Performing 2D Hamiltonian-Exchange method for Biological Free Energy Calculations on Petascale Supercomputer

Yun Lyna Luo, Wei Jiang, Benoît Roux*

Argonne Leadership Computing Facility, Argonne National Laboratory

Department of Biochemistry and Molecular Biology
University of Chicago

Why replica-exchange:

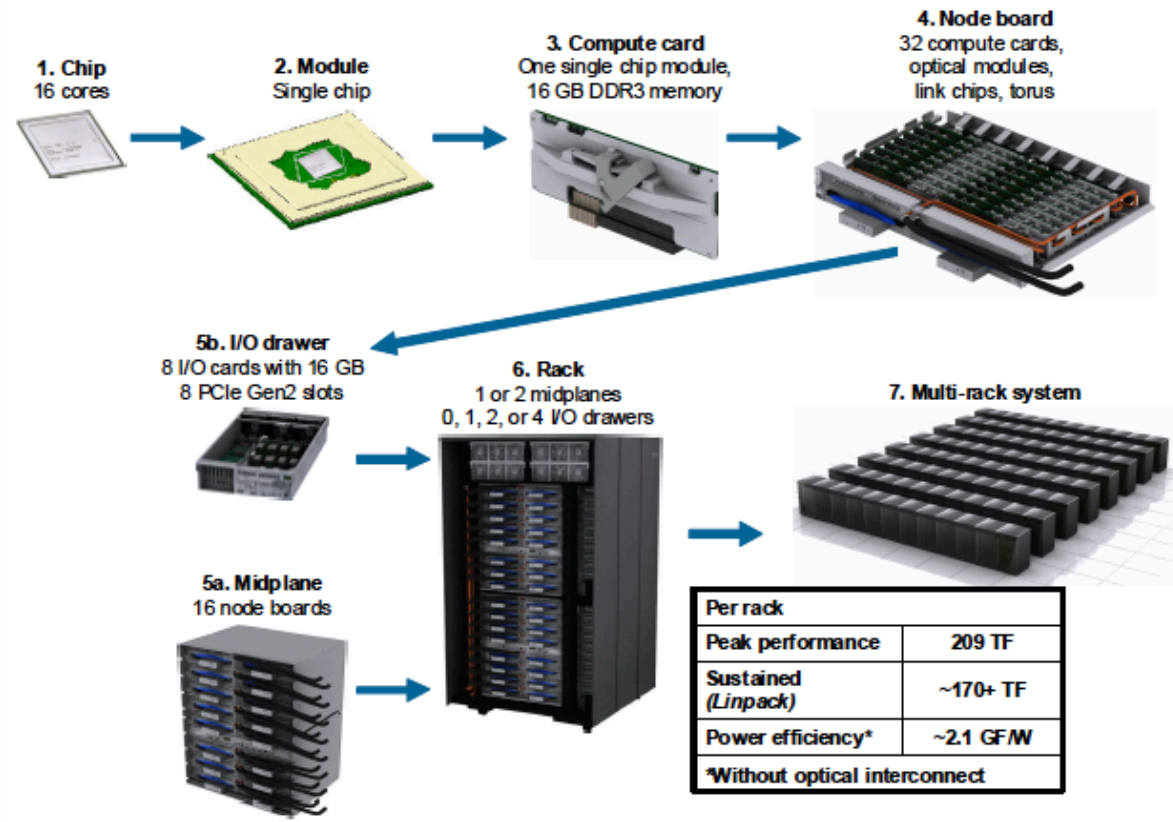
- Suitable for massively distributed supercomputer.
- Free energy calculations for complex biomolecular system.

Top500 and Green500

2004
Blue Gene/L

2007 Blue Gene/P
557 Teraflops
40*1024*4=163840
cores

2012 Blue Gene/Q
10 Petaflops
40*1024*16=655360
cores



Exascale computing
(10^18 operations/second)
over the next 5-10 years

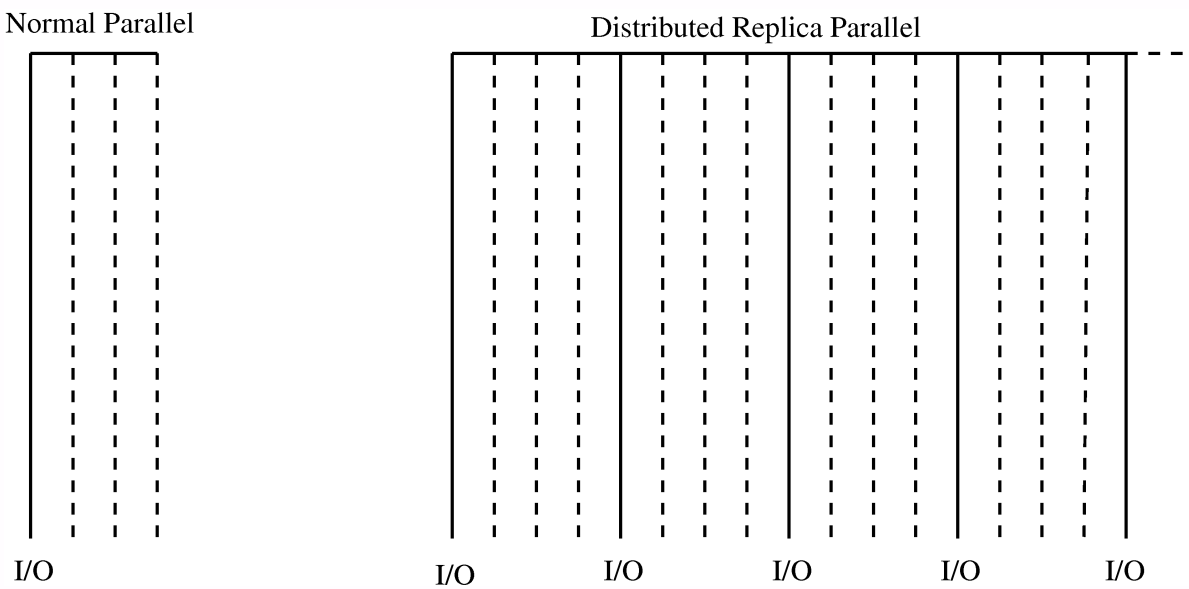
Hamiltonian-Exchange: Umbrella sampling (US) and FEP

$$\exp\left[\frac{1}{k_B T}\left(E(\lambda_m,\mathbf{r}_m)+E(\lambda_n,\mathbf{r}_n)-E(\lambda_m,\mathbf{r}_n)-E(\lambda_n,\mathbf{r}_m)\right)\right] \geq \text{random}(0,1)$$

Distributed Replica (REPDSTR) module

Multi-dimensional REMD on CHARMM

From Tcl script driven REMD to MPI-based REMD on NAMD

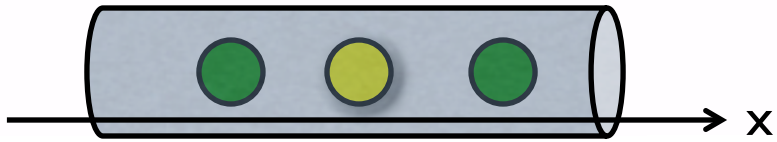


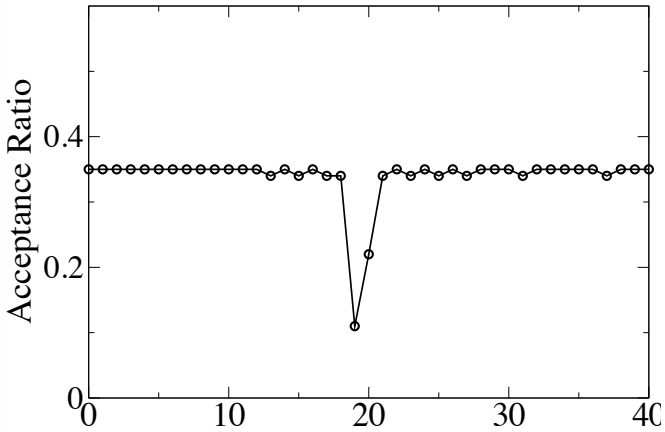
Postprocessing:
OpenMP or POSIX Thread weighted histogram analysis on shared memory nodes

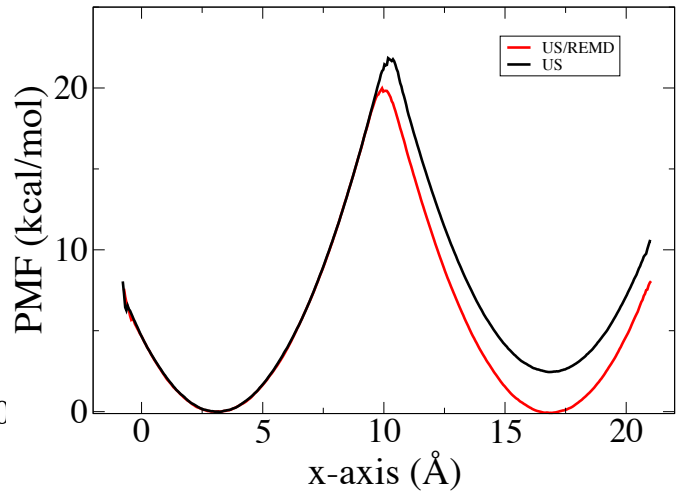
Jiang W, Hodoscek M, Roux.JCTC, 2009, 5, 2583-2588.

