The gb1p Analysis of GPU-GBMV2/SA-MSES Simulation

Xiping Gong

04/01 - 04/30/2020

Summary of what you have done since your last report

1. Check GPU-GBMV2/SA-MSES MMTSB implementation and use it to simulate the gb1p system

by manually calculating the exchange probability of MMTSB replica exchange simulation, the calculated delta values are exactly same as the MMTSB calculations.

I simulated the gb1p system that has been studied a lot for MSES and MD standard simulations.

2. Do the trajectory analysis and calculate the properties of gb1p system

using the MMTSB tool to extract the trajectories and the properties of interest include the temperature, energies, RMSD and native contact fractions of AT and CG systems, and hydrogen bond number of AT system.

3. Develop molecular dynamic-based Metropolis-Hastings algorithms for solving the trap problem

The motivation/goals of what you have done

The goal is to reproduce the [CPU simulations](https://onlinelibrary.wiley.com/doi/full/10.1002/jcc.24734), and to evaluate the capabilities of MSES-GBMV2/SA to accelerate the conformational transitions of disordered proteins.

Analysis/conclusion of the results: progress and remaining problems

1. gb1p beta-hairpins experimental data

Some experimental data on the folded rate at 298K has been studied and thus available to us. Table 1 shows the folded rates of some beta-hairpins. Previous CPU result can reproduce the stability of three beta-hairpins, gb1m3 > gb1p > gb1m1.

Experimental data

Table 1. Hairpin Sequences and Fold Populations

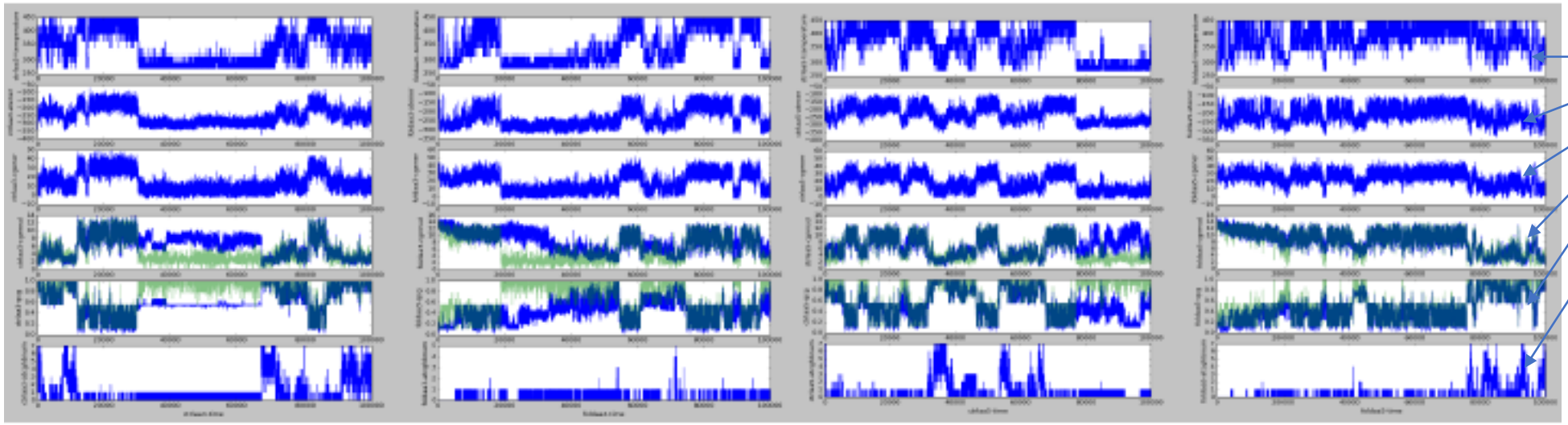
		% folded (298 K)	T_m (°C)
GB1p	GE**W**T**Y**DDATKT**F**T**V**TE	ca. 30%	
GB1m1	GE**W**T**Y**DDATKT**A**T**V**TE	6 ± 6%	
GB1m2	GE**W**T**Y**NPATGK**F**T**V**TE	74 ± 5%	47 ± 2
GB1m3	KK**W**T**Y**NPATGK**F**T**V**QE	86 ± 3%	60 ± 2
trpzip4	GE**W**T**W**DDATKT**W**T**W**TE		70
HP5W4	KK**W**T**W**NPATGK**W**T**W**QE	>96%	85
HP5W	KK**Y**T**W**NPATGK**W**T**V**QE	92 ± 2%	66
HP5F	KK**Y**T**W**NPATGK**F**T**V**QE	82 ± 4%	53 ± 5
HP5A	KK**Y**T**W**NPATGK**A**T**V**QE	21 ± 10%	

