Published in final edited form as:

Methods Mol Biol. 2018 ; 1777: 101–119. doi:10.1007/978-1-4939-7811-3_5.

Replica Exchange Molecular Dynamics: A Practical Application Protocol with Solutions to Common Problems and a Peptide Aggregation and Self-Assembly Example

Ruxi Qi¹, Guanghong Wei², Buyong Ma³, and Ruth Nussinov^{4,5}

¹Department of Physics, State Key Laboratory of Surface Physics, Key Laboratory for Computational Physical Sciences (MOE), Fudan University, Shanghai, P.R. China.

²Department of Physics, State Key Laboratory of Surface Physics, Key Laboratory for Computational Physical Sciences (MOE), Fudan University, Shanghai, P.R. China. ghwei@fudan.edu.cn.

³Basic Science Program, Leidos Biomedical Research, Inc. Cancer and Inflammation Program, National Cancer Institute, Frederick, MD, USA.

⁴Basic Science Program, Leidos Biomedical Research, Inc. Cancer and Inflammation Program, National Cancer Institute, Frederick, MD, USA. nussinov@helix.nih.gov.

⁵Sackler Institute of Molecular Medicine, Department of Human Genetics and Molecular Medicine Sackler School of Medicine, Tel Aviv University, Tel Aviv, Israel. nussinov@helix.nih.gov.

Abstract

Protein aggregation is associated with many human diseases such as Alzheimer's disease (AD), Parkinson's disease (PD), and type II diabetes (T2D). Understanding the molecular mechanism of protein aggregation is essential for therapy development. Molecular dynamics (MD) simulations have been shown as powerful tools to study protein aggregation. However, conventional MD simulations can hardly sample the whole conformational space of complex protein systems within acceptable simulation time as it can be easily trapped in local minimum-energy states. Many enhanced sampling methods have been developed. Among these, the replica exchange molecular dynamics (REMD) method has gained great popularity. By combining MD simulation with the Monte Carlo algorithm, the REMD method is capable of overcoming high energy-barriers easily and of sampling sufficiently the conformational space of proteins. In this chapter, we present a brief introduction to REMD method and a practical application protocol with a case study of the dimerization of the 11–25 fragment of human islet amyloid polypeptide (hIAPP(11–25)), using the GROMACS software. We also provide solutions to problems that are often encountered in practical use, and provide some useful scripts/commands from our research that can be easily adapted to other systems.

Keywords

Replica exchange method; Molecular dynamics simulations; Free energy landscape; Protein aggregation; Human islet amyloid polypeptide; GROMACS; REMD

1 Introduction

1.1 Replica Exchange Molecular Dynamics Study of Peptide Aggregation and Self-Assembly

Proteins and peptides can self-assemble into β -sheet rich filaments that form naturally in bacteria and abnormally in tissues of patients with neurodegenerative diseases. Such assemblies also have remarkable mechanical properties with potential nanotechnological applications [1, 2]. The pathological self-assembly of amyloidogenic proteins is related to many human diseases such as Alzheimer's disease (AD), Parkinson's disease (PD), and type II diabetes (T2D) [3–5]. Revealing the mechanism of protein self-assembly is essential for the development of effective inhibitors against amyloidosis and designing new biological materials. Computational approaches are powerful tools to study the early self-assembly step of protein aggregation, which is challenging to characterize experimentally due to the fast aggregation process and the transient nature of oligomers. Conventional molecular dynamics (MD) simulations can hardly explore the complete free energy surface of biomolecules as the simulated system could be easily trapped in local free-energy-minimum conformations. Thus, enhanced sampling methods have been developed in recent years [6, 7], such as metadynamics [8] and replica exchange molecular dynamics (REMD) [9]. Here we focus on the REMD method, a highly practical and efficient algorithm to study peptide selfassembly.

The REMD method was first introduced in the study of biomolecules by Okamoto and coworkers [9]. It is a hybrid method by combining MD simulations with the Monte Carlo algorithm. In REMD simulations, several copies (replicas) of the same system are simulated in parallel using MD simulations at different temperatures or at the same temperature but using different Hamiltonians. Swapping between neighboring replicas is periodically attempted with a probability given by the Metropolis criterion (Fig. 1). This method generates a generalized ensemble of the simulated system. In this way, REMD is capable of overcoming high-energy barriers easily and sampling conformational space sufficiently, which allows exploration of the free energy landscape of protein aggregates. Furthermore, the parallel feature of the REMD method makes it suitable for the highly parallel computer cluster resources that are often available nowadays.

In this chapter, we present a brief introduction to REMD method and a practical application to study the initial dimerization step of the 11–25 fragment of human islet amyloid polypeptide (hIAPP(11–25)) aggregation [10]. We describe the detailed protocol for performing a REMD simulation and data analysis, and discuss the problems that are often encountered in REMD simulations. We also provide some useful scripts/commands from our research experience that can be easily adapted to other biomolecular systems.

