Accelerating Convergence of Free Energy Calculation with Hamiltonian Replica Exchange

Wei Jiang

2017 Computational Biophysics Workshop, Sept 25-29th

Outline

1. Multiple Copy Algorithm of NAMD
Aims & Implementation
User interface
Popular applications

2. Hysteresis Minimization
Free Energy Perturbation with λ-Exchange

3. Overcome Hidden Barrier with REST2
REST2 Algorithm & Implementation
Sampling enhancement application of REST2
Free Energy Perturbation/H-REMD
FEP/REST2
FEP/REMD/REST2

4. Solvent Sampling Enhancement with REST2
Solvent inaccessible region or Buried pocket

5. Overcome Hidden Barrier of Umbrella Sampling with REST2
US/REMD
US/REMD/REST2

Intelligent sampling with Multiple Copy Algorithms

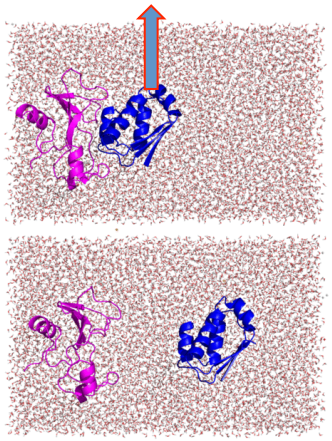
'**Problem** decomposition' ⇒ Many weakly coupled trajectories (Divide-and-conquer)

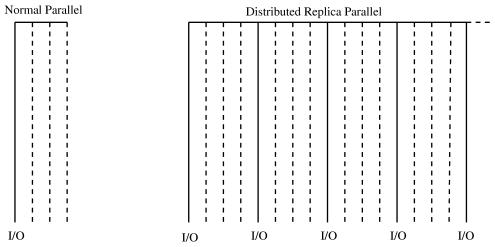
Each trajectory ⇒ Accelerated molecular dynamics with biased terms

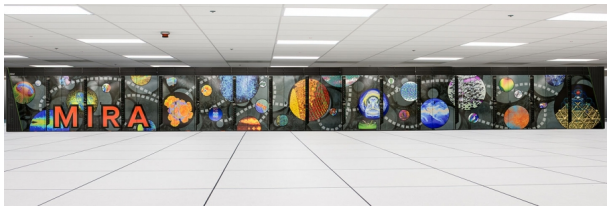
Periodic inter-trajectory communication ⇒ Optimal sampling efficiency

Number of trajectories ⇒ Controlled with acceptance ratio and replica travel

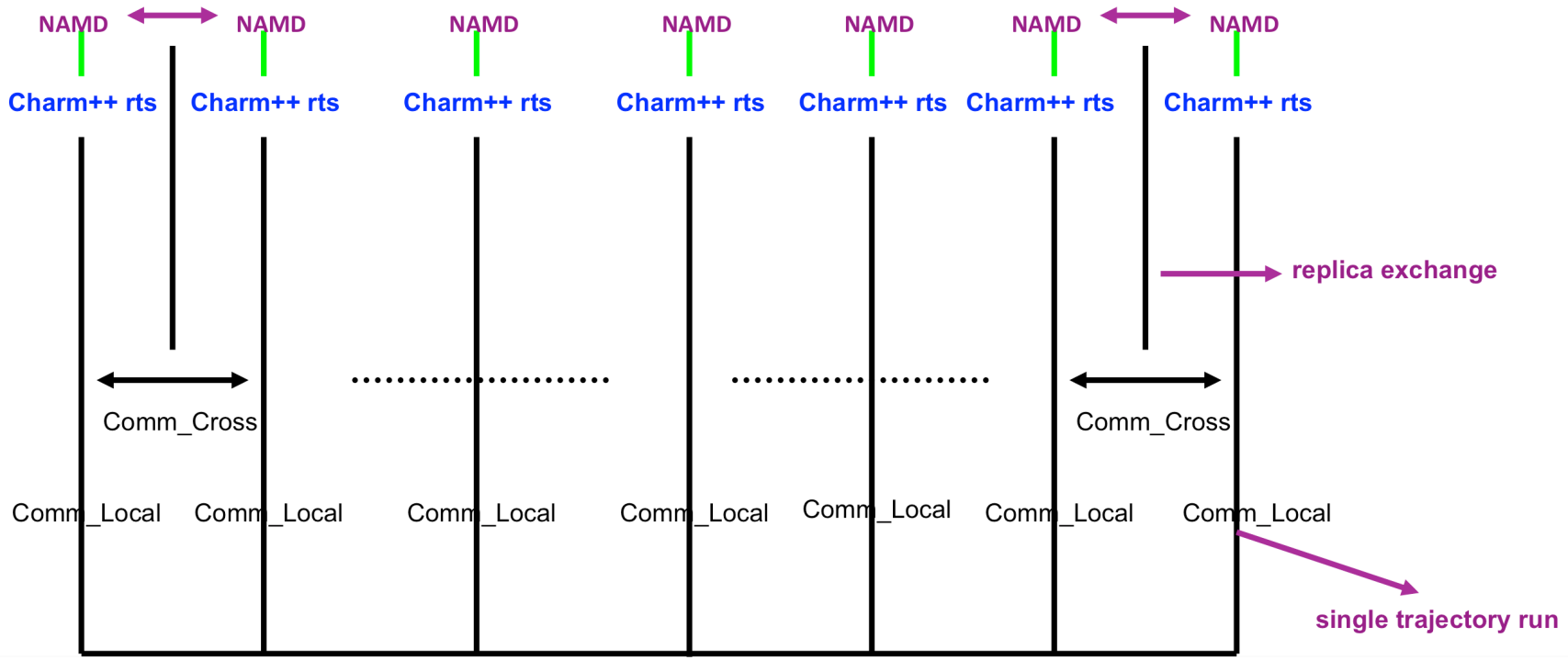
Quantitative info ⇒ Free energy, transition path, reaction rate, protein folding/unfolding







Scalable Multiple Copy Algorithms in NAMD



Multiple Copy Algorithm(MCA) : Coupling multiple trajectories to characterize/accelerate complex molecular processes on massively distributed computer

MCA instances: REST2, T-REMD, AMD/REMD, FEP/REMD, US/REMD, String method, Multi-MetaDynamics, FFM ……

Communication enabled Tcl scripting interface by which user can arbitrarily design **any** MCA or accelerated sampling algorithm

Major Sampling Difficulties and Solutions in Free Energy Calculations

Hysteresis

Reaction coordinates exchange along reaction path

Enhance window overlapping

Optimizing positions of windows along reaction path

Doesn't overcome large time scale problem

Hidden barrier

Orthogonal to reaction path

Construction of barrier flattening potential

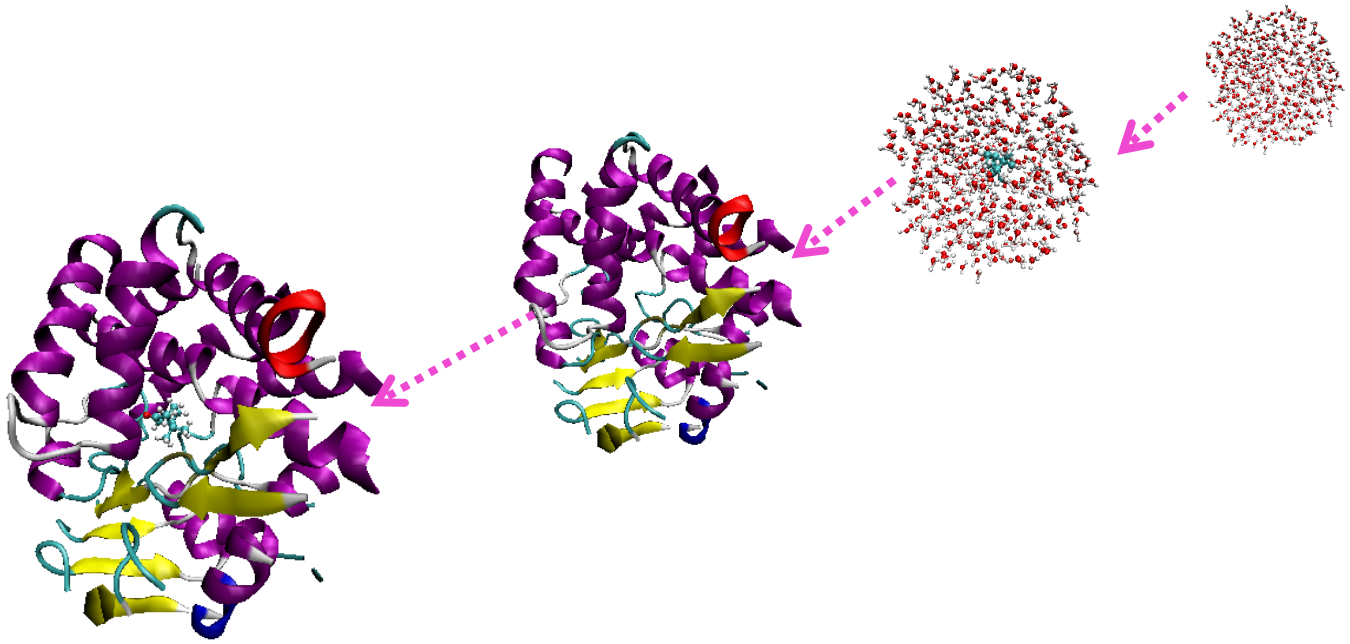
In MCA frame -> Extra boosting windows -> Multi-dimensional

Solvent sampling

Monte Carlo -> Detailed balance->poor efficiency

Alternative ?

