Biosc (2012).

Applications

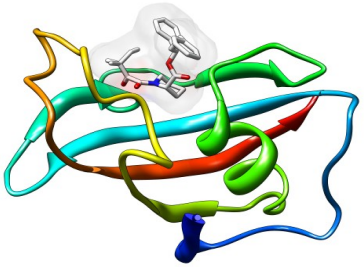
Lysozyme



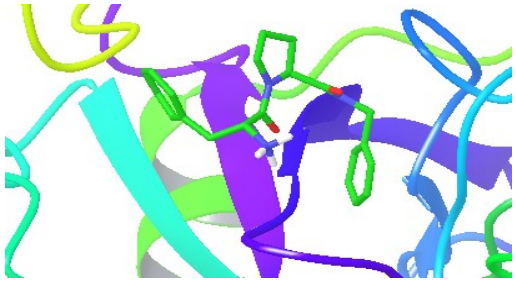
Trypsin



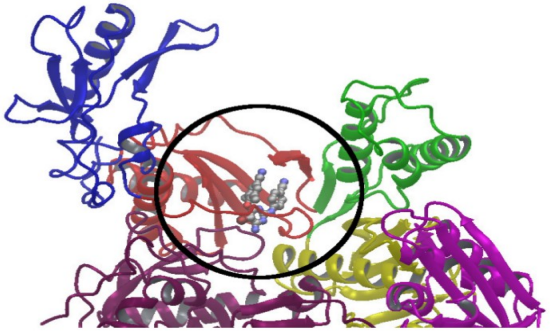
FKBP



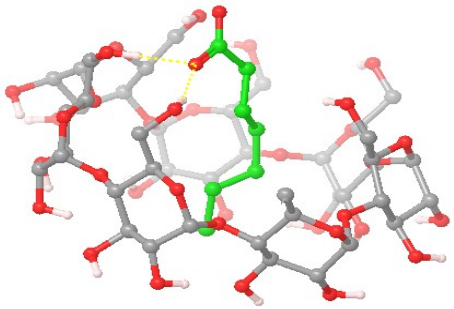
Thrombin



HIV-RT



Host/Guest



- Good agreement with experimental affinities
- Model captures main physical effects of binding
- Suitable for studying free energy methodologies and conformational sampling

Gallicchio, Lapelosa, Levy, JCTC (2010); Gallicchio & Levy, Curr. Op. Struct. Biol. (2011); Gallicchio & Levy, Adv. Prot. Chem. (2011); Lapelosa, Gallicchio, Levy, JCTC (2012). Gallicchio, Levy J. Comp. Aid. Mol. Design (2012); Gallicchio, Mol. Biosc (2012).

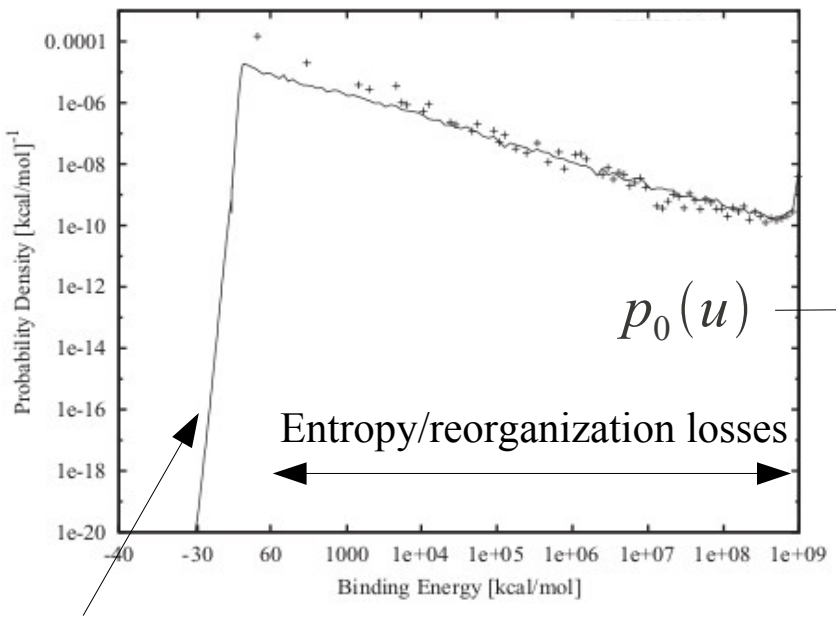
Binding Energy Distributions Perspective

$p_\lambda(u)$: binding energy probability distribution at λ

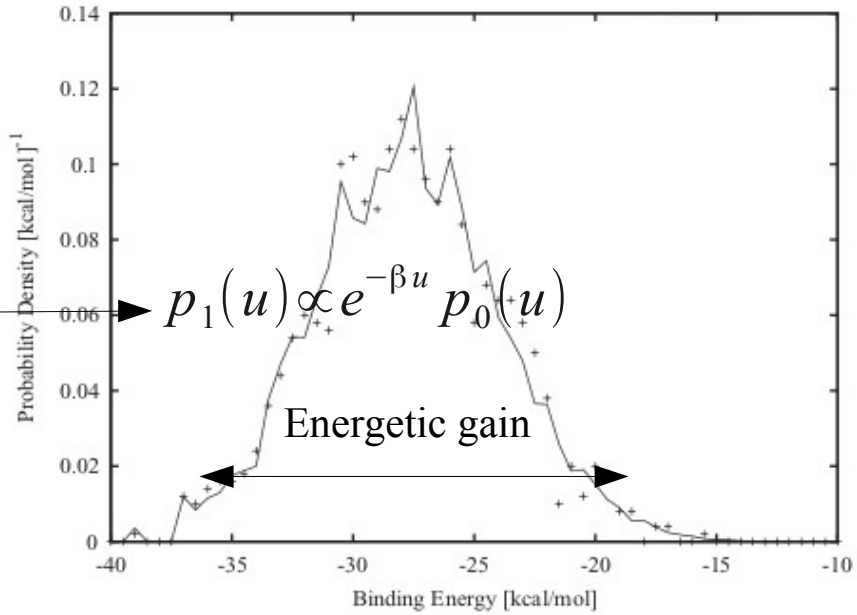
$$e^{-\beta \Delta G^\circ_{AB}} = K_{AB} = C^\circ V_{site} \int du\, e^{-\beta u}\, p_0(u) \qquad \text{(PDT)}$$

$p_0(u)$ (uncoupled state) completely encodes binding free energy

λ = 0 (uncoupled)