1.2 REMD Method

The methodology of REMD was presented in detail in a previous study [9]. Here, we provide a brief description. Considering a system with N particles of mass $m_k (k = 1, \dots, N)$, with coordinates $q \equiv (q_1, \dots, q_N)$ and momenta $p \equiv (p_1, \dots, p_N)$, the Hamiltonian of the system is

$$H(q, p) = K(p) + V(q)$$

where $K(p)$ is the kinetic energy $\left(K(p) = \sum_{k=1}^N \frac{p_k^2}{2m_k}\right)$, and $V(q)$ is the potential energy.

In the canonical ensemble at temperature T , the probability of finding the system in state $x \equiv (q, p)$ is

$$\rho_B(x, T) = \exp[-\beta H(q, p)]$$

where $\beta = 1/k_B T$, and k_B is Boltzmann constant.

The simulation system has the generalized ensemble, where M non-interacting copies (replicas) of the above system are distributed at M different unique temperatures $T_m (m = 1, \dots, M)$. The replica $i (i = 1, \dots, M)$ corresponds to the temperature

$$T_m (m = 1, \dots, M).$$

Let

$$X = (x_i^{[i(1)]}, \dots, x_i^{[i(M)]}) = (x_{m(1)}^{[i]}, \dots, x_{m(M)}^{[i]})$$

denotes the state of the generalized ensemble, and X is specified by the M sets of coordinates $q^{[i]}$ and momenta $p^{[i]}$ of N particles in replica i at temperature T_m :

$$x_m^i \equiv (q^{[i]}, p^{[i]})_m$$

Since each replica is independent from others, the probability of the generalized ensemble in state X is

$$\begin{aligned} \rho_{\text{REM}}(X) &= \exp \left\{ - \sum_{i=1}^M \beta_{m(i)} H(q^{[i]}, p^{[i]}) \right\} \\ &= \exp \left\{ - \sum_{m=1}^M \beta_m H(q^{[i(m)]}, p^{[i(m)]}) \right\} \end{aligned}$$

where $i(m)$ and $m(i)$ are connected by

$$\begin{cases} i = i(m) \\ m = m(i) \end{cases}$$

The simulation tends to exchange replica i (at temperature T_m) and j (at temperature T_n):

$$X = (\dots, x_m^{[i]}, \dots, x_n^{[j]}, \dots) \rightarrow X' = (\dots, x_m^{[j]}, \dots, x_n^{[i]}, \dots)$$

or in detail,

$$\begin{cases} x_m^i \equiv (q^{[i]}, p^{[i]})_m \rightarrow x_m^{[j]} \equiv (q^{[j]}, p^{[j]})_m \\ x_n^j \equiv (q^{[j]}, p^{[j]})_n \rightarrow x_n^{[i]} \equiv (q^{[i]}, p^{[i]})_n \end{cases}$$

Note that exchanging temperature between two replicas is equivalent to exchanging two configurations at adjacent temperatures. In this REMD protocol, the momenta should also be rescaled,

$$\begin{cases} p^{[i]} \equiv \sqrt{\frac{T_n}{T_m}} p^{[i]} & p^{[j]} \equiv \sqrt{\frac{T_m}{T_n}} p^{[j]} \end{cases}$$

This scheme rescales uniformly the momenta of all particles to keep at a proper temperature, and furthermore, provides a neat way to satisfy the detailed balance condition by canceling out the kinetic energy terms, as shown below.

For transition probability $w(X \rightarrow X')$, the detailed balance condition requires

$$\rho_{\text{REM}}^{(X)} w(X \rightarrow X') = \rho_{\text{REM}}^{(X')} w(X' \rightarrow X)$$

Thus, we have

$$\begin{aligned} \frac{w(X \rightarrow X')}{w(X' \rightarrow X)} &= \exp \left\{ -\beta_m \left[K(p^{[j]}) + V(q^{[j]}) \right] - \beta_n \left[K(p^{[i]}) + V(q^{[i]}) \right] \right. \\ &\quad \left. + \beta_m \left[K(p^{[i]}) + V(q^{[i]}) \right] + \beta_n \left[K(p^{[j]}) + V(q^{[j]}) \right] \right\} \\ &= \exp \left\{ -\beta_m \frac{T_m}{T_n} K(p^{[j]}) - \beta_n \frac{T_n}{T_m} K(p^{[i]}) + \beta_m K(p^{[i]}) + \beta_n K(p^{[j]}) \right. \\ &\quad \left. - \beta_m \left[V(q^{[j]}) - V(q^{[i]}) \right] - \beta_n \left[V(q^{[i]}) - V(q^{[j]}) \right] \right\} = \exp(-\Delta) \end{aligned}$$

$$\text{where } \Delta \equiv [\beta_n - \beta_m] (V(q^{[i]}) - V(q^{[j]})).$$

This can be satisfied by introducing the Metropolis criterion [11]:

$$w(X \rightarrow X') \equiv w(x_m^{[i]} | x_n^{[j]}) = \min(1, \exp(-\Delta))$$

2 Materials

2.1 GROMACS-4.5.3 [12]

The REMD simulation can be performed using popular MD simulation packages including GROMACS [12], AMBER [13], CHARMM [14], NAMD [15], etc. For this case study we use GROMACS-4.5.3. The implementation in other GROMACS versions is straightforward. When using other simulation packages, we recommend referring to the documents or tutorials provided by the official sites of these packages.