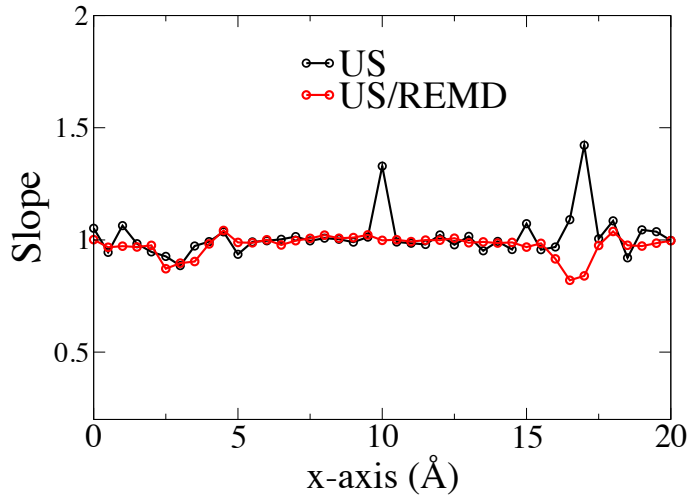
PMF/REMD Toy model analysis:

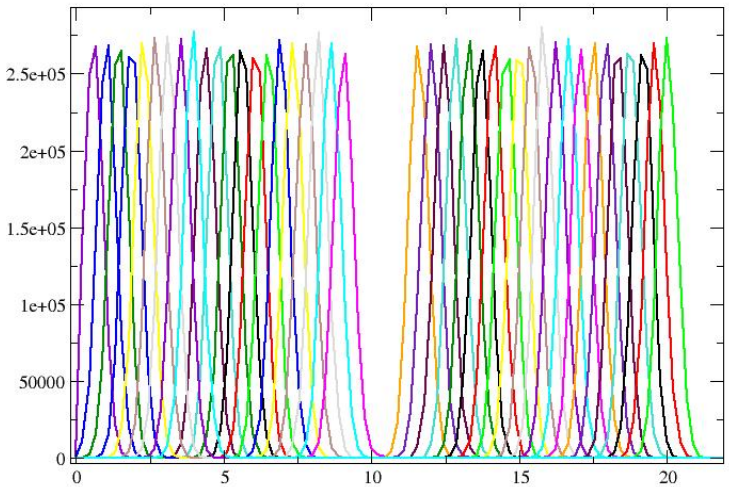
necessity but not sufficiency ?

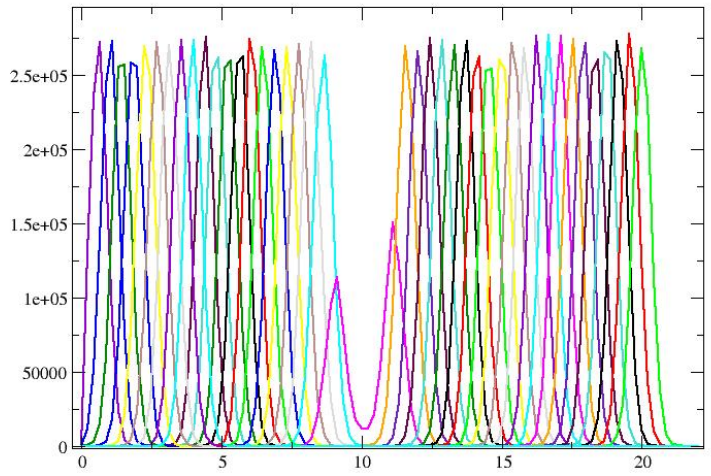












Binding Free Energy of Calbindin D_{9k}

$$k_{eq} = \frac{\int_{\text{site}} d1 \int dX e^{-\beta U}}{\int_{\text{bulk}} d1 \delta(r_1 - r_1^*) \int dX e^{-\beta U}}$$

$$= \frac{\int_{\text{site}} d1 \int dX e^{-\beta U}}{\int_{\text{site}} d1 \int dX e^{-\beta [U + u_a]}}$$

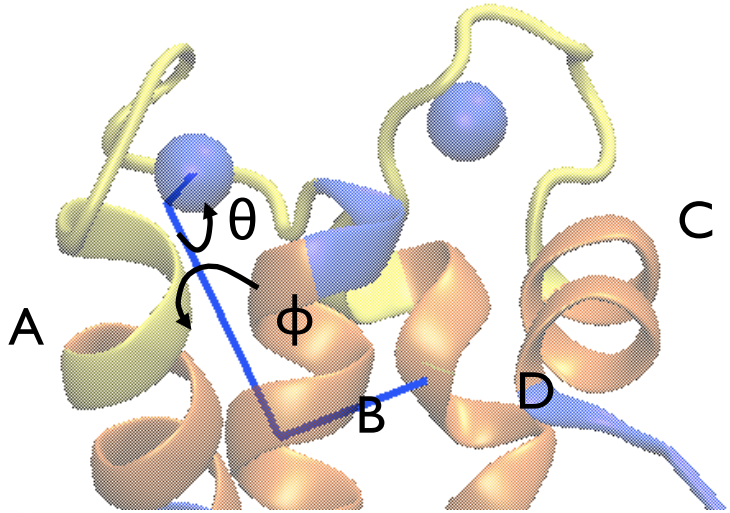
$$\times \frac{\int_{\text{site}} d1 \int dX e^{-\beta [U + u_a]}}{\int_{\text{bulk}} d1 \delta(r_1 - r_1^*) \int dX e^{-\beta U}}$$

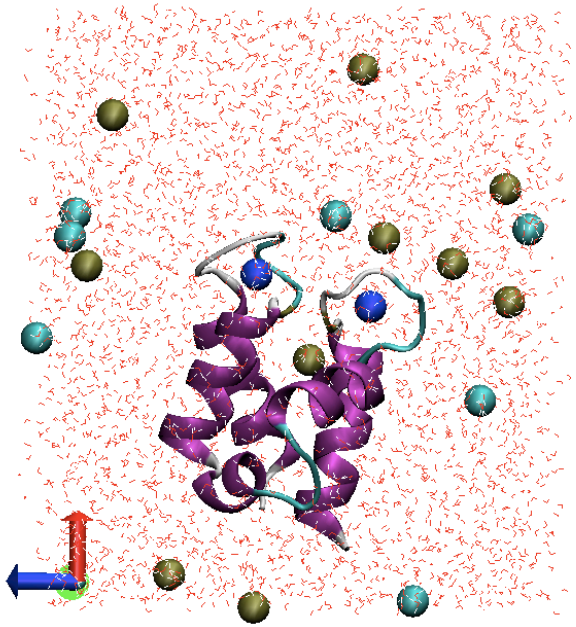
$$= e^{-\beta(-\Delta G_a^{\text{site}})} * I^* S^*$$

$$I^* = \int_{\text{site}} dr \; e^{-\beta [w(r) - w(r^*)]}$$

$$S^* = (r^*)^2 \int_0^{\pi} sin(\theta) d\theta \int_0^{2\pi} d\phi \; e^{-\beta u_a(\theta, \phi)}$$