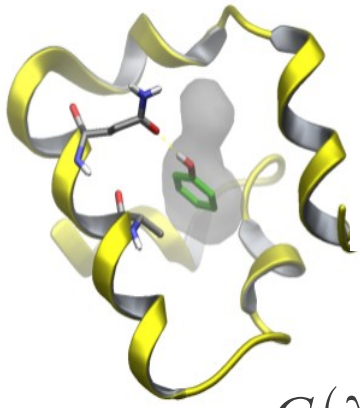


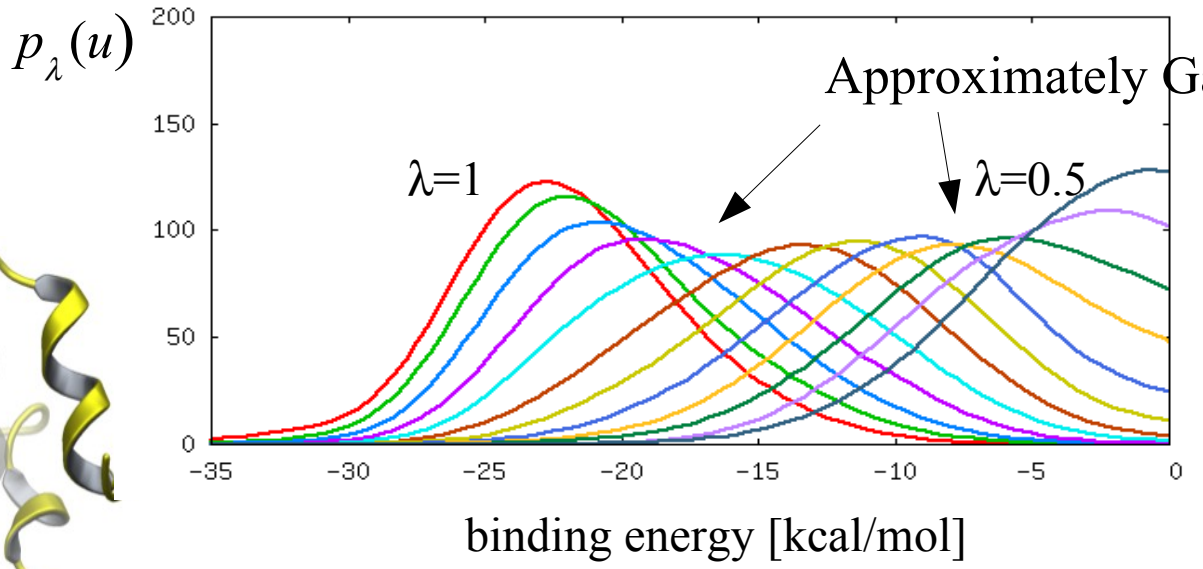
λ = 1 (coupled)

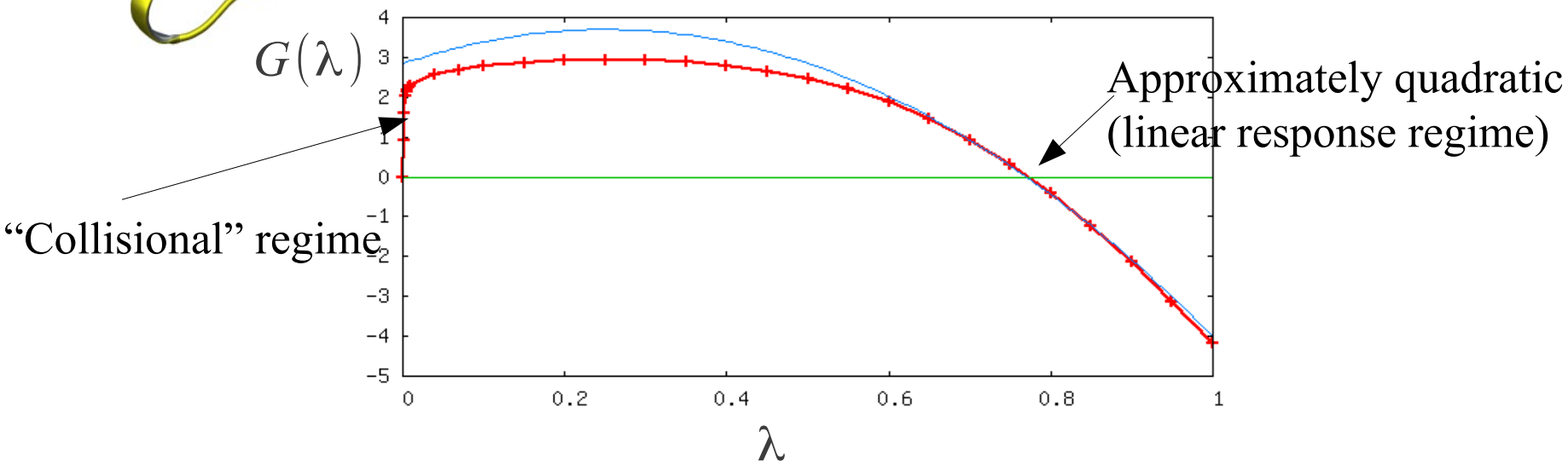


Magnitude and shape of $p_0(u)$ low energy tail is critical.

Fast-Converging Systems Follow Linear Response

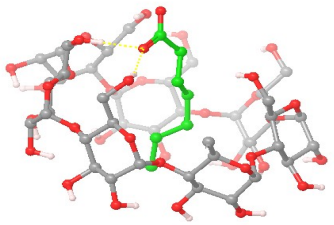






Slow-Converging Systems often have Complex Binding Energy Distributions

[Wickstrom et al., in preparation]

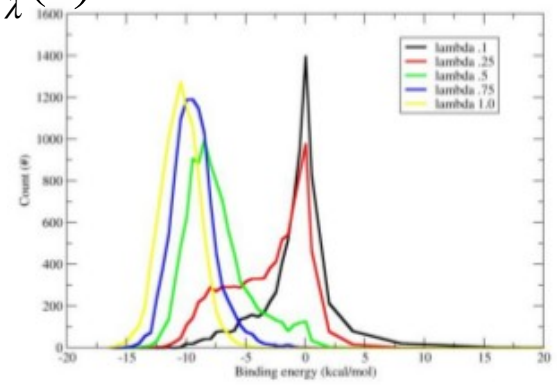


96 β-cyclodextrin complexes

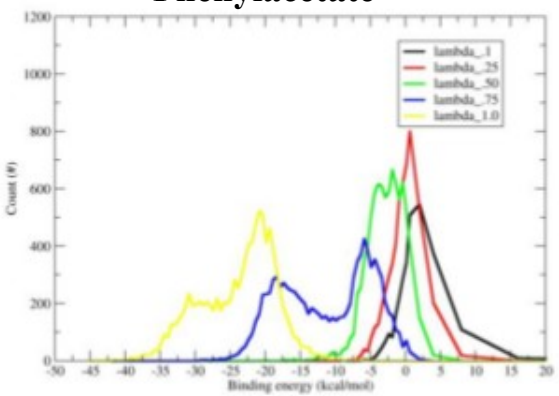
It is necessary to convergence of $p_\lambda(u)$ at all λ's