Ref. **Enhanced Hairpin Stability through Loop Design: The Case of the Protein G B1 Domain Hairpin**

Analysis/conclusion of the results: progress and remaining problems

Six properties are studied, including temperature,
energies, RMSD and native contact fractions of AT and CG systems, and hydrogen bond number of AT system

GPU-MSES/GBMV2/SA, T/H REX: CG parameter scale=0.8



Properties:
1) Temperature
2) Atener
3) Cgener
4) Atrmsd (blue) + cgrmsd (green)
5) Qat (blue) + qcg (green)
6) hbnum

Foldaa3-time: unfolded starting structure and 3rd replica, each is 2ps.

Conclusions
1) CG drives AT very well when coupling is there, the stronger the coupling is (by looking at all conditions), the bigger driving force is. (See next slide, MD simulations)
2) REX simulations have a trapped state (when couple=0, the CG is folded but AT is unfolded)
3) Hbnum is a hard indicator to monitor the convergence
4) It seems that the temperature exchange determines the structural motions, which requires the temperature exchanges very frequently

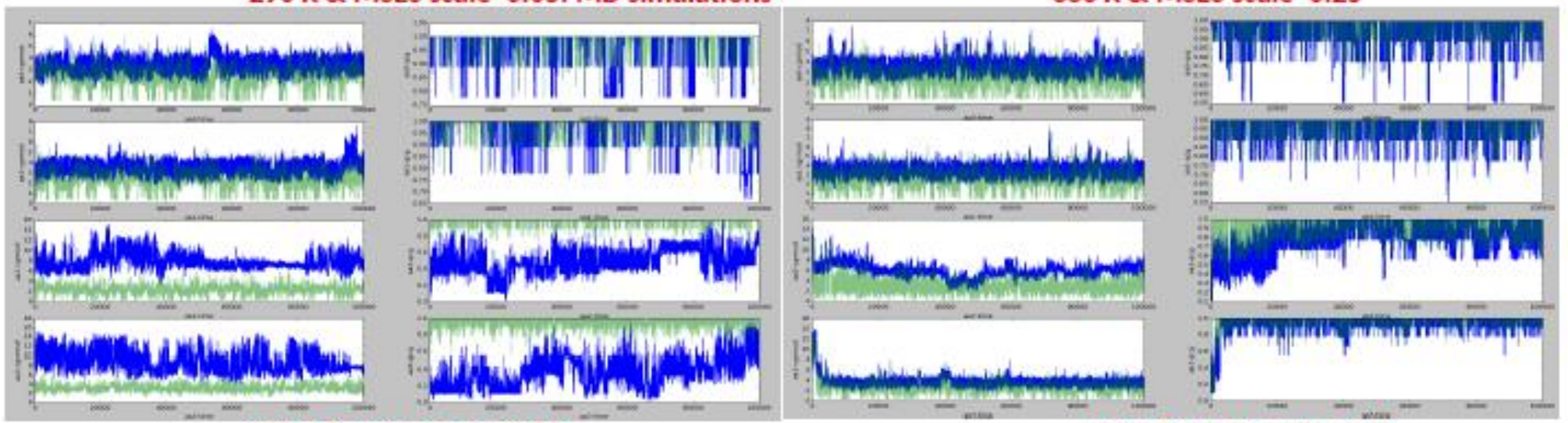
Optimize the conditions to allow the frequent temperature exchange

Analysis/conclusion of the results: progress and remaining problems

The following is standard MD simulations, the goal is to explore the effect of MSES scale factor on the AT and CG coupling simulations. It was found that the bigger the scale factor is, the stronger the coupling is.