$$\Delta G^0_{\text{bind}} \equiv -k_{\text{B}} T ln[k_{\text{eq}} C^0]$$





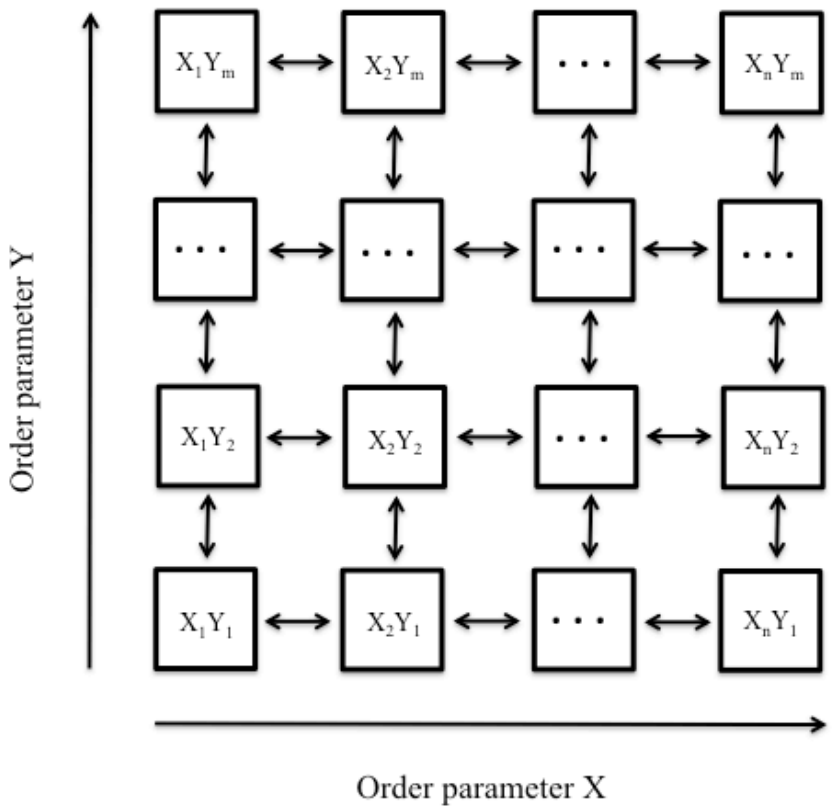
Woo H. Roux B., PNAS, 2005, 19, 6825

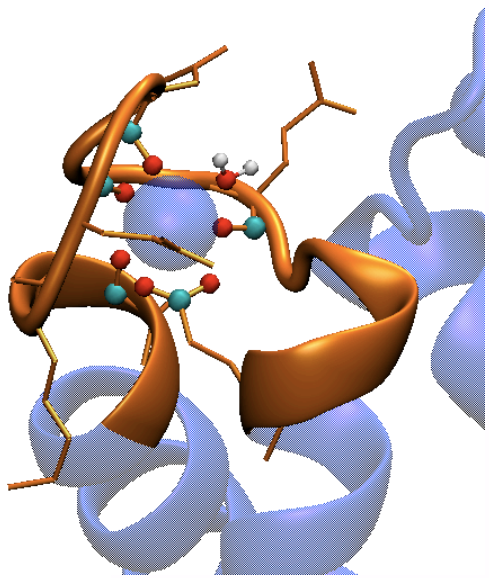
2D US/H-REMD

Order list of replicas is defined by $k_1(x-p_1)^2+k_2(y-p_2)^2$ $\{p_1^i, k_1^i, p_2^i, k_2^i\}$

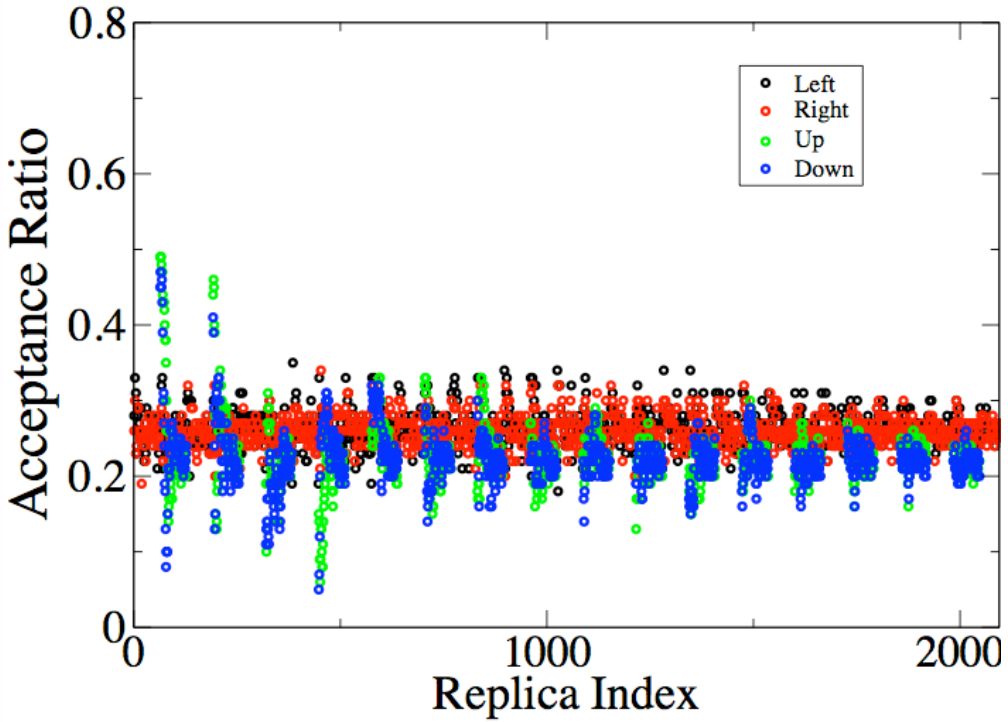
$$P(p_1^i \rightarrow p_1^j, p_2^l) = \min\left\{1, e^{-\left[U(p_1^i,p_2^l,\mathbf{r}_{i,l})+U(p_1^j,p_2^l,\mathbf{r}_{j,l})-U(p_1^j,p_2^l,\mathbf{r}_{j,l})-U(p_1^i,p_2^l,\mathbf{r}_{i,l})\right]/k_BT}\right\}$$

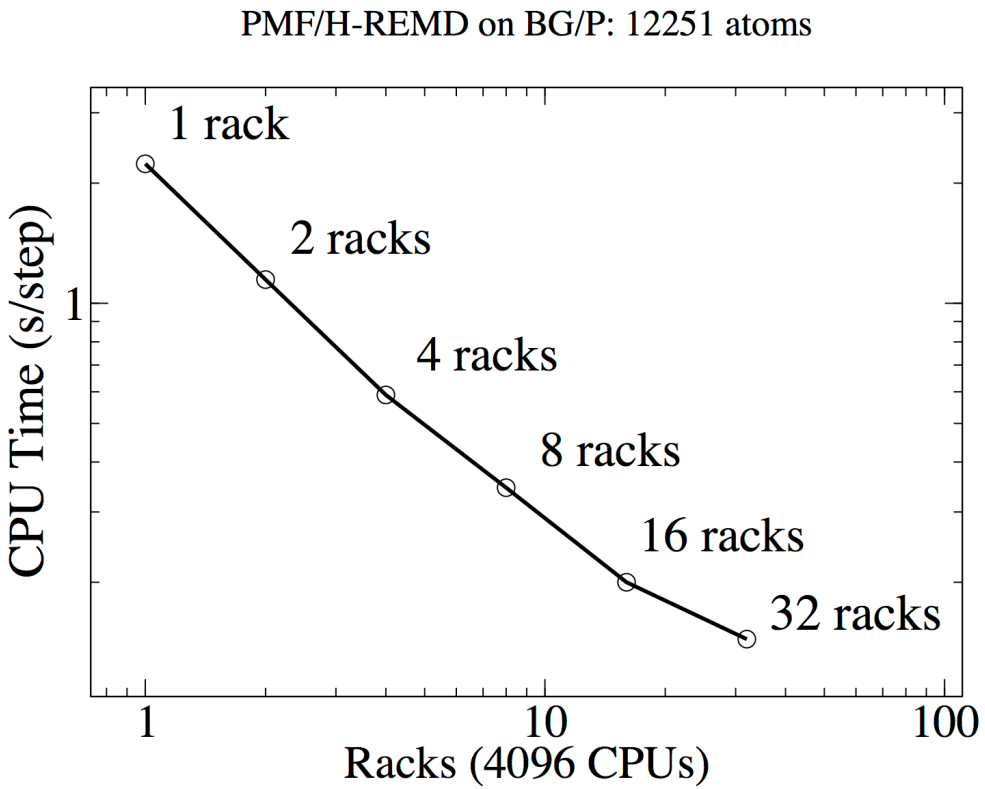
$$P(p_2^m \rightarrow p_2^n, p_1^k) = \min\left\{1, e^{-\left[U(p_1^k,p_2^m,\mathbf{r}_{k,m})+U(p_1^k,p_2^n,\mathbf{r}_{k,n})-U(p_1^k,p_2^m,\mathbf{r}_{k,n})-U(p_1^k,p_2^n,\mathbf{r}_{k,m})\right]/k_BT}\right\}$$