Molecular recognition With Free Energy Perturbation



$$K_b = \frac{\int_{\text{site}} d(\mathbf{L}) \int d\mathbf{X} \ \exp[-U/kT]}{\int_{\text{bulk}} d(\mathbf{L}) \ \delta(\mathbf{r}-\mathbf{r}') \int d\mathbf{X} \ \exp[-U/kT]}$$

$$U(s,\xi,\lambda,\lambda_r) = U_0 + U^{rep}(s) + \xi U^{dis} + \lambda U^{elec} + \lambda_r u_r$$

Theoretical and algorithmic foundation for relative FEP
Long alchemical path
Complex barrier landscape demanding sampling/FF

$$U(s=0,\xi=0,\lambda=0,\lambda_r=1) \rightarrow U(s=1,\xi=0,\lambda=0,\lambda_r=1)$$
$$U(s=0,\xi=0,\lambda=0,\lambda_r=1) \rightarrow U(s=1,\xi=0,\lambda=0,\lambda_r=1)$$
$$U(s=1,\xi=1,\lambda=0,\lambda_r=1) \rightarrow U(s=1,\xi=1,\lambda=1,\lambda_r=1)$$
$$U(s=1,\xi=1,\lambda=1,\lambda_r=1) \rightarrow U(s=1,\xi=1,\lambda=1,\lambda_r=0)$$

Quick Application of FEP/λ-REMD

a) FEP/λ-REMD scheme

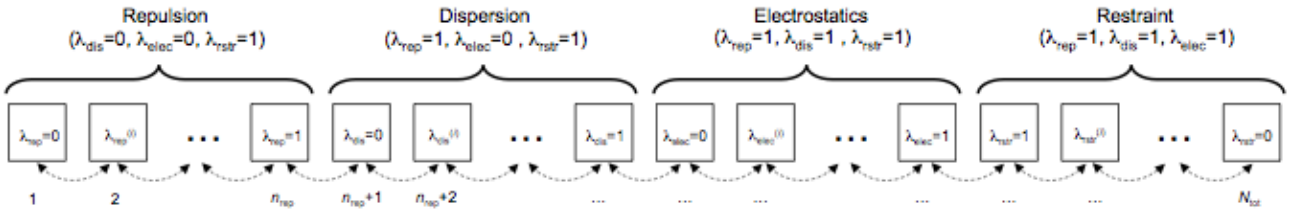


Table 1. Hydration Free Energy and the individual components for TIP3

Prod.	Rep. exchange	ΔG_{rep}	ΔG_{disp}	ΔG_{elec}	ΔG	Expt.
40 ps	0	4.79 ± 0.11	-2.81 ± 0.03	-8.09 ± 0.07	-6.12 ± 0.14	-6.3
	1 /1000 steps	5.10 ± 0.16	-2.87 ± 0.01	-8.20 ± 0.12	-5.97 ± 0.23	
	1/100 steps	5.11 ± 0.15	-2.87 ± 0.02	-8.13 ± 0.08	-5.89 ± 0.18	
100 ps	0	5.12 ± 0.10	-2.88 ± 0.01	-8.20 ± 0.05	-5.95 ± 0.11	
	1/1000 steps	5.11 ± 0.06	-2.87 ± 0.01	-8.21 ± 0.07	-5.97 ± 0.12	
	1/100 steps	5.09 ± 0.07	-2.88 ± 0.01	-8.21 ± 0.06	-6.00 ± 0.12	

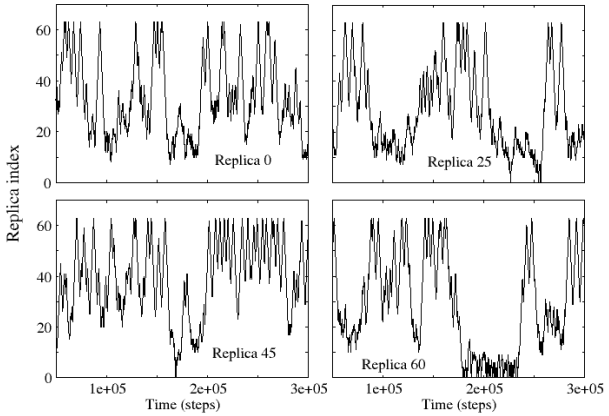


Table 2. Hydration Free Energy and Individual Components for Benezene

Prod.	Rep. exchange	ΔG_{rep}	ΔG_{disp}	ΔG_{elec}	ΔG	Expt.
40 ps	0	13.46 ± 0.47	-12.63 ± 0.18	-1.88 ± 0.04	-1.05 ± 0.45	-0.87
	1 /1000 steps	14.41 ± 0.31	-13.07 ± 0.06	-1.89 ± 0.06	-0.55 ± 0.29	
	1/100 steps	14.45 ± 0.39	-13.01 ± 0.07	-1.85 ± 0.05	-0.41 ± 0.39	
	1/10 steps	14.67 ± 0.45	-13.07 ± 0.07	-1.90 ± 0.10	-0.30 ± 0.50	
100 ps	0	14.47 ± 0.20	-13.06 ± 0.06	-1.87 ± 0.04	-0.45 ± 0.19	
	1/1000 steps	14.50 ± 0.21	-13.06 ± 0.04	-1.86 ± 0.06	-0.42 ± 0.18	
	1/100 steps	14.49 ± 0.11	-13.03 ± 0.05	-1.86 ± 0.03	-0.41 ± 0.13	
	1/10 steps	14.49 ± 0.13	-13.03 ± 0.08	-1.86 ± 0.07	-0.41 ± 0.15	

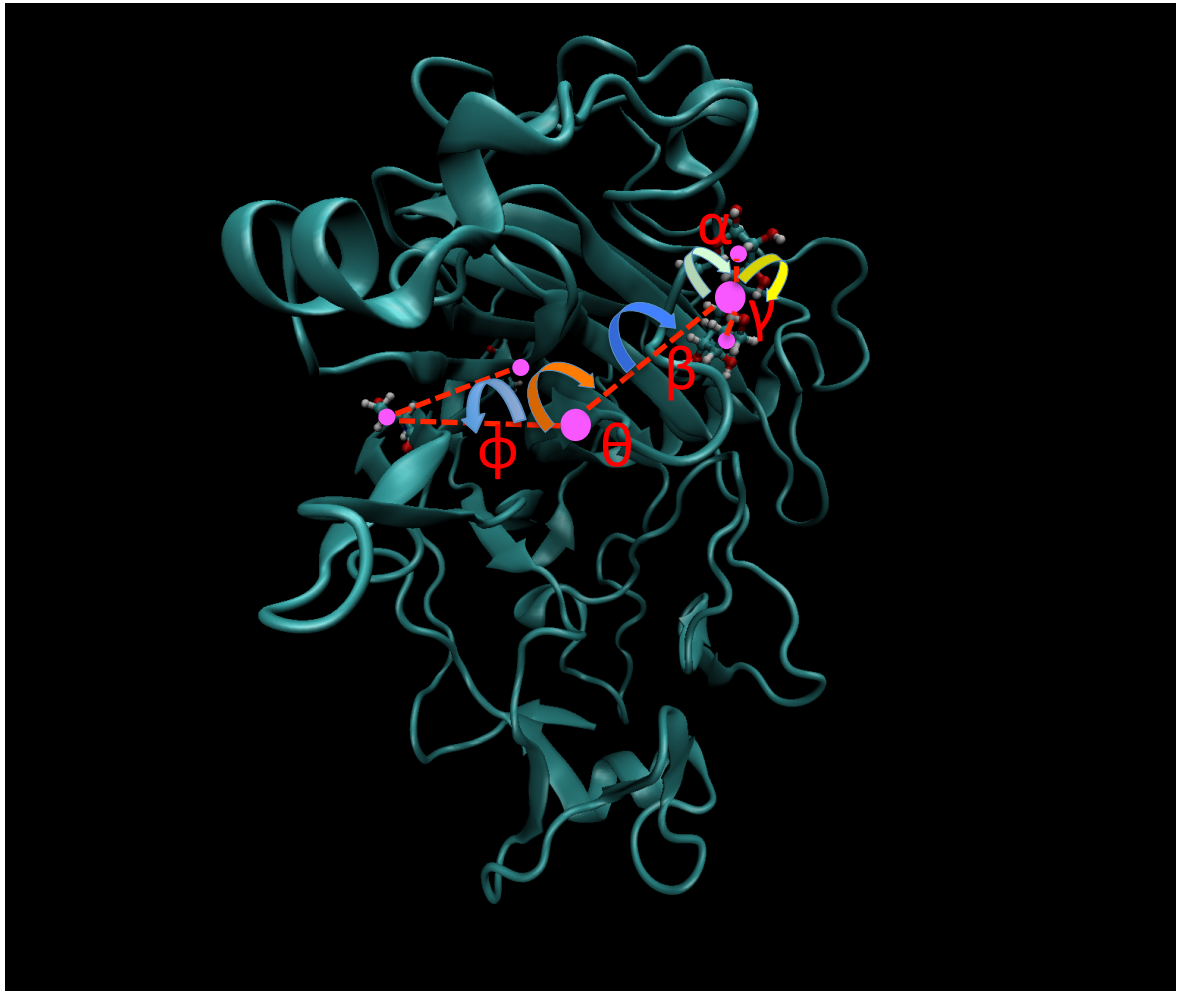
Hysteresis Minimization with High Frequency λ exchange