2.2 High Performance Computing (HPC) Cluster

The REMD simulations are highly parallel and resource demanding. HPC cluster installed with the GROMACS-4.5.3 and a standard message passing interface (MPI) library are required. Typically two cores per replica can give nice productivity in a HPC cluster equipped with Intel Xeon X5650 CPUs or above.

2.3 Visual Molecular Dynamics (VMD)

In this protocol, the versatile package VMD [16] is used for molecular modeling and structure visualization. VMD supports computers running MacOS X, Unix, or Windows systems, and is distributed free of charge.

2.4 Linux Shell

Some shell scripts coded in Bash are provided for data preparation and file processing. These scripts can run on a normal Linux distribution or a virtual Linux environment such as Cygwin on Windows.

3 Methods

Here we illustrate how to perform a REMD study from scratch using the GROMACS-4.5.3 and VMD packages via a case study of hIAPP(11–25) dimer system at a constant pressure and temperature (*NPT*) ensemble. The peptide with amino acid sequence of RLANFLVHSSNNFGA is capped by an acetyl (CH_3CO) group at the N-terminus and a NH_2 group at the C-terminus, as done experimentally [17]. Note that the REMD algorithm was initially developed for canonical ensemble (*NVT*, constant volume and constant temperature). It can be easily adapted to the *NPT* ensemble with the Hamiltonian of the system as

$$H(q, p) = K(p) + V(q) + PV$$

where P and V are the pressure and volume of the system, respectively. The contribution of volume fluctuations to the total energy is negligible [18].

The REMD method combines the replica exchange algorithm with multiple MD simulations. Thus, the initial step of the REMD simulation is identical to that of the conventional MD simulation. In the final step, a series of parallel MD replicas at different temperatures are created. Here we illustrate the major processes of carrying out a REMD

study without too much details, unless there are some key issues needing further clarification.

3.1 Construct an Initial Configuration of hIAPP(11–25) Dimer

The first step in performing a REMD study is to construct a starting configuration (in some cases, more than one initial configurations are needed) (*see Note 1*).

1. Use VMD TCL console to construct a configuration in which the two hIAPP(11–25) molecules have a vertical chain distance of 1.4 nm, i.e., with no atomic contacts (*see Note 2*).

To construct the desired configuration, one can construct a configuration by manipulating the molecules via modifying and combining the following TCL commands (# comments):

```
set chainA [atomselect 0 "all"] # Select chain A
set chainB [atomselect 1 "all"] # Select chain B
measure center $chainA # Measure the center of mass of chain A
measure center $chainB
$chainA moveby {0 0 5} # Move chain A by 5 Angstrom in Z direction
$chainA move [transaxis z 30 deg] # Rotate chain A by 30 degree around Z axis
$chainA writpdb chainA.pdb
$chainB writpdb chainB.pdb
```

2. Merge the two chains into one pdb file, e.g., dimer.pdb, by Linux Shell commands, such as cat, and then use the Gromacs command *editconf* to convert it to proper pdb format (Fig. 2).
3. Use *pdb2gmx* command tool to create the coordinate and topology files.

```
pdb2gmx -f dimer.pdb -o dimer.gro -p dimer.top -inter -ignh
```

3.2 Choose a Simulation Box

1. After we created an initial configuration, now we need to solvate the molecule in a water box by the following command.

```
editconf -f dimer.gro -o dimer_box.gro -c -box 5 5 5
```

Note that the box should be big enough so that the molecule will not interact with its neighboring mirror (i.e., within the potential distance cutoff) under periodic boundary condition (PBC). Beware that the protein molecule rotates in solution. As a general rule of thumb, the rotational correlation time τ_c (time taken to rotate by one radian) of a monomeric protein in solution in nanoseconds is approximately 0.6 multiplying its molecular weight in kDa [19], which means that a 90° rotation of a protein takes time $\sim [0.94 \times \text{molecular weight (in kDa)}]$ ns. For a globular protein system a more spherical box shape such as

dodecahedron is highly recommended, which saves as much as 29% water molecules (*see Note 3*).