$p_\lambda(u)$

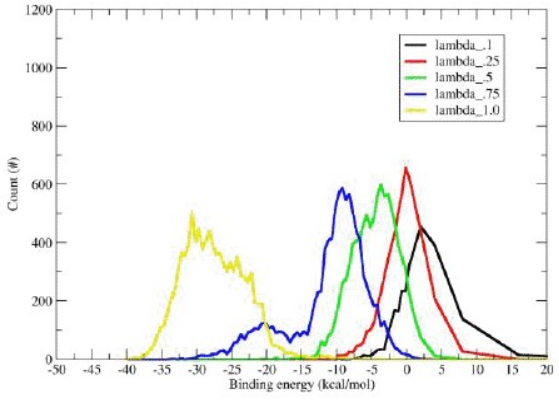
Benzene



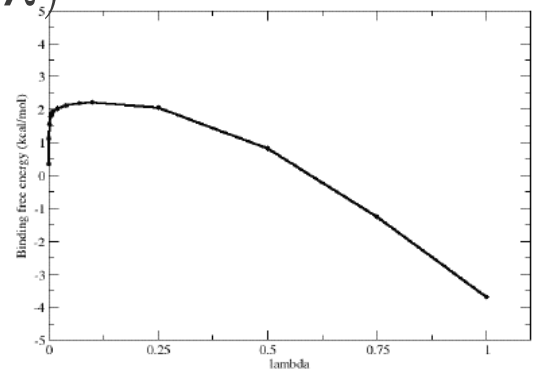
Phenylacetate

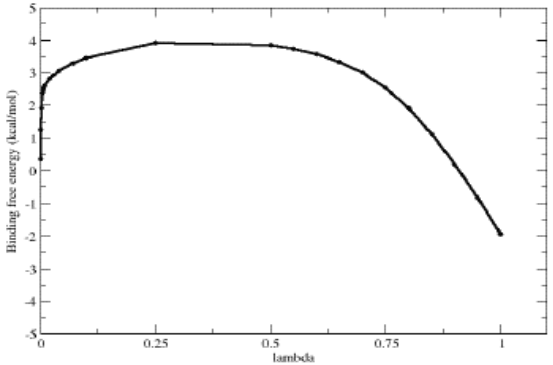


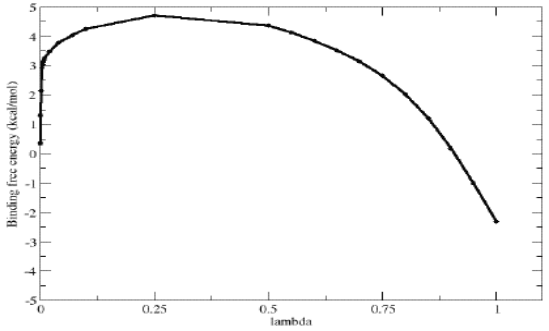
Nt-Boc-L-alanine



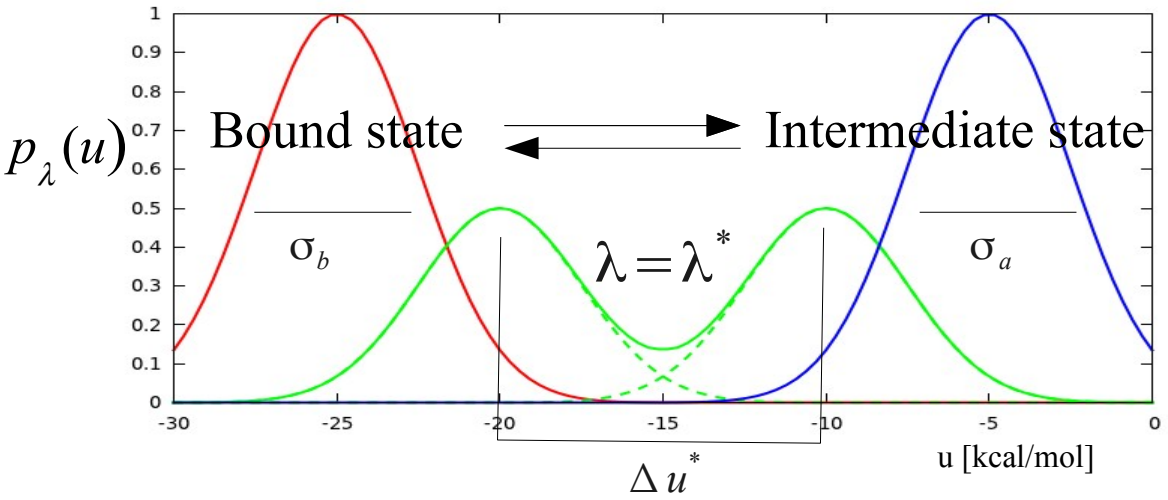
$G(\lambda)$







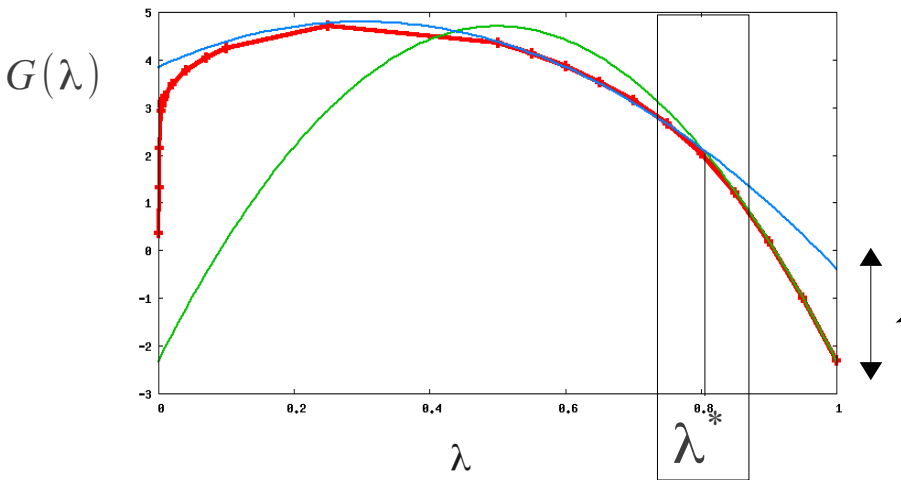
Two-States Linear Response
Phase Transition Model