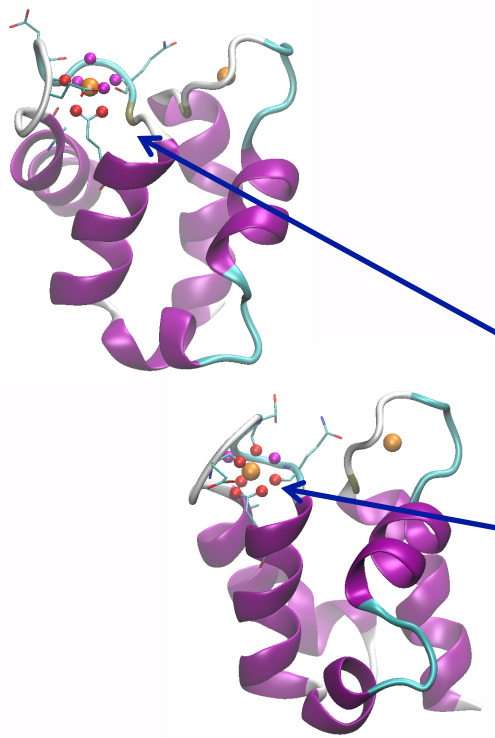


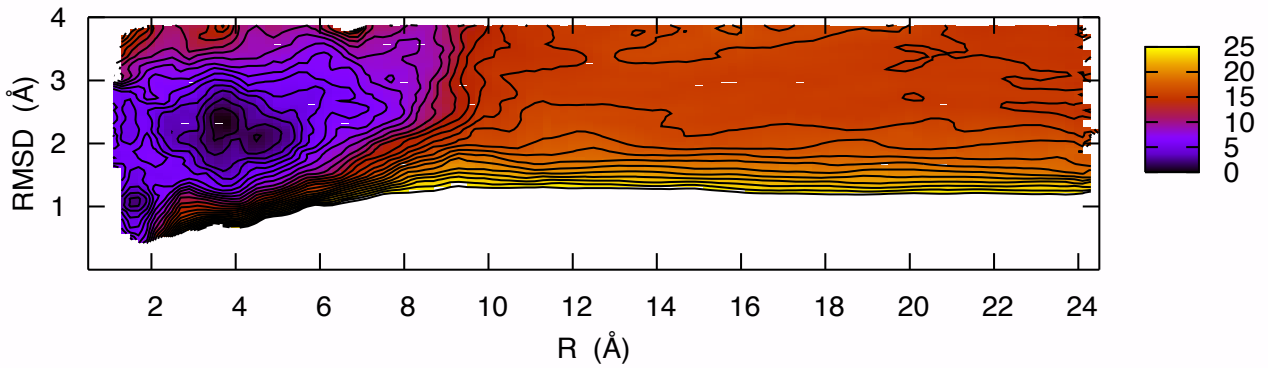


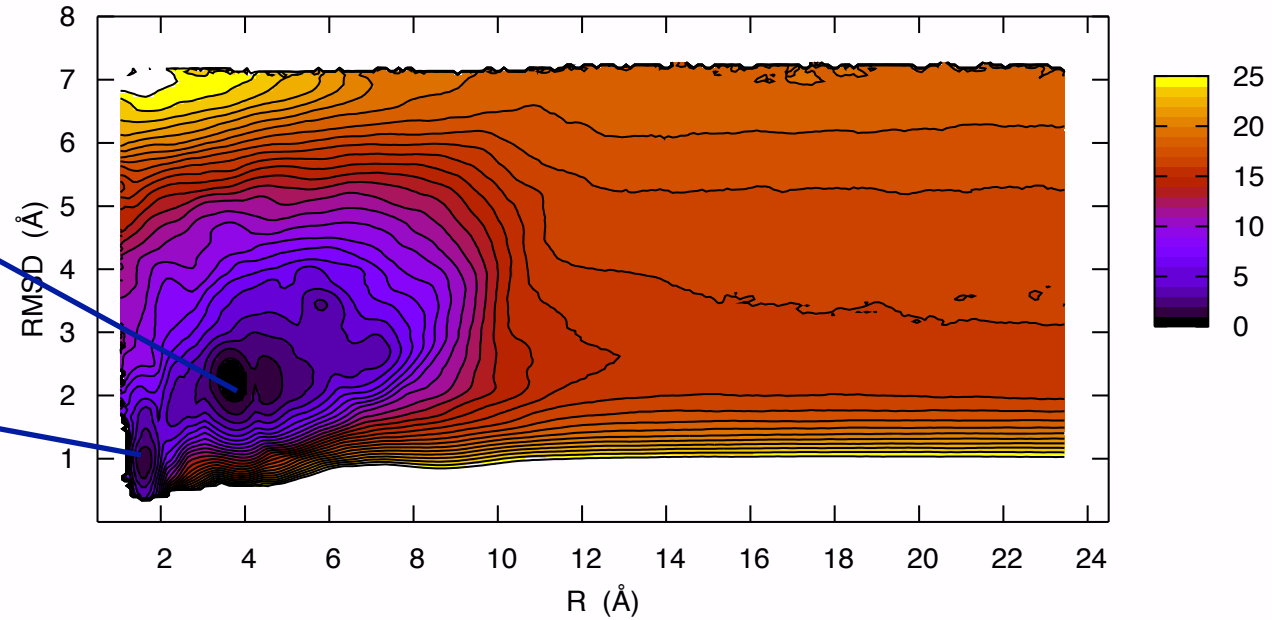
Acceptance Ratio & Strong Scaling











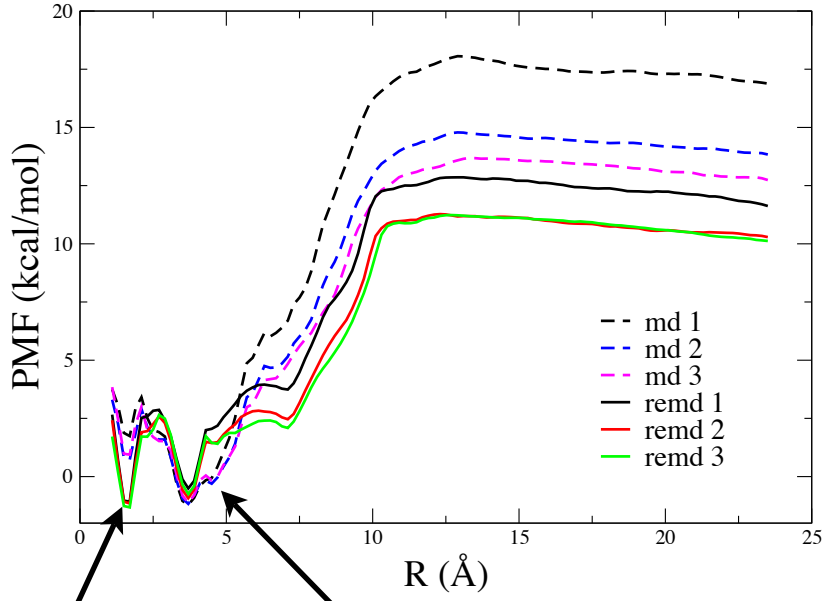
$$I^* = \int_{\text{site}} dr\ e^{-\beta[w(r)-w(r^*)]} = 1.35 \times 10^{10}\ \text{Å}$$

$$S^* = (r^*)^2 \int_0^{\pi} sin(\theta) d\theta \int_0^{2\pi} d\phi\ e^{-\beta u_a(\theta,\phi)} = 15.35\ \text{Å}^2$$

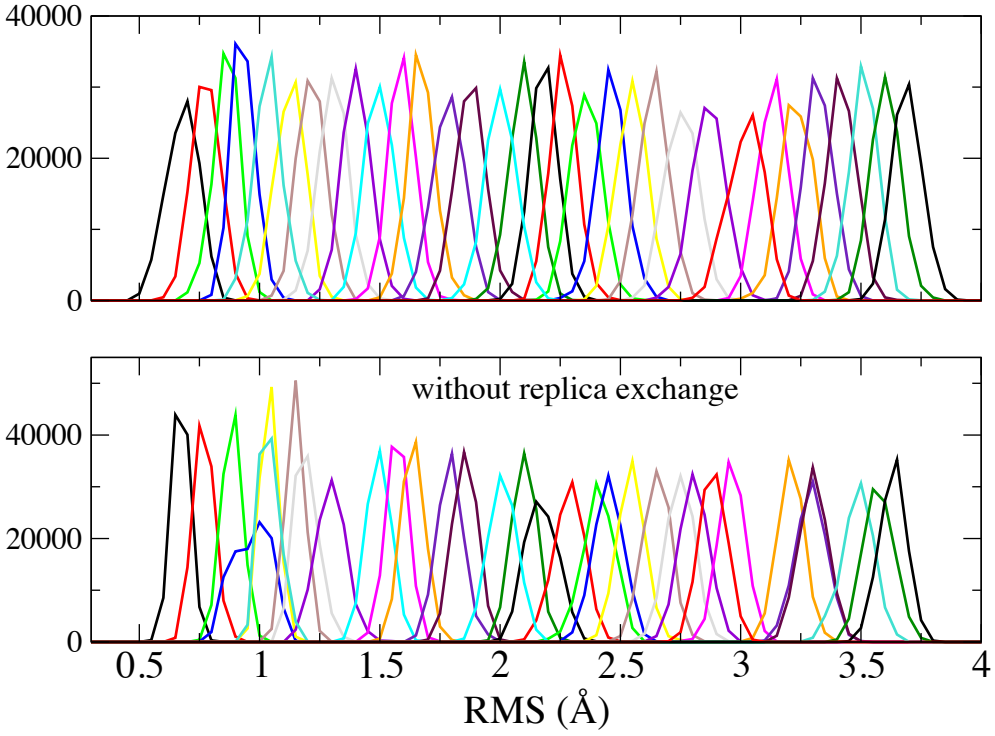
$$\begin{aligned} \Delta G^0_{\text{binding}} &= -k_B T \ln(I^* S^* C^0) - \Delta G_a^{\text{site}} \\ &= -11.1\ \text{kcal/mol} \end{aligned}$$

$$C^o = 1/1661\ \text{Å}^{-3} \qquad \Delta G_a^{\text{site}} = 0.19\ \text{kcal/mol}$$