3.3 Do Energy Minimization and Equilibrium

1. After creating a simulation box, we need to do the energy minimization and equilibrium in both vacuum and solvent. The following commands do the job.

```
grompp -f em.mdp -c dimer_box.gro -p dimer.top -o dimer_ - em_vac.tpr
mdrun -v -deffnm dimer_em_vac

genbox -cp dimer_em_vac.gro -cs spc216.gro -p dimer.top -o dimer_sol.gro
grompp -f em.mdp -c dimer_sol.gro -p dimer.top -o dimer_em_sol.tpr
mdrun -v -deffnm dimer_em_sol
```

2. Now add counter ions and extra salt ions to get a neutral system usually with a final salt concentration of 0.1 M.

```
grompp -f em.mdp -c dimer_em_sol.gro -p dimer.top -o dimer_ions.tpr
genion -s dimer_ions.tpr -p dimer.top -o dimer_ions.gro -neutral -conc 0.1
Prompt selection: SOL

grompp -f em.mdp -c dimer_ions.gro -p dimer.top -o dimer_em_ions.tpr
mdrun -v -deffnm dimer_em_ions
```

3. Equilibrate the system in NVT ensemble first and then in NPT ensemble to further relax the system and reduce the possibility of system exploding in production run afterwards.

```
grompp -f NVT.mdp -c dimer_em_ions.gro -p dimer.top -o dimer_nvt.tpr
mdrun -v -deffnm dimer_nvt

grompp -f NPT.mdp -c dimer_nvt.gro -p dimer.top -o dimer_npt.tpr
mdrun -v -deffnm dimer_npt
```

3.4 Set Temperature Distribution and Create Replicas

Now that we have a well-equilibrated system, it's time to construct replicas for the REMD simulation. A key step in the REMD method is to set a proper temperature distribution, which should give a sufficiently high temperature and similar acceptance ratio between all adjacent replica pairs over the entire temperature range [9, 20]. In addition, for efficiency each replica should spend equal amounts of time at each temperature [21]. The upper limit of the temperature distribution should be high enough so that no trapping in a local minimum-energy state occurs. However, this is not straightforward to verify even though random walks among replicas are observed. Typically one can examine the selection of the highest temperature by comparing the energy distribution with a classical MD run at an expected temperature, which should not produce a significant deviation. Previous studies

showed that the maximum temperature should be just above the temperature where the activation free energy vanishes and any replicas higher than that would decrease the efficiency due to non-Arrhenius behavior [22, 23]. In the REMD study of hIAPP(11–25) dimer, 453 K was chosen and verified as a sufficient temperature upper limit. The lower limit temperature is usually near the temperature of highest interest (mostly at the physiological temperature of 310 K for the study of peptide aggregation).

Temperatures that are distributed exponentially usually provide a uniform acceptance ratio in the limit of constant heat capacity [24]. Different algorithms have been developed to further optimize this distribution. Here we recommend an algorithm derived by Patriksson and van der Spoel [20] where the energy distributions are assumed to be Gaussian. A temperature distribution can be provided once the number of atoms in the system, upper and lower temperature limits, and a desired exchange probability are given. No further prior knowledge about energies or temperatures is needed.

1. Generate the temperature distribution ranging from 310 to 453 K using the web server (<http://folding.bmc.uu.se/remd>). It should be noted that the temperature differences between adjacent replicas should not be too large in order to obtain sufficient potential energy overlaps between the two replicas, thus the acceptance ratio between them is not too small (as a rule of thumb, 0.2~0.3 is good). A short (e.g., 1 ns) REMD test is practicable to check the acceptance ratio.
2. Generate the replicas of the hIAPP(11–25) dimer system. With a selected temperature distribution, the replicas can be generated using the following Bash script `create_replicas.sh` (with instructions on how to prepare files in # comments):

```
#-----Begin-----
#!/bin/bash

# First prepare the file temperature.dat with your temperature distribution
arranged in two columns as:

# 0 310.0
# 1 312.84
# 2 315.7
# ...

# Then in remd.mdp file, specify any temperature value with the “TEMP” string,
e.g.

# ref_t = TEMP TEMP

# gen_temp = TEMP

# and specify the remd.mdp file directory here.

mdpFile=/DIRECTORY/TO/YOUR/REMD.MDP

# Execute this script by:
```



```

# ./create_replicas.sh < temperature.dat

while read replica temp
do
sed -e 's/TEMP/'$temp'/g' $mdpFile > rep_${replica}.mdp

# Option 1. Generate *.tpr with single starting structure, modify it if needed
grompp -f rep_${replica}.mdp -c dimer_npt.gro -p dimer.top -o dimer_${replica}.tpr

# Option 2. Generate *.tpr with 4 (or any other number) starting structures
(should be named as dimer_npt_0.gro dimer_npt_1.gro dimer_npt_2.gro
dimer_npt_3.gro etc), modify it if needed

# grompp -f rep_${replica}.mdp -c dimer_npt_${replica%4}). gro -p dimer.top -o
dimer_${replica}.tpr

done

rm \#* mdout.mdp

#-----End-----

```

3. Run a 1-ns REMD simulation to check whether the acceptance ratio between two adjacent replicas is within 0.2~0.3. Adjust the temperature limits and/or the exchange probability to regenerate the desired temperature distribution.