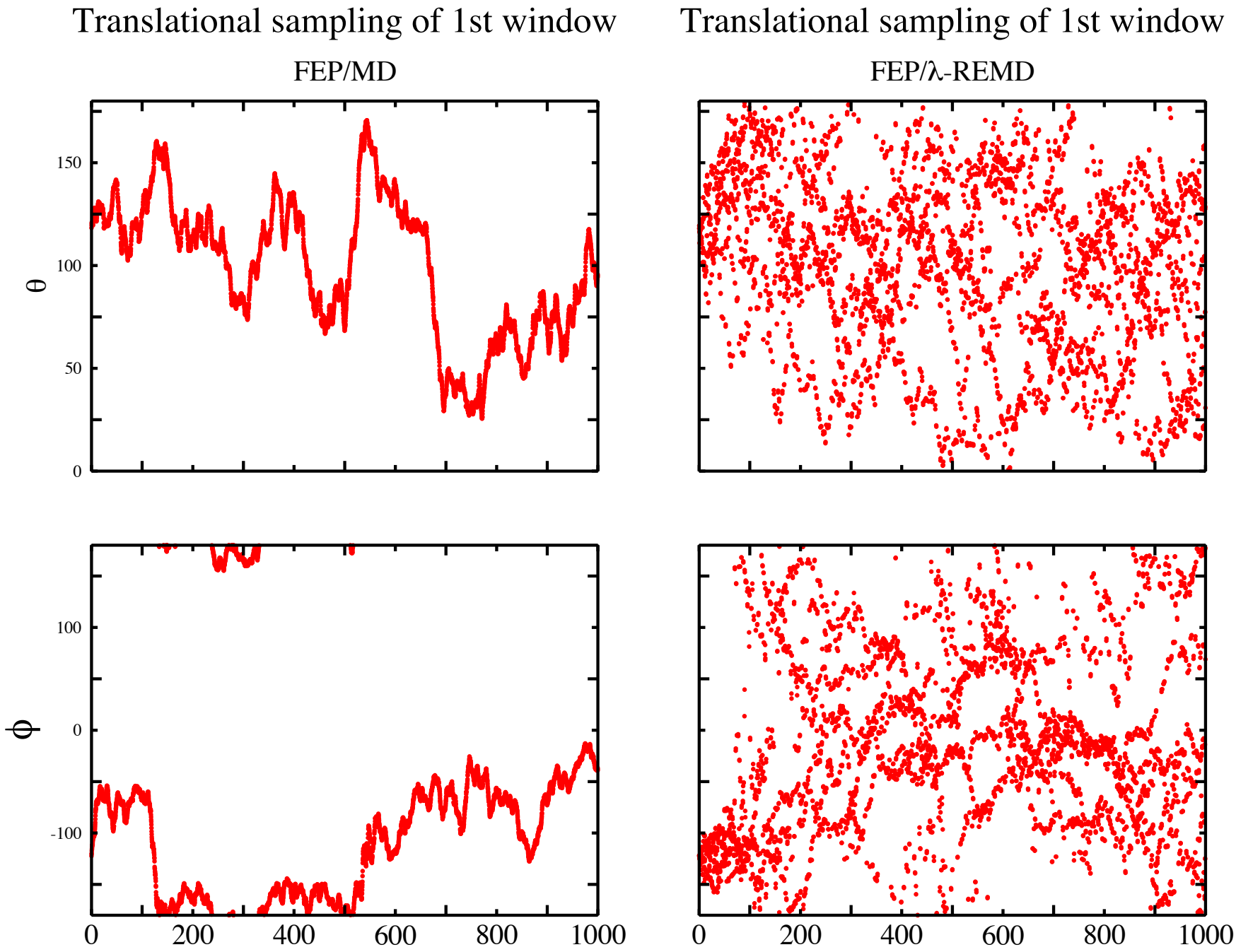
Sampling difficulties of ligand's translational and orientational degrees of freedom relative to target

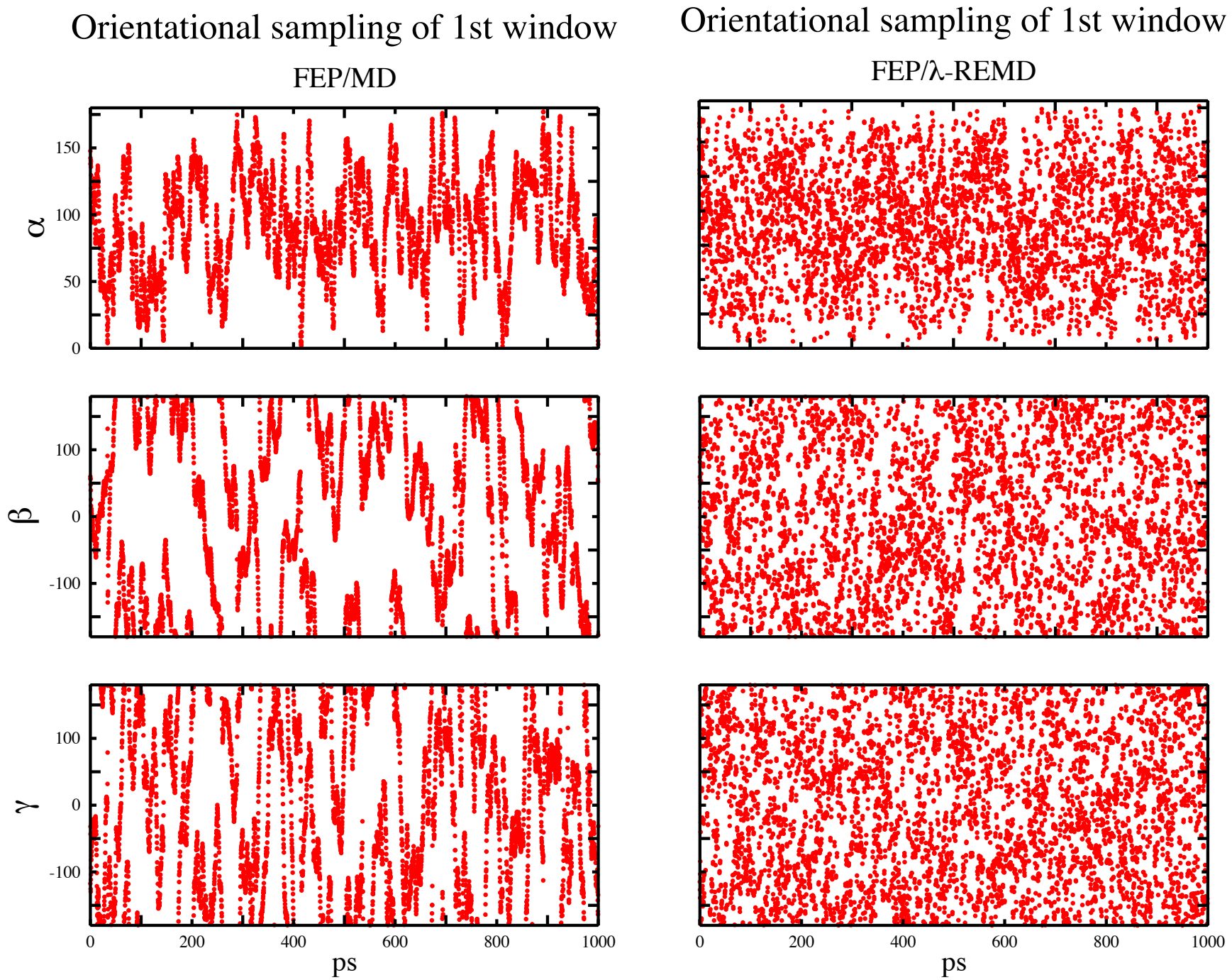
Severe sampling issues arises in the initial switching-on, i.e, 1st or last window.