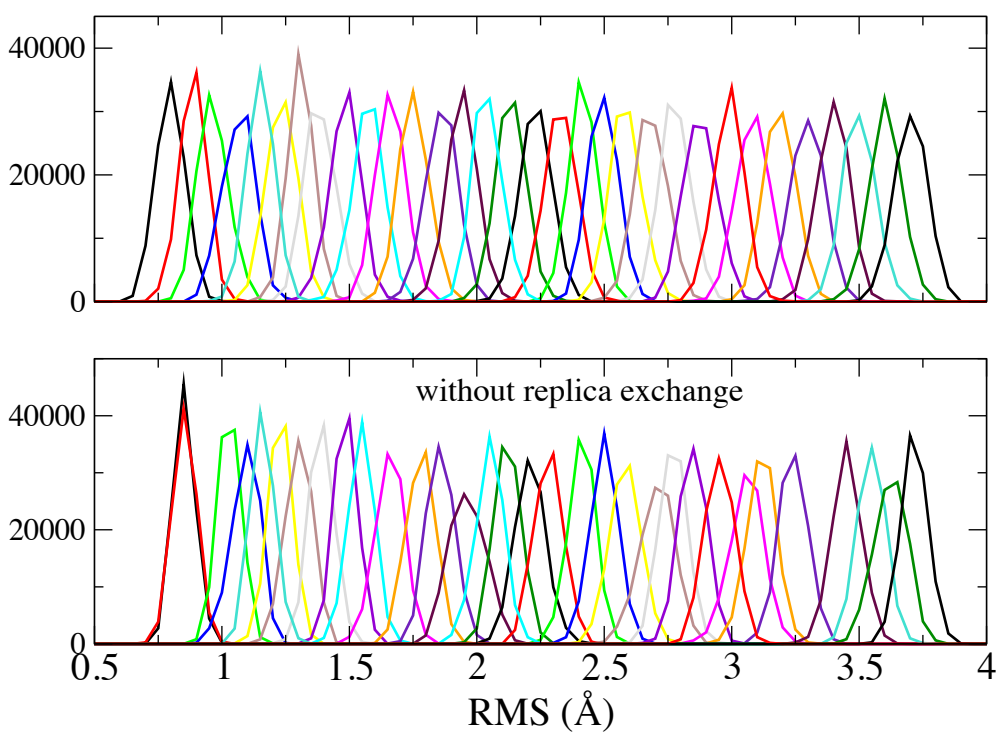
(Experimental value -9.4 kcal/mol)



Histogram of RMS replicas at distance 1.5 Å with REMD



Histogram of RMS replicas at distance 4.25 Å with REMD



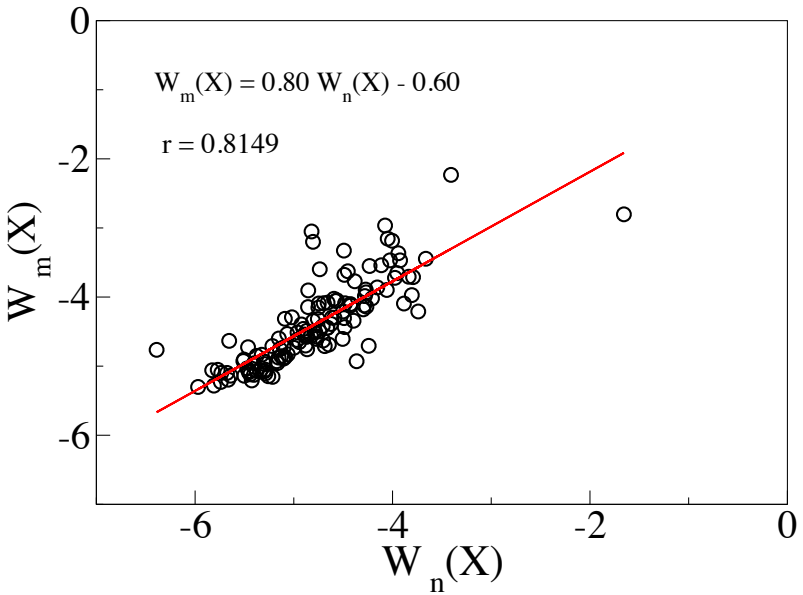
Analysis of PMF convergence

Slope $W_m = MW_n + B$

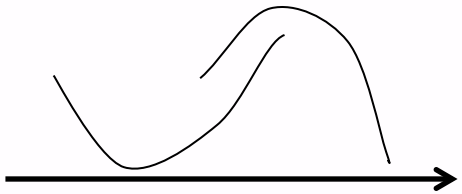
$$M_{nm} = \frac{\left(\sum_{\mathbf{x}} p_{nm}(\mathbf{x})W_n(\mathbf{x})W_m(\mathbf{x})\right) - \left(\sum_{\mathbf{x}} p_{nm}(\mathbf{x})W_n(\mathbf{x})\right)\left(\sum_{\mathbf{x}} p_{nm}(\mathbf{x})W_m(\mathbf{x})\right)}{\left(\sum_{\mathbf{x}} p_{nm}(\mathbf{x})W_n(\mathbf{x})W_n(\mathbf{x})\right) - \left(\sum_{\mathbf{x}} p_{nm}(\mathbf{x})W_n(\mathbf{x})\right)\left(\sum_{\mathbf{x}} p_{nm}(\mathbf{x})W_n(\mathbf{x})\right)}$$

$$p_{nm}(\mathbf{x}) = \frac{\rho_n(\mathbf{x})\rho_m(\mathbf{x})}{\sum_{\mathbf{x}} \rho_n(\mathbf{x})\rho_m(\mathbf{x})}$$

Variencs: $$\sigma^2_{nm} = \left(\sum_{\mathbf{x}} p_{nm}(\mathbf{x})W_n(\mathbf{x})W_n(\mathbf{x})\right) - \left(\sum_{\mathbf{x}} p_{nm}(\mathbf{x})W_n(\mathbf{x})\right)^2 + \left(\sum_{\mathbf{x}} p_{nm}(\mathbf{x})W_m(\mathbf{x})W_m(\mathbf{x})\right) - \left(\sum_{\mathbf{x}} p_{nm}(\mathbf{x})W_m(\mathbf{x})\right)^2$$



