Accelerating Convergence of Free Energy Calculation with Replica Exchange Solute Tempering (REST2)

Wei Jiang

Computational Science Division, Argonne National Laboratory

2018 Computational Biophysics Workshop, Sept 10-14th

Outline

1. Multiple Copy Framework of NAMD
Aims & implementation
Popular applications

2. Hysteresis Minimization
λ-Exchange (λ-REMD)

3. Overcome Hidden Barrier with REST2
REST2 Algorithm & Implementation
Straight applications of REST2
Free Energy Perturbation/H-REMD
FEP/REST2
FEP/λ-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/REST2

Intelligent sampling with Multiple Copy (Trajectory) Algorithms

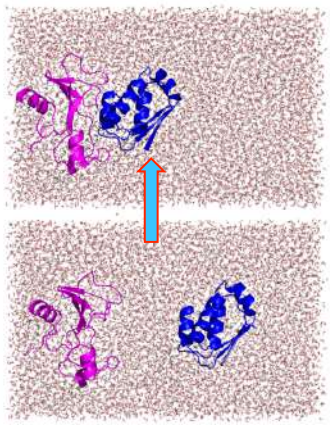
'**Problem** decomposition' ⇨ Many weakly coupled trajectories (Divide-and-conquer)

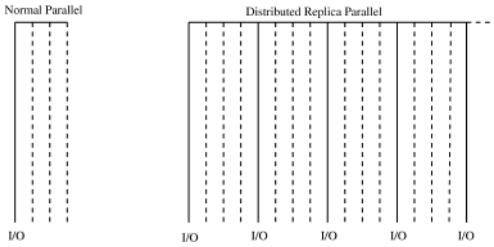
Each trajectory ⇨ molecular dynamics with biased terms

Periodic inter-trajectory communication ⇨ Optimal sampling efficiency

Number of trajectories ⇨ Controlled with acceptance ratio and replica travel

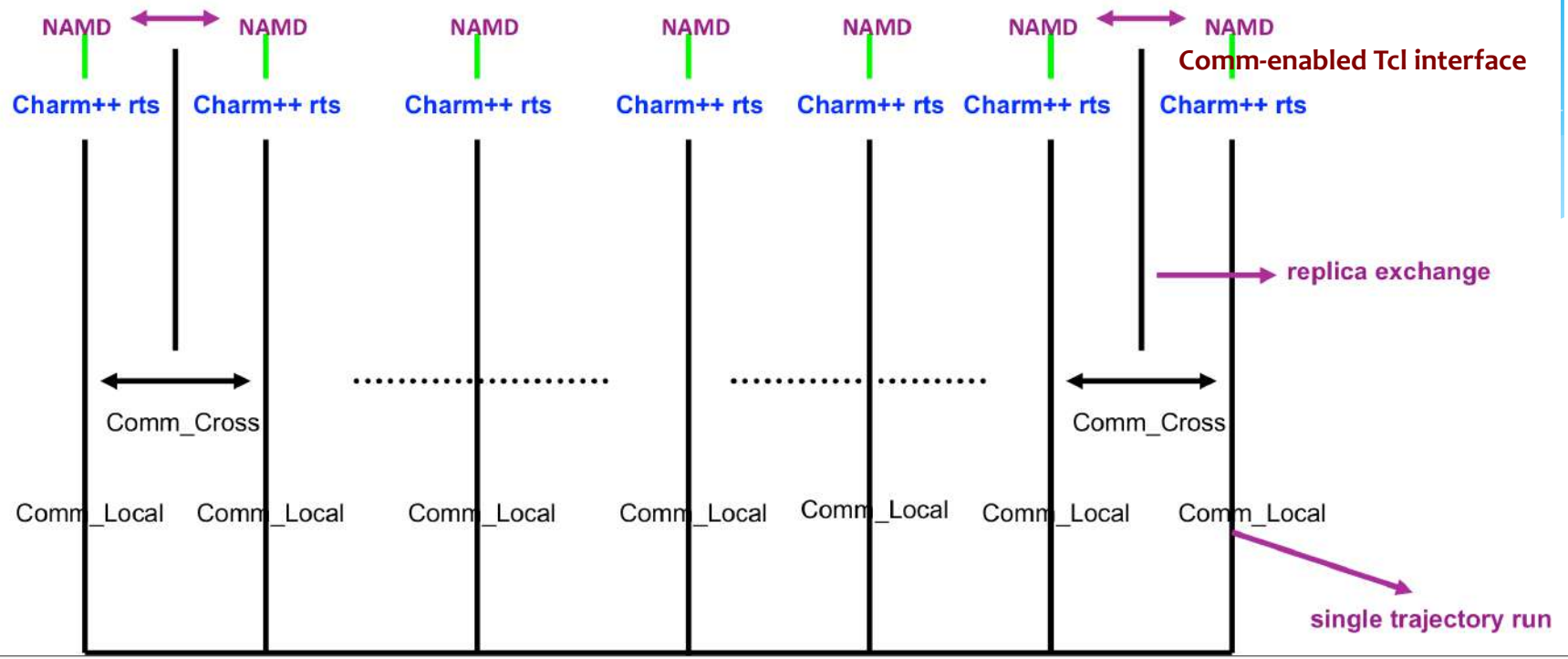
Quantitative info ⇨ Free energy, transition path, reaction rate, protein folding/unfolding





Concurrent tasks

Scalable Multiple Copy Framework 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

Wei Jiang, James Phillips etc, Computer Physics Communications, 2014, 185, 908-916

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

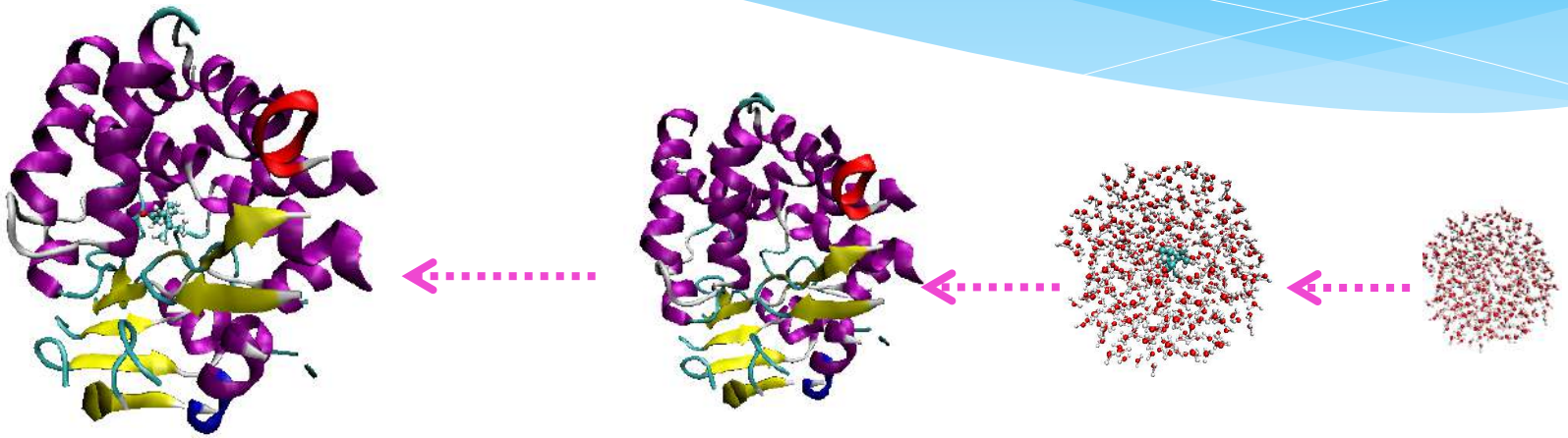
Monte Carlo -> Detailed balance->poor efficiency

Alternative ?