Absolute FEP:
Geometry restraints
Translational (r,θ,ϕ)
Orientation (α,β,γ)

Relative FEP: Single topology (restraint)




Translational sampling of 1st window
FEP/MD
Translational sampling of 1st window
FEP/λ-REMD
θ
150
100
50
0
ϕ
100
0
-100
0
200
400
600
800
1000


Orientational sampling of 1st window
FEP/MD
Orientational sampling of 1st window
FEP/λ-REMD
α
β
γ
150
100
50
0
100
0
-100
100
0
-100
0
200
400
600
800
1000
ps
0
200
400
600
800
1000
ps

Co-product of λ exchange: Simple Overlap Sampling

Without λ exchange:

WHAM

BAR (m-BAR)

With λ exchange:

Better overlapped windows and correlated data

Instant output of bi-direction potential energies

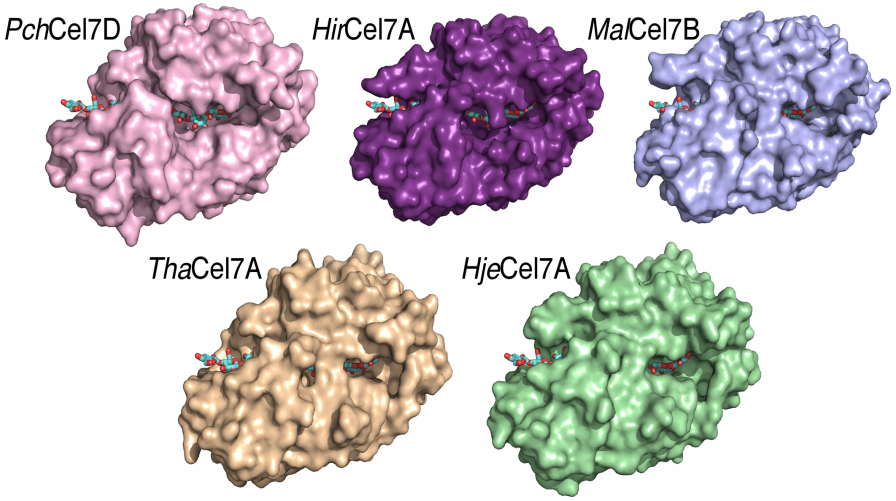
V(λ,X1) V(λ+Δλ,X2) V(λ,X2) V(λ+Δλ,X1)

SOS is a handy choice -> identical result with WHAM and BAR

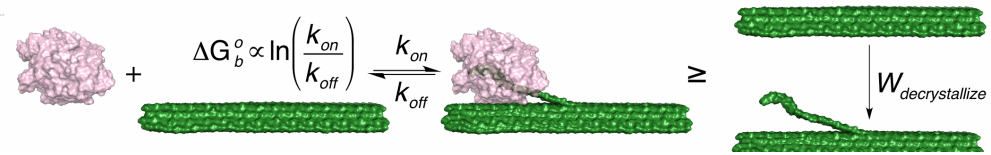
$$\exp(-\beta\Delta A) = \frac{\left\langle \exp(-(V(\lambda+\Delta\lambda, X2) - V(\lambda, X1))/(2.0 * RT)) \right\rangle_0}{\left\langle \exp((V(\lambda, X2) - V(\lambda+\Delta\lambda, X1))/(2.0 * RT)) \right\rangle_1}$$

Receive result in < 5s

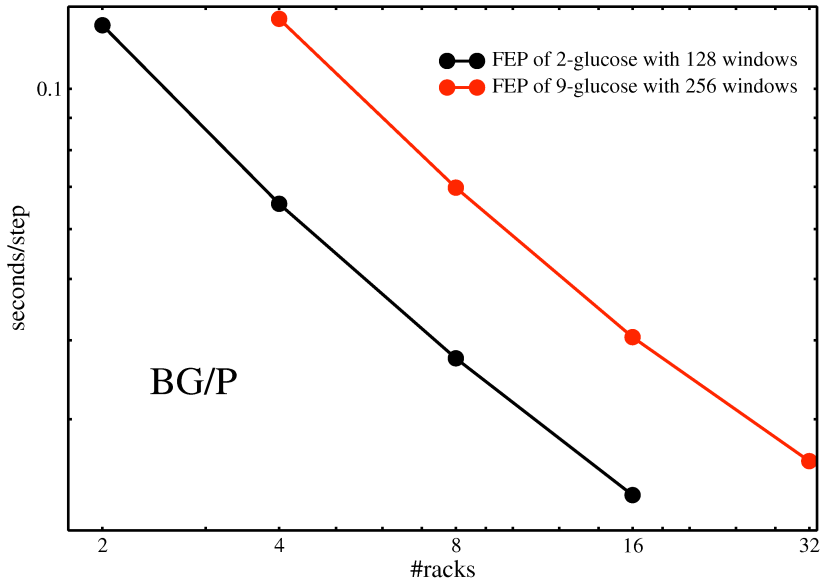
Thermodynamics of Binding Biomass to Cellulases

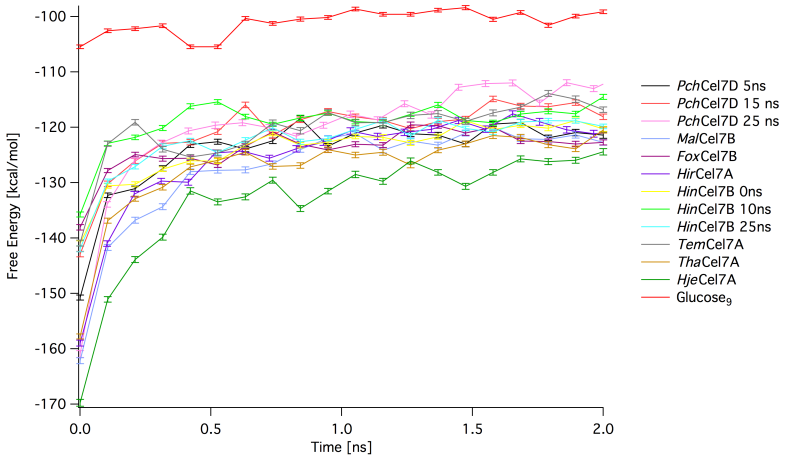


Fungal Family 7 glycoside hydrolase (GH)s hypothesized to act processively on cellulose crystalline microfibrils. Each cellulase exhibits the same characteristic fold along with attached loop domains forming the tunnel-shaped active site. The active sites encompass the cello-oligomer ligand to varying degrees. Of the five enzymes, PchCel7D has the most open of the active site tunnels, while HjeCel7A has the most enclosed active site tunnel.



The **absolute binding free energy** is defined as the free energy change between a polysaccharide chain (of n monosaccharide units, where n is the chain length required to saturate the GH binding sites) and the enzyme-substrate complex in the catalytically-active complex. We hypothesize that this quantity, ΔG_b°, is directly correlated to processivity and to suggest the morphology dependence of cellulose attack by enzymes.



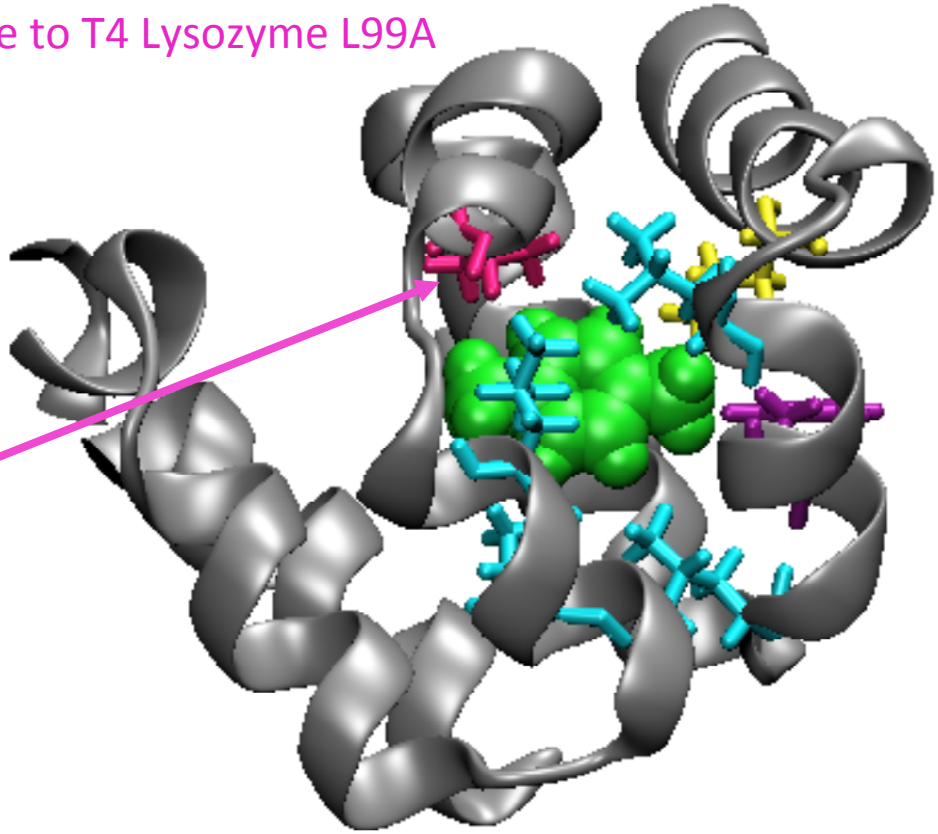


Christina M. Payne, Wei Jiang, Michael R. Shirts, Michael E. Himmel, Michael F. Crowley and Gregg T. Beckham, J. Am. Chem. Soc. 2013, 135, 18831-18839

Kinetically Trapped Conformations in Free Energy calculations

Problems arise when large structural reorganizations happen
Hidden barriers orthogonal to reaction path->Kinetically trapped
Beyond timescale of typical FEP or US/MD trajectory
Efficient flattening potential and quick implementation wanted!

