# Random walk in orthogonal space to achieve efficient free-energy simulation of complex systems

by **Lianqing Zhenga, Mengen Chena, and Wei Yang**

Presented by Michael J. Schnieders
to the Ren Lab
May 2, 2011

# Introduction

- Goal: Compute the free energy difference between 2 states
  1. Alchemical
     - Path between physical end-states can be unphysical
     - Define a parameter $\lambda$ that controls state
     - Example: protein-ligand binding free energy
       - At $\lambda = 1$ the protein and ligand interact
       - At $\lambda = 0$ the protein and ligand do not interact
       - Intermediate $\lambda$ values can specify unphysical potentials
         - Softcore van der Waals and/or electrostatics
         - No ligand electrostatics
  2. Conformational

# The Usual Approaches

- Sample at end-states and alchemical intermediate states
- Calculate the free energy between adjacent states using:
  1. Free Energy Perturbation (FEP) from the Zwanzig equation:

$$\Delta G(A \to B) = G_B - G_A = -k_B T \ln \left\langle \exp\left(-\frac{E_B - E_A}{k_B T}\right)\right\rangle_A$$

     - Bennett-Acceptance Ratio (BAR) optimally use trajectories A *and* B
     - Multi-state BAR (MBAR) optimally use *all* trajectories

  1. Thermodynamic Integration (TI)

$$F(\lambda = 1) - F(\lambda = 0) = \int_{\lambda=0}^{\lambda=1} d\lambda \left\langle \frac{\partial U(\lambda)}{\partial \lambda} \right\rangle_\lambda$$

# Caveats / Questions

1. Must have overlap between the partition functions of neighboring intermediate states
   - Rule of thumb: perturbation steps can be a few kT
   - Larger perturbations mean less overlap and slower convergence
   - **How do we know *a priori* the set of λ to choose?**

2. All important configurations must be sampled at eachλ
   - There can be large free energy barriers
     - Protein side chain rotamers (or even backbone rearrangements)
   - **How do we get over barriers to converge quickly?**

3. Hamiltonian Lagging
   - **If our system is equilibrated at λ=1, how long does equilibration take upon switching to λ= 0?**

# Lambda Dynamics by Kong and Brooks (1996)

- Treat λ as a dynamic variable
  - Known as an *extended system* approach
  - λ has a mass (M), velocity, etc.
- Force on the λ particle from the "real system"
  - $F_\lambda = -dU(X, \lambda) / d\lambda$
- Propogate λ using Langevin dynamics

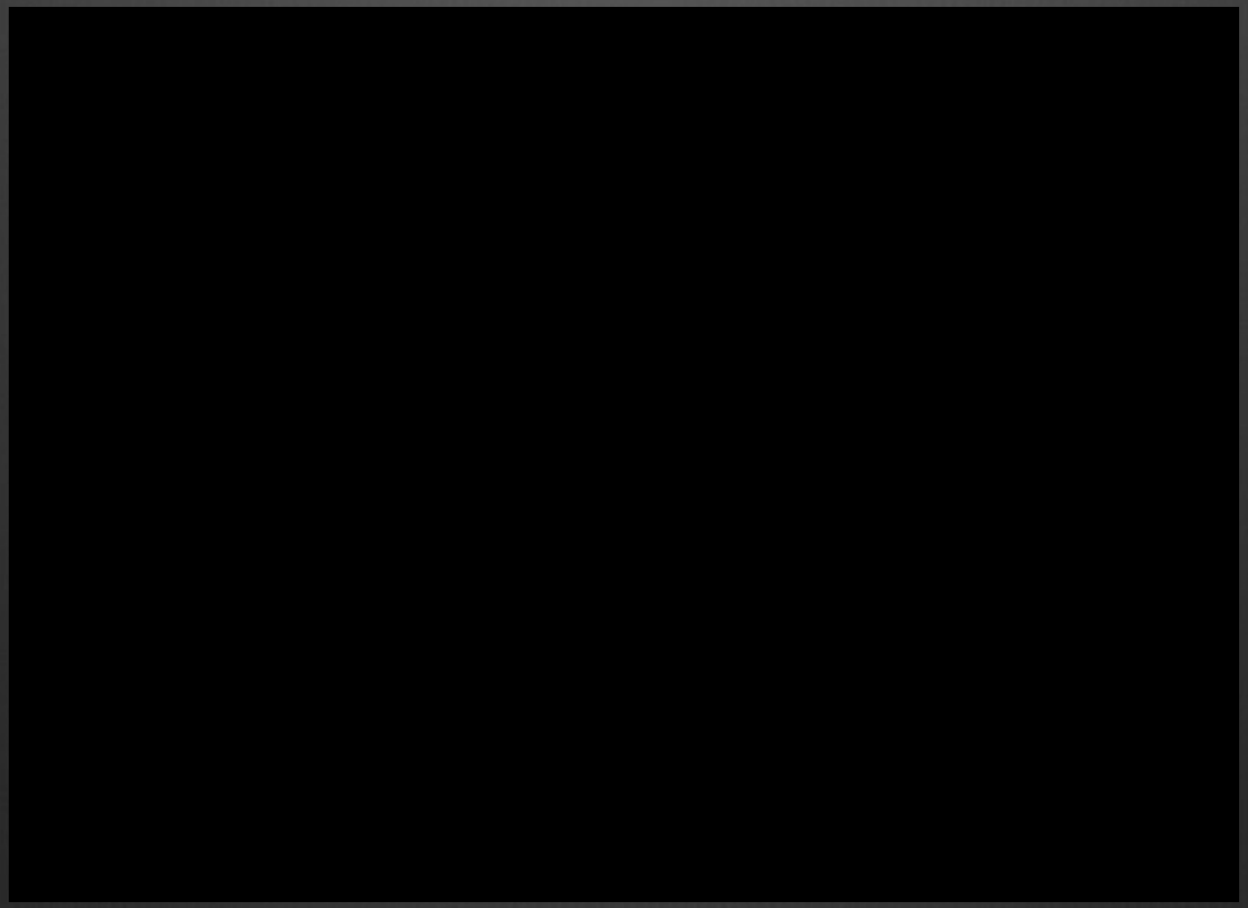
$$M\ddot{X} = -\nabla U(X) - \gamma M\dot{X} + \sqrt{2\gamma k_B T M} R(t),$$