At the transition $\lambda=\lambda^*$ equal populations of bound and intermediate states

Difficult to pinpoint λ^*

$$G(\lambda)=G(\lambda^*)-kT\ln\left[\frac{1}{2}e^{\beta^2\sigma_a^2(\lambda-\lambda^*)^2/2-\beta u_a^*(\lambda-\lambda^*)}+\frac{1}{2}e^{\beta^2\sigma_b^2(\lambda-\lambda^*)^2/2-\beta u_b^*(\lambda-\lambda^*)}\right]$$

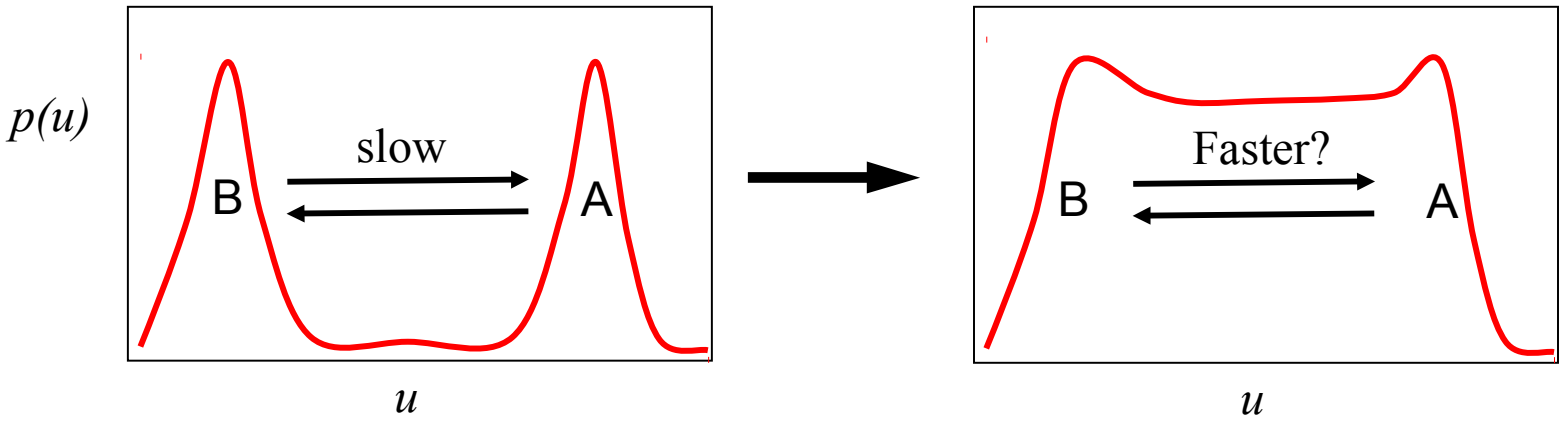


An error of $\delta\lambda^*$ in transition point leads to an error in the binding free energy of:

$$\delta\Delta G(\lambda)=\delta\lambda^*\Delta u^*$$

The stronger the transition the slower the convergence

Biased Sampling to "Tunnel" through Phase Transitions



It's hard to converge relative population of A and B states

Orthogonal Space Random Walk approach?

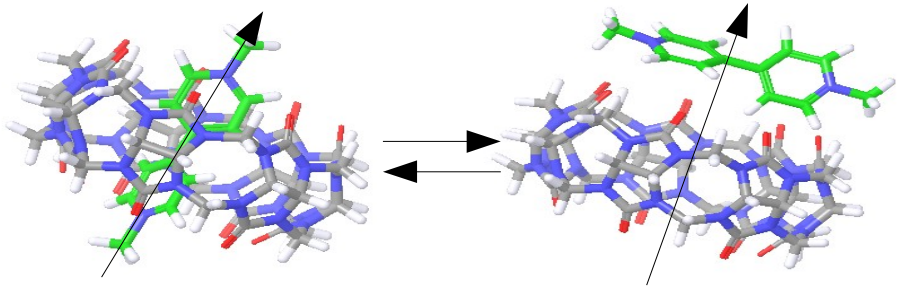
Bias sampling along $\frac{\partial U}{\partial \lambda} = u$ (binding energy)

Biasing potentials: $\omega_\lambda(u) = \frac{k_\lambda}{2}\left(u - u_\lambda^*\right)^2$

Zheng, Chen, Yang. PNAS 105:20227 (2008)
Kim, Straub. JCP 133:154101 (2010)

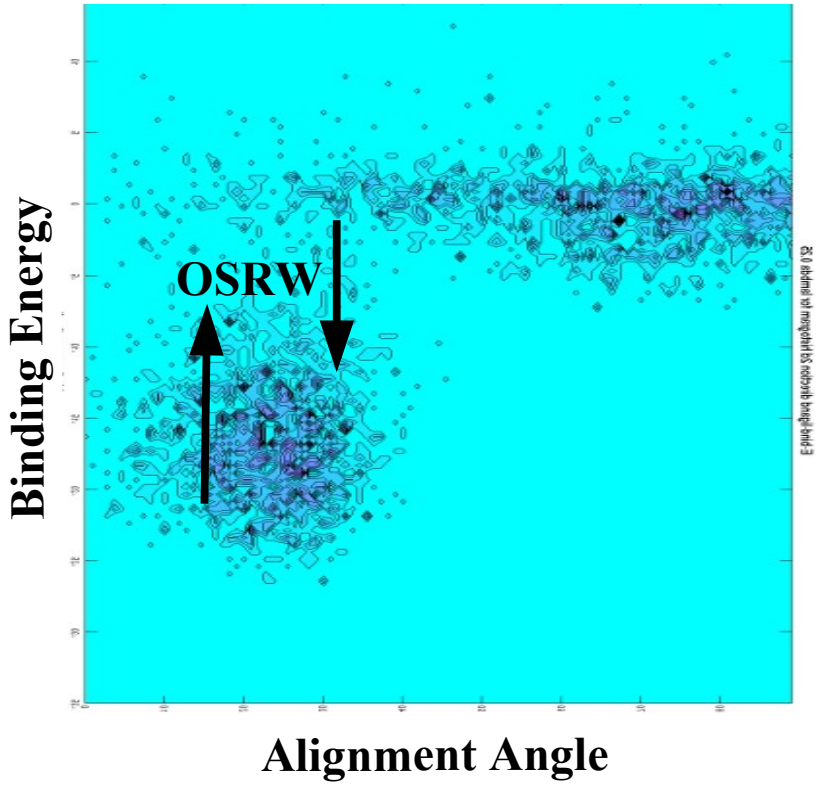
Promising Results for Order-Disorder Transitions