Binding of large aromatic molecule to T4 Lysozyme L99A



Valine 111 *gauche*$^-$ → *trans*

Replica Exchange Solute Tempering (REST2)

All replicas are run at the same temperature but the potential energy for each replica is scaled differently;
Lowering energy barrier of small group atoms -> significantly higher efficiency than traditional temperature exchange

$$E_m^{\mathrm{REST2}}(X) = \frac{\beta_m}{\beta_0} E_{ss}(X) + \sqrt{\frac{\beta_m}{\beta_0}} E_{sw}(X) + E_{ww}(X)$$ -> parameter rescaling

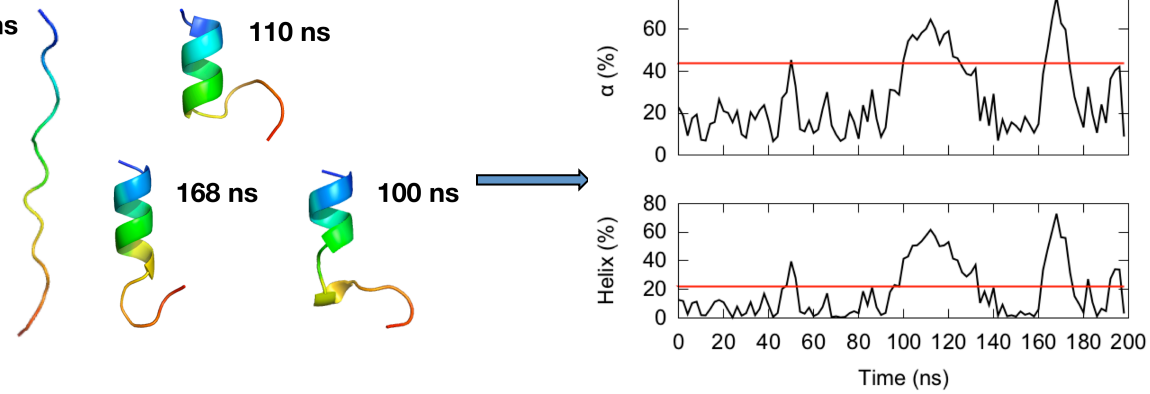
$$\Delta_{mn}(\mathrm{REST2}) = \left(\beta_m - \beta_n\right)\left[\left(E_{ss}(X_n) - E_{ss}(X_m)\right) + \frac{\sqrt{\beta_0}}{\sqrt{\beta_m} + \sqrt{\beta_n}}\left(E_{sw}(X_n) - E_{sw}(X_m)\right)\right]$$

Replica exchange solute tempering:
Adjustable flattening potential;
Straightforward to implement, multiple versions;
The most popular Hamiltonian exchange method.

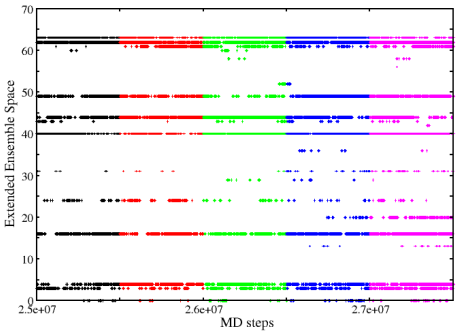
REST2 in NAMD:
Generic implementation -> free end user preparing customized input files.
Parameter exchange -> high frequency exchange attempt
Communication master -> Tcl script
Ready to employ along with other free energy methods.

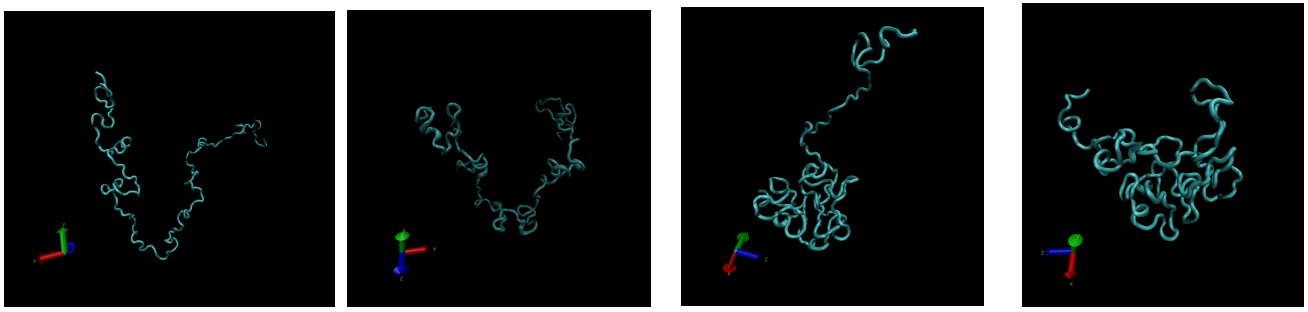
Protein Folding-Unfolding Transitions with REST2

Peptide folding-unfolding, **explicit** solvent, 16 replica, effective temperature range 300 – 600K
Acceptance ratio: 50% >> T-REMD



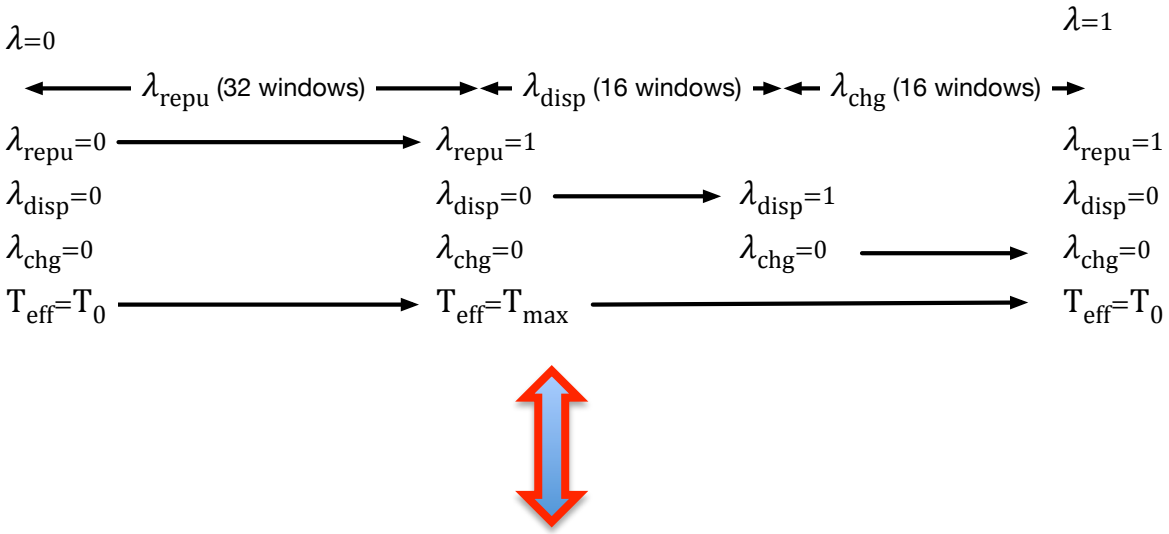
Large protein folding-unfolding, **explicit** solvent, 64 REST2 replicas, 60% acceptance ratio with exchange attempt frequency 1/20 steps





0 ns 12 ns 24 ns 50 ns

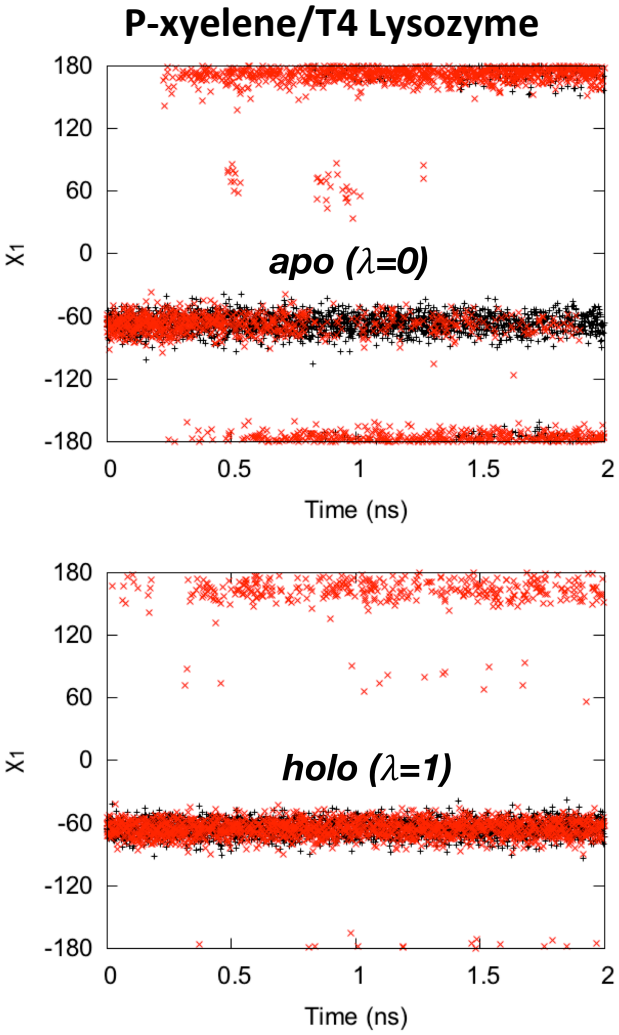
FEP/REST2 (Schordinger Version)