  - First term is $F_\lambda$
  - Second is friction
  - Third is random

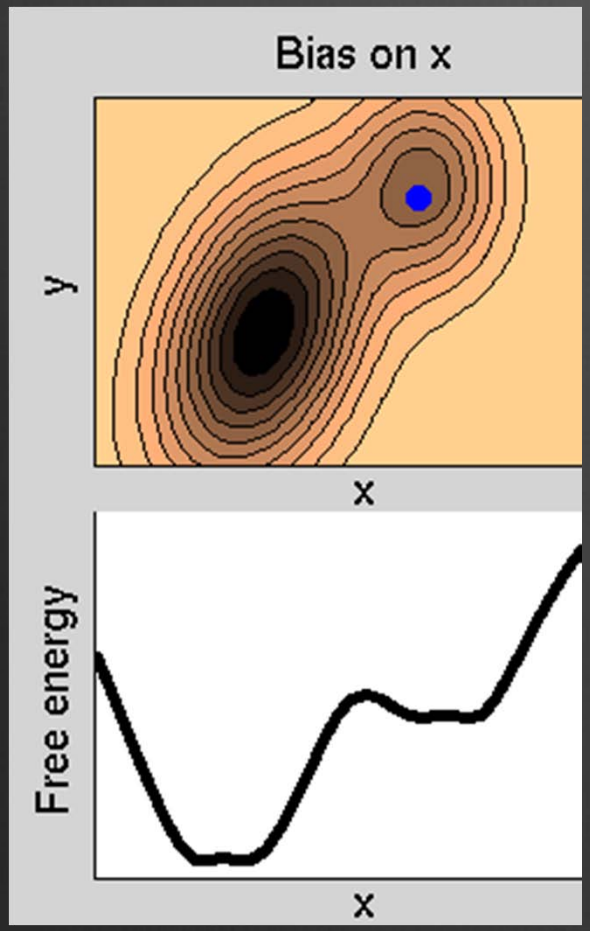
# Metadynamics by Laio and Parrinello (2002)

- Imagine λdynamics for the solvation of Na+ → Mg++
  - Force on the λ particle will push it to Mg++ state
  - Little (or no) sampling of Na+
- Solution – add a positive Gaussian to the total potential every ~100 steps
  - Gaussians accumulate for low free energy values of λ
  - Eventually, all λ values are sampled equally
    - **A Random Walk in λ!**
- At the end, the sum of the Gaussians provides the negative image of the free energy

# Example Courtesy of Prof. Alessandro Laio



# Example 2 Courtesy of Prof. Alessandro Laio




- 2D potential energy surface U(x,y)
- Order parameter $\lambda$=x
- Bias G($\lambda$,t) is be built up until G($\lambda$,t) + F($\lambda$) becomes flat
- Then -G($\lambda$,t) = F($\lambda$) + C
- What if there are large barriers along the y-coordinate?

# Orthogonal Space Random Walk

- Construct a 2D bias along **both** $\xi$ *and* $F_\xi$

$$G(\xi, F_\xi) = \sum_{t_i} h \exp\left( - \frac{|\xi - \xi(t_i)|^2}{2w_1^2} \right) \exp\left( - \frac{|F_\xi - F_\xi(t_i)|^2}{2w_2^2} \right)$$

- The free energy derivative at $\xi=\xi'$ is the ensemble average of $F_\xi$ at $\xi'$

$$\left. \frac{dG}{d\xi} \right|_{\xi'} = \langle F_\xi \rangle_{\xi'} = \frac{\sum F_\xi \, exp[\beta G(\xi, F_\xi)] \delta(\xi - \xi')}{\sum exp[\beta G(\xi, F_\xi)] \delta(\xi - \xi')}$$

- Using the TI formula, the free energy difference between $\xi_i$ and $\xi$is

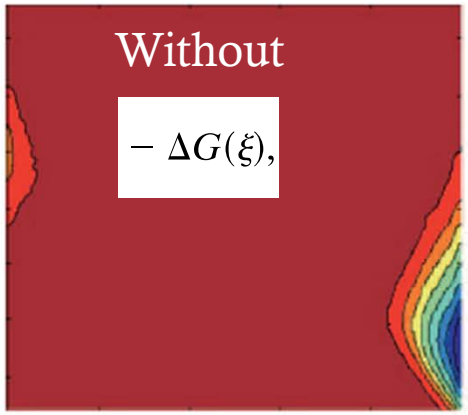
$$\Delta G(\xi) = \int_{\xi_i}^{\xi} \left. \frac{dG}{d\xi} \right|_{\xi'} d\xi'$$