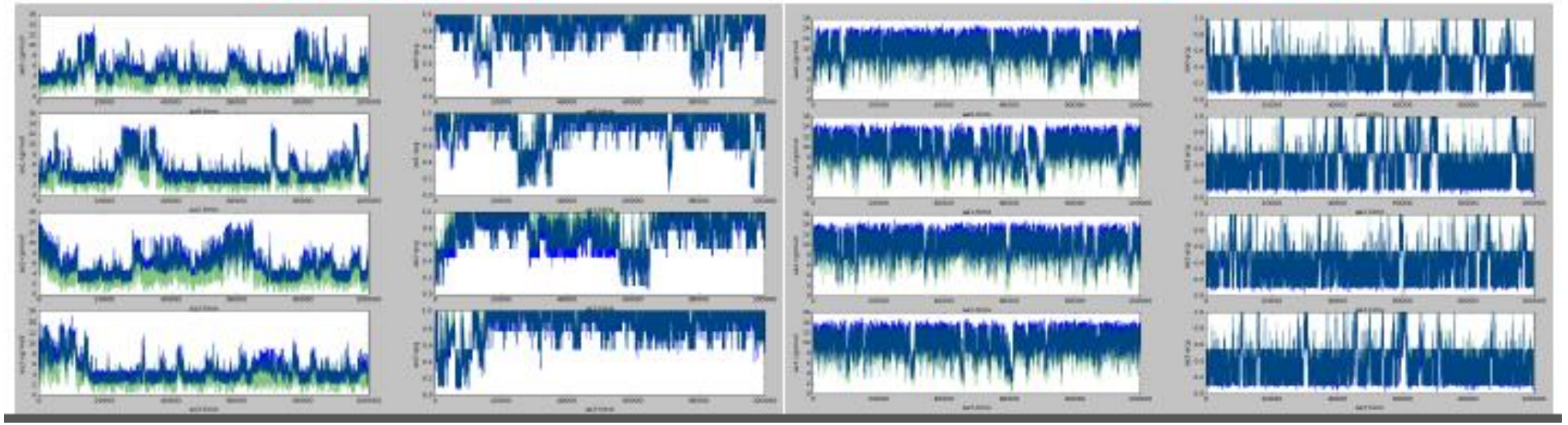
290 K & MSES scale=0.05: MD simulations

336 K & MSES scale=0.25



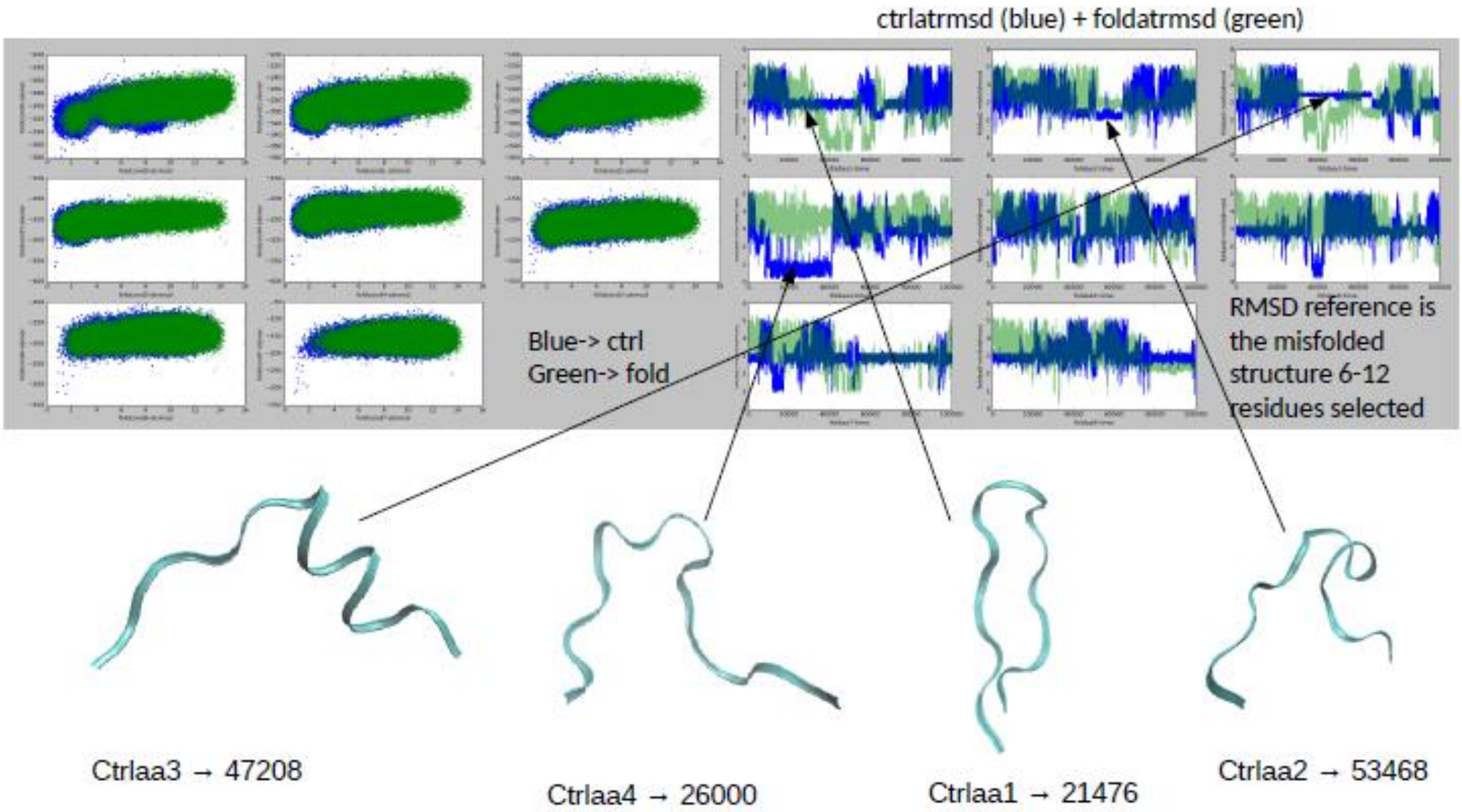