Bound



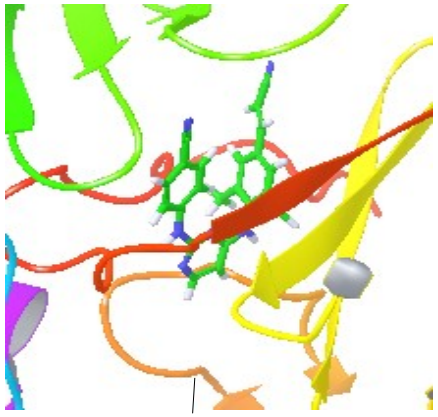
Unbound

Distribution of alignment angle vs. binding energy



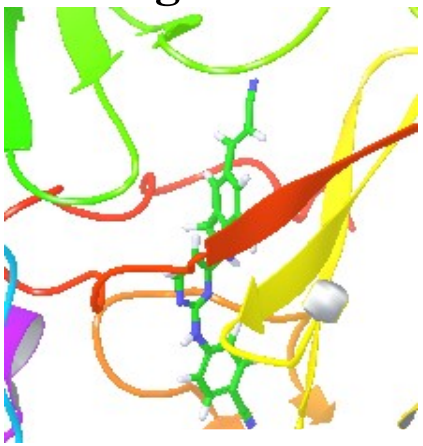
HIVRT/TMC278: Slow Conformational Equilibration