# Orthogonal Space Random Walk (OSRW)

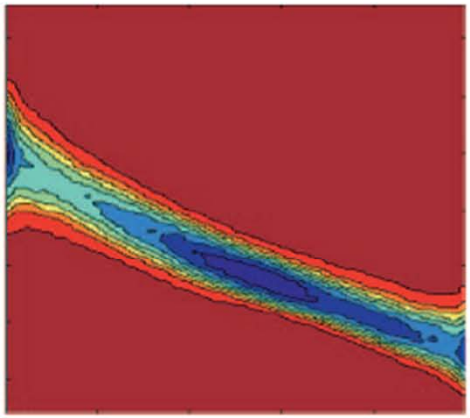
- Total OSRW potential:

$$U = U_r(\xi) + U_e + G(\xi, F_\xi) - \Delta G(\xi), \qquad [5]$$

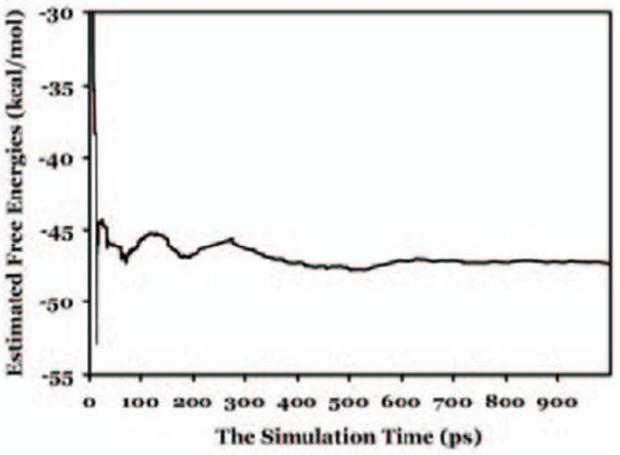
- Transformation of leucine ➔ asparagine ~0.6 nsec




(a)



(b)

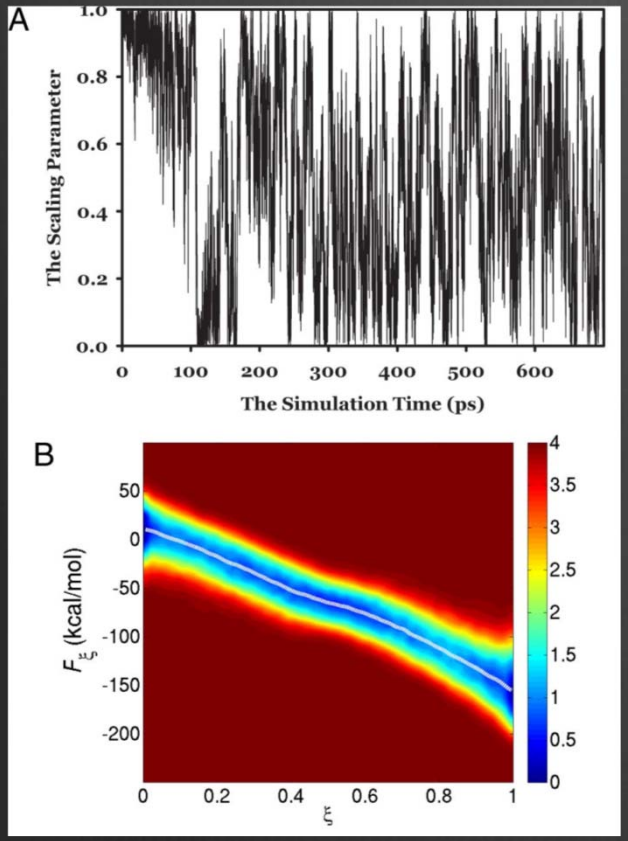



(c)

# Applications of OSRW

- Transformation of leucine ➔ asparagine
- pKa of value of a buried ionizable residue: Asp-66 of the V66D variant of staphylococcal nuclease (SNase)
- PMF for separation of sodium and chloride in water
- Redox Potentials (QM/MM)
- Tautomerization (QM/MM)

# Transformation of leucine ➔ asparagine



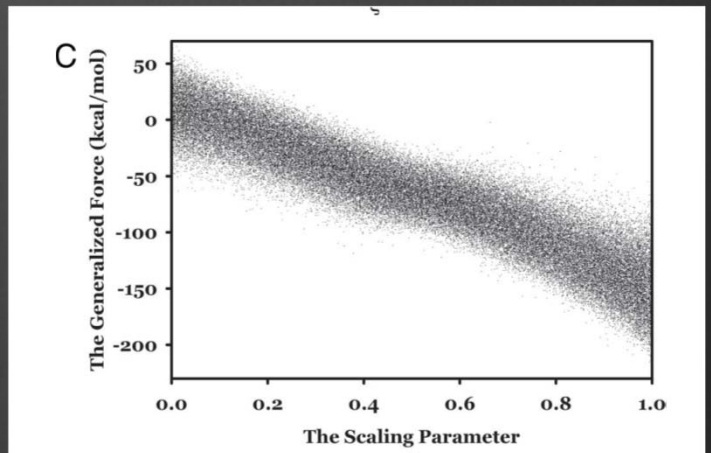




**Fig. 1.** The behaviors of the OSRW simulation on the solution system in model study 2. (*A*) The time-dependent scaling parameter change. (*B*) The free-energy surface along the orthogonal space produced by 300-ps OSRW simulation; in the middle, the white dash curve shows the estimated free-energy derivatives. (*C*) The distribution of the samples collected during the OSRW simulation in the orthogonal space.