3.5 Running the Simulation

1. Now, run the REMD simulation by the following command (suppose we have 44 replicas):

```
mdrun -s dimer_.tpr -multi 44 -replex 1000 -maxh 240 -noappend
```

For data safety and subsequent ease-of-coding, we recommend using the parameters `–noappend` to save the output data separately, and `-maxh` to gracefully terminate REMD jobs before reaching the HPC cluster wall time limit (e.g., 10 days) that can easily corrupt data. To extend or restart a normally terminated/ crashed job, checkpoint (.cpt) files are needed. However, data outputs including .cpt files are not always saved at the same pace among all replicas (*see Note 4*). A practical trick to avoid corrupting a .cpt file is to use the parameter `–cpt`, e.g., 60 (min) and set long enough time intervals of writing out .cpt file.

The `-replex 1000` was used in our REMD simulation, which means we make the exchange trial between a pair of replicas every 1000 integration steps (with an integration time step of 2 fs). Different swap time settings can affect the sampling efficiency of REMD simulation. Details on this issue are discussed in **Note 5**.

3.6 Data Analysis

3.6.1 Convergence Check of REMD Simulation

The first step to analyze the data after finishing REMD run is to check the convergence of the simulation, as physical variables can only be correctly calculated based on a converged sampling. There are several ways to assess the convergence (Figs. 3, 4, and 5), including checking physical properties between two independent time intervals and checking whether a replica spans full temperature space multiple times by tracing a selected replica with simulation time.

1. Monitor the time evolution of hIAPP(11–25) dimer replicas by tracing one replica (replica 1) in temperature space.

The time evolution of replica 1 shows that the peptide visited sufficiently the entire temperature space (Fig. 3).

2. Calculate the percentage of dwell time at each temperature for each of the 44 replicas.

Figure 4 demonstrates each replica samples all of the 44 temperatures through the 300-ns simulation time, indicating sufficient sampling. The overall standard deviations also show an appropriate sampling of the REMD simulation. For rigorous results, we need to check the distributions of some physical variables within two or more independent time intervals. For example, in our hIAPP(11–25) dimer system, the convergence of the REMD run was checked using the probability distributions of R_g generated within two independent time intervals: 50–175 ns and 175–300 ns. As seen from Fig. 5, the two probability distribution curves overlap very well, indicating a good convergence of the REMD simulations.

3. Compare simulation results with experimental data if available. A comparison of simulation with experimental results is necessary if experimental data are available. A good agreement between them can validate the reliability of the simulation results. As in our hIAPP(11–25) monomer REMD simulation, we calculated the H^α chemical shifts of each amino acid residue using SPARTA + [25], and compared it with solution NMR data [17]. As shown in Fig. 6a, b, our simulations essentially reproduced the experimental H^α chemical shifts. The calculated H^α secondary chemical shifts were also in agreement with solution NMR experimental observations, with a Pearson correlation coefficient of 0.57. These results demonstrate that our REMD simulations reasonably converged.

3.6.2 Fix the Periodic Boundary Condition (PBC) Problem

Simulations are normally performed under PBC. Molecules are often seen in a bond-broken state if they cross the PBC boundaries. We need to restore all atoms into the same unit cell to avoid bond broken state. Generally the following steps can do the job.

1. Make a group in.ndx file using make_ndx to select a group of atoms or molecules as center, which will make the whole system gathered into one unit cell. Then do the following command:

```
trjconv -f traj.xtc -o center.xtc -n index.ndx -center -boxcenter rect
```

2. trjconv -f center.xtc -o inbox.xtc -ur compact -pbc atom.
3. trjconv -f inbox.xtc -s topol.tpr -o final.xtc -pbc mol.

Then we need to check the final.xtc with VMD. If there are still molecules that diffuse away (often encountered with multi-peptides system), repeat the **steps 1~3** and use a different center group in the **step 1**.

3.6.3 Calculate Chain- Independent RMSD for Multi-Peptide System

It is often useful to cluster all the conformations sampled in the REMD simulation, either for characterizing the conformational ensemble or for further use in other analyses (such as Markov State Model (MSM)) [26, 27]. However, a common issue frequently encountered with the RMSD-based clustering is the exchange effect of topologically identical peptide chains in the coordinate files. As more typically shown in our previous REMD study on A β (16–22) octamer system [28], the eight chains are not covalently connected, and their relative spatial positions in one conformation may be different from those of another. The RMSD value calculated directly using the coordinate file from the REMD run may not be the smallest and cannot be used for structure similarity comparison. In this situation, the permutation of all chains should be considered, although this may drastically increase the computational cost. While there is no such issue with single peptide/protein, ignoring the chain exchange effect in multi-peptide/protein system will cause errors in RMSD calculation.