388 K & MSES scale=0.60

450 K & MSES scale=1.00



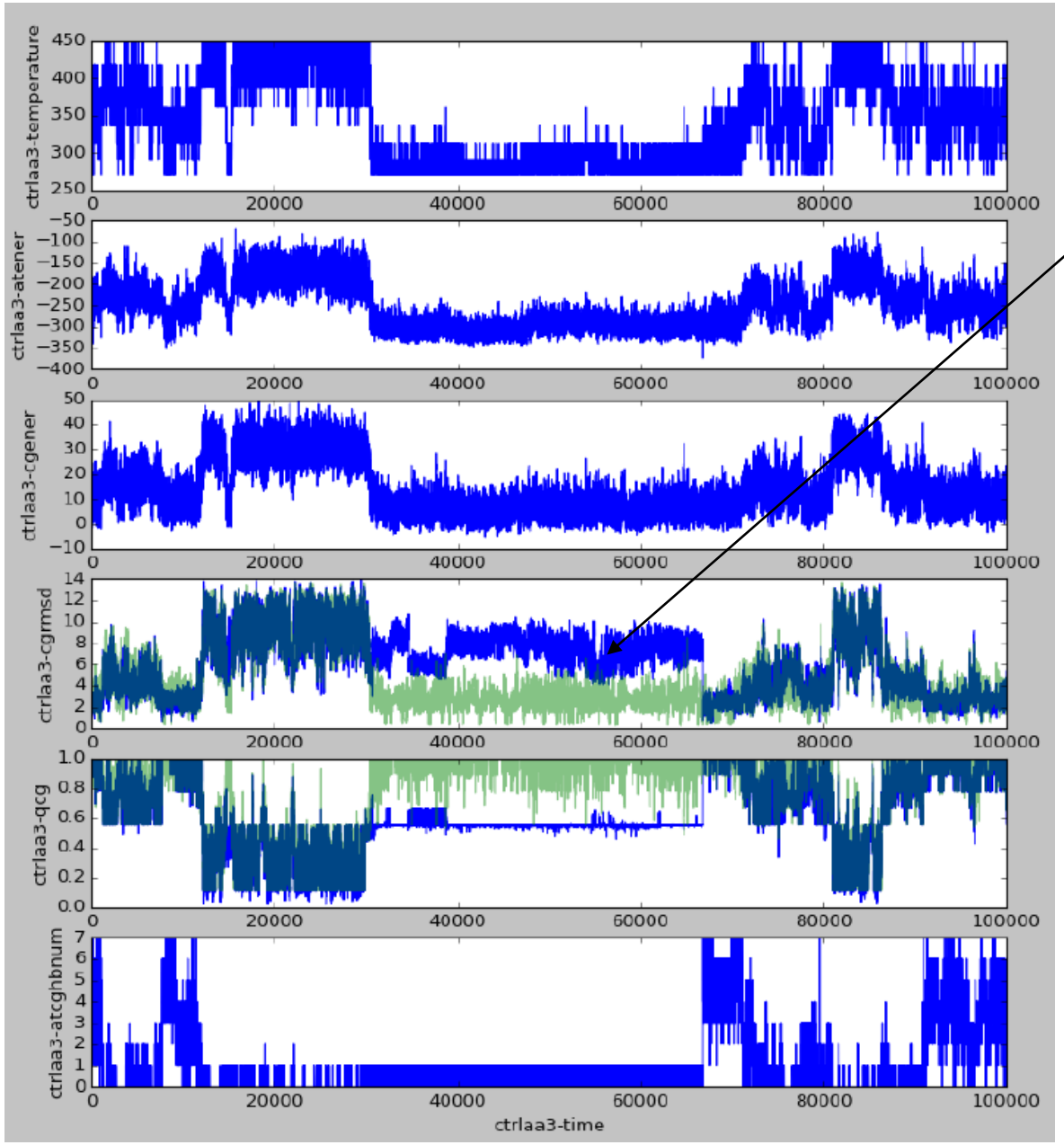
Analysis/conclusion of the results: progress and remaining problems