# pKa of Asp-66 in SNase

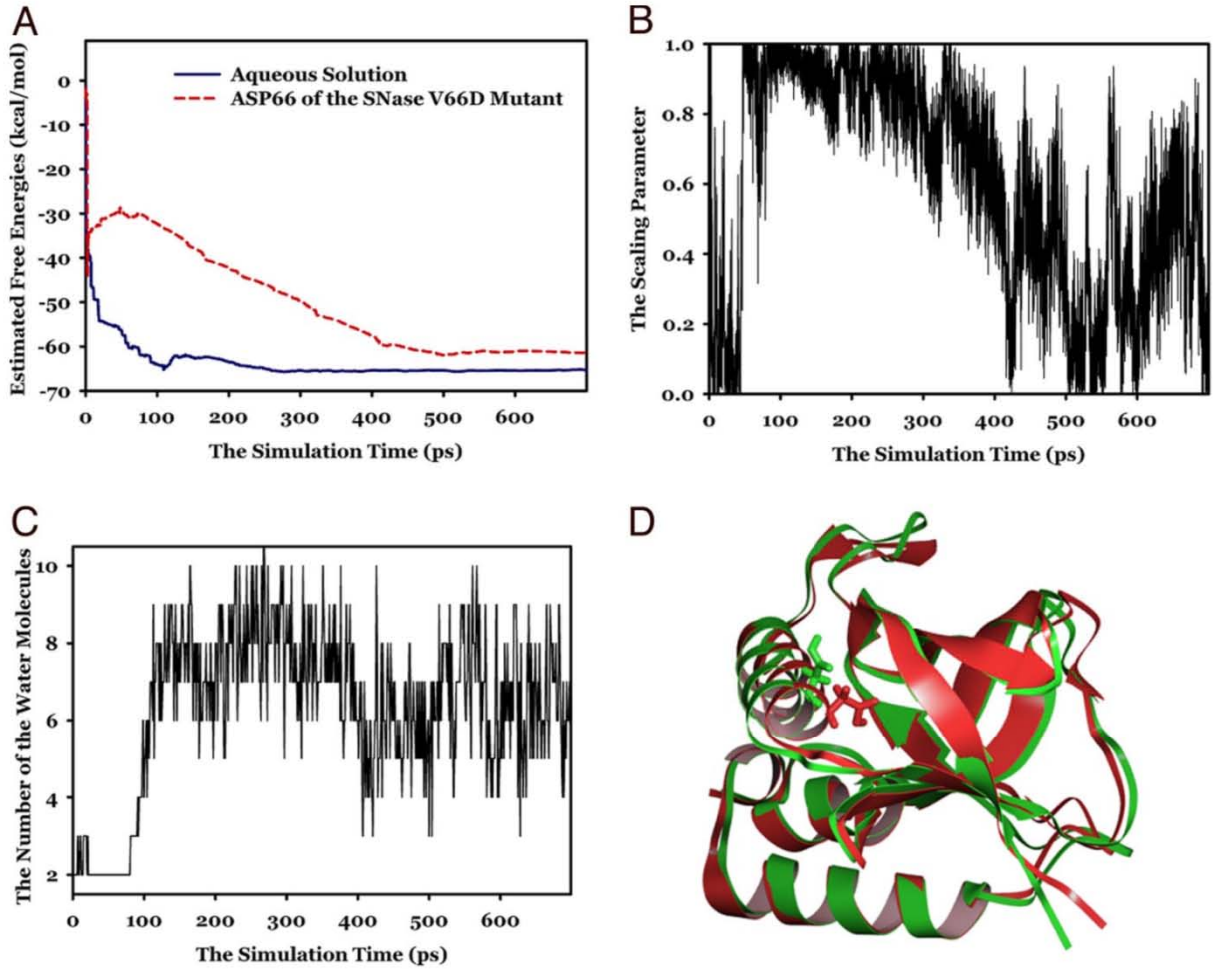



**Fig. 2.** The free-energy simulation results and the behaviors of the OSRW simulation on the protein system in model study 2. (*A*) The time-dependent free-energy values. Dashed lines show the results on the protein (the SNase mutant) system; solid lines show the results on the solution system. (*B*) The time-dependent scaling parameter changes. (*C*) The time-dependent curve on the number of the water molecules that are exposed to Asp-66. (*D*) The structures of the SNase V66D mutant at the beginning and the end of the OSRW simulation.