3.6.4 Calculate the Connectivity Length to Characterize the Degree of Peptide Aggregation

Oligomerization is the initial process of protein aggregation and fibrillation, in which a quantity capable of capturing the degree of aggregation would be very useful. We have introduced a mathematical parameter—connectivity length (CL) to characterize the ordering of the β -sheet-rich oligomers in our previous study [29]. This method can also be used to characterize the extent of aggregation. CL is defined as the sum of the square root of the sizes of all oligomers (or β -sheet sizes in the characterization of oligomer ordering). For example, for an octa-peptide system, the resulting two-tetramer aggregate (4,4) will give a $CL = \sqrt{4} + \sqrt{4} = 4$; while another arrangement (1,3,2,2) (i.e., a system consisting of one monomer, one trimer, and two dimers) will give a $CL = \sqrt{1} + \sqrt{3} + \sqrt{2} + \sqrt{2} = 5.56$. Therefore, CL reflects the extent of oligomerization (or the ordering the oligomers). The smaller the CL is, the stronger the peptide oligomerization becomes (or the more ordered the oligomer is).

3.6.5 Construct the Free Energy Landscape

REMD simulations sample the conformational space more efficiently than conventional MD simulations, making it quite suitable to investigate peptide aggregation. Here, taking the

hIAPP(11–25) dimer system as an example, we illustrate the procedure of free energy landscape construction.

1. Choose two variables that can reflect the essential information of the system as order parameters.
2. Calculate the free energy of different states based on binned-state-statistics using $-RT \ln P(x, y)$, where R , T , P are, respectively, the gas constant, temperature, and probability of specific state.

For the hIAPP(11–25) dimer (Fig. 5), the two-dimensional free energy surface is constructed using $-RT \ln [P(R_g, \text{RMSD})]$, where $P(R_g, \text{RMSD})$ is the probability of a conformation having a certain value of R_g and RMSD. Here, R_g and RMSD denote the radius of gyration of the dimer and the root mean square of deviation of the configuration referenced to the structure E in Fig. 5 of [10]. Five minimum energy basins are observed in free energy landscape. The corresponding representative snapshots show hIAPP(11–25) dimers mostly adopt disordered conformations and to a lesser extent antiparallel β -sheet conformations. However, caution should be taken when explaining the result of a free energy landscape, as the information reflected by just two reaction coordinates could be incomplete.

3. Use different reaction coordinates to construct the free energy landscape.

To optimize the free energy analysis, the landscape based on different reaction coordinates should be plotted. For the hIAPP (11–25) dimer study, the free energy landscape based on R_g and Hydrogen bond was also constructed and are in good agreement with the one described here (Fig. 7).

3.6.6 Tips on REMD Data Manipulation

REMD simulation generates large data set, often in hundreds of gigabytes. Data transfer and manipulation should be taken carefully. We recommend a Linux tool *rsync* for data upload/download, which is better than the typical *scp* tool (see **Note 6**). In addition, a practical trick to save space in GROMACS REMD simulation is to convert the *.trr* format into the *.xtc* format, which significantly reduces the space occupation by removing the velocities and compressing the data, but still keeps a sufficient resolution (see **Note 6**).

3.6.7 Trajectory Visualization with VMD

There are versatile applications in VMD for MD trajectory visualization and analysis. When using VMD, a novice user may encounter the issue that the secondary structure doesn't change properly with trajectory animation. A TCL script can solve the problem (see **Note 7**).

All-atom water REMD is one of the most rigorous simulation methods currently used to explore the free energy surface of a protein system. With the proper system size and physical conditions, REMD can provide structure insights into various biomolecules. As an example here we illustrate the simulation of the peptide oligomer system, hIAPP(11–25), to obtain clues of peptide assemblies and aggregation mechanism. Our extensive REMD simulations

revealed the conformational distribution feature and the α -Helix to β -Sheet transition mechanism of the hIAPP(11–25) system.

4 Notes

1. In REMD simulations, multiple starting configurations are recommended.
2. Although the *editconf* tool implemented in GROMACS package can also do the same thing, the VMD TCL commands are more powerful.
3. For a dodecahedron simulation box, the system visualization in VMD will be shown as a triclinic box, which can be transferred to a normal dodecahedron view by the following GROMACS command:


```
trjconv -f tri.xtc -s md.tpr -pbc mol -ur compact -o dodeca.xtc
```
4. Due to the parallel feature of REMD simulation jobs, the .cpt files can be easily corrupted if all jobs are not terminated properly (e.g., by an abrupt job killing), so that the .cpt files of all replicas don't match when attempting a job continuation or restart. To check whether a .cpt file is damaged or not, use the following command:


```
gmxcheck -f state.cpt
```

A batch checking in advance for all checkpoint files can be performed by the following Shell script, which can avoid queuing-time waste in HPC cluster when .cpt files are actually corrupted.

```
#!/bin/bash
for i in $(seq 0 43) #Replica 0~43
do
if gmxcheck -f state$i\_prev.cpt &>/dev/null
then
echo $i.ok
else
echo $i.no
fi
done
```
5. Generally it has been assumed that for sampling efficiency a long enough relaxation time is needed for each replica's MD run after a swap. There has been a long debate on the exchange attempt frequency in REMD. Some argue that the time interval for exchange of conformations between adjacent replicas should not be shorter than the potential energy autocorrelation time [30, 31]. Others argue that the sampling efficiency increases with the increasing exchange-attempt frequency [32, 33]. Combining computational efficiency considerations and

REMD sampling, a rule-of-thumb attempt exchange time is 500~1500 integration steps (with a typical integration time step of 2 fs for biomolecules).