Bound



slow

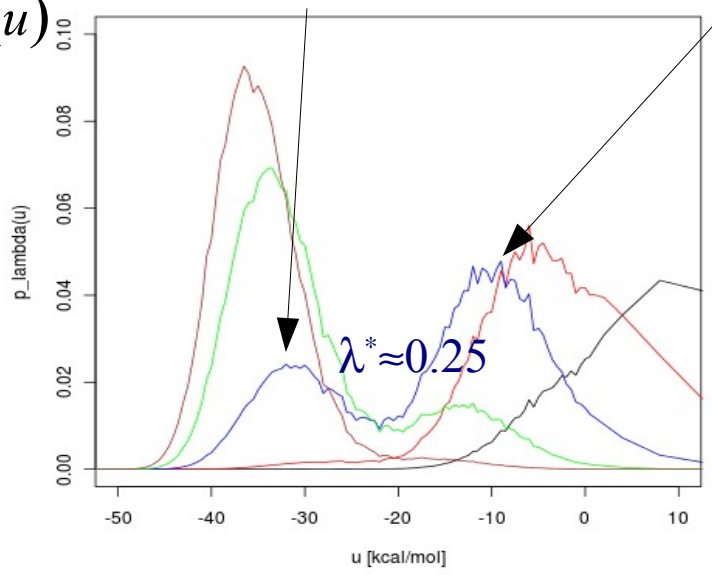
Binding Intermediate



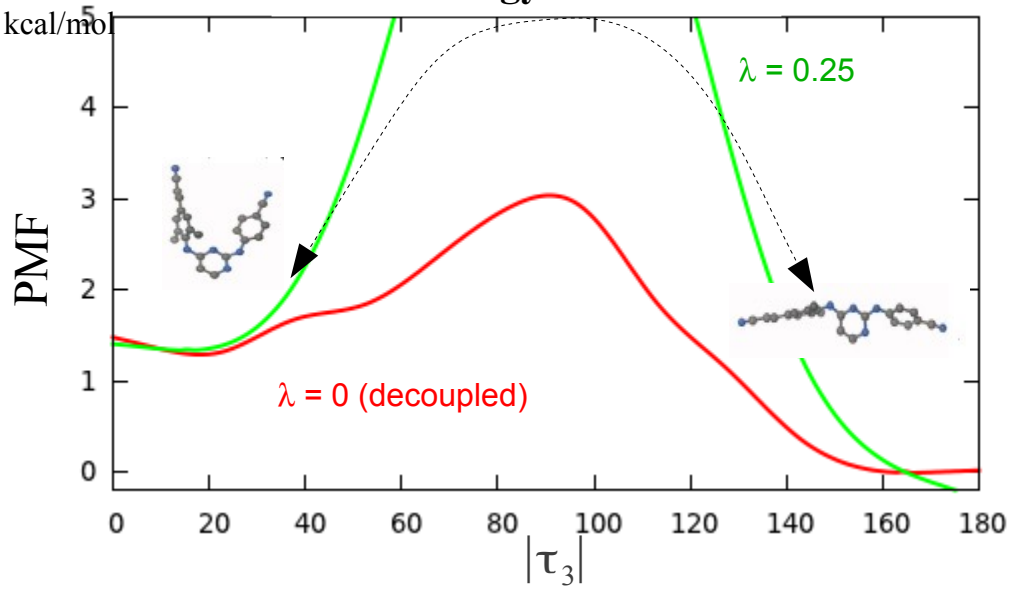
Conformational rearrangement

Bimodal distributions slow equilibration

$p_\lambda(u)$

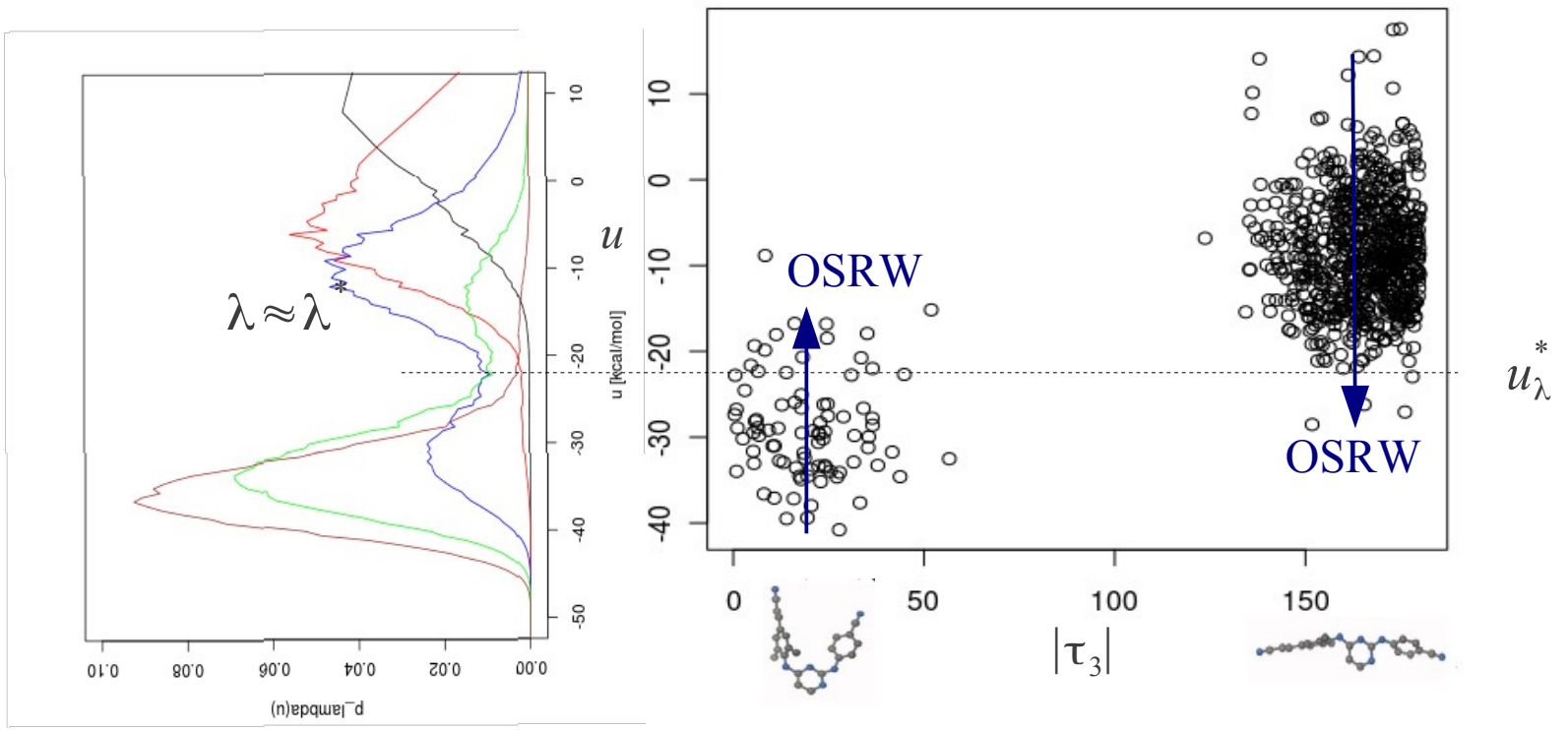


Large Potential Energy Barrier



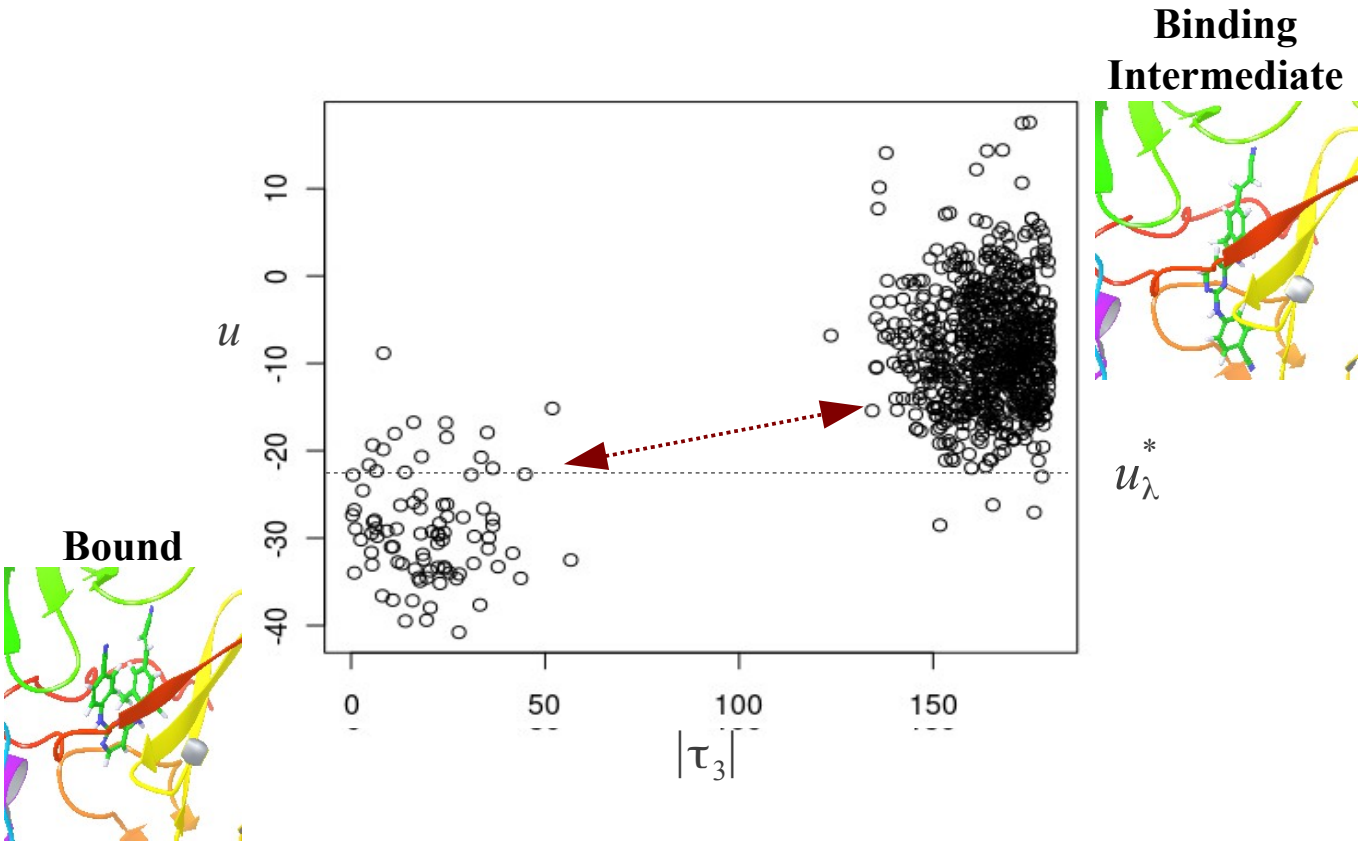
HIVRT/TMC278: OSRW Test

Unsurprisingly, OSRW doesn't help in this case.