# PMF for NaCl in Water OSRW vs. Metadynamics

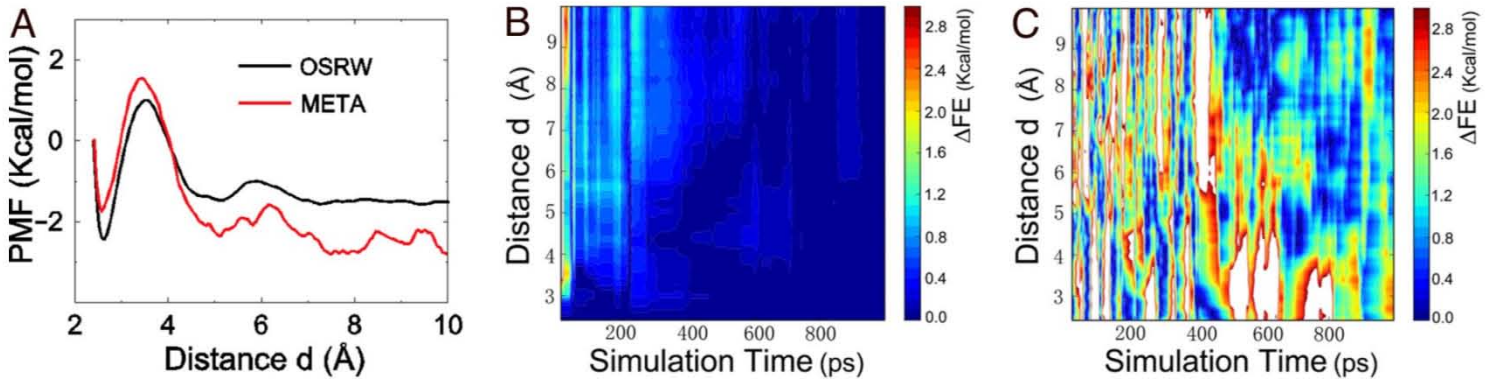
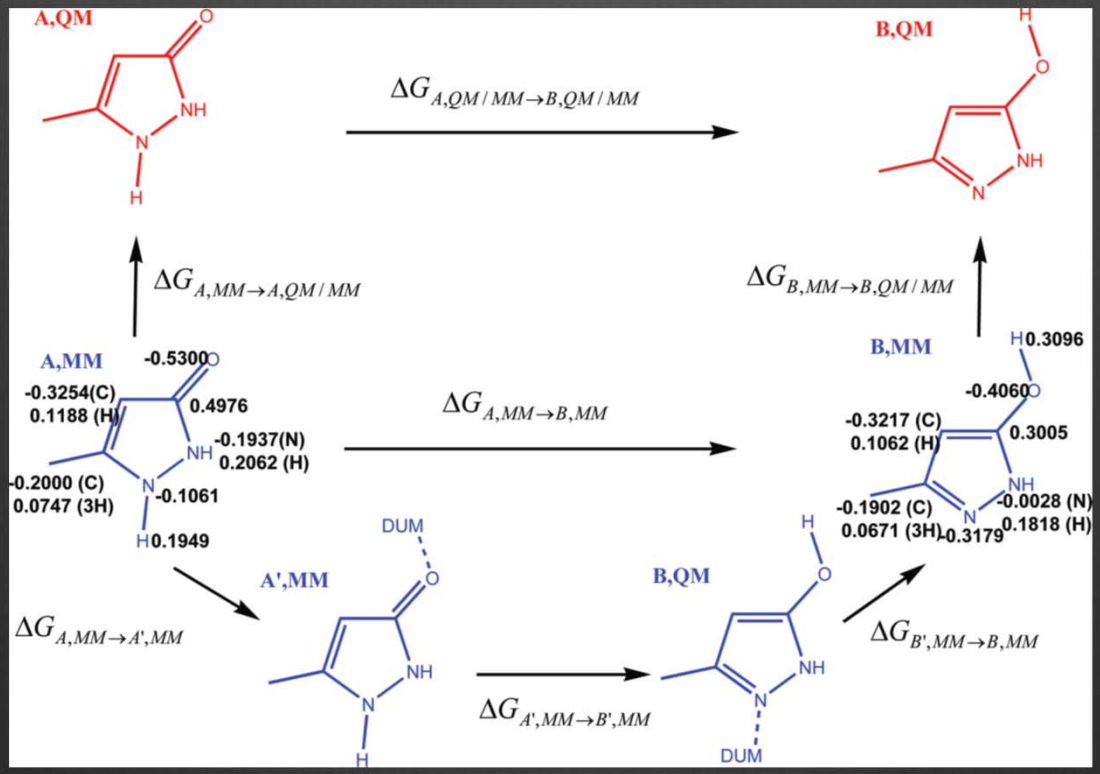



**Fig. 3.** The free-energy simulation results in model study 3. (*A*) The distance-dependent potential of mean forces produced by 1-ns OSRW simulation (black) and 1-ns metadynamics simulation (red). (*B*) The time-dependent deviation of the estimated free-energy profiles produced by the OSRW simulation from the target result. (*C*) The time-dependent deviation of the estimated free-energy profiles produced by the metadynamics simulation from the target result.

# Further Development of OSRW

- Zheng, L., M. Chen, et al. (2009). "Simultaneous escaping of explicit and hidden free energy barriers: Application of the orthogonal space random walk strategy in generalized ensemble based conformational sampling." *The Journal of Chemical Physics* **130**(23): 234105.
- Lee, S., M. Chen, et al. (2010). "Sampling Long Time Scale Protein Motions: OSRW Simulation of Active Site Loop Conformational Free Energies in Formyl-CoA:Oxalate CoA Transferase." *Journal of the American Chemical Society* **132**(21): 7252-7253.
- Min, D., L. Zheng, et al. (2010). "Practically Efficient QM/MM Alchemical Free Energy Simulations: The Orthogonal Space Random Walk Strategy." *Journal of Chemical Theory and Computation* **6**(8): 2253-2266
- AMOEBA in 2011 (Ohhhh Yeah!)
  - Binding Free Energy
  - AMOEBA-assisted Crystallography