Large length/time scale target structure change

Exceptionally long trajectory

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 FE

Long reaction 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

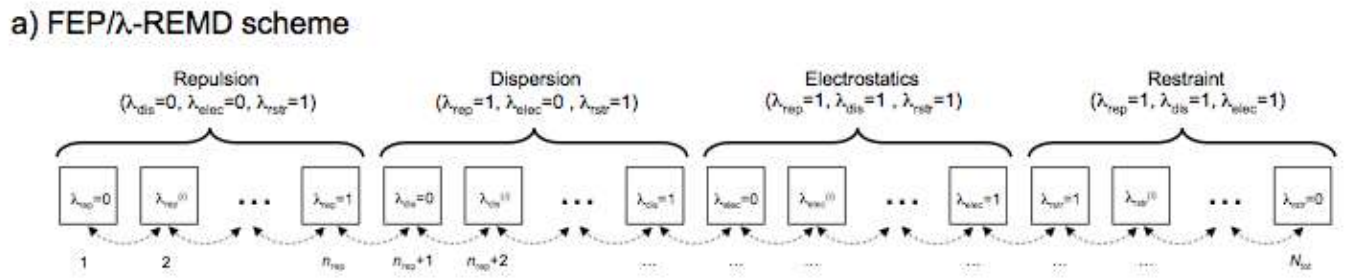
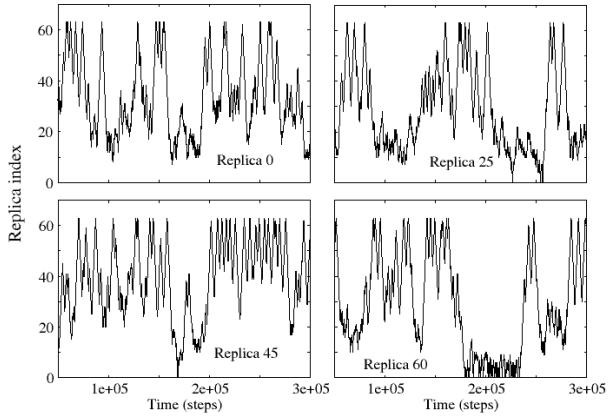


Table 1. Hydration Free Energy and the individual components for TIP3

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	



Wei Jiang, Milan Hodoscek, Benoit Roux, *J. Chem. Theory Comput., Letter, 2009, 5, 2583*

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

Co-product of λ exchange: Simple Overlap Sampling

Without λ exchange:

WHAM

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

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

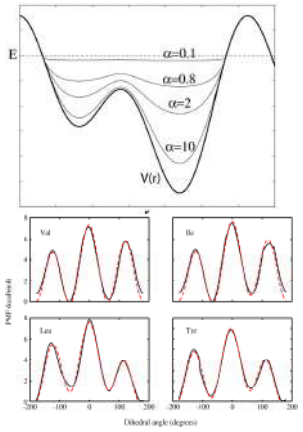
Receive result in < 5s

λ-Exchange: NOT a real sampling enhancement

Automatic bi-directional -> minimize hysteresis
Merely an improved computational protocol
No acceleration mechanism introduced each replica

What is a **REAL** sampling enhancement replica exchange?

Boosting mechanism overcoming energy barrier
Boosts kinetic energy (T-REMD) $K=\Sigma 1/2mV^2$
Lowers potential energy barrier (Accelerated MD)
Lowers free energy barrier (Pre-fitted PMF)



Why Replica Exchange Solute Tempering (REST2)

T-REMD:
#replicas proportional to square root of #atoms
Works only for small/medium size system. <30K atoms
Unable to select interested degrees of freedom

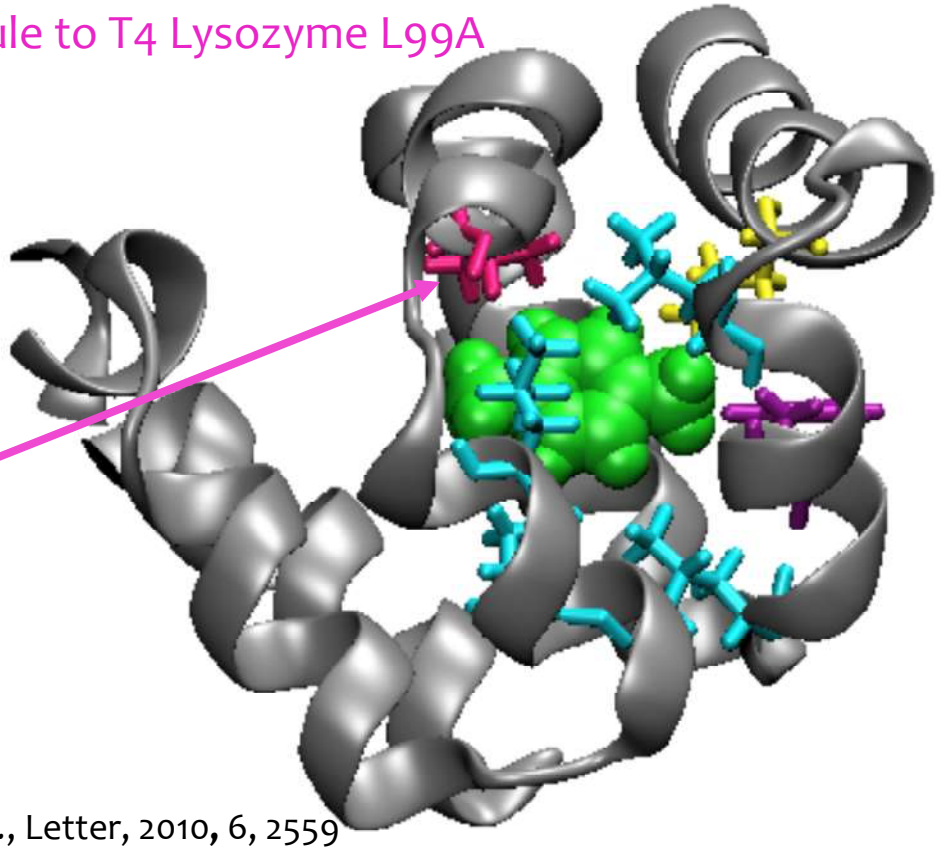
Accelerated MD:
No selection of interested degrees of freedom;
Magic choice of flattening strength

REST2:
User selects interested degrees of freedom;
Biasing strength corresponds to an effective T_{eff} -> exp(-V/kT)
Enhances energy overlap between neighboring replicas

Kinetically Trapped Conformations in Free Energy calculation

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*

Wei Jiang, Benoit Roux, J. Chem. Theory Comput., Letter, 2010, 6, 2559

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

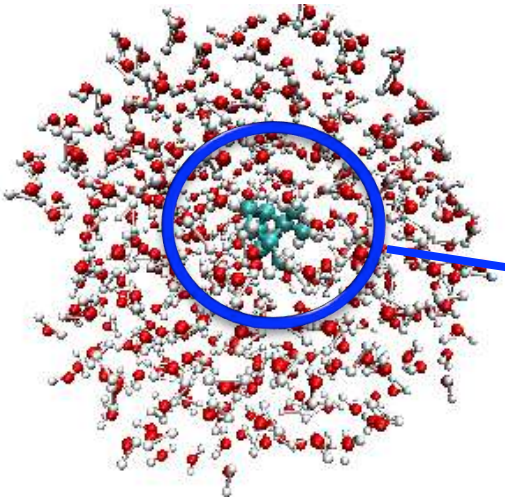
$$E_m^{\mathrm{REST2}}(X) = \frac{\beta_m}{\beta_0} E_{ss}(X) + \sqrt{\frac{\beta_m}{\beta_0}} E_{sw}(X) + E_{ww}(X)$$ -> parameter rescaling

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

Replica exchange solute tempering:
High transferability;
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.

