Today's lecture: methods for facilitating equilibration and sampling in complex, frustrated, or slow-evolving systems

Difficult-to-simulate systems

Practically speaking, one is always limited in the length and time scales accessible from simulation. The former relates to the maximum system size that can be used (in terms of number of atoms), and the latter to the length of the run that can be performed (in terms of number of MD integration or MC steps).

If the intrinsic, physical time scale for some system of interest is beyond that accessible from simulation, we cannot investigate it using straightforward dynamical methods like molecular dynamics because it will be challenging to equilibrate the system. However, if we are only interested in thermodynamic and not kinetic properties, then we can construct artificial dynamics in the system that accelerate convergence towards equilibrium. Such methods inevitably involve a Monte Carlo component.

What makes a system **difficult to sample** in an equilibrium sense? Typically, these systems have **frustrated** or **rugged energy landscapes**. This means that the underlying potential energy surface contains many deep minima separated by high energy barriers. At low temperatures, it is challenging for the system to move between the relevant low-energy regions because it must surmount these barriers—a rare event process that takes time.

A wide range of methods have been designed to accelerate equilibration in simulations of systems with rugged energy landscapes. Here, we will only review a small selection of these. Some kinds of systems that might fall into this category include:

- systems with strong electrostatic interactions
- fluids of strongly orientation-dependent interactions (e.g., dipoles, hydrogen bonds)
- macromolecular and polymeric systems
- biomolecular systems (proteins, lipids)
- self-organizing or self-structuring systems (micelles, bilayers)
- supercooled liquid and amorphous systems at very low temperature

© M. S. Shell 2009

Simulated tempering

The **simulated tempering** algorithm is equivalent to a canonical Monte Carlo simulation in which the temperature changes randomly. That is, there are two kinds of moves:

- **energy fluctuation moves** – single-particle displacements, orientational displacements, etc. The usual canonical acceptance criteria, at the current temperature, are used.
- **temperature fluctuation moves** – the temperature is periodically increased or decreased by fixed amounts so that the temperature can acquire one of a fixed set of values $T_{\min}, T_{\min} + \Delta T, T_{\min} + 2\Delta T, \ldots, T_{\max}$.

These temperature fluctuations help the system equilibrate by allowing it to fluctuate to higher temperatures where it can more readily cross energy barriers.

In these simulations, the microstate is now a function of both the configurations and the temperature:

$$\wp(\mathbf{r}^N, T) \propto e^{-\beta U(\mathbf{r}^N)}$$

What is the acceptance criterion for a temperature change? We want all temperatures to appear with equal probability. That is, if we measured the distribution of temperature over the simulation run, we would want it to be roughly uniform. To do this, we introduce a weighting function in temperature:

$$\wp(\mathbf{r}^N, T) \propto e^{-\beta U(\mathbf{r}^N)+\eta(T)}$$

This means that the acceptance criterion for a random temperature perturbation $T_2 = T_1 + \Delta T$ is given by

$$\begin{aligned} P_{12}^{\text{acc}} &= \min\left[1, e^{-(\beta_2-\beta_1)U+\eta(T_2)-\eta(T_1)}\right] \\ &= \min\left[1, e^{-U\Delta\beta+\Delta\eta}\right] \end{aligned}$$

The distribution of temperatures is found by integrating this joint probability over configurations:

$$\begin{aligned} \wp(T) &\propto \int \wp(\mathbf{r}^N, T) d\mathbf{r}^N \\ &= Z(T, V, N) e^{\eta(T)} \\ &= e^{\eta(T)-\beta A(T)} \end{aligned}$$

To achieve a uniform distribution in temperature, we demand that this expression is constant. This means that we should choose:

$$\eta(T) = \beta A(T)$$

We do not know the free energies at the different temperatures a priori. Instead, we must iteratively determine them over the course of a simulation. Many methods can be used to do this, including:

- **flat histogram methods** –the flat histogram coordinate is the temperature
- **multiple histogram-reweighting** – we maintain a histogram of energies at each temperature and periodically use the reweighting equations to determine the relative free energies for each temperature
- **free energy perturbation** – we use, for example, Bennett's method between adjacent temperatures to compute free energy differences at periodic intervals in the simulation

How do we compute properties at a temperature of interest from this approach? We can use the WHAM-based reweighting equation applied to all temperatures and all trajectories. We first compute a configurational weight according to

$$w_i = \frac{e^{-\beta U_i}}{\sum_{j=1}^{J} e^{\beta_j A_j - \beta_j U_i}}$$

Here, i is an index over simulation observations. The inverse temperature β corresponds to an *arbitrary* reweighting temperature T. One should choose $T_{\min} \leq T \leq T_{\max}$ for good statistical accuracy.

For an arbitrary observable X, we can then compute averages and distributions using

$$\langle X \rangle = \frac{\sum_i w_i X_i}{\sum_i w_i}$$
$$\wp(X) \propto \sum_i w_i \delta_{X_i, X}$$

Replica exchange

The simulated tempering approach is useful for facilitating equilibrium because it allows the system to explore multiple temperatures. It has two drawbacks, however:

- We need to compute the free energies $A(T_j)$ at every temperature in order to properly sample all temperatures.
- We need to wait for the system to traverse the entire temperature range many times in order to accumulate good statistics at each. This can be a very long time if we have many temperatures.

An alternative approach, which has become the method of choice for sampling challenging systems is the **replica exchange** method. It overcomes these limitations by having the following general construction:

- J simulations (**"replicas"**) of the same system are performed simultaneously at different temperatures T_j.
- Each simulation is evolved independently, either through MD or MC methods, at the corresponding temperature.
- At set intervals, **replica swap** moves are performed between adjacent temperature replicas. In a swap move, the instantaneous configurations are exchanged between the two temperatures.

Swap move acceptance criterion

The replica exchange simulation performs a Markov chain in the entire J-system ensemble. We need to determine the acceptance criterion for performing swap moves. To do that, we need to determine the total probability of one microstate in the entire ensemble. Here, a microstate is the list of all of the positions in each of the replicas, $\mathbf{R} = \left(\mathbf{r}_1^N, \mathbf{r}_2^N, \ldots, \mathbf{r}_J^N\right)$. Since the replicas do not interact (there are no energy terms between atoms in different replicas), we can write

$$\wp(\mathbf{R}) = \prod_j \wp_j\left(\mathbf{r}_j^N\right)$$

Using canonical probabilities in each replica,

$$\wp(\mathbf{R}) = \prod_j \frac{e^{-\beta_j U\left(\mathbf{r}_j^N\right)}}{Z_j}$$

Consider a swap move between two temperatures 1 and 2. Initially the configuration in temperature 1 is $\mathbf{r}_1^N$ and in temperature 2 is $\mathbf{r}_2^N$. The acceptance criterion for the move stems from the detailed balance equation,

$$P_{12,\text{swap}}^{\text{acc}} = \min\left[1, \frac{\wp(\mathbf{R}_2)}{\wp(\mathbf{R}_1)}\right]$$

Here, $\mathbf{R}_2$ corresponds to the set of configurations in which $\mathbf{r}_1^N$ and $\mathbf{r}_2^N$ are transposed to different temperatures. We have:

$$\wp(\mathbf{R}_1) = \frac{e^{-\beta_1 U(\mathbf{r}_1^N)}}{Z_1} \times \frac{e^{-\beta_2 U(\mathbf{r}_2^N)}}{Z_2} \times \prod_{j=3}^{J} \frac{e^{-\beta_j U(\mathbf{r}_j^N)}}{Z_j}$$

$$\wp(\mathbf{R}_2) = \frac{e^{-\beta_1 U(\mathbf{r}_2^N)}}{Z_1} \times \frac{e^{-\beta_2 U(\mathbf{r}_1^N)}}{Z_2} \times \prod_{j=3}^{J} \frac{e^{-\beta_j U(\mathbf{r}_j^N)}}{Z_j}$$