# Tautomerization Free Energy Cycle



# Tautomerization Free Energy Results

***Table 2.*** Free Energy Components for the Calculation of $\Delta G_{A,QM/MM \rightarrow B,QM/MM}$ in Model Study 2[a]

| | $\Delta G_{A',MM \rightarrow B',MM}$ (kcal/mol) | $\Delta G_{A,MM \rightarrow B,MM}$ (kcal/mol) | $\Delta G_{A,MM \rightarrow A,QM/MM}$ (kcal/mol) | $\Delta G_{B,MM \rightarrow B,QM/MM}$ (kcal/mol) | $\Delta G_{A,QM/MM \rightarrow B,QM/MM}$ (kcal/mol) |
|---|---|---|---|---|---|
| charge set 1 | 6.72 | 6.70 | −10 761.54 | −10 771.91 | −3.67 |
| charge set 2 | 1.90 | 1.88 | −10 794.92 | −10 800.55 | −3.75 |

[a] $\Delta G_{A,MM \rightarrow B,MM}$ was estimated on the basis of eq 12; on the basis of the VBA analysis, $\Delta G_{A,MM \rightarrow A',MM} + \Delta G_{B',MM \rightarrow B',MM}$ is estimated to be −0.02 kcal/mol. $\Delta G_{A,QM/MM \rightarrow B,QM/MM}$ was estimated on the basis of eq 5. The experimental value of $\Delta G_{A,QM/MM \rightarrow B,QM/MM}$ is −3.40 kcal/mol, which was measured on the basis of the $pK_a$ value difference between two target molecules.

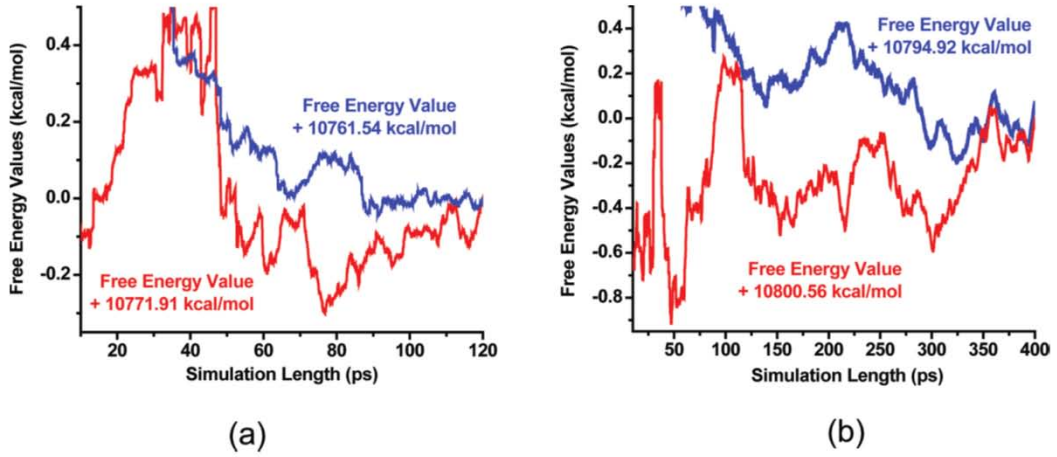
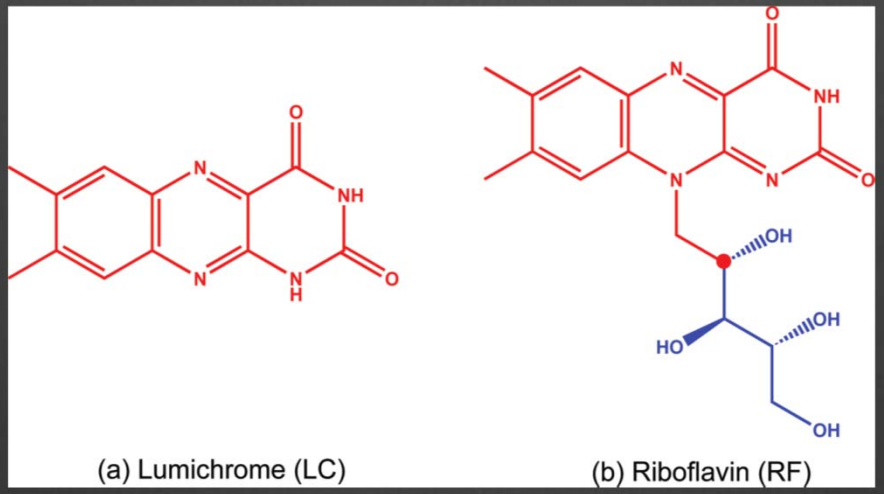



***Figure 7.*** Time-dependent predicted free energy values in model study 2. The free energy results from the simulations where the charge set 1 of the MM parameters was employed are shown in a, and the free energy results from the simulations where the charge set 2 of the MM parameters was employed are shown in b. The free energy results on molecule A are shown in red, and the free energy results on molecule B are shown in blue. To show the results clearly, the values from each simulation are shown in the shifted fashion; for instance, the red line in Figure 6a shows the change of the free energy value plus 10771.91 kcal/mol.