Replica Exchange Solute Tempering (REST2)



Surrounding region feels parameter scaling

q_1 is rescaled, $F_1=F_2=\lambda*q_1*q_2$

Decouples timescale of ‘solute’ and surroundings.
Not a real barrier flattening method
Parameter scaling influences un-tempered particles
Modify hydrophobic/hydrophilic properties -> oversampling
Exchange attempt frequency, highest effective temperature and selection of heated region

Basic Usage and Tips of REST2 in NAMD

Works on CPU and GPU

ssFile myfile.pdb # pdb format file, generated with VMD plugin
ssCol O # with value '1.0' for REST2 atoms
soluteScaling on # REST2 On
soluteScalingFactor 0.9 # default scaling keyword
soluteScalingFactorCharge 0.8 # override soluteScalingFactor
soluteScalingFactorVdw 0.7 # overrides soluteScalingFactor
soluteScalingAll off # only dihedral/improper terms are tempered

soluteScalingFactorCharge and soluteScalingFactorVdw:
fine-grained implementation
improving acceptance ratio
minimizing nonequilibrium effect of REST2
Flexible for different systems. Ie, **soluteScalingFactorCharge** membrane system.

SoluteScalingFactor Setup Per Replica

soluteScalingFactor setup per replica

```
proc replica_sptscale { i } {
 global num_replicas num_replicasa num_replicasb num_replicasc min_temp max_temp

 if { $i < $num_replicasa } {
  set temp [expr ($max_temp * exp( log(1.0*$min_temp/$max_temp)*(1.0*$i/($num_replicasa-0.0)) ) )]
   return [ expr $min_temp/$temp ]
 } elseif { $i >= $num_replicasa && $i < [expr $num_replicasa+$num_replicasb] } {
   return 1.0                                                    # Modifying soluteScalingFactor
 } else {
   set temp [expr ($min_temp * exp( log(1.0*$max_temp/$min_temp)*(1.0*($i-$num_replicasa-$num_replicasb+1.0)/
($num_replicasc-0.0)) ) )]
   return [ expr $min_temp/$temp ]
 }
}
```

```
proc setup_parameters { ID } {
 global num_replicas restart_root
 soluteScalingFactor  [replica_sptscale  $ID]  # REST2
set IDN [expr ($ID + 1)]
 if { $IDN < $num_replicas } {  # soluteScalingFactor for each Replica

    set Lambda [replica_lambda $ID]
    set Lambda2 [replica_lambda $IDN]
    alchLambda    $Lambda                  # Free energy perturbation
    alchLambda2   $Lambda2
 } else {
    alchLambda    1.0
    alchLambda2   1.0
 }
}
```

Master config file

```
set num_replicas 16
set min_temp 300 ; # physical temperature
set max_temp 900 ; # highest temperature where parameters
 of selected region is rescaled by 1/3 (300/900)
set TEMP 300
set steps_per_run 100 ; # 0.2 ps #replica exchange frequency
set num_runs 100 ; #total steps  steps_per_run * steps_per_run
# num_runs should be divisible by runs_per_frame *
#frames_per_restart
set runs_per_frame 10 ; # 5 ps per frame
set frames_per_restart 10 ; # 1000 ps per restart
set namd_config_file "aaqaa3_rest2_base.namd"
set output_root "output_spt_aaqaa3/%s/rest2" ; # directories must exist
```

Replica Exchange - Communication Enabled Tcl

Potential Energy Exchange

```
if { $replica(index) < $replica(index.$swap) } {
  set POTENTIAL2 [replicaRecv $replica(loc.$swap)]
 }
 if { $replica(index) > $replica(index.$swap) } {
  replicaSend $POTENTIAL $replica(loc.$swap)
 }
 if { $replica(index) > $replica(index.$swap) } {
  set POTENTIAL2 [replicaRecv $replica(loc.$swap)]
 }
 if { $replica(index) < $replica(index.$swap) } {
  replicaSend $POTENTIAL $replica(loc.$swap)
 }
 if { $replica(index) != $replica(index.$swap) } {
  set replica(ParamID) $replica(index.$swap)
  setup_parameters $replica(ParamID)
```

NAMD source tree: lib/replica/REST2/rest2_remd.namd

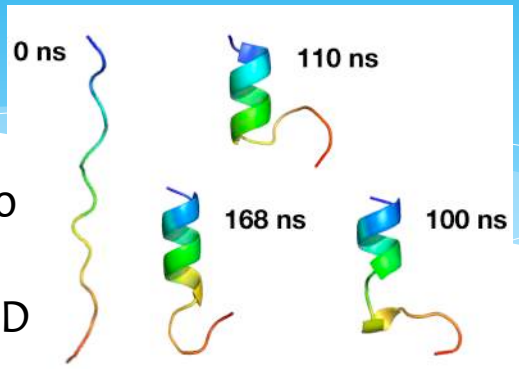
soluteScalingFactor Exchange

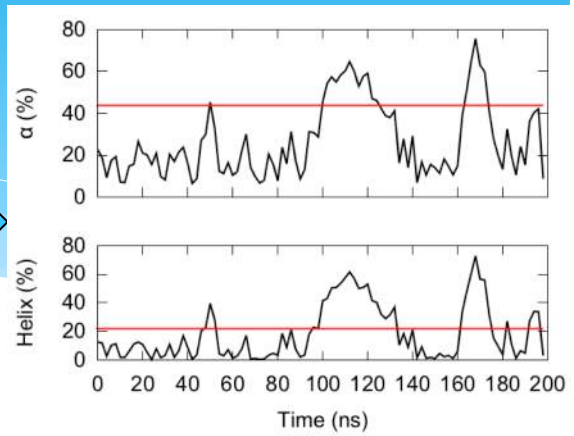
```
if { $replica(index) < $replica(index.$swap) } {
    set BOLTZMAN 0.001987191
    set delta [expr ($POTENTIAL_NEW + $POTENTIAL_NEW2 - $POTENTIAL -
$POTENTIAL2)/($BOLTZMAN * $TEMP)]
    set doswap [expr $delta < 0. || exp(-1. * $delta) > rand()]
    replicaSend $doswap $replica(loc.$swap)
    puts $sos_history_file "$i_step $replica(index) $replica(index.$swap) $TEMP
$POTENTIAL $POTENTIAL_NEW $POTENTIAL2 $POTENTIAL_NEW2 $doswap"
    if { $doswap } {
     puts stderr "EXCHANGE_ACCEPT $replica(index) $replica(index.$swap) RUN
$i_run"
     incr replica(exchanges_accepted)
    }
    incr replica(exchanges_attempted)
  }
  if { $replica(index) > $replica(index.$swap) } {
    set doswap [replicaRecv $replica(loc.$swap)]
    puts $sos_history_file "$i_step $replica(index) $replica(index.$swap) $TEMP
$POTENTIAL $POTENTIAL_NEW $POTENTIAL2 $POTENTIAL_NEW2 $doswap"
  }
 }
```

Protein Folding-Unfolding Transitions with REST2

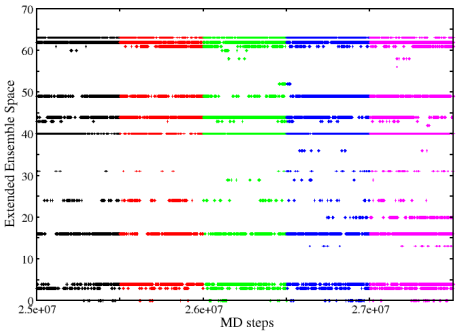
Peptide folding-unfolding, **explicit** solvent, 16 replica, effective temperature range 300 – 600K