Variation big

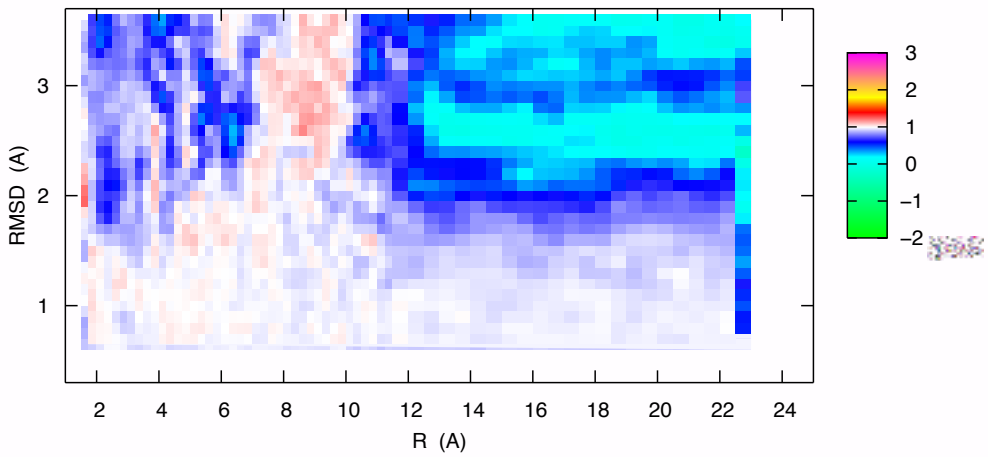


Variation small

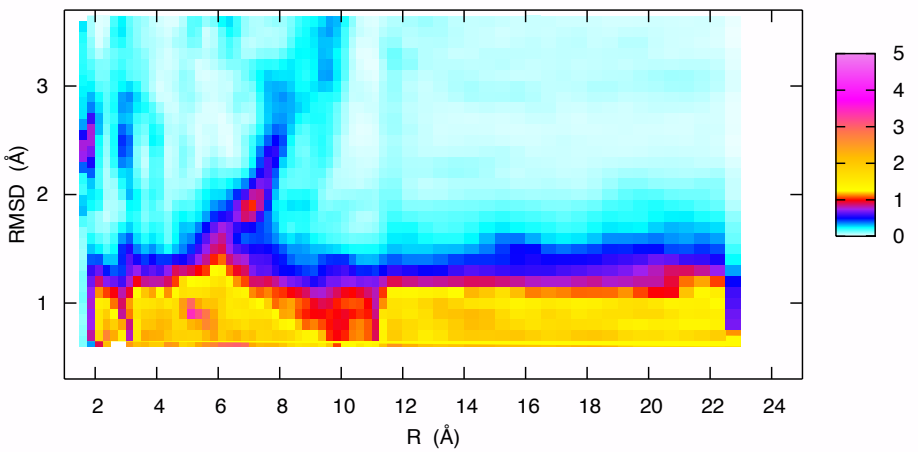
Analysis of PMF convergence: US/REMD *vs.* US