# Redox of Lumichrome and Riboflavin



# Redox Convergence

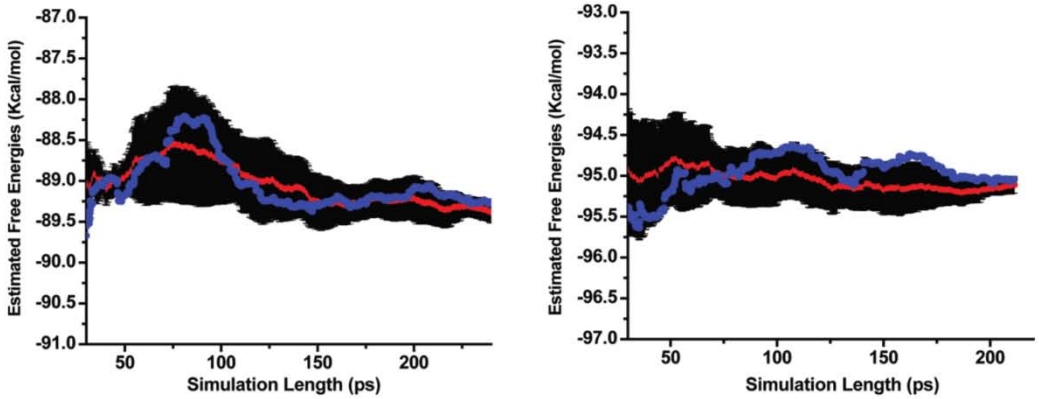
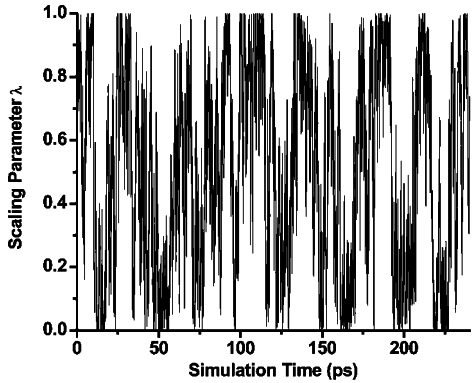
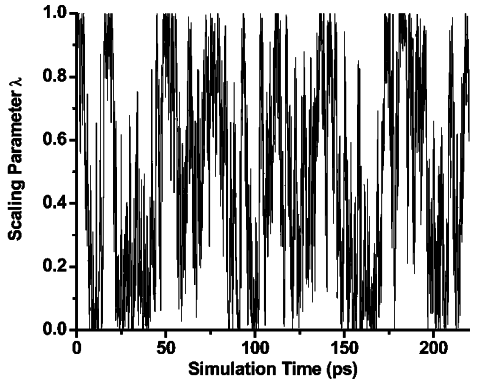



***Figure 3.*** Time-dependent estimated reduction potentials on lumichrome (a) and riboflavin (b). The averages of seven simulation results are shown by the red lines. On each system, the results of a randomly picked free energy simulation are shown by the blue dots, and the statistical errors among seven simulation results are shown around their average results.

# Scaling Parameter Sampling




(a) Lumichrome (LC)




(b) Riboflavin (RF)

*Figure 4.* Time-dependent scaling parameter changes of the representative runs on lumichrome (a) and riboflavin (b).