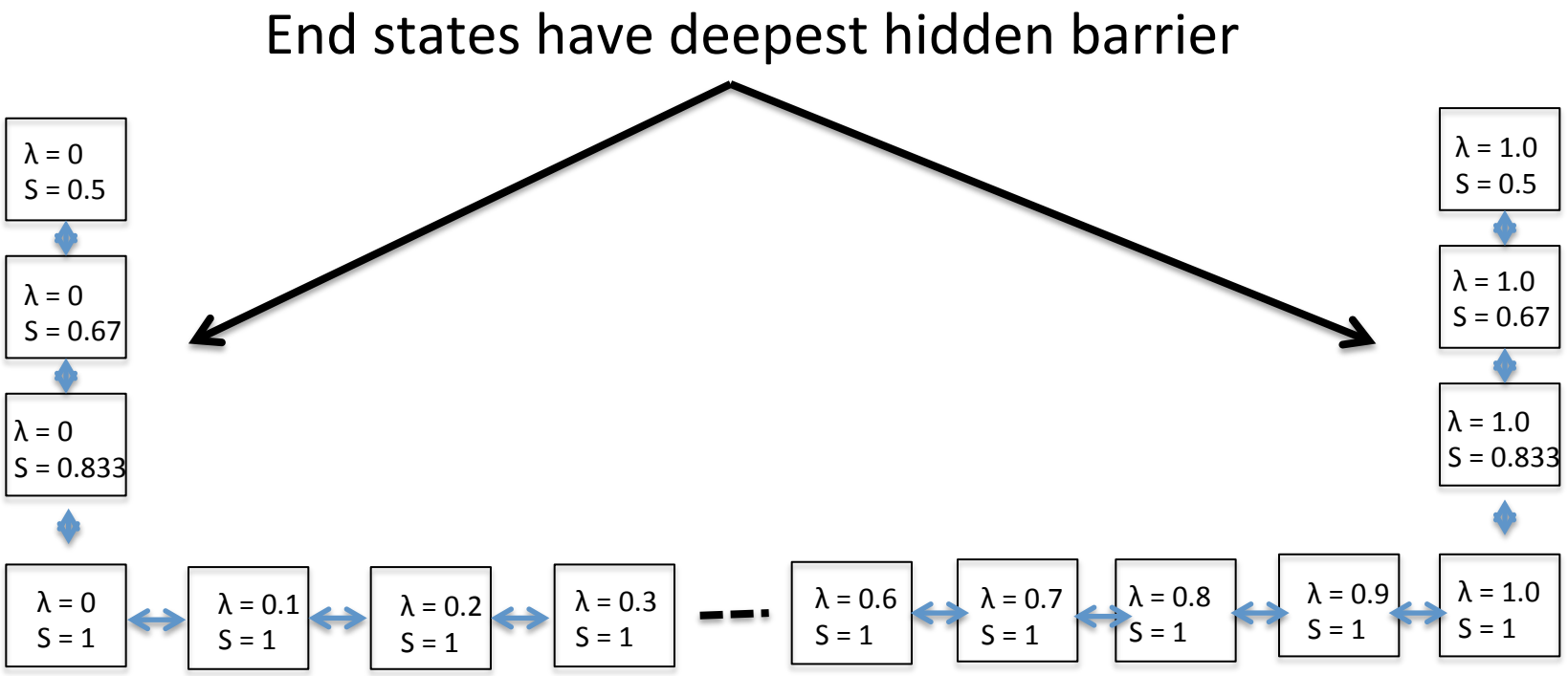
Cheap solution and easy implementation

Thermodynamic axis is contaminated by the brutal mixing of REST2 and FEP

Carefully controlled heated region minimizes nonequilibrium effects.



Orthogonal Implementation of FEP/REMD/REST2

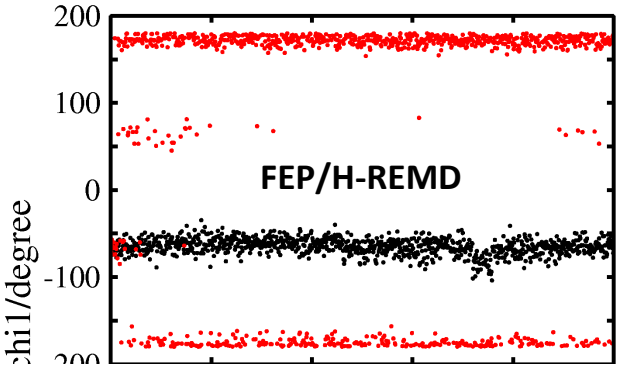


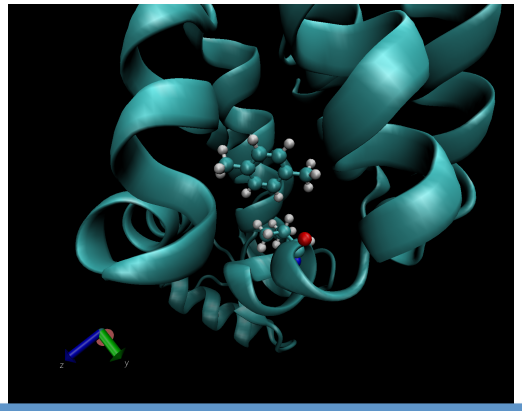
Separation of λ-REMD and REST2, leaving FEP as it is

REST2 windows adjustable with size of heated region

Need slightly more parallel computing resource

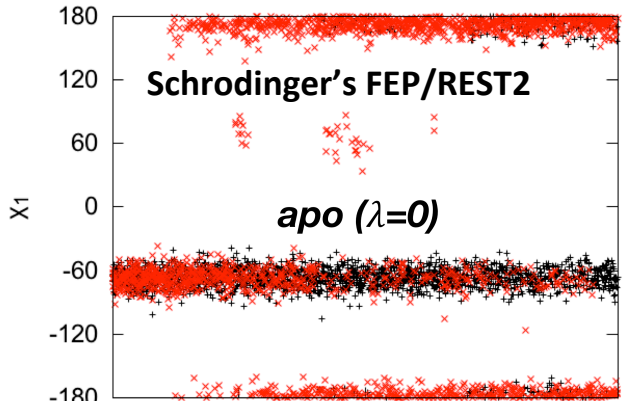
Comparison with FEP/H-REMD and Schrodinger's FEP/REST2

