

Chem 860. Lecture 14, 15

Monte Carlo: Parallel tempering; Landau-Wang (DOS)

March 11, 2009

1 Sampling with advanced Monte Carlo methods

Previous discussions have emphasized that sampling is a major issue in molecular simulations. With molecular dynamics, which follows the **natural** dynamical evolution of systems, it is difficult (though not impossible - e.g., umbrella sampling if you know what to bias, or transformed potential that is smoother based on Tsallis statistics) to surmount barriers and sample different regions efficiently. With Monte Carlo, by contrast, one can develop innovative techniques because the process is stochastic. All we have to make sure of is that when we analyze results, we are using the correct probability distribution.

Many versions of MC have been developed over the years. We mainly discuss two approaches: (i). Parallel tempering or replica exchange, which combines samplings at different temperatures as well as parallel computing technology. (ii). Density of state Monte Carlo (or known as “Landau-Wang approach”, which samples energy space with uniform distribution).

1.1 Parallel tempering/Replica Exchange (RP)

The RP approach takes advantage of simulations with multiple temperatures (in principle can be multiple Hamiltonians as well). We construct an extended ensemble that contains multiple copies of the system, each copy is simulated at a different temperature. In addition, we occasionally swap systems; e.g., swap a high- T with a low- T system. Since the high- T simulation can sample the configuration space very efficiently, the swap helps the low- T simulation to explore the landscape much better than an isolated low- T simulation. To make the swap get accepted with reasonable probability, we can choose a range of temperatures.

The only thing we have to figure out is the acceptance criterion, i.e., *acc* function. The detailed balance condition for a pair of states is, assuming that α is taken as symmetric,

$$\begin{aligned} & \mathcal{N}(\mathbf{i}, \beta_i) \mathcal{N}(\mathbf{j}, \beta_j) \times \alpha \times acc[(\mathbf{i}, \beta_i; \mathbf{j}, \beta_j) \rightarrow (\mathbf{i}, \beta_j; \mathbf{j}, \beta_i)] \\ &= \mathcal{N}(\mathbf{i}, \beta_j) \mathcal{N}(\mathbf{j}, \beta_i) \times \alpha \times acc[(\mathbf{i}, \beta_j; \mathbf{j}, \beta_i) \rightarrow (\mathbf{i}, \beta_i; \mathbf{j}, \beta_j)] \end{aligned}$$

Simple rearrangements give then,

$$\begin{aligned} \frac{acc[(\mathbf{i}, \beta_i; \mathbf{j}, \beta_j) \rightarrow (\mathbf{i}, \beta_j; \mathbf{j}, \beta_i)]}{acc[(\mathbf{i}, \beta_j; \mathbf{j}, \beta_i) \rightarrow (\mathbf{i}, \beta_i; \mathbf{j}, \beta_j)]} &= \frac{\mathcal{N}(\mathbf{i}, \beta_j)\mathcal{N}(\mathbf{j}, \beta_i)}{\mathcal{N}(\mathbf{i}, \beta_i)\mathcal{N}(\mathbf{j}, \beta_j)} \\ &= \exp[-(\beta_i - \beta_j)(U(\mathbf{i}) - U(\mathbf{j}))] = \exp[-\Delta\beta_{ij}\Delta U_{ij}] \end{aligned}$$

Therefore, we can choose the *acc* function as,

$$acc[(\mathbf{i}, \beta_i; \mathbf{j}, \beta_j) \rightarrow (\mathbf{i}, \beta_j; \mathbf{j}, \beta_i)] = \min\{1, \exp[-\Delta\beta_{ij}\Delta U_{ij}]\}$$

Few points to note:

- Swap is computationally inexpensive (no need to recompute any energy).
- Swap does not perturb Boltzmann distribution - so simple to accumulate result.
- Some technical issues: # of replica ($\sqrt{N_{dof}}$) and temperature distribution (see, e.g., JCP,116, 9058, 2002); frequency of exchange (see JCP, 128, 024103, 2008)
- Swap can occur in more than temperature - can be chemical potential, polymer length, form of Hamiltonian etc.
- There is a serial version known as “simulated tempering”, which takes a random walk in T space with a single simulation.
- As expected, any dynamics info is lost.