# λ vs. dU/dλ

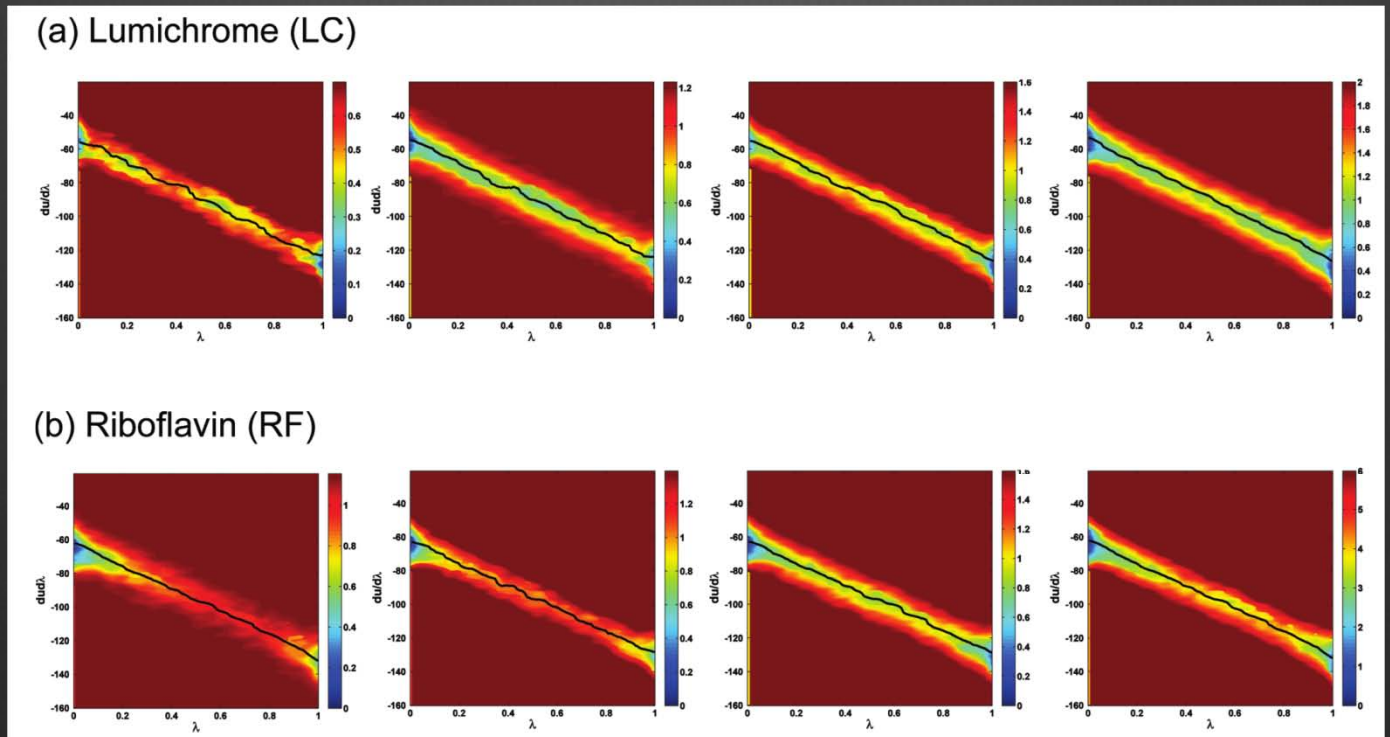



***Figure 5.*** Contour plots on the estimated free energy potential $G'_o(\lambda,\partial U_o/\partial\lambda)$ (colored contours) and the plots on the $\lambda$-dependent free energy derivatives (black lines). The depth of each estimated free energy potential $G'_o(\lambda,\partial U_o/\partial\lambda)$ contour plot reveals the activation level in the $\partial U_o/\partial\lambda$ direction. These results were obtained from the representative simulations on lumichrome (a) and riboflavin (b). On each system, the results from the 50 ps simulation are shown as the first, the results from the 100 ps simulation are shown as the second, the results from the 150 ps simulation are shown as the third, and the results from the 200 ps simulation are shown as the fourth.

# Comparison of OSRW* to TI Using 11 λ States for Redox

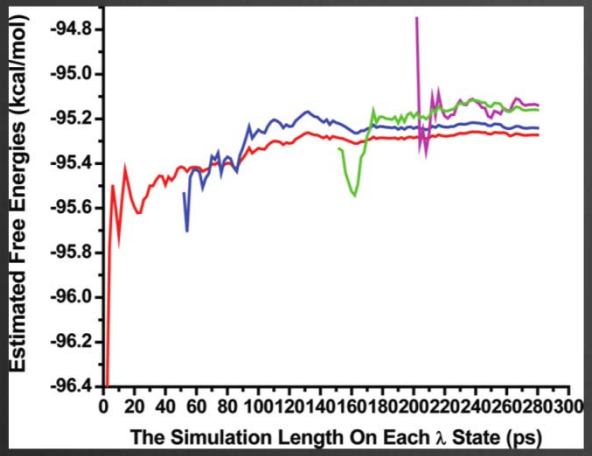



***Figure 6.*** Free energy simulation results from the comparison study on the riboflavin model system. The lines show the changes of the estimated free energies with the increase of the simulation length on each single $\lambda$ state. The red line shows the time-dependent results based on the samples without the removing of any equilibration data. The blue line shows the time-dependent results when the first 50 ps samples in each $\lambda$ state simulation are removed. The green line shows the time-dependent results when the first 150 ps samples in each $\lambda$ state simulation are removed. The pink line shows the time-dependent results when the first 200 ps samples in each $\lambda$ state simulation are removed.

- OSRW required a single ~200 psec simulation
- TI requires ~11 ~250 psec simulations = 2.75 nsec

# OSRW for AMOEBA in FFX

- Note: requires $dU/d\lambda$, $d^2U/d\lambda^2$ and $dU/dX/d\lambda$

✓ Softcore van der Waals

- D. Jiao, P.A. Golubkov, T. Darden, P. Ren (2008)

$$U_{ij} = \lambda^n \varepsilon_{ij} \frac{1.07^7}{[\alpha(1-\lambda)^2 + (\rho + 0.07)^7]} \cdot \left( \frac{1.12}{\alpha(1-\lambda)^2 + \rho^7 + 0.12} - 2 \right)$$

✓ Softcore permanent multipole real space PME electrostatics

$$r' = \sqrt{r^2 + \alpha \cdot (1-\lambda)^2}$$

# OSRW for AMOEBA in FFX

- Polarization – Thole damping so softcore is not necessary
  - Compute $U_{\text{pol}}$ for protein+ligand, protein alone and ligand alone
  - Advantage: ligand electrostatics are not turned off

$$U_{pol}(\lambda) = \lambda \cdot U_{pol}(1) + (1-\lambda) \cdot \left( U_{pol}^{(p)} + U_{pol}^{(l)} \right)$$

$$\frac{\partial U_{pol}(\lambda)}{\partial \lambda} = U_{pol}(1) - \left( U_{pol}^{(p)} + U_{pol}^{(l)} \right)$$

$$\frac{\partial^2 U_{pol}(\lambda)}{\partial \lambda^2} = 0$$

$$\frac{\partial^2 U_{pol}(\lambda)}{\partial x \partial \lambda} = \frac{\partial U_{pol}^{(p+l)}}{\partial x} - \left( \frac{\partial U_{pol}^{(p)}}{\partial x} + \frac{\partial U_{pol}^{(l)}}{\partial x} \right)$$

# OSRW for AMOEBA in FFX

- The biasing energy/forces are implemented
  - THANKS to Wei Yang!

$$U = U_r(\xi) + U_e + G(\xi, F_\xi) - \Delta G(\xi), \qquad [5]$$

- Applications
  - Binding free energies
  - Solvation free energies
  - AMOEBA-assisted X-ray crystallography optimization

# Conclusions of OSRW

- Key problem addressed: **Hamiltonian Lagging** of the system behind the order parameter
- Approach: Random walk in both the order parameter space ($\lambda$) and generalized force space ($dU/d\lambda$)
  - Synchronization of order parameter moves and conformational relaxation