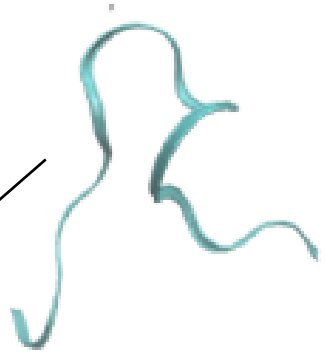
Calculate the RMSD versus potential energy in AT system. The goal is to check whether the trapped state can be found in this simulations. The results are that many trapped states can be founded in the MSES replica exchange simulations, including the state Xiaorong found in her TREX simulations.



Similar trapped states can be found in CPU simulations, please look at the "*thmsesgb1psimulations.pdf* " document for more details.

The existing problem: T/H-MSES-REX often goes to trapped states (in uncoupled condition)

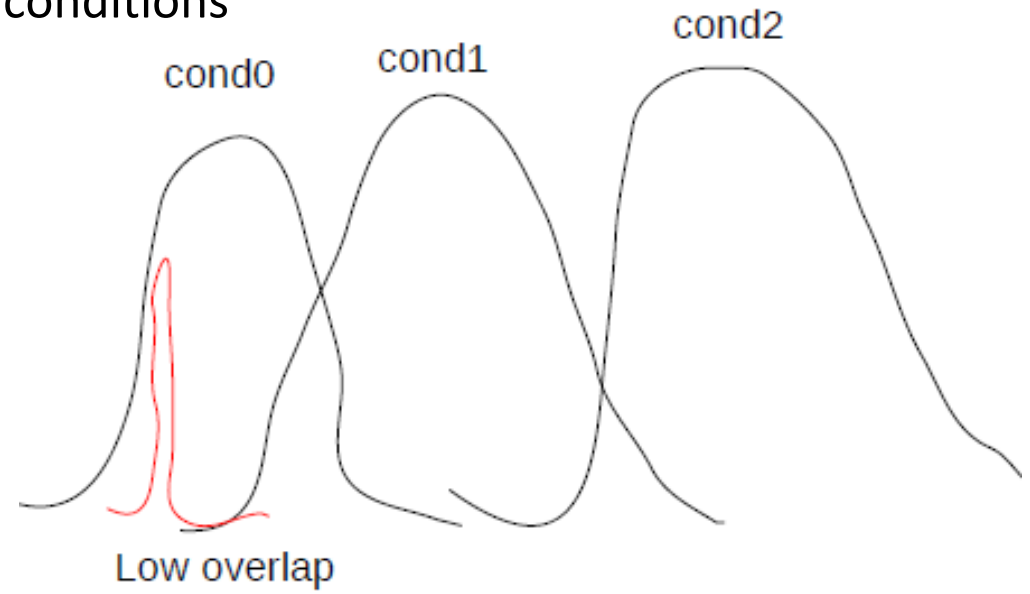




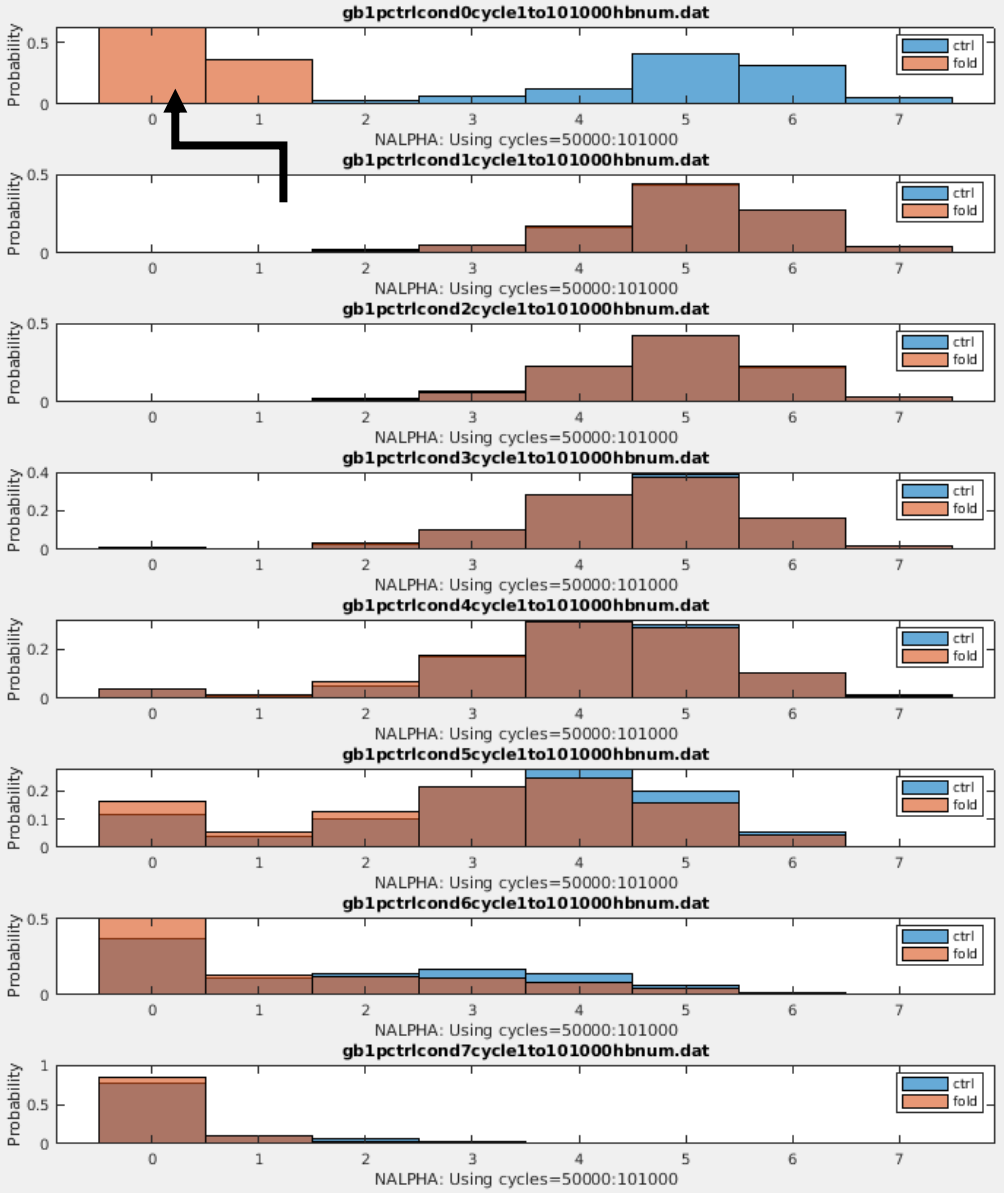
Trapped state

Explanation

These trapped states: CG is a folded state and AT is a meta-state. It causes a very shape and stable energy distribution, resulting in a low overlap with other conditions



Is that possible to calculate the unbiased probability of any properties only from the biased simulations?