US/REMD

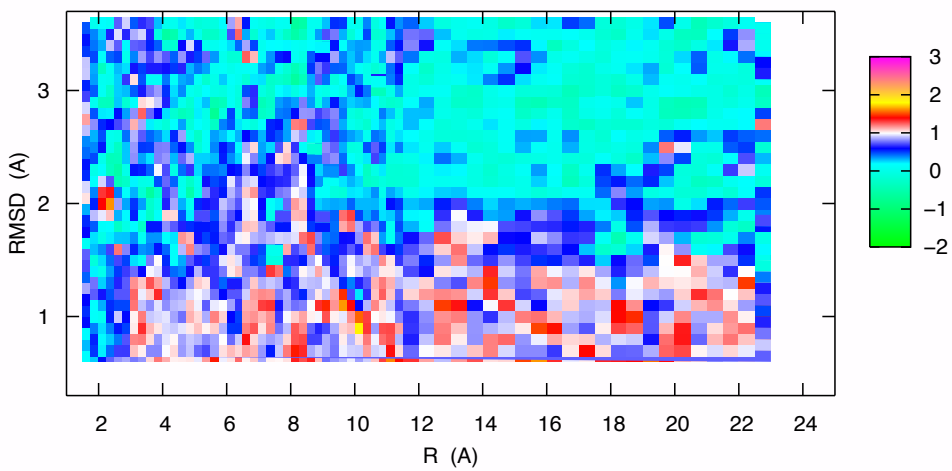
Slope

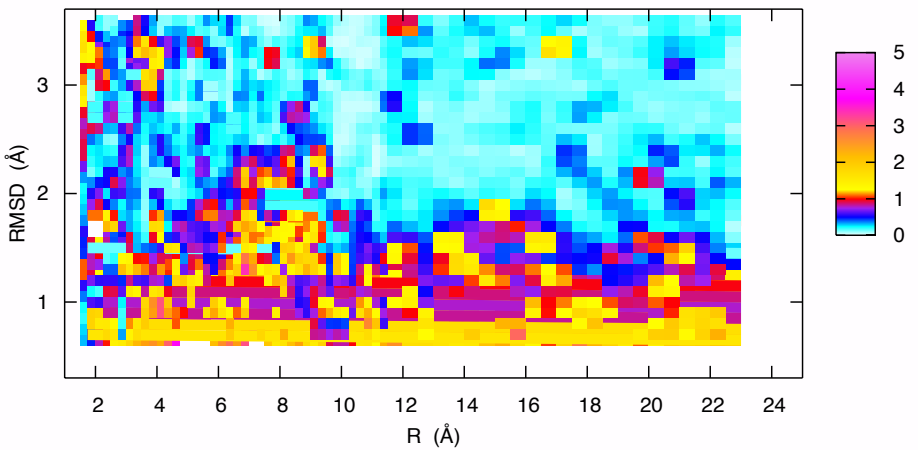


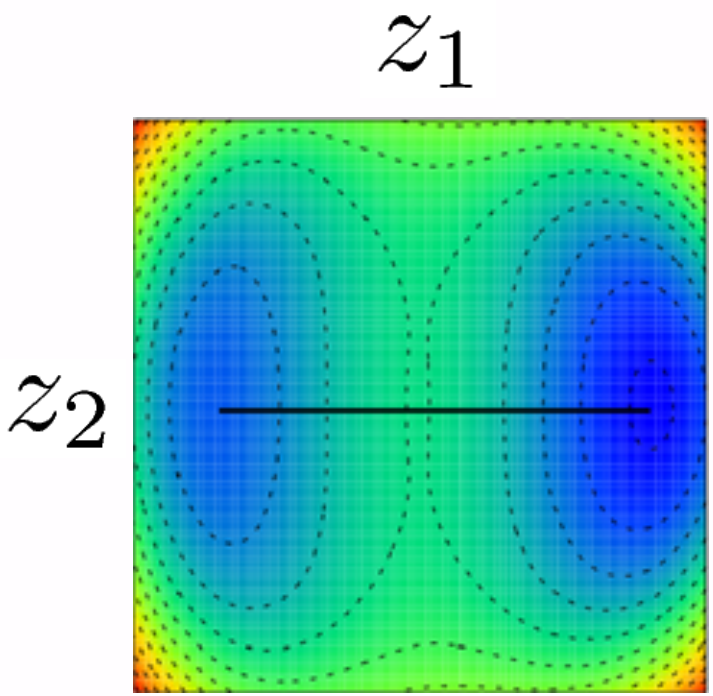
Variations

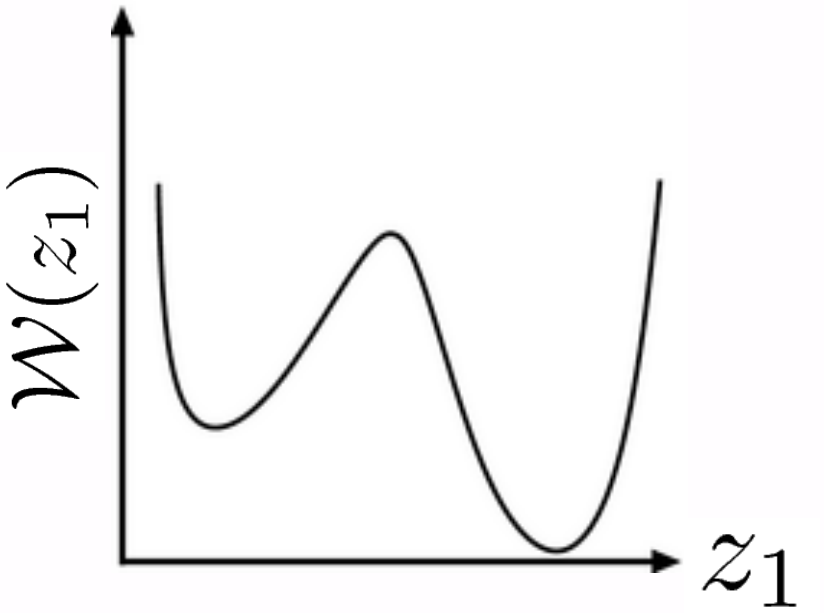


US

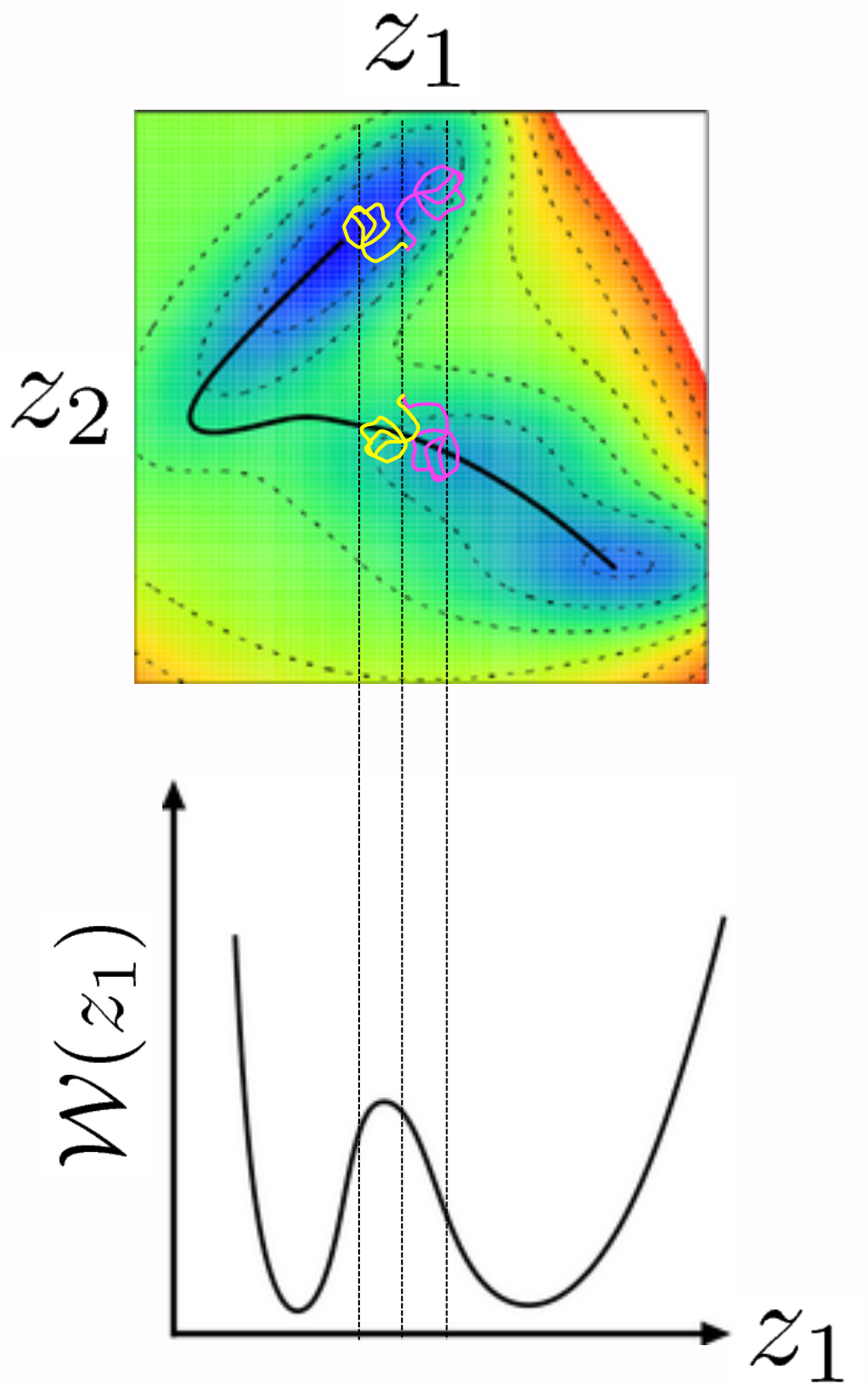








Trivial umbrella sampling in 1D



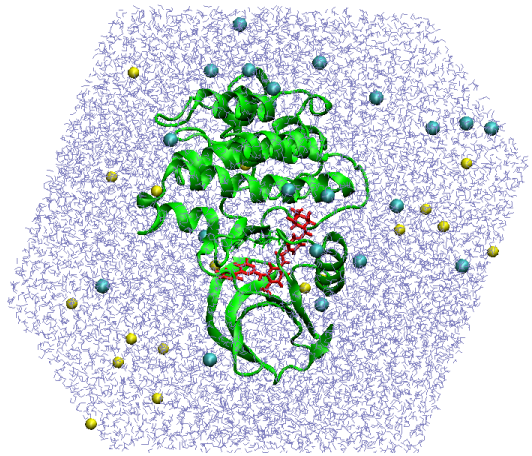
Sampling problems when all the collective variables relevant to the dynamical transition pathway are not included

Slowly varying orthogonal degrees of freedom

Degeneracy of Z_I

Binding Free Energy of Gleevec in Abl, c-Kit, Lck, and c-Src

$$U(\lambda_{rep},\lambda_{dis},\lambda_{elec},\lambda_{rstr}) = U_0 + U_{rep}(\lambda_{rep}) + \lambda_{dis}U_{dis} + \lambda_{elec}U_{elec} + \lambda_{rstr}U_{rstr}$$



FEP/λ-REMD scheme

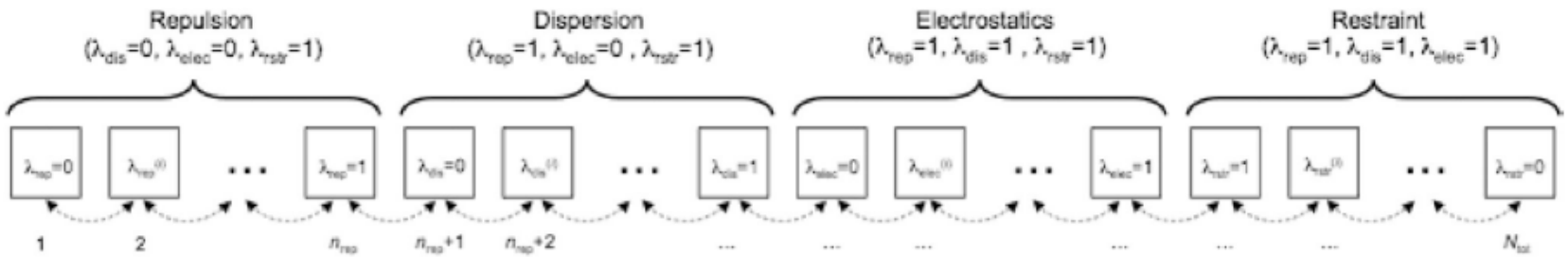
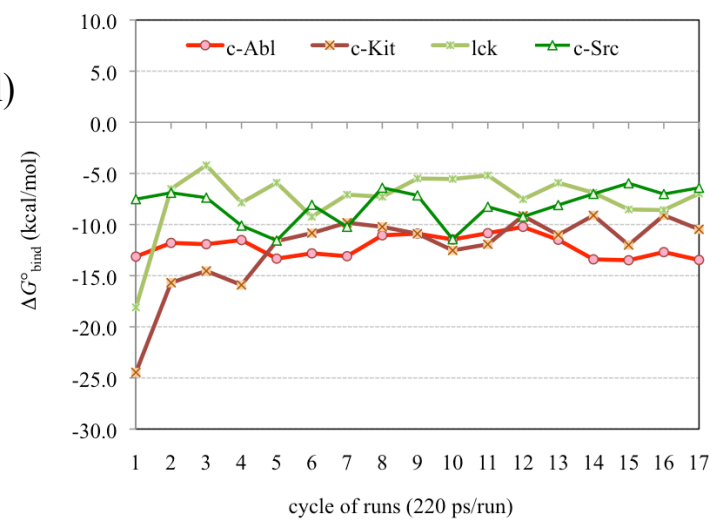
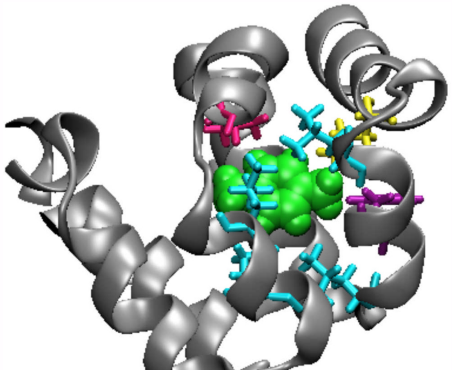


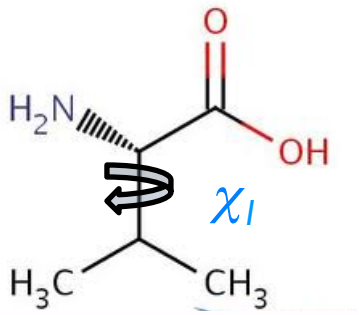
Table: Absolute Binding Free Energy of Gleevec With Src-family kinases (unit: kcal/mol)

	$\Delta G°_{bind}$	ΔG_{expt}	
		ref. 6	*ref. 7*
bulk → Abl	–11.8	–10.8	–8.2
bulk → c-Kit	–**1**.4	––	–7.4
bulk → Lck	–7.3	–8.7	–6.9
bulk → c-Src	–6.8	–6.2	–4.1