1.2 Density of state Monte Carlo

In canonical MC, energy barriers high compared to $k_B T$ are difficult to overcome. A powerful alternative is to actually perform sampling that has uniform distribution in energy levels (instead of exponential as in Boltzmann!)- in such a way, we can sample a wide range of energies and therefore overcome barrier easily. If we can estimate the density of state, $\Omega(E)$ - which measures the degeneracy (number of microstates) of energy levels, then we can calculate any statistical quantities. After all, $\Omega(E)$ and the canonical partition function is related by a Laplacian transformation,

$$Q(N, V, T) = \int dE \Omega(E) \exp(-\beta E)$$

The canonical distribution is simply

$$P(E) = \Omega(E) \exp(-\beta E) / Q$$

which can be used to compute any ensemble average of observable. We note that it is difficult to use traditional canonical MC to estimate $\Omega(E)$ through sampling $P(E)$, because the sampling is poor for high energy and regions separated by high barriers.

The Landau-Wang algorithm (*Phys. Rev. Lett.*, **86**, 2050, 2001) is a method that converges to $\Omega(E)$ within a given tolerance. Imagine that we have a set of energy levels, $\{E_i\}$, and we will perform a random walk in energy space such that the histogram $H(E)$ is uniform

or “flat”. Such a random walk can be constructed in the similar way as the canonical MC discussed earlier, i.e., make a move according to $\alpha(o \rightarrow n)$, then accept the move with a given rule specified by $acc(o \rightarrow n)$. If we choose α to be symmetric, then the acc function takes the following form to satisfy detailed balance,

$$acc(o \rightarrow n) = \min(1, \frac{\mathcal{N}(n)}{\mathcal{N}(o)})$$

where the distribution \mathcal{N} in our case is inversely proportional to the density of state (because we want to have **uniform** distribution in *energy levels*); i.e.,

$$acc(o \rightarrow n) = \min(1, \frac{\Omega(o)}{\Omega(n)})$$

The only problem is that we don’t know $\Omega(E)$ to start with (if we can indeed generate a flat histogram, then we have the correct $\Omega(E)$)! So we have to iterate, which is an essential contribution from Landau-Wang.

We initialize $H(E) = 0$ and $\Omega(E) = 1$. Then we perform MC moves according to (uniform) random displacements and the acceptance rule established above (based on the current guess of $\Omega(E)$). Throughout the simulation, we update both quantities based on the energy level (E_i) after the move (whether accepted or rejected!):

$$H(E_i) = H(E_i) + 1$$

$$\Omega(E_i) = f\Omega(E_i)$$

where f is taken as a large number to start with (e.g., e^1).

The update is continued until the histogram $H(E)$ is reasonably “flat”. At this point, $\Omega(E)$ has converged with error proportional to $\ln f$. To improve the result, we move on to the next iteration; we reset histogram to be zero, and reset f as

$$f \rightarrow \sqrt{f}$$

The iteration stops until f is reasonably close to be 1 (e.g., $\ln f < 10^{-8}$).

In practice, there are issues like

- How flat is flat for $H(E)$? (A minimum number of hits for each bin)
- Range of energy (E_{min}, E_{max})
- Unique functional form for f update?
- Accumulation of $\Omega(E)$ (relative value, which is O.K. for most purposes)

We emphasize that Landau-Wang does not rigorously satisfy detailed balance - because acceptance criterion is changing during the iteration. But in reality, the deviation is very small especially if f is close to 1.

For Landau-Wang application to proteins, see some recent work of de Pablo and N. Rathore. Also, Landau-Wang and MD combination: Phys. Rev. E 75, 066706 (2007)