MBAR technique? J. Chem. Phys. 129, 124105 (2008)

$$\hat{f}_i = -\ln \sum_{j=1}^{K} \sum_{n=1}^{N_j} \frac{\exp[-u_i(\boldsymbol{x}_{jn})]}{\sum_{k=1}^{K} N_k \exp[\hat{f}_k - u_k(\boldsymbol{x}_{jn})]}, \qquad (11)$$

Unfortunately, it needs not only converged biased samples but also unbiased samples, suggesting it could not work for our case.

However, it is still feasible to use some other reweighting techniques to obtain the unbiased samples from biased simulations!

Ideas from accelerated molecular dynamic simulations

From an aMD simulation, the ensemble average, $\langle A \rangle$, of an observable, $A(\mathbf{r})$, can be calculated using the following reweighting procedure:

$$\langle A \rangle = \frac{\int d\mathbf{r}\, A(\mathbf{r}) \exp(-\beta V(\mathbf{r}))}{\int d\mathbf{r} \exp(-\beta V(\mathbf{r}))} \quad (3)$$

$$= \frac{\int d\mathbf{r}\, A(\mathbf{r}) \exp(-\beta V(\mathbf{r}))}{\int d\mathbf{r} \exp(-\beta V^*(\mathbf{r}))} \Big/ \frac{\int d\mathbf{r} \exp(-\beta V(\mathbf{r}))}{\int d\mathbf{r} \exp(-\beta V^*(\mathbf{r}))} \quad (4)$$

$$= \frac{\int d\mathbf{r}\, A(\mathbf{r}) \exp(\beta \Delta V(\mathbf{r})) \exp(-\beta V^*(\mathbf{r}))}{\int d\mathbf{r} \exp(-\beta V^*(\mathbf{r}))} \Big/ \frac{\int d\mathbf{r} \exp(\beta \Delta V(\mathbf{r})) \exp(-\beta V^*(\mathbf{r}))}{\int d\mathbf{r} \exp(-\beta V^*(\mathbf{r}))} \quad (5)$$

$$= \frac{\langle A(\mathbf{r}) \exp(\beta \Delta V(\mathbf{r})) \rangle^*}{\langle \exp(\beta \Delta V(\mathbf{r})) \rangle^*}, \quad (6)$$

in which $\beta = 1 = k_B T$, where k_B is the Boltzmann constant, T is the temperature, and $\langle \ldots \rangle$ and $\langle \ldots \rangle^*$ represent the ensemble average in the original (unbiased) and the aMD (biased) ensembles, respectively.

Comput Sci Discov. **2011**, 4(1)
J. Chem. Theory Comput. **2014**, 10, 2677−2689.

This states that the unbiased average of any observable can be estimated from a biased MD simulation, which is exactly what we need.
However, it still has the “energetic noise” problem.

The probability distribution P(A) along a reaction coordinate A(r)?

$$P(A_i) = \left\langle \delta\left[A(\mathbf{r}) - A_i\right] \right\rangle = \frac{\int d\mathbf{r}\, \delta\left[A(\mathbf{r}) - A_i\right] \exp\left(-\beta V(\mathbf{r})\right)}{\int d\mathbf{r} \exp\left(-\beta V(\mathbf{r})\right)}$$

$$= \frac{\left\langle \delta\left[A(\mathbf{r}) - A_i\right] \exp\left(\beta^* V^*(\mathbf{r}) - \beta V(\mathbf{r})\right) \right\rangle^*}{\left\langle \exp\left(\beta^* V^*(\mathbf{r}) - \beta V(\mathbf{r})\right) \right\rangle^*}$$

$$= \frac{P^*(A_i) \left\langle \exp\left(\beta^* V^*(\mathbf{r}) - \beta V(\mathbf{r})\right) \right\rangle_k^*}{\sum_{k=1}^{M} P^*(A_k) \left\langle \exp\left(\beta^* V^*(\mathbf{r}) - \beta V(\mathbf{r})\right) \right\rangle_k^*} = \frac{P^*(A_i) \left\langle e^{\Delta S(\mathbf{r})} \right\rangle_k^*}{\sum_{k=1}^{M} P^*(A_k) \left\langle e^{\Delta S(\mathbf{r})} \right\rangle_k^*}.$$

Hard to estimate

My solution: combining Metropolis-Hastings algorithms and MD simulations and achieving an optimal reweighting technique.