6. The Linux tool *rsync* is powerful for data backup and/or synchronization with the feature of breakpoint-resume, which can be conveniently used for transferring large REMD data. Simply use the following command to download data:

```
rsync -vaPz user@server-address:/Directory/of/data./Local/ directory
```

Note: caution with the parameter *-u*, which can improperly skip a partly transferred file.

We provide the following Shell script for converting all.trr files in a current directory to.xtc files, which can then be transferred by *rsync*. A secure deletion of the converted .trr files is executed to avoid the surge of disk occupation.

```
#---Begin---

#!/bin/bash

echo -e "\e[32mListing trr files...\e[0m"
echo ""
ifls *.trr
then
echo ""
echo -e "\e[32mNow converting all.trr to.trr.xtc\e[0m"
echo -e "\e[33mPlease wait\e[0m"
echo -e "\e[33m...\e[0m"
for i in `ls *.trr`
do
if
trjconv -f $i -o $i.xtc &>trr2xtc.out
then
rm $i
else
echo -e "\e[31m***Fatal error: Oh Gosh! Something bad happens! Please check
the output file trr2xtc.out\e[0m"
break
fi
done
echo ""
```

```

echo -e "\e[5m\e[32mComplete.\e[0m"

else

echo ""

echo -e "\e[31m***Error: Failed listing trr file. Please check the directory.\e[0m"

fi

#---End---

```

7. Type the following command in VMD TCL console to get a dynamic secondary structure visualization.

```

source ss.tcl

The content of ss.tcl is:

#---Begin---

trace variable vmd_frame(0) w pvar

proc pvar {name element op} {

mol ssrecalc top

}

#---End---

```

Acknowledgments

This project has been funded in whole or in part with Federal funds from the National Cancer Institute, National Institutes of Health, under contract number HHSN261200800001E. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government. This research was supported (in part) by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research.

References

1. Paparcone R, Cranford SW, Buehler MJ (2011) Self-folding and aggregation of amyloid nanofibrils. *Nanoscale* 3:1748–1755 [PubMed: 21347488]
2. Knowles TP, Buehler MJ (2011) Nanomechanics of functional and pathological amyloid materials. *Nat Nanotechnol* 6:469–479 [PubMed: 21804553]
3. Kelly JW (1998) The alternative conformations of amyloidogenic proteins and their multi-step assembly pathways. *Curr Opin Struct Biol* 8:101–106 [PubMed: 9519302]
4. Rochet J-C, Lansbury PT Jr (2000) Amyloid fibrillogenesis: themes and variations. *Curr Opin Struct Biol* 10:60–68 [PubMed: 10679462]
5. Clark A, Lewis CE, Willis AC, Cooper GJS, Morris JF, Reid KBM, Turner RC (1987) Islet amyloid formed from diabetes-associated peptide may be pathogenic in type-2 diabetes. *Lancet* 330:231–234
6. Mitsutake A, Mori Y, Okamoto Y (2013) Enhanced sampling algorithms. *Methods Mol Biol* (Clifton, NJ) 924:153–195
7. Barducci A, Pfauendtner J, Bonomi M (2015) Tackling sampling challenges in biomolecular simulations. *Methods Mol Biol* (Clifton, NJ) 1215:151–171