Plugging these probabilities into the acceptance criterion,

$$P_{12,\text{swap}}^{\text{acc}} = \min\left[1, e^{\beta_1 U(\mathbf{r}_1^N) + \beta_2 U(\mathbf{r}_2^N) - \beta_1 U(\mathbf{r}_2^N) - \beta_2 U(\mathbf{r}_1^N)}\right]$$

Notice that the partition functions in the denominator cancel. This means that we do not need to know the free energies in each temperature when evaluating the acceptance criterion. It is because we swap systems at two temperatures, rather than perturb a single system's temperature, that we no longer need to know the free energy (as compared to simulated tempering).

Simplifying the above expression,

$$\boxed{P_{12,\text{swap}}^{\text{acc}} = \min\left[1, e^{\Delta\beta\Delta U}\right] \quad \text{where } \Delta\beta = \beta_2 - \beta_1\,, \Delta U = U(\mathbf{r}_2^N) - U(\mathbf{r}_1^N)}$$

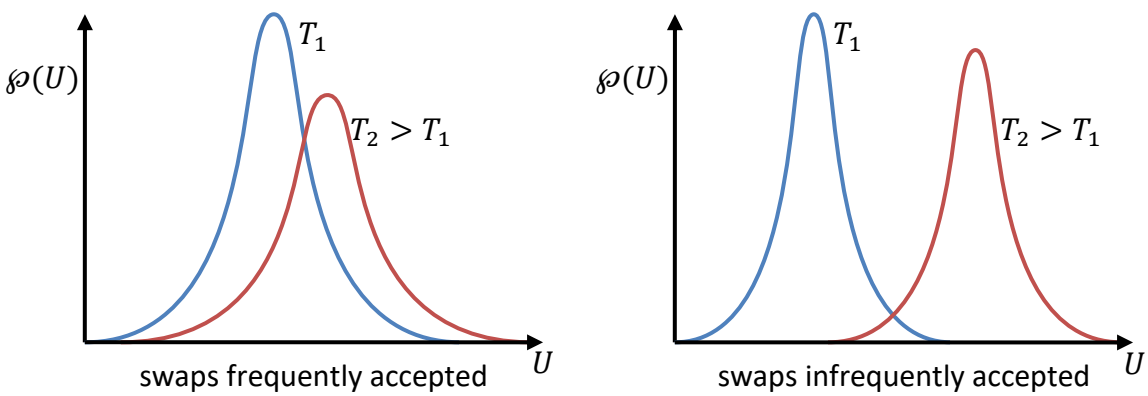
Notice that we are required to compute the potential energy difference between the instantaneous configurations in the two temperature replicas.

Swap moves are performed like any other MC moves: the move is proposed, the acceptance criterion is computed, a random number is drawn, and it is decided whether or not to perform the move. Unlike many other MC moves, swap moves are very inexpensive to perform: they only require the current energies in each temperature, which is typically maintained throughout the simulation anyways.

Considerations for acceptance ratios and the temperature schedule

Consider a swap between two temperatures $T_2 > T_1$. We expect that the configuration drawn from the higher temperature will have a higher energy. That is, $U(\mathbf{r}_2^N) > U(\mathbf{r}_1^N)$. Thus the quantity $\Delta\beta\Delta U$ will most often be negative, which will make the acceptance probability small. For very large temperature differences, this quantity becomes even more negative. The result is that the temperatures must be spaced close enough together in order to achieve a good rate of accepted swaps.