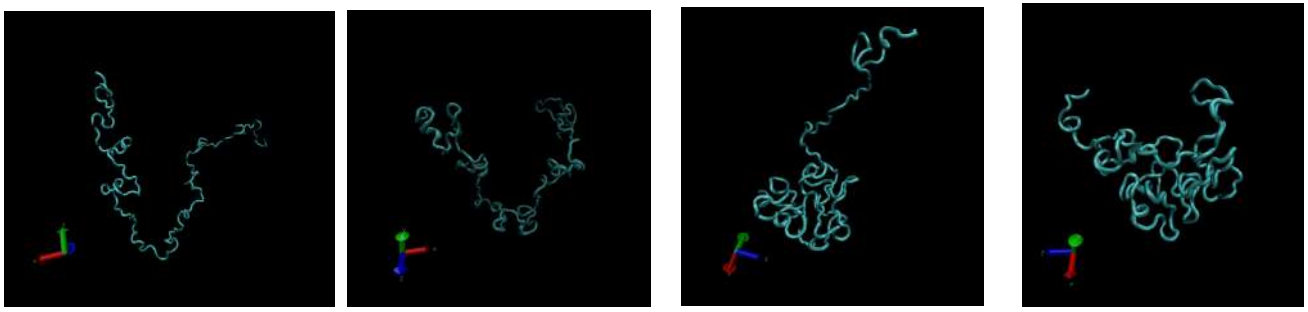
Acceptance ratio: 50% >> T-REMD





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

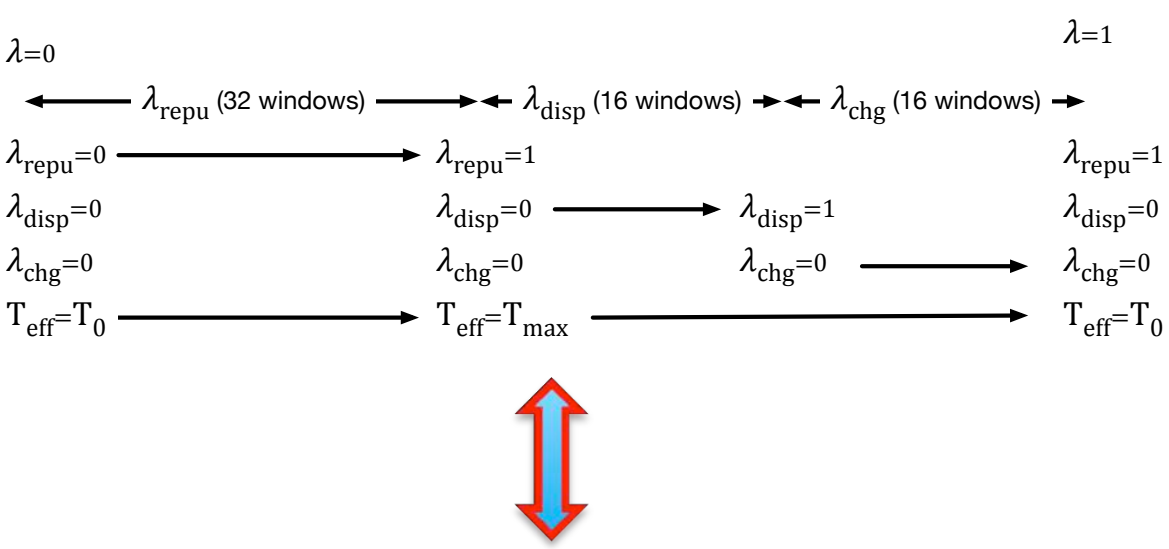




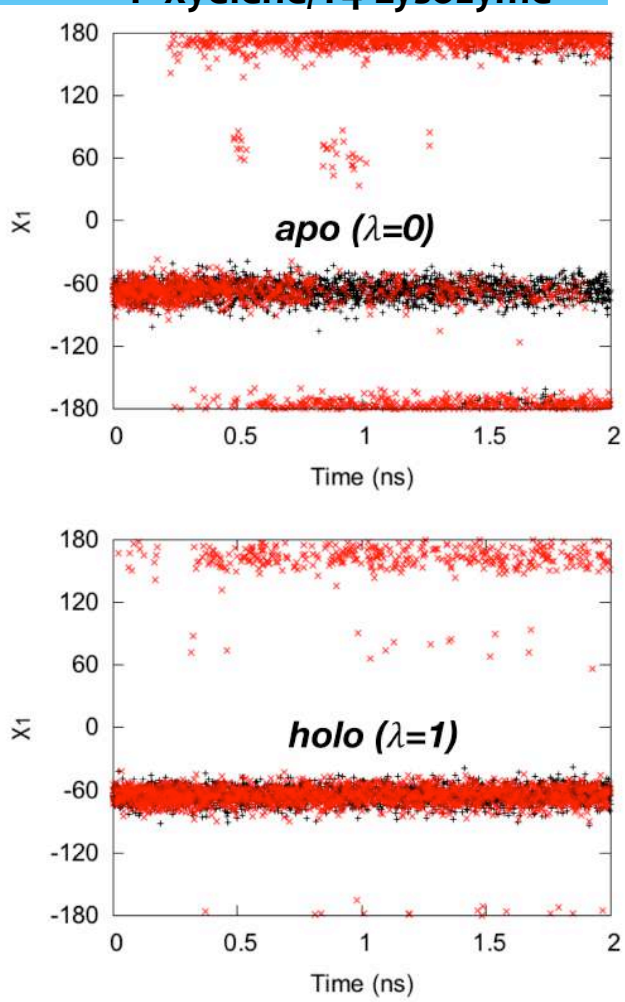
0 ns 12 ns 24 ns 50 ns

Sunhwan Jo, Wei Jiang, *Computer Physics Communications*, 2015, 197, 304-311

FEP/REST2 (Schordinger Version)