8. Laio A, Parrinello M (2002) Escaping free-energy minima. *Proc Natl Acad Sci USA* 99:12562–12566 [PubMed: 12271136]
9. Sugita Y, Okamoto Y (1999) Replica-exchange molecular dynamics method for protein folding. *Chem Phys Lett* 314:141–151
10. Qi R, Luo Y, Ma B, Nussinov R, Wei G (2014) Conformational distribution and alpha-helix to beta-sheet transition of human amylin fragment dimer. *Biomacromolecules* 15:122–131 [PubMed: 24313776]
11. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953) Equation of state calculations by fast computing machines. *J Chem Phys* 21:1087–1092
12. Berendsen HJC, van der Spoel D, van Drunen R (1995) GROMACS: a message-passing parallel molecular dynamics implementation. *Comput Phys Commun* 91:43–56
13. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J Am Chem Soc* 117:5179–5197
14. Brooks B, Bruccoleri R, Olafson B, States D, Swaminathan S (1983) Karplus M: CHARMM: a program for macromolecular energy minimization and dynamics calculations. *J Comput Chem* 4:187–217
15. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kalé LV, Schulten K (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* 26:1781–1802 [PubMed: 16222654]
16. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. *J Mol Graph* 14:33–38 [PubMed: 8744570]
17. Liu G, Prabhakar A, Aucoin D, Simon M, Sparks S, Robbins KJ, Sheen A, Petty SA, Lazo ND (2010) Mechanistic studies of peptide self-assembly: transient α -helices to stable β -sheets. *J Am Chem Soc* 132:18223–18232 [PubMed: 21138275]
18. Seibert MM, Patriksson A, Hess B, van der Spoel D (2005) Reproducible polypeptide folding and structure prediction using molecular dynamics simulations. *J Mol Biol* 354: 173–183 [PubMed: 16236315]
19. Cavanagh J, Fairbrother WJ, Palmer AG, Skelton NJ (1996) Protein NMR spectroscopy: principles and practice. *Psychol J*
20. Patriksson A, van der Spoel D (2008) A temperature predictor for parallel tempering simulations. *Phys Chem Chem Phys* 10:2073–2077 [PubMed: 18688361]
21. Rathore N, Chopra M, de Pablo JJ (2005) Optimal allocation of replicas in parallel tempering simulations. *J Chem Phys* 122:024111 [PubMed: 15638576]
22. Zheng W, Andrec M, Gallicchio E, Levy RM (2007) Simulating replica exchange simulations of protein folding with a kinetic network model. *Proc Natl Acad Sci* 104: 15340–15345 [PubMed: 17878309]
23. Nymeyer H (2008) How efficient is replica exchange molecular dynamics? An analytic approach. *J Chem Theory Comput* 4:626–636 [PubMed: 26620937]
24. Kofke DA (2002) On the acceptance probability of replica-exchange Monte Carlo trials. *J Chem Phys* 117:6911–6914
25. Shen Y, Bax A (2010) SPARTA+: a modest improvement in empirical NMR chemical shift prediction by means of an artificial neural network. *J Biomol NMR* 48:13–22 [PubMed: 20628786]
26. Bowman GR, Huang X, Pande VS (2009) Using generalized ensemble simulations and Markov state models to identify conformational states. *Methods (San Diego, Calif)* 49:197–201
27. Qiao Q, Qi R, Wei G, Huang X (2016) Dynamics of the conformational transitions during the dimerization of an intrinsically disordered peptide: a case study on the human islet amyloid polypeptide fragment. *Phys Chem Chem Phys* 18:29892–29904 [PubMed: 27759128]
28. Li H, Luo Y, Derreumaux P, Wei G (2011) Carbon nanotube inhibits the formation of β -sheet-rich oligomers of the Alzheimer's amyloid- β (16–22) peptide. *Biophys J* 101: 2267–2276 [PubMed: 22067167]

29. Lu Y, Derreumaux P, Guo Z, Mousseau N, Wei G (2009) Thermodynamics and dynamics of amyloid peptide oligomerization are sequence dependent. *Proteins* 75:954–963 [PubMed: 19089954]
30. Periole X, Mark AE (2007) Convergence and sampling efficiency in replica exchange simulations of peptide folding in explicit solvent. *J Chem Phys* 126:014903 [PubMed: 17212515]
31. Abraham MJ, Gready JE (2008) Ensuring mixing efficiency of replica-exchange molecular dynamics simulations. *J Chem Theory Comput* 4:1119–1128 [PubMed: 26636365]
32. Sindhikara DJ, Emerson DJ, Roitberg AE (2010) Exchange often and properly in replica exchange molecular dynamics. *J Chem Theory Comput* 6:2804–2808 [PubMed: 26616081]
33. Sindhikara D, Meng Y, Roitberg AE (2008) Exchange frequency in replica exchange molecular dynamics. *J Chem Phys* 128:024103 [PubMed: 18205439]

Fig. 1.
Illustration of replica exchange molecular dynamics (REMD) method

Fig. 2.
The initial conformation of hIAPP(11–25) dimer

Fig. 3. Convergence check of the REMD run of hIAPP(11–25) dimer. The time evolution of replica-1 in temperature space

Fig. 4.

The percentage of dwell time at each temperature for all of the 44 replicas. The overall standard deviations are shown in bars. With a certain exchange probability, a good sampling efficiency requires that the time a replica spends at every temperature site should be similar. So that the unbiased random walk of a replica through the whole temperature space is guaranteed

Fig. 5. Convergence check of the REMD run of hIAPP(11–25) dimer. Probability distributions of Rg within two different time intervals of 50–175 ns and 175–300 ns are overlapped very well

Fig. 6.
Comparison between calculated and experimental chemical shifts of hIAPP(11–25) monomer along with representative REMD-generated α -helical and β -structures at 310 K.
(a) H^α chemical shifts (δ), (b) H^α secondary chemical shifts ($\Delta\delta$)

Fig. 7. Free energy surface (in kcal/mol) of hIAPP(11–25) dimer at 310 K as a function of radius of gyration (R_g) and backbone RMSD. The reference structure for RMSD calculation is structure E given in Fig. 5 of [10]. The free energy surface contains five minimum energy basins, centered at (R_g , RMSD) values of (0.88 nm, 1.06 nm), (0.91 nm, 0.96 nm), (0.88 nm, 0.9 nm), (0.99 nm, 0.64 nm), and (1.12 nm, 0.12 nm). Representative structures in each minimum-energy basin along with their probabilities: 14.98% (A), 6.24% (B), 6.24% (C), 14.41% (D), and 2.25% (E)