The central idea is using the biased MD simulations to generate the conformational samples and screening out the unbiased conformations from biased samples by the Metropolis-Hastings algorithm.

Reviewing the Metropolis-Hastings idea

$$P_{\mathrm{MH}}(x \mid y)\pi(y) = P_{\mathrm{MH}}(y \mid x)\pi(x)$$

Markov chain

$$P_{\mathrm{MH}}(x \mid y) = q(x \mid y)\alpha(x \mid y),$$

$$\alpha(y \mid x) = \min\left[\frac{\pi(y)q(x \mid y)}{\pi(x)q(y \mid x)}, 1\right],$$

Metropolis-Hasting criterion

Acceptance probability

Proposal distribution could be any function

Molecular dynamic-based Metropolis-Hastings (MDMH): Methods

1) Setting the proposal distribution to a biased distribution.

$$q\left(x \mid y\right) = \pi^{*}\left(x\right)$$

The acceptance probability can be written in a Metropolis-like way,

$$\alpha\left(y \mid x\right) = \min\left[\frac{w\left(y\right)}{w\left(x\right)}, 1\right], \quad \text{Eq. (3)}$$

$$w\left(x\right) = \pi\left(x\right) / \pi^{*}\left(x\right).$$

- It crosses out the normalized terms of biased and unbiased distribution;
- The importance of unbiased representative rare states will be enhanced by biased simulations that only sample a few of them.

2) For the canonical NVT ensembles (similar descriptions for NPT, μVT ensembles), the density distribution is exponentially proportional to the of total energy of system, . Therefore, one general biased Hamiltonian, H*(p, r) can be described as follows,

$$\beta^{*} H^{*}\left(\mathbf{p}, \mathbf{r}\right) = \beta H\left(\mathbf{p}, \mathbf{r}\right) + \log\left(w\left(\mathbf{p}, \mathbf{r}\right)\right). \quad \text{Eq. (4)}$$

Biased Hamiltonian

log-weight distribution

Molecular dynamic-based Metropolis-Hastings (MDMH): Algorithms

0. Initialized an unbiased configuration and momentum $\mathbf{x} = (\mathbf{r}_0 \,\&\, \mathbf{p}_0)$ for your molecular system.

1. Proposal a conformational candidate ($\mathbf{y}$) with a configuration and momentum, $\mathbf{r}$ & $\mathbf{p}$, from a converged MD simulation using a biased Hamiltonian.

2. Calculate the log-like weight $\log\left(w\left(\mathbf{x}\right)\right)$ and $\log\left(w\left(\mathbf{y}\right)\right)$, and then compute log-like acceptance probability $\log\left(\alpha\left(\mathbf{y}\mid\mathbf{x}\right)\right)=\log\left(w\left(\mathbf{y}\right)\right)-\log\left(w\left(\mathbf{x}\right)\right)$ shown in Eq. 3 and 4.

3. Accept or reject this biased candidate

- generate a random value $u\in U\left(0,1\right)$;
- if $\log\left(\alpha\left(\mathbf{y}\mid\mathbf{x}\right)\right)>$ log (u): accept this biased candidate is unbiased, $\mathbf{y} \rightarrow \mathbf{x}$

  else: reject this biased candidate and keep current unbiased candidate, $\mathbf{x} \rightarrow \mathbf{x}$

4. Repeat steps 1-3, until the unbiased ($\mathbf{x}$) and biased ($\mathbf{y}$) simulations sufficiently sample the representative conformational space.

Molecular dynamic-based Metropolis-Hastings (MDMH): Significance

- The MDMH algorithm works for either the unnormalized low-dimensional probability along one reaction coordinate or the unnormalized high-dimensional probability of conformations.
-
- It is a general reweighting technique, which is suitable not only for biased simulations (e.g., accelerated MD simulations), but also for the replica exchange simulations including unbiased and biased conditions (e.g., temperature or Hamiltonian replica exchange simulations)

Future plan and why

The future plan is trying to implement the MDMH algorithm and use the existing biased simulations data to generate the unbiased conformations.