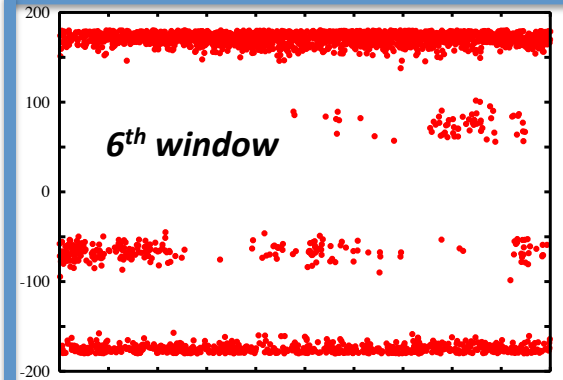


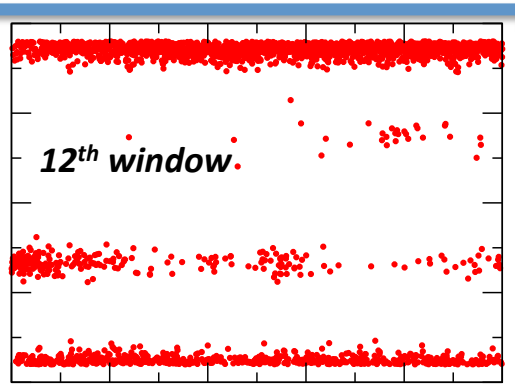


P-xyelene/T4 Lysozyme

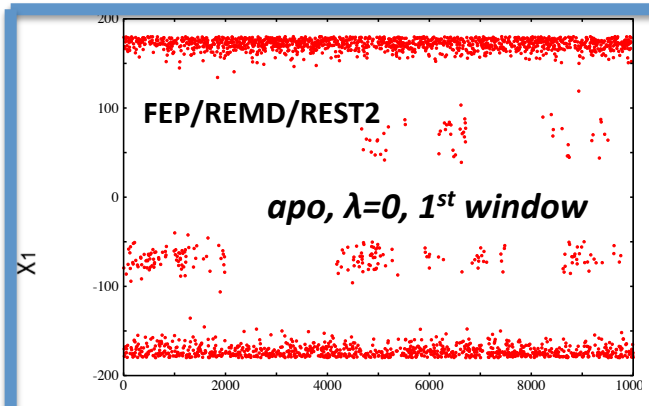
-5.5 kcal/mol vs exp -4.7 kcal/mol

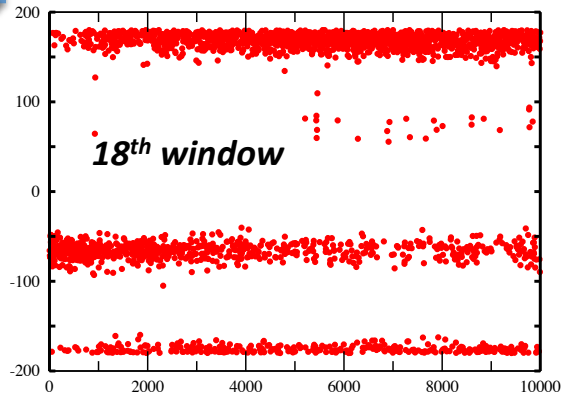


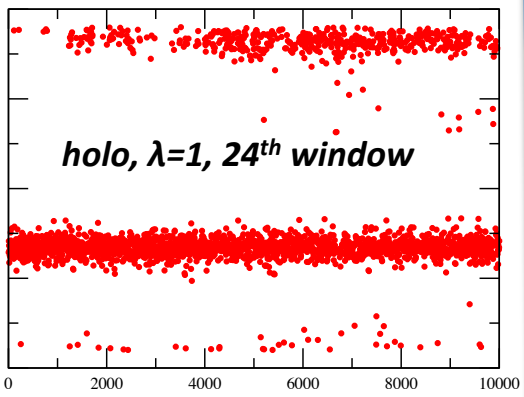




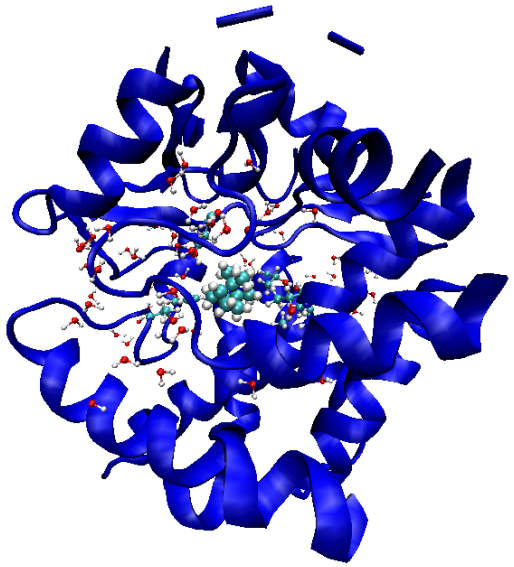
Reasonable evolution from apo to holo state







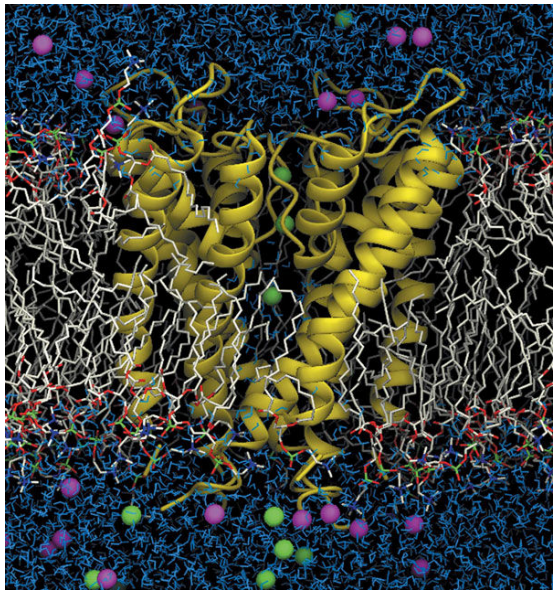
Solvent acceleration with simulated annealing REST2



Camphor/P450 Binding Complex

Buried binding pocket

Interior polar residues



KcSA Ion Channel

Large cavity

Solvent configuration sampling

Demanding sampling of solvent!!!

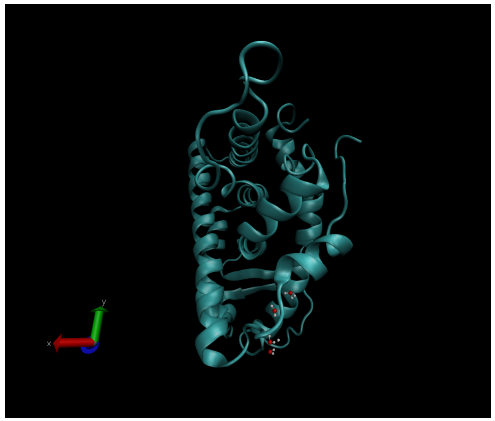
Many solvent configurations needed

Monte Carlo method is too slow and doesn't match MD trajectory on-the-fly

Temperature replica exchange doesn't work efficiently with explicit solvent

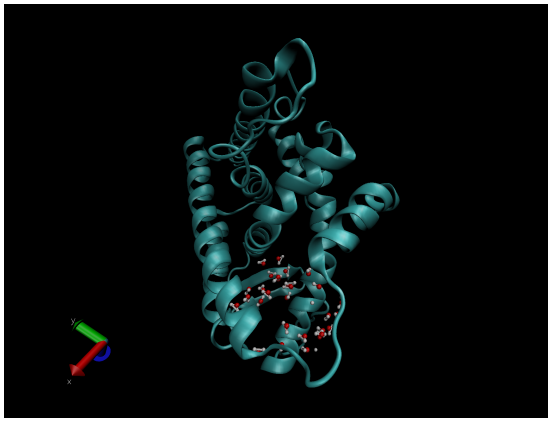
Hybrid of Simulated Annealing and REST2 for Solvent acceleration

Effective simulated annealing (SA) can replace solvent temperature replica exchange
Re-scale potential energy: (1) scale potential energy of solvent with others fixed; (2)or only scale protein atoms at buried binding pocket
Good SA schedule of solvent remove bad steric interaction
Periodic SA during FEP -> SA – FEP – SA – FEP …….. Analogous to CHARMM FEP/GCMC