P-xyelene/T4 Lysozyme

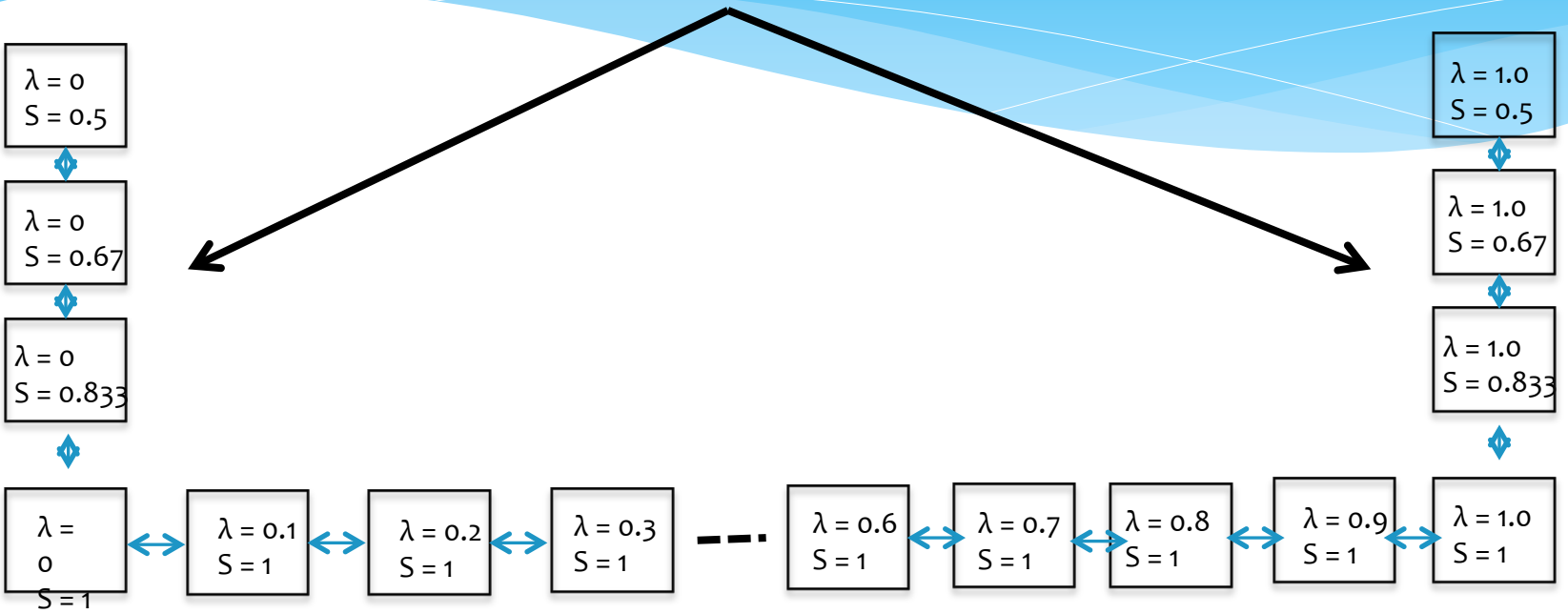


Cheap solution and easy implementation
Thermodynamic axis is contaminated by the brutal mixing of REST2 and FEP
Carefully controlled heated region minimizes nonequilibrium effects.

Sunhwan Jo, Wei Jiang, *Computer Physics Communications*, 2015, 197, 304-311

Orthogonal Implementation of FEP/REMD/REST2

End states have deepest hidden barrier

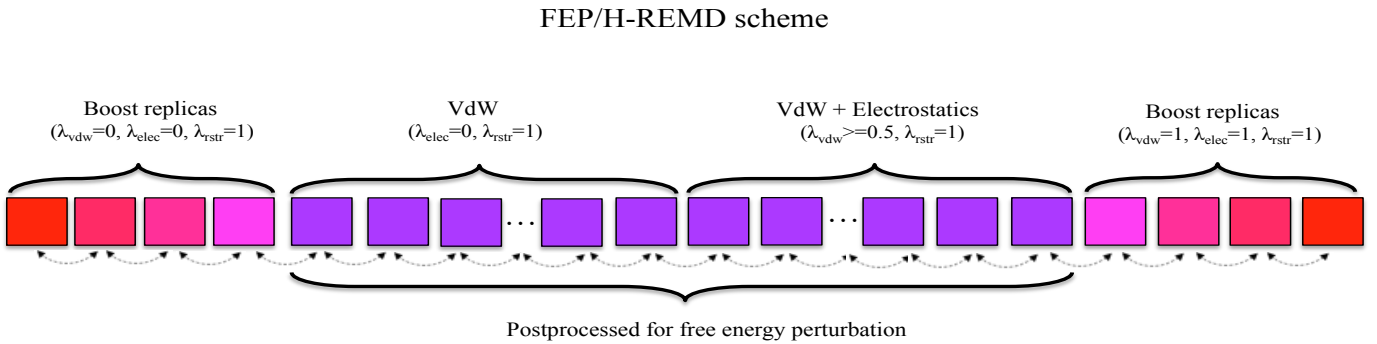


Separation of λ-REMD and REST2, leaving FEP as it is

REST2 windows adjustable with size of heated region

Need slightly more parallel computing resource

Reduced 1D FEP/H-REMD with REST2 in NAMD



Metropolis MC

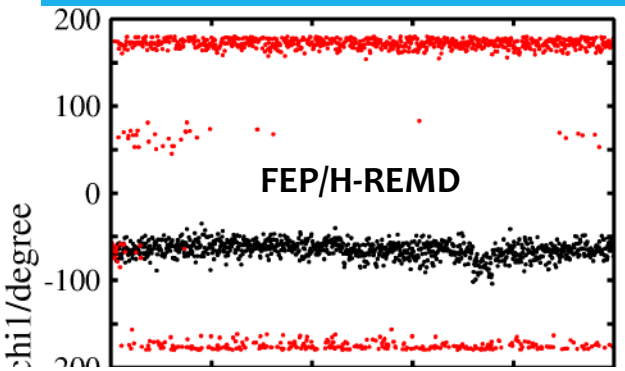
$$P(\lambda_i, b_i \rightarrow \lambda_j, b_j) = \min\left\{1, e^{-\left[U(\lambda_i, b_i, \mathbf{r}_i) + U(\lambda_j, b_j, \mathbf{r}_j) - U(\lambda_i, b_i, \mathbf{r}_j) - U(\lambda_j, b_j, \mathbf{r}_i)\right]/k_B T}\right\}$$

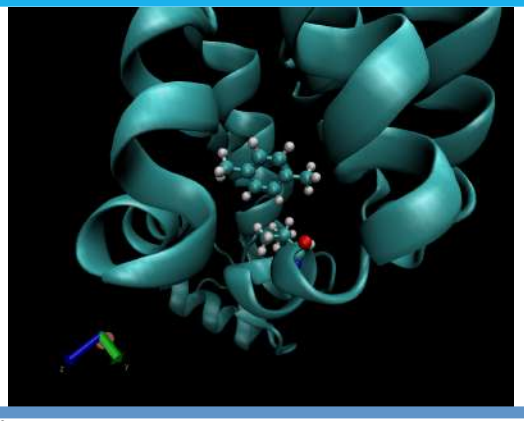
Two end states involve large time scale

High frequency exchange

Accelerated conformations travel through whole path

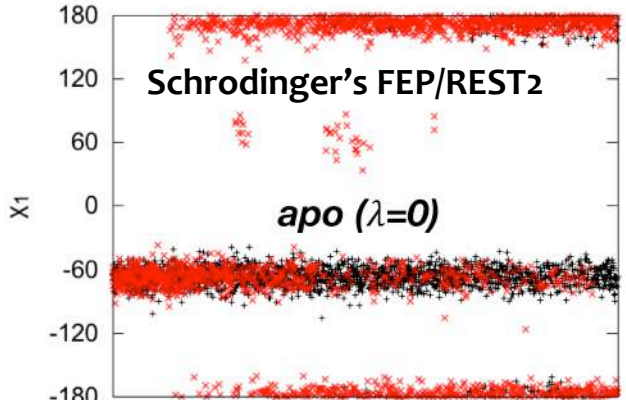
FEP/REMD/REST2 and Schrodinger's FEP/REST2