We can think about this problem in terms of the distributions $\wp_1(U), \wp_2(U)$ between neighboring temperatures. If there is a substantial overlap, we will have a high frequency of swap move acceptance:



These considerations play into the way in which we pick the temperatures in our replica exchange simulation?:

- $T_{\min}$ - typically we pick a minimum temperature to be the temperature of interest that is difficult to simulate
- $T_{\max}$ - one chooses a high temperature where free energy barriers can be crossed, but not so high as to require many intermediate temperatures
- J - the number of temperatures is usually chosen so as to achieve ~50% acceptance of swap moves between adjacent replicas
- T_j - for a system with a constant heat capacity, it can be shown that a constant rate of acceptance between adjacent temperatures corresponds to an exponential distribution in temperature, equivalently, a power law in replica number. Thus, one normally picks

$$T_j = T_{\min}\left(\frac{T_{\max}}{T_{\min}}\right)^{\frac{j}{J-1}}$$

Scaling with system size

As the system size grows, the distribution in energy at a given temperature becomes increasingly narrow with respect to the average energy, as $1/\sqrt{N}$. Without overlap between the energy fluctuations of adjacent temperatures in the replica exchange scheme, few swap moves will be accepted. Thus, as the system size increases, more and more intermediate temperature replicas are needed to achieve a 50% acceptance ratio. This makes the method challenging to apply to very large systems.

Evaluating properties and distributions

Rigorously, all of the configurations at each temperature (regardless of whether they swapped in from other temperatures) converge to a canonical distribution at that temperature. Thus, we

could compute the averages of properties at each temperature T_j by simply averaging over the configurations at that temperature. A **trajectory** is therefore considered the evolution of configurations *at a given temperature*.

However, we can also compute an average at any *arbitrary* temperature between the minimum and maximum in our simulation by using a reweighting approach. When a replica exchange simulation is performed, we can collect histograms or histories of each potential energy visited in each temperature, U_{ij}. We can then use the reweighting equations to compute the free energies and configurational weights at each temperature. With the computed configurational weights, we can express the average of any property as a weighted sum over all temperature trajectories:

$$w_{ij} = \frac{e^{-\beta U_{ij}}}{\sum_{l=1}^{J} e^{\beta_l A_l - \beta_l U_{ij}}}$$

$$\langle X \rangle = \frac{\sum_{j=1}^{J} \sum_{i=1}^{n} w_{ij} X_{ij}}{\sum_{j=1}^{J} \sum_{i=1}^{n} w_{ij}}$$

$$\wp(X) \propto \sum_{j=1}^{J} \sum_{i=1}^{n} w_{ij} \delta_{X_{ij}=X}$$

Implementation

To maintain the different replicas, one typically uses a **parallel computing scheme**, in which one computer node (or processor) is assigned to each replica. When swaps are performed, all nodes stop evolving the system in time and a master or head node sorts through and makes the swap moves. This involves parallel communication between the different nodes and the head node.

At each interval where swaps are performed, any number of swaps could be attempted. Typically, one chooses the number of swap attempts to be the number of temperature replicas. The swaps can be attempted in serial order, from lowest/highest to highest/lowest temperature, or in random order.

In a MC simulation, the evolution of the system in between swaps can be governed by any number of MC moves that accomplish changes in the potential energy, such as single particle displacements.

In a MD simulation, the evolution in between swaps can be performed using short MD trajectories. These moves can be considered hybrid MC/MD moves with good energy conservation such that they are always accepted. After a round of swap moves, one picks random velocities for each atom at every temperature j from a Boltzmann distribution at the corresponding T_j. These

velocities are used to start the short MD trajectory before another round of swap moves is performed.

An alternative approach in MD would be to **rescale the velocities** in the configurations to the new temperature after a swap move. This approach was first derived by Sugita and Okamoto [1999]. Here, random new velocities are *not* picked at any time. Instead, after swapping configurations, the momenta are scaled by a factor $\sqrt{T_{\text{new}}/T_{\text{old}}}$, where T_{new} is the new temperature into which the configuration was swapped and T_{old} is the temperature from which it came. This approach requires a thermostat to be used during the MD trajectories. While the velocity rescaling approach is more frequently used in the literature than the hybrid approach (with random resampling), it is not necessarily more efficient and either approach is valid.