1 ns

Estrogen Receptor

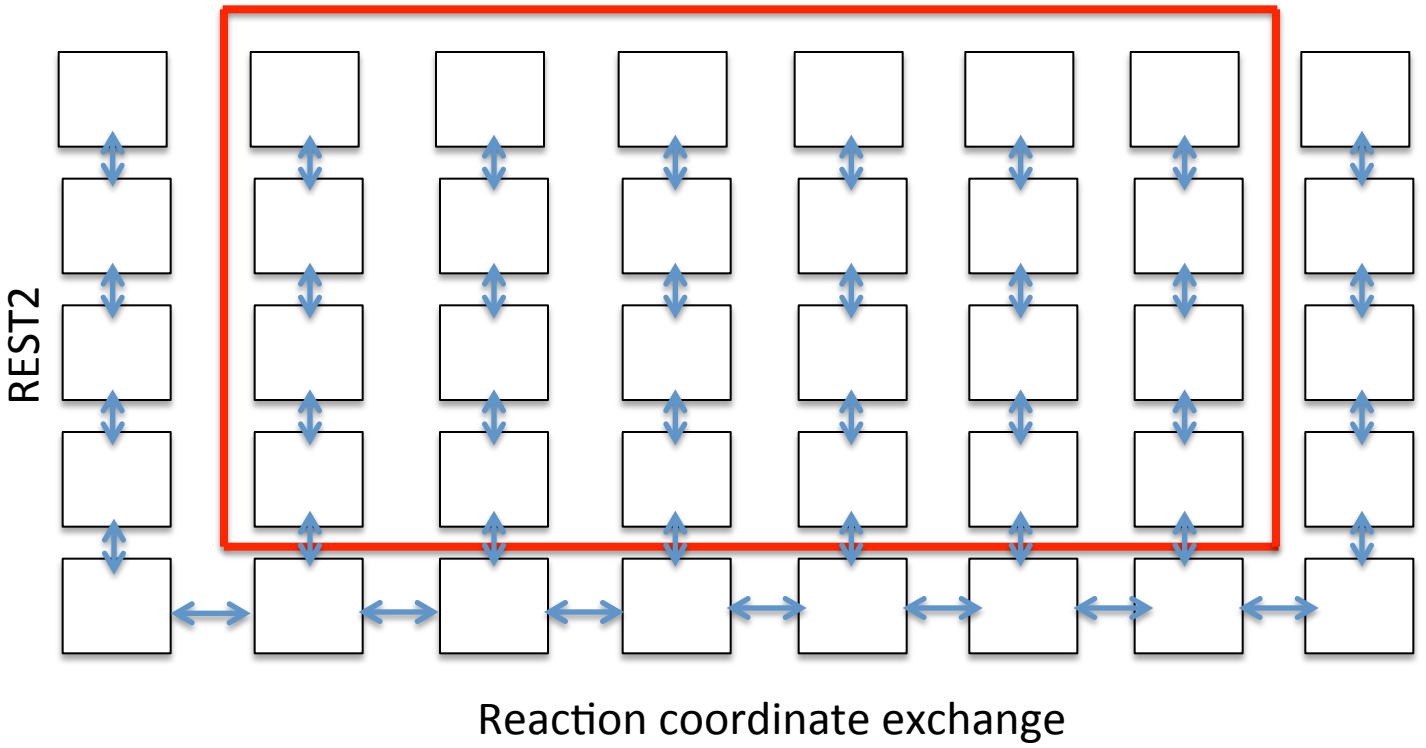


100 ps (scaling factor 0.75) + 100ps 300K

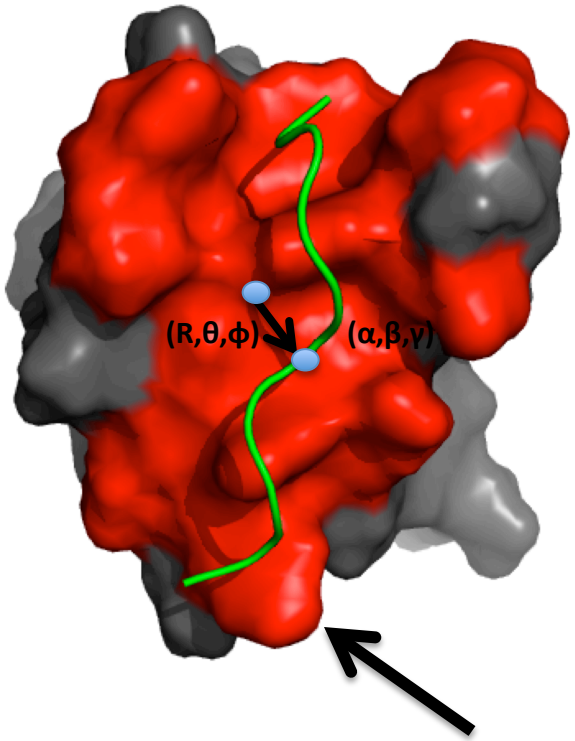
		EXP (kcal/mol)	FEP/SA/REST2
1	DES	-13.2	-14.0
2	OHT	-12.7	-13.5
3	EST	-12.3	-13.3
4	E1T	-10.7	-11.5
5	TAM	-10.6	-11.3
6	NAF	-9.4	-10.6
7	NOR	-8.7	-9.80

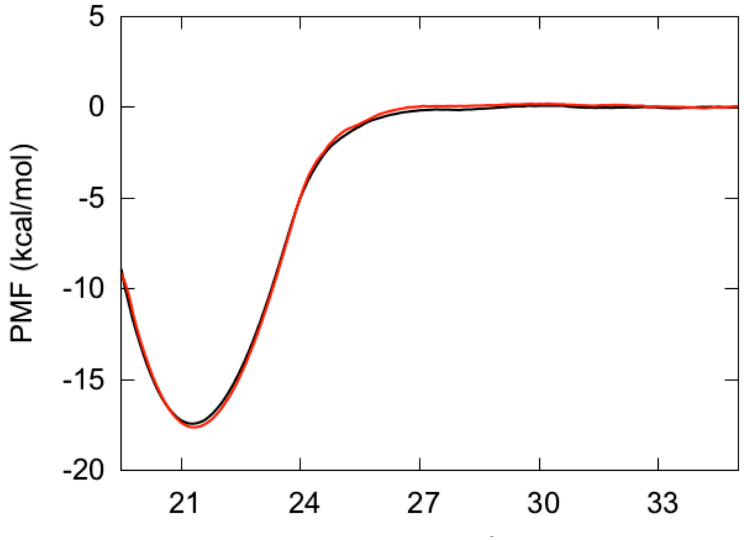
Combination of Umbrella Sampling and REST2

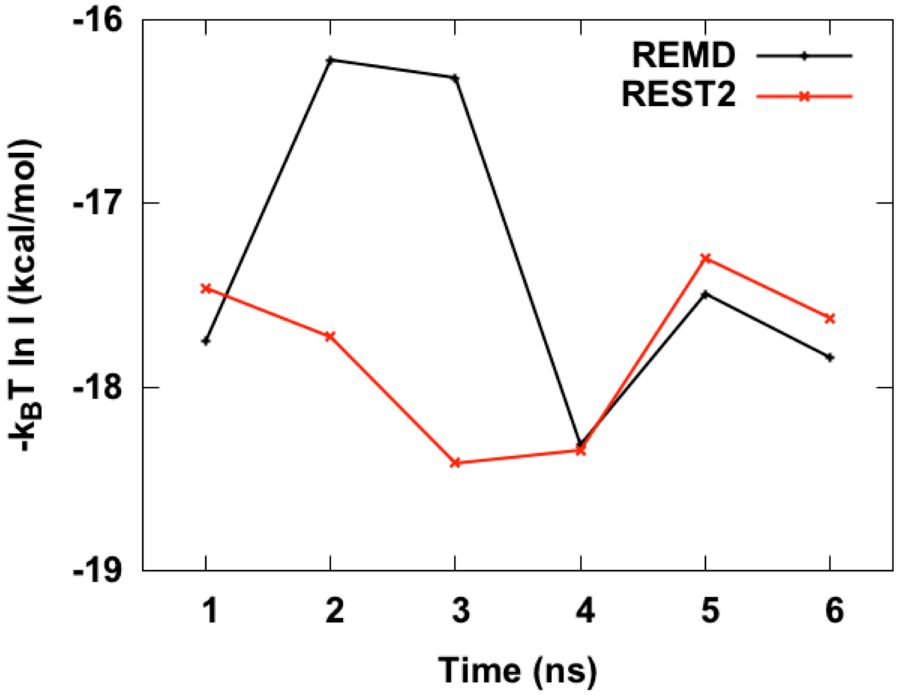
REST2 overcomes Hamiltonian lagging
Multidimensional Hamiltonian exchange scheme
Umbrella biases are exchanged in one axis
REST2 in another axis



Binding of Peptide to SH3 domain of Kinase

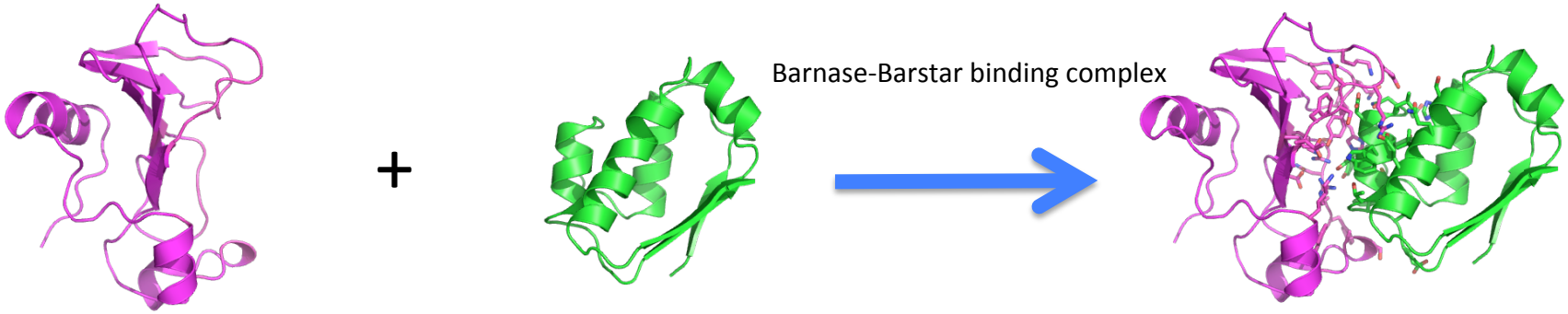


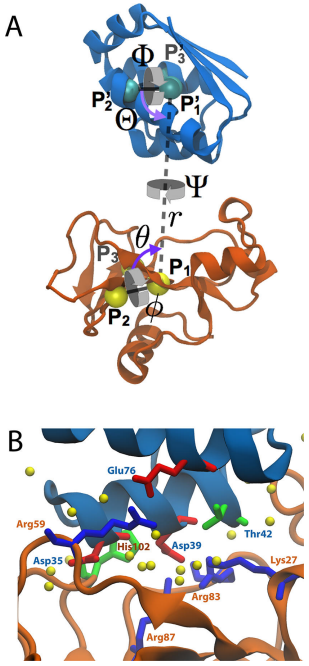




REMD and REMD/REST2
32 and 128 replicas
side chain atoms that are within 4 Å from the peptide were selected for tempering

Quantifying Protein-Protein Binding Energy and Entropy





6 collective variables were used to constraint orientation (**Θ**,**Φ**, and **Ψ**) and translation **(r,** θ**,** ϕ**)**

Barnase binding interface is plastic and selected as tempering

255 umbrella windows × 8 REST2 replicas = 2040 replicas

Barstar-Barnase Binding Entropy

	WT	ΔH	-TΔS
ΔΔGBs,c	0.2	-5.0	3.8
ΔΔGBn,c	0.6	-7.3	7.9
ΔΔGBs,res	3.5	-8.3	9.5
ΔΔGBn,res	6.8	-5.7	13.0
ΔΔGorient	4.5	1.2	3.5
-kBT log(S*I*C°)	-35.0	16.1	-47.3
ΔGBind	-19.6 ± 0.6	-9.1 ± 10.5	-9.6 ± 10.5
ΔGexp	-19.0	-19.3	0.3

Summary

(1) Restraint free absolute binding free energy calculation with high frequency λ exchange

(2) Straightforward for relative FEP (hybrid single/dual topology)

(3) Hybrid Simulated Annealing and REST2 accelerating solvent sampling

(4) Multidimension umbrella sampling with REST2

(5) Self adaptive biasing force and MetaDynamics with REST2