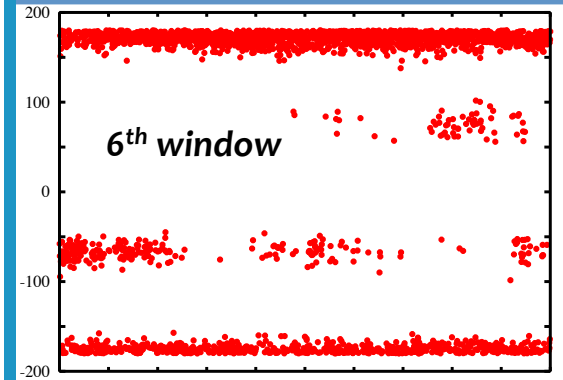


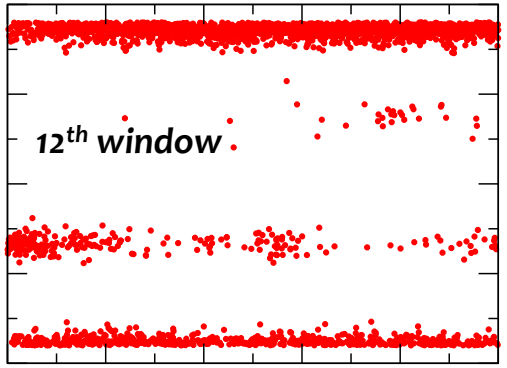


P-xyelene/T4 Lysozyme

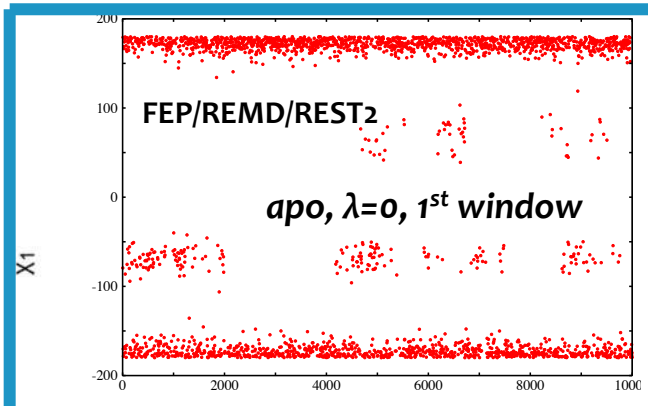
-5.5 kcal/mol vs exp -4.7 kcal/mol

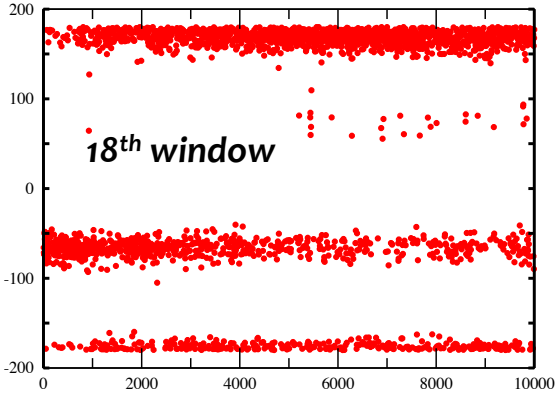


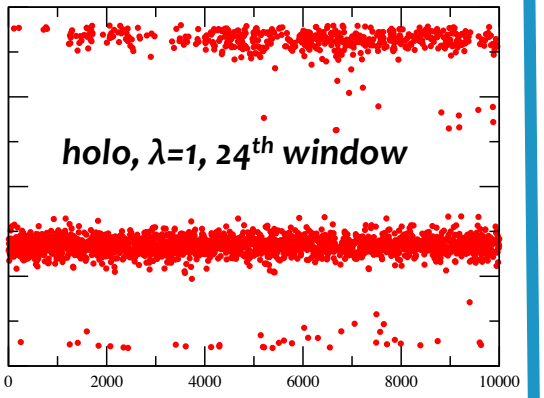




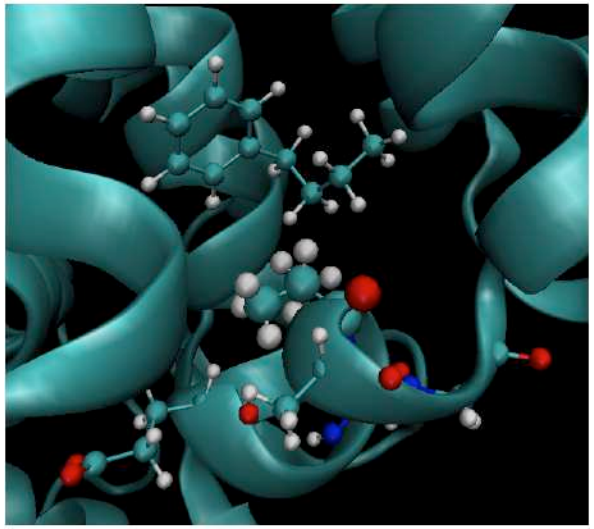
Reasonable evolution from apo to holo state

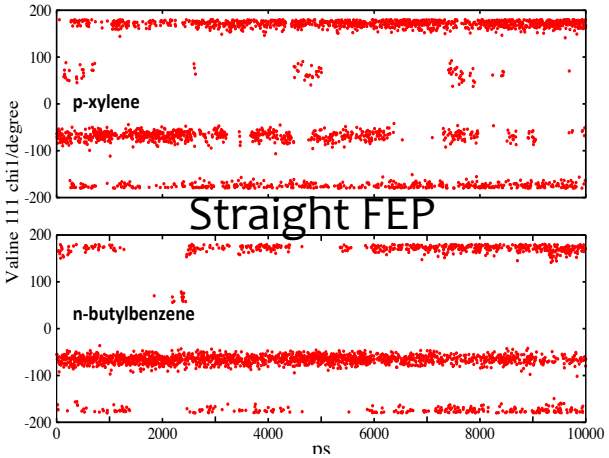




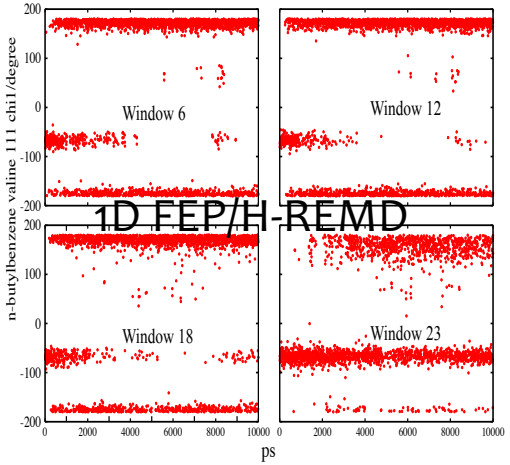


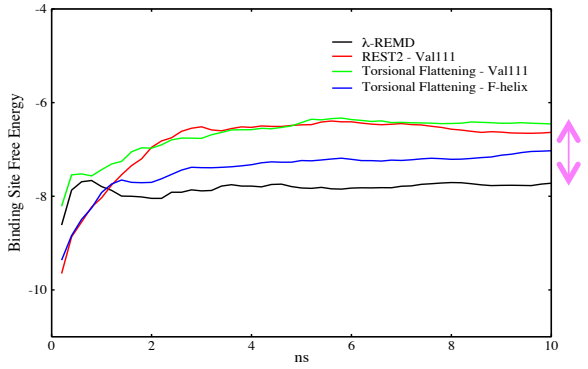
1D FEP/H-REMD to N-Butylbenzene/T4 Lysozyme





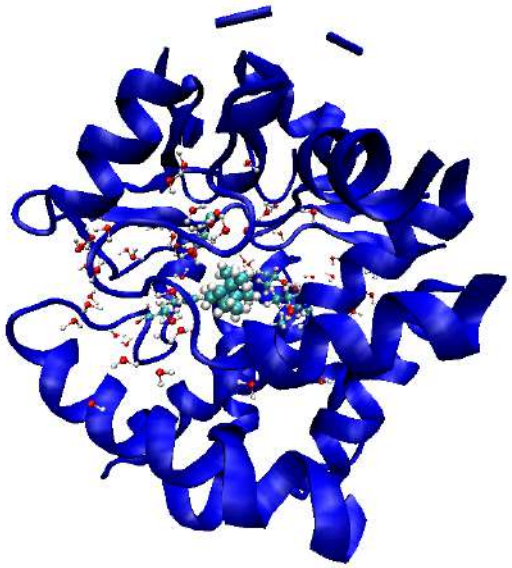
Hamiltonian lagging





Accelerated convergence

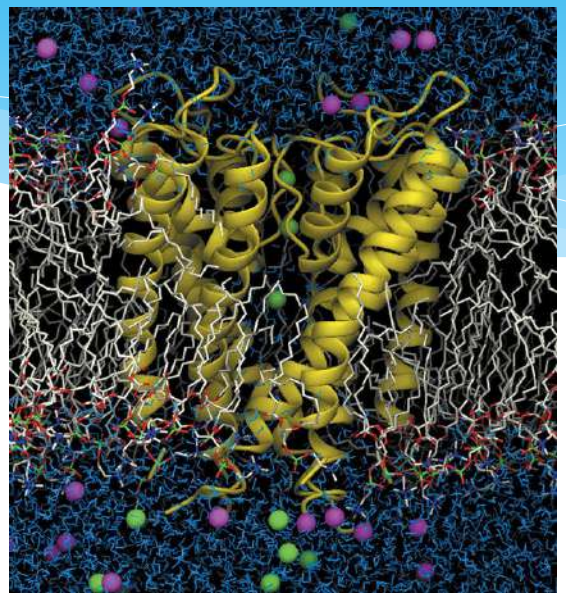
Solvent acceleration with simulated annealing REST2