Biasing in the *u*-direction does not enhance transitions along conformational direction.

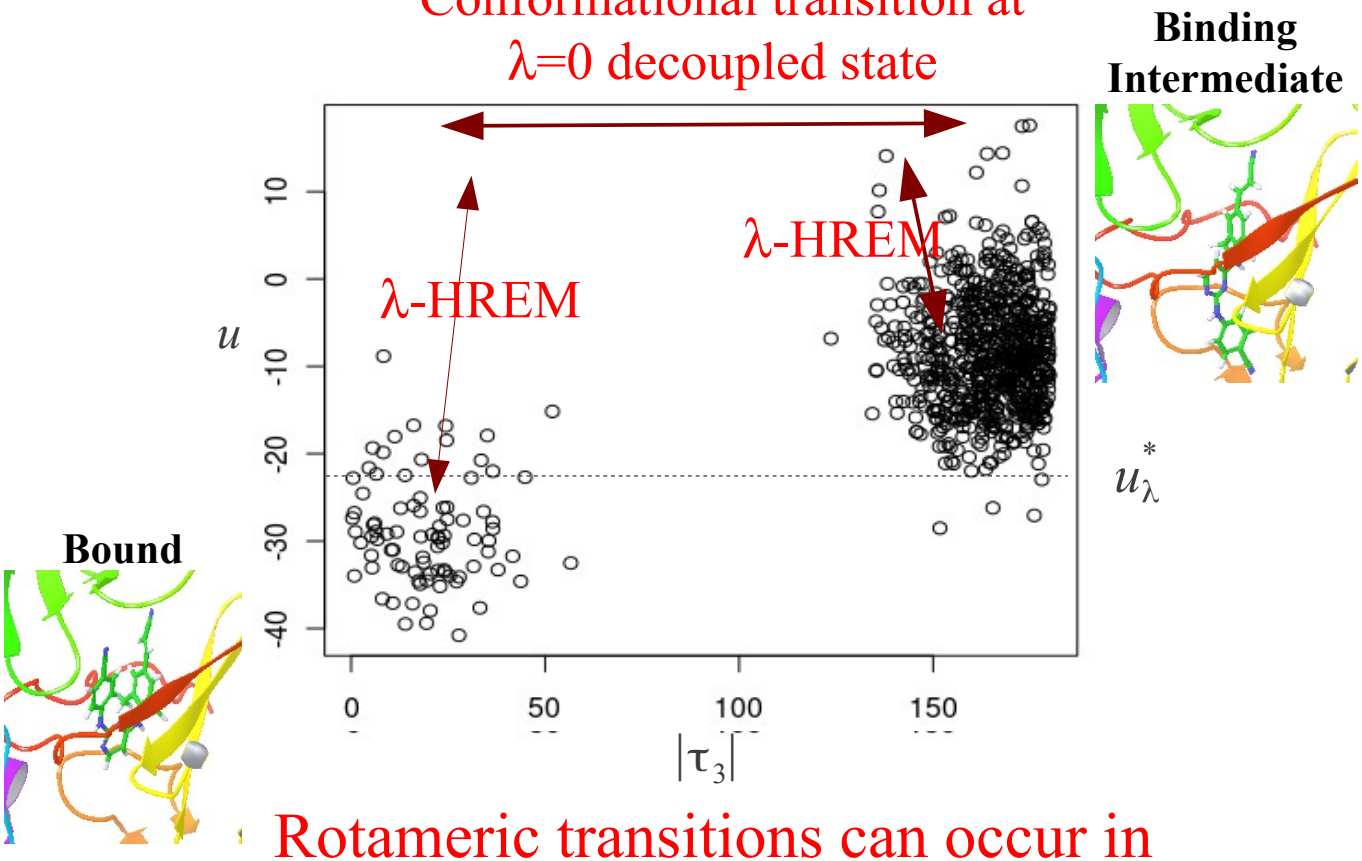
An Alchemical Route?



Direct path impeded by ligand-receptor clashes

An Alchemical Route?

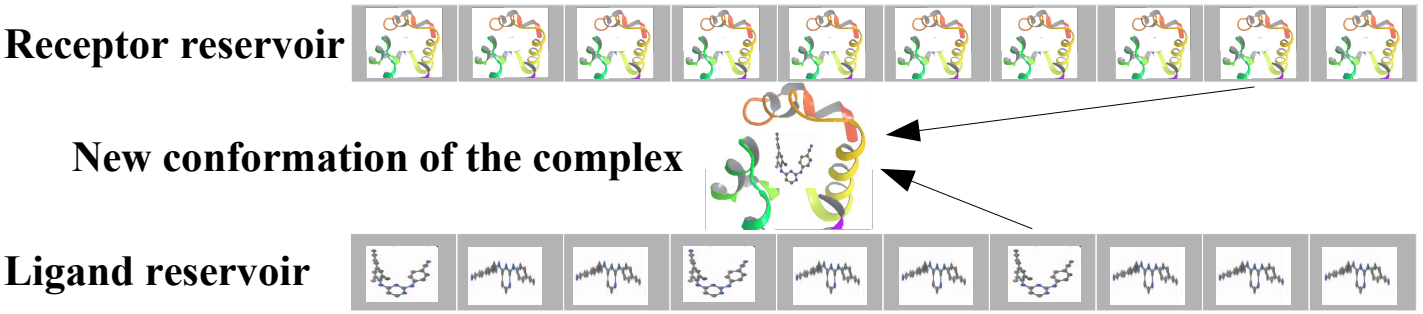
Conformational transition at
λ=0 decoupled state



Rotameric transitions can occur in absence of ligand-receptor interactions

Can we speed it up?