Variants

Here, we considered a replica exchange simulation where the replicas differed in temperature. This is often also called **parallel tempering** for its similarity to simulated tempering. However, we are not limited to temperature. One can have replicas differ in chemical potential or pressure if the individual simulations are GCMC or NPT simualtions, respectively. **Two-dimensional replica exchange** methods allow replicas to differ in *both* temperature and chemical potential or pressure.

Another way to facilitate sampling is to modify the potential energy function itself. Here, one wants to perturb the energy function so that the underlying energy landscape is smoother and easier to sample. For example, one systematically scale the partial charges from zero to their full values as one moves from replica to replica. This approach is often called **Hamiltonian exchange.** The appropriate form of the acceptance criterion is derived as before, but with different potential energy functions in each simulation U_j:

$$\boxed{P_{12,\text{swap}}^{\text{acc}} = \min\left[1, e^{\beta_1 U_1(\mathbf{r}_1^N)+\beta_2 U_2(\mathbf{r}_2^N)-\beta_1 U_1(\mathbf{r}_2^N)-\beta_2 U_2(\mathbf{r}_1^N)}\right]}$$

In evaluating this acceptance criterion, one must evaluate the cross-energies $U_2(\mathbf{r}_1^N)$ and $U_1(\mathbf{r}_2^N)$ between two replicas with each swap move, since the energy function is no longer constant between them.

In all of these cases, we can use multiple histogram reweighting techniques to express the average or distribution of any property as a weighted average over all temperatures j and trajectory configurations i.

Extended ensemble molecular dynamics

One way to enhance the exploration of phase space in MD at lower temperatures is bias the ensemble probabilities to artificially broad fluctuations in potential energy. This enables the system to hop over higher-energy barriers more frequently. Ultimately, the behavior of the system under normal conditions can be recovered by standard reweighting procedures. We have already discussed this approach in the context of MC simulations. Now, we consider an MD implementation.

We add to our MD simulation a weighting function that biases the configurational probabilities, in the same way that we did during our discussion of biased sampling:

$$U^{\mathrm{w}}(\mathbf{r}^N) = U(\mathbf{r}^N) - k_B T \eta\big(U(\mathbf{r}^N)\big)$$

Here, $\eta(U)$ is a weighting function with dimensionless units, and it only depends on the potential energy. The equations of motion are derived as before. For an atom i:

$$\begin{aligned}
\mathbf{f}_i^{\mathrm{w}} &= -\frac{dU^{\mathrm{w}}}{d\mathbf{r}_i} \\
&= -\frac{dU}{d\mathbf{r}_i} + k_B T \frac{d\eta(U)}{dU}\frac{dU}{d\mathbf{r}_i} \\
&= \mathbf{f}_i \left(1 - k_B T \frac{d\eta(U)}{dU}\right)
\end{aligned}$$

In other words, the force on each atom in the weighted ensemble is scaled by a term involving the derivative of the weighting function *evaluated* at the energy of the current configuration. If the weighting function is a constant, no scaling occurs.

In order for this approach to work, the weighting function must be a **continuous** function. One typically uses splines or other mathematical constructs so that an analytical derivative can be computed.

How do we determine the weighting function? We can fit it to estimates of the probability distribution function. In the canonical ensemble, we have the relationship

$$\frac{\wp(U)}{\wp^{\mathrm{w}}(U)} \propto e^{-\eta(U)}$$

If we want the weighted distribution to be approximately flat, we can choose

$$\eta(U) = -\ln \wp(U) + \mathrm{const}$$

Notice that we will have to fit the weighting function to discrete histogram calculation, since η must be continuous. It is for this reason that it becomes difficult to evaluate η to high statistical

accuracy using MD techniques. Still, broadening the energy distribution even a little bit may speed equilibration in MD simulations.