Camphor/P450 Binding Complex

Buried binding pocket

Interior polar residues



KcSA Ion Channel

Large cavity

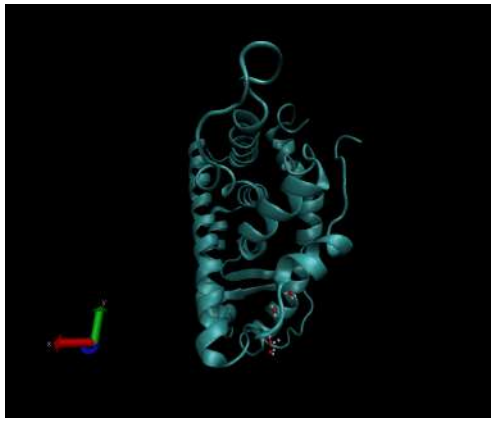
Solvent configuration sampling

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

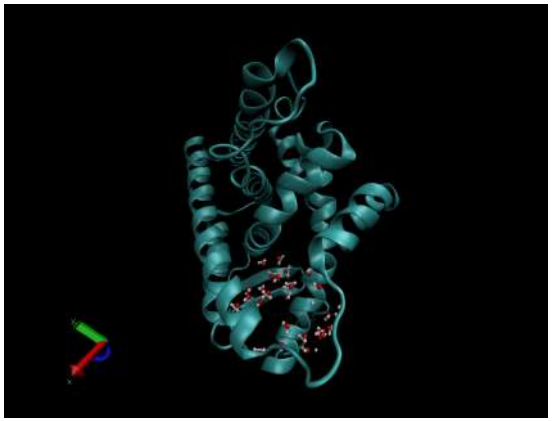
Hybrid of Simulated Annealing and REST2 for Solvent acceleration

Effective simulated annealing (SA) can replace solvent temperature replica exchange
Re-scale potential energy: (1) scale elec energy of solvent with others fixed.
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

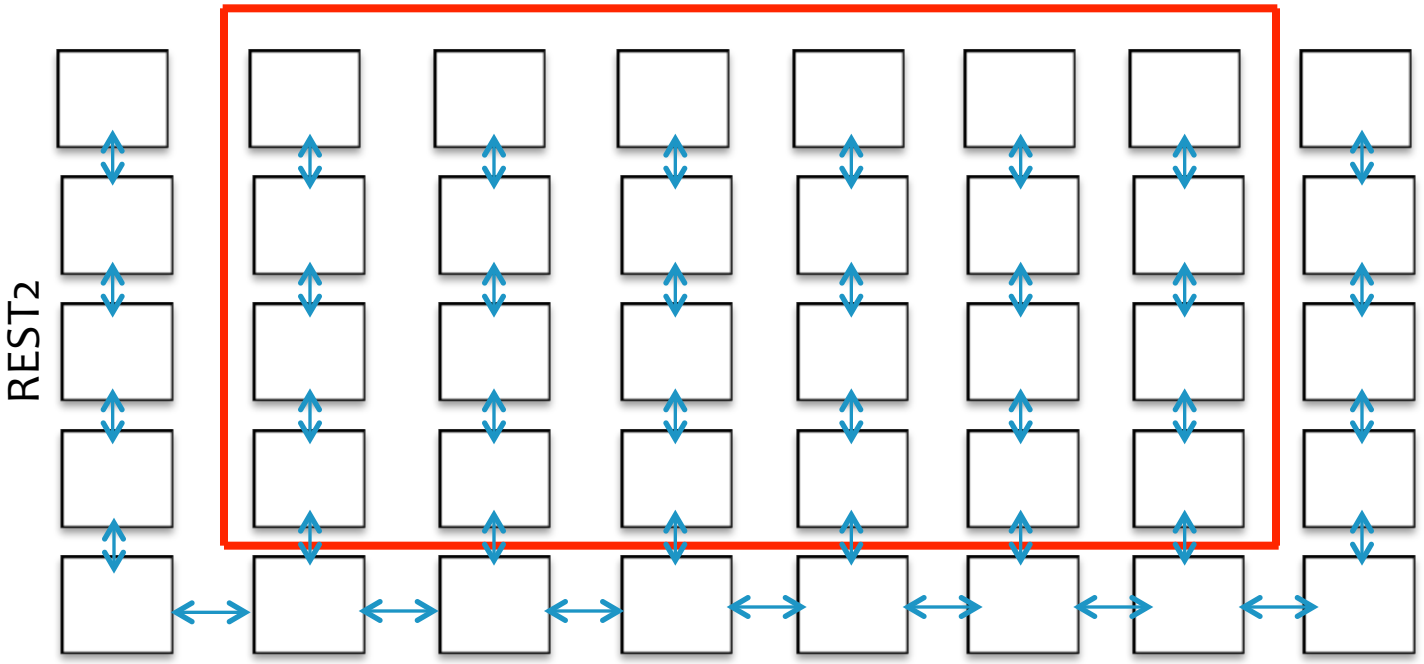


100 ps (scaling factor 0.75) + 100ps 300K

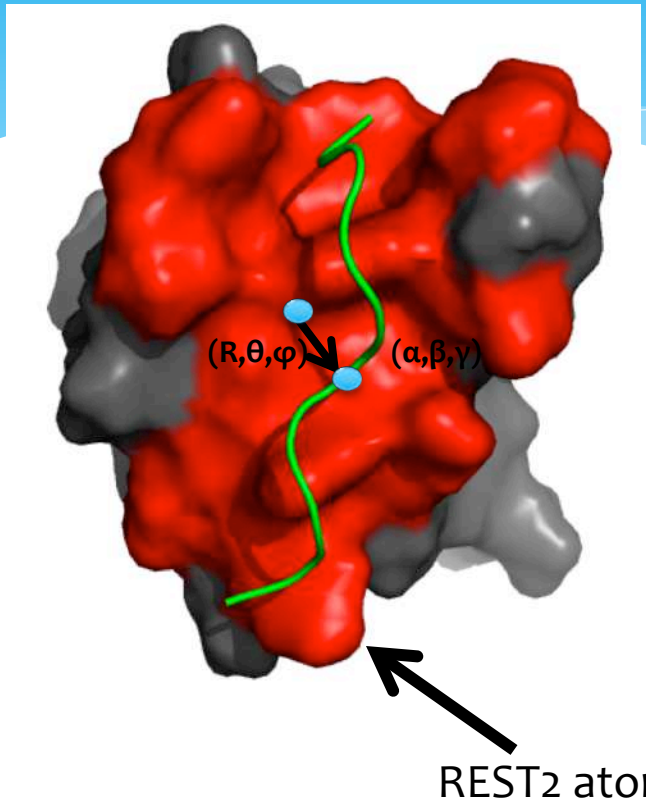
		EXP (kcal/mol)	FEP/SA/REST2
1	DES	-13.2	-14.0
2	OHT	-12.7	-13.5
3	EST	-12.3	-13.3
4	E1T	-10.7	-11.5
5	TAM	-10.6	-11.3
6	NAF	-9.4	-10.6
7	NOR	-8.7	-9.80