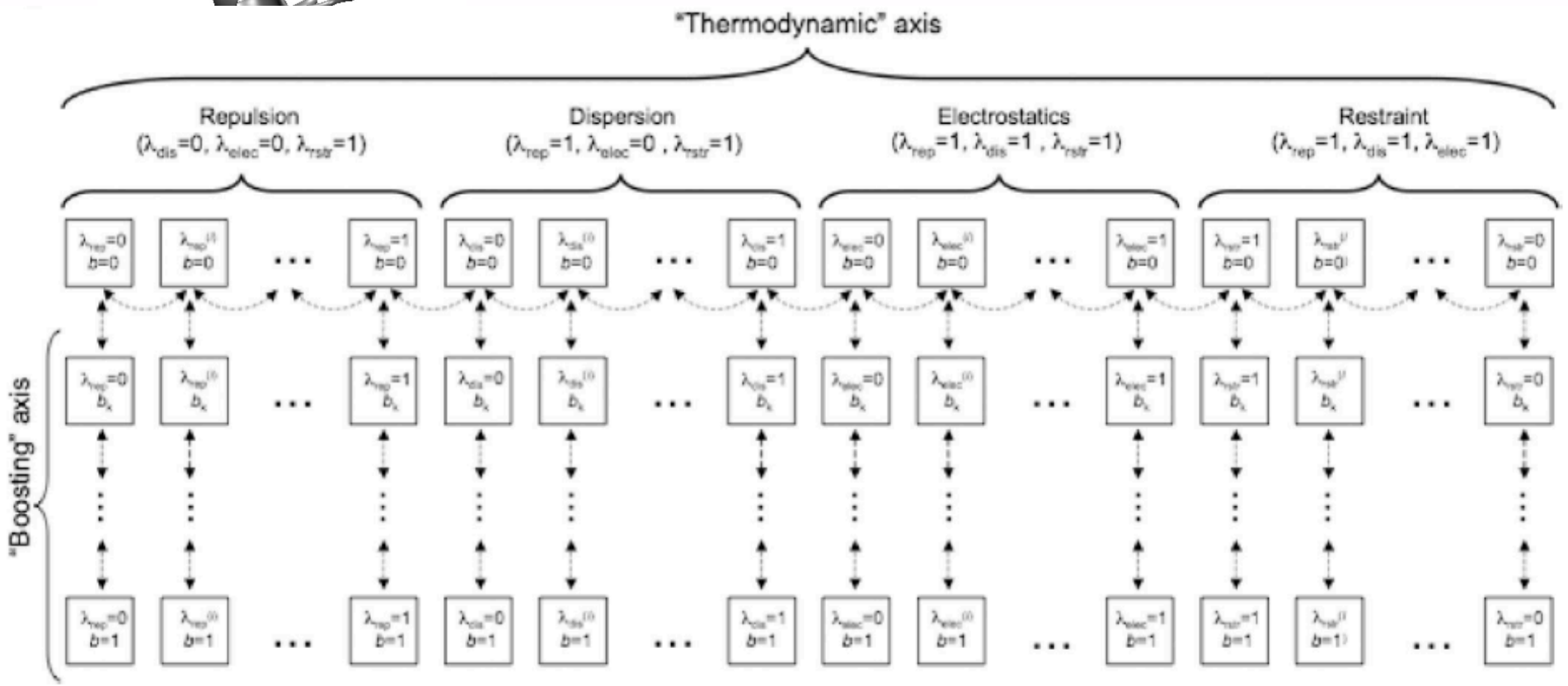
2-Dimensional FEP/H-REMD





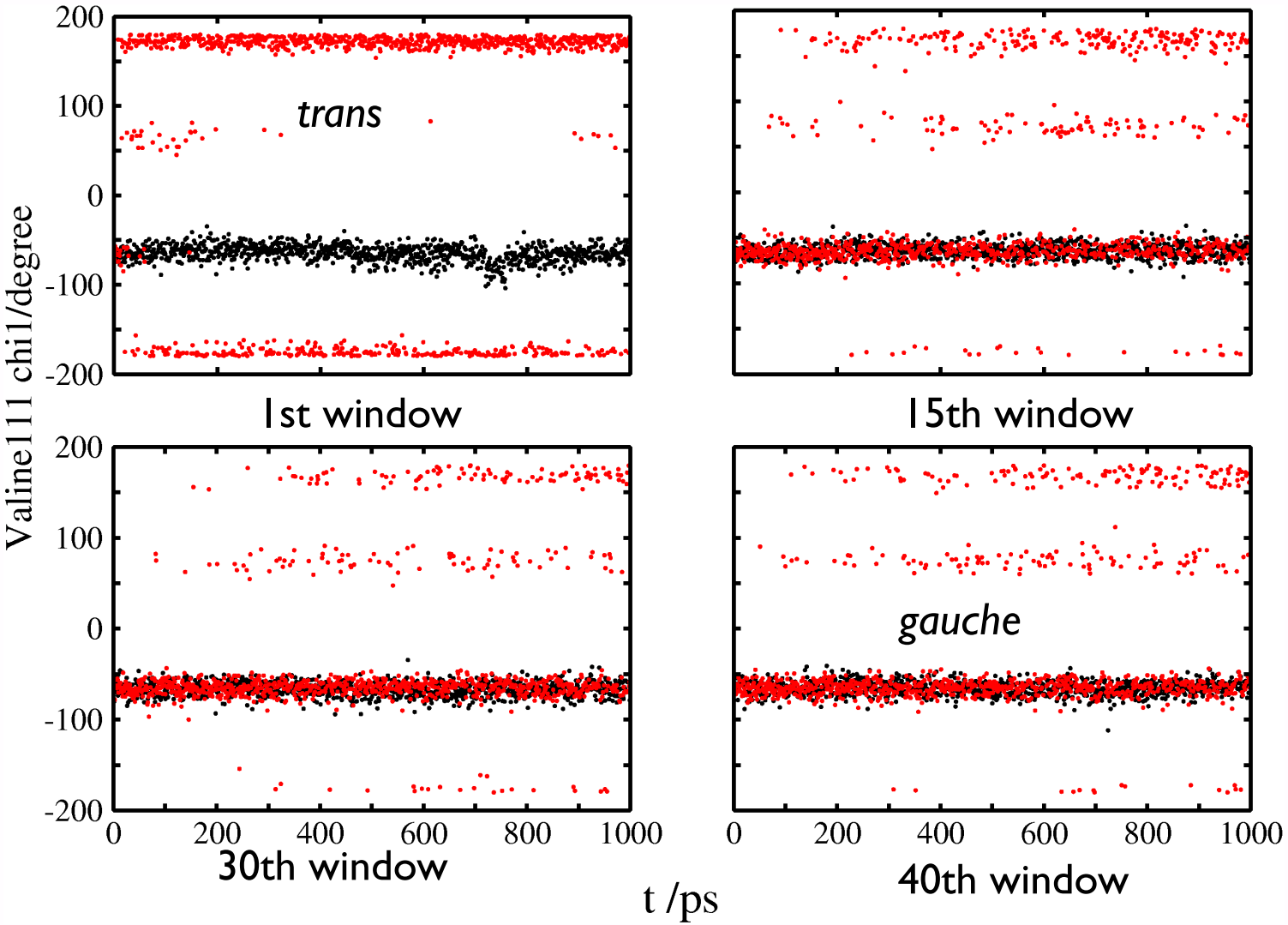
Binding of p-xylene to T4 lysozyme

$$V(\phi) = \sum_{n=1}^{3} K_n(1 + \cos(n(\phi - \phi_{\min})))$$

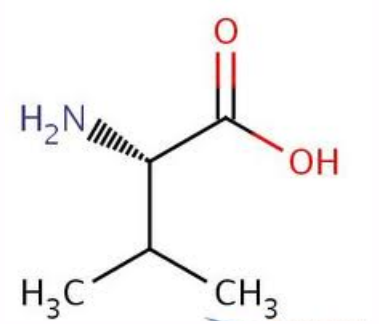


Jiang W, Roux . JCTC, 2010, 6, 2559-2565.

Enhanced χ_1 Sampling by FEP/H-REMD



- FEP/λ-REMD
- FEP/H-REMD



$$\Delta G_{\text{bind}} = -5.1 \text{ kcal/mol}, \ (\text{exp} : -4.7 \text{kcal/mol})$$

Conclusions and Perspectives

MPI level REMD is implemented in CHARMM and NAMD (including threaded version), suitable for petascale supercomputing platform

Two case studies prove the enhance sampling and extremely scalability of two free energy calculation methods coupled with REMD.
2D US/H-REMD: Calcium ion binding Calbindin
2D FEP/H-REMD: *p*-xylene binding T4 Lysozyme

Undergoing Improvements:

- Irregular shape multi-dimensional US/REMD by assigning an array of neighboring replicas
- String methods or coarse grain transition path search to define several initial path, and expanding replicas along the path way.
- Performing PCA to determine few order parameters out of high dimensional space

Argonne NATIONAL LABORATORY

NAMD Scalable Molecular Dynamics

Benoit Roux
Wei Jiang
Yen-lin Lin
Lei Huang
Luca Maragliano
Roux Group members

Klaus Schulten
James Phillips

cecam Centre Européen de Calcul Atomique et Moléculaire

Chris Chipot
Tony Lelièvre
Andrew Pohorille

ANL Leadership Computing Facility

The Early Science Program (ESP)

U.S. DEPARTMENT OF ENERGY