Conformational Reservoirs at Decoupled State (λ=0)



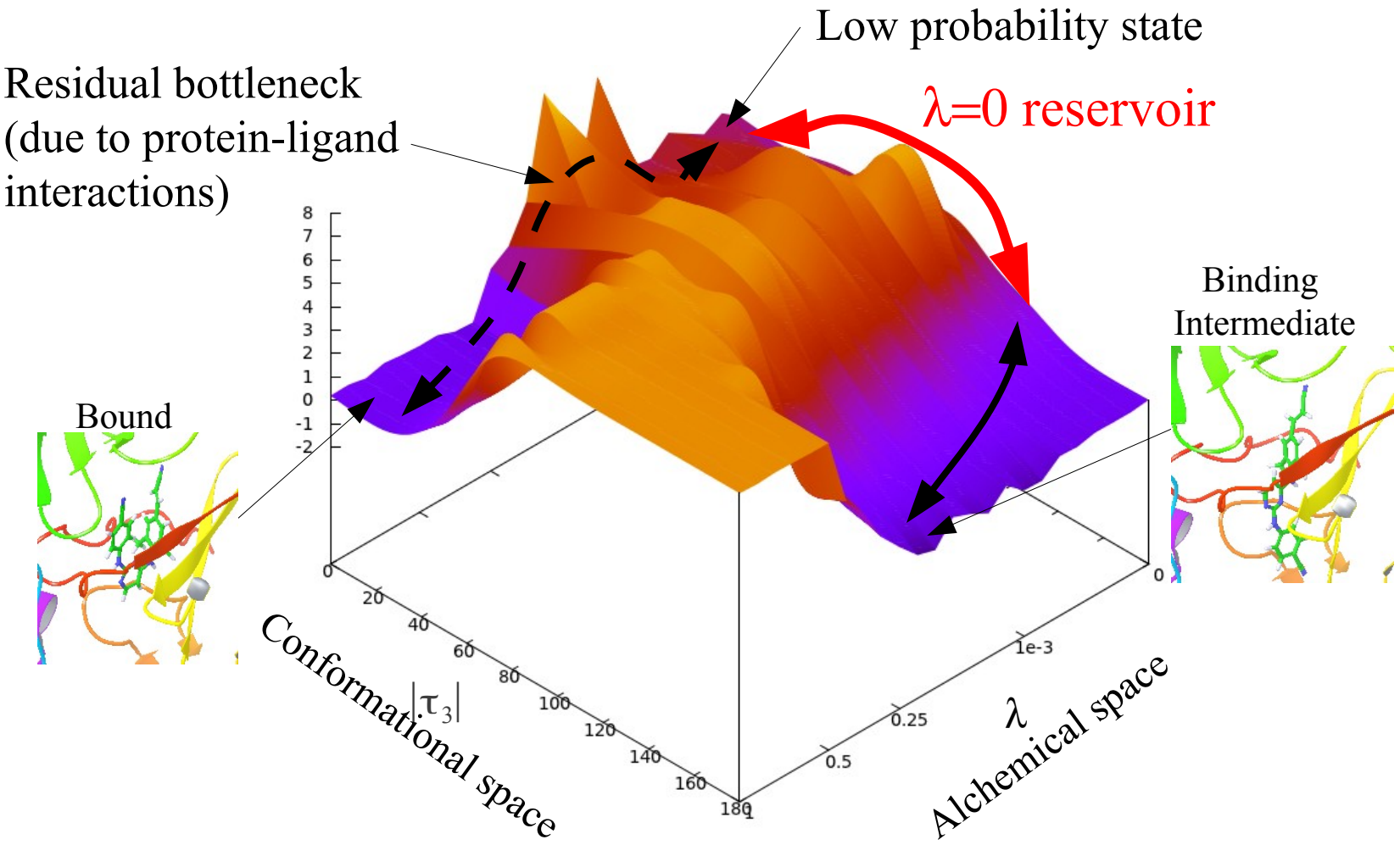
At each exchange ligand and receptor conformations are combined randomly
Ligand is placed in a random position and orientation within Vsite

- Preparatory temperature-RE calculations for ligand and receptor separately
- Store room temperature conformations in conformational repositories
- Queries from repositories replace MD replica at λ=0
- Fully automated
- Calculations for ligand reservoirs are inexpensive
- Receptor reservoir needs to be computed only once for a ligand series
- Combinatorially large number of complex conformations
- Independent sampling of translational/orientational d.o.f.'s of the ligand

Reservoirs provide conformational diversity
Sampling not affected by free energy barriers

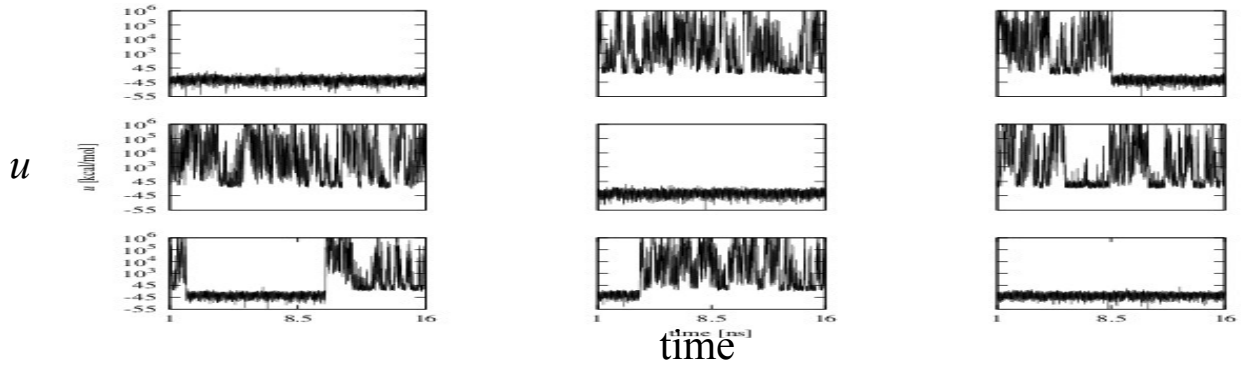
New Conformational Transition Pathways

λ-HREM and reservoirs open conformational communication channels by moving in λ-space



Still Painfully Few Transitions

Binding energy transitions = transition events



- Significantly less than one transition event per replica per ns
- Slow and uneven progress towards convergence
- Convergence nevertheless

		$\Delta G_b(\text{calc})$	$\Delta G(\text{restr})$	ΔG_b°
TMC278	Unrestrained + Reservoir	-11.37±0.67	0	-11.37±0.67
	Restrained	-14.96±0.60	3.21±0.21	-11.75±0.64

OK!

kcal/mol

Gallicchio, Mol. Biosc (2012)

Conclusions

- Unrestrained alchemical calculations are potentially more reliable and easier to automate
- But achieving sufficient conformational sampling is challenging.
- Convergence hinges on achieving equilibrium at all λ-states
- Order-disorder binding transitions can be accelerated by OSRW approaches
- λ-hopping and reservoirs can accelerate conformational transitions by routing the system through alchemical space.
- A lot more work is needed to make these techniques generally applicable and automated.

Acknowledgements

- Ronald Levy
- Peng He
- Lauren Wickstrom
- Mauro Lapelosa (now at Drexel with Cameron Abrams)
- NIH, NSF (research support + computing support)