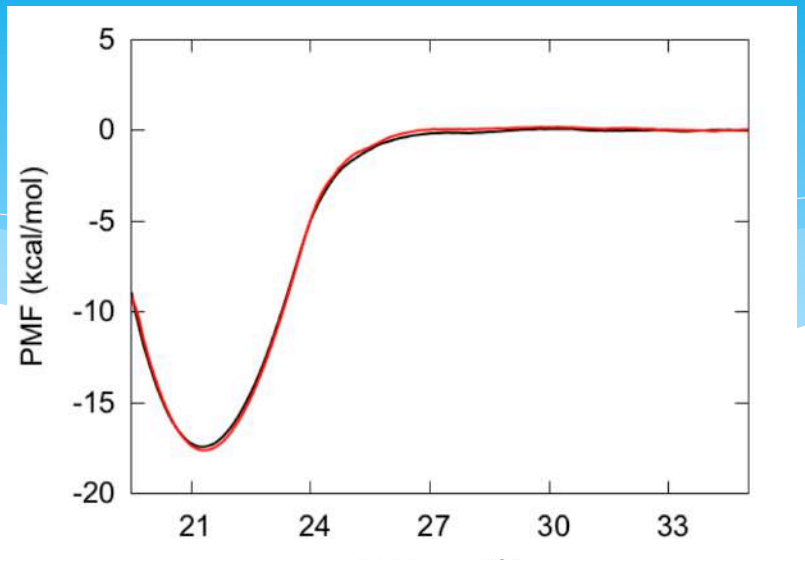
Combination of Umbrella Sampling and REST2

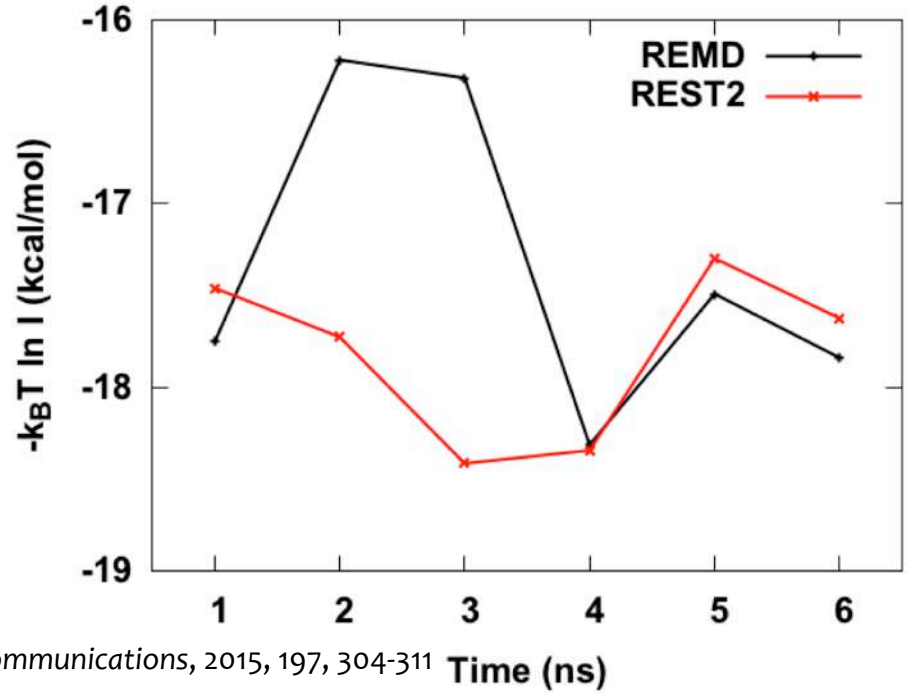
REST2 overcomes Hamiltonian lagging
Multidimensional Hamiltonian exchange scheme Umbrella biases are exchanged in one axis
REST2 in another axis



Binding of Peptide to SH3 domain of Kinase



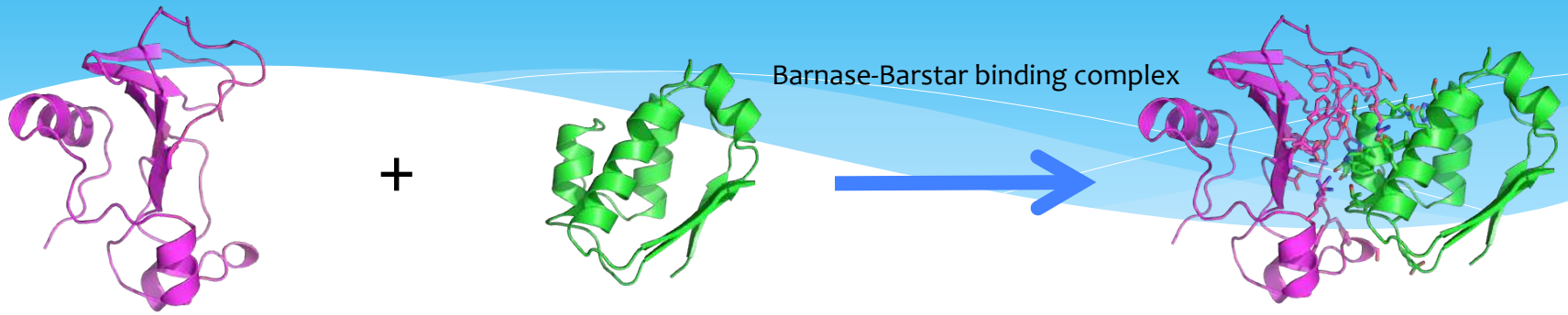


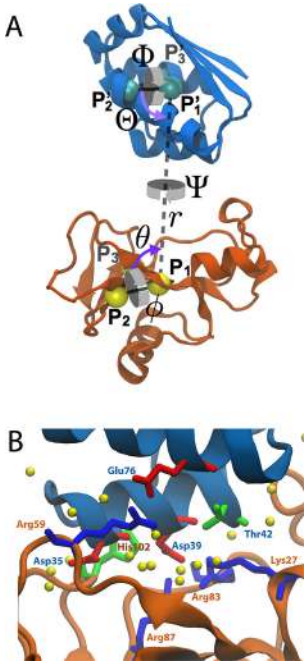


REMD and REMD/REST2
32 and 128 replicas
side chain atoms that are within 4 Å from the peptide were selected for tempering

Sunhwan Jo, Wei Jiang, *Computer Physics Communications*, 2015, 197, 304-311

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) Straightforward usage on CPU and GPU

(2) High transferability

(3) Potential oversampling -> carefully monitored

(4) Fine grained tempering -> specific problem

(5) Hybrid Simulated Annealing and REST2 accelerating solvent sampling

(6) Free energy calculation with REST2

Sample REST2 usage

Download NAMD 2.13b1

https://www.ks.uiuc.edu/Development/Download/download.cgi?PackageName=NAMD

cd lib/replica/REST2

Single replica test case of REST2: aaqaa3_rest2_test.namd

A REST2 example with parallel computing: rest2_remd.namd, aaqaa3_rest2_base.namd, init.conf

Submission of REST2 task:

… namd2 +replicas 16 init.conf --source rest2_remd.namd +stdout > output_rest2/%d/job0.%d.log > test.out

You need mk directory 'output_rest2' and 16 subdirectories 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 inside 'output_rest2'