

分子建模与模拟导论

王延颢

2012年11月10日

8. 分子模拟基本概念

全原子分子模拟是把原子看作质点,在有限温度下让原子在原子间相互作用力作用下进行运动,从而对实际体系进行计算机上的模拟。因为原子内的核子和电子自由度都被质点的自由度代替,分子模拟相对于第一性计算而言大大简化了计算量,更着重于有限温度下体系的性质,可以观察到物质微观快速运动的细节,是联系理论与实验的桥梁。

8.1. 统计物理基本假设

统计物理是研究多体体系的集合行为的理论。目前平衡态统计物理理论比较成熟。对于一个给定的热力学系统, **势能面** (potential energy surface) 是指由不同构型形成的势能的集合, **系综** (ensemble) 是指系统在给定宏观条件下所有状态的集合。统计物理的两个基本假设是等几率原理与各态历经。

等几率原理 (principle of equal weights): 一个热力学体系有相同的几率访问具有相同能量的每一个微观态 (注意: 不是能量的等几率! 一个能量一般会对应很多微观态)。由等几率原理推导得出在温度恒定的系综下, 系统处在第 j 个状态的几率服从 Boltzmann 分布

$$P_j = \exp(-\beta E_j) / Z \quad (8.1)$$

其中配分函数 (partition function)

$$Z = \sum_i \exp(-\beta E_i) \quad \beta = \frac{1}{k_B T} \quad (8.2)$$

而系统处于能量为 E 的几率为

$$P(E) = g(E) \exp(-\beta E) / Z \quad (8.3)$$

其中 $g(E)$ 为能量为 E 的状态数。

各态历经 (ergodicity): 只要系统演化无穷长时间, 总有几率历经势能面上的所有点。即在极限情况下, 系综平均和时间平均是等价的。

系综平均: 蒙特卡罗模拟 (Monte Carlo, MC) 基于对系统的系综进行重要性采样, 从

而一个热力学量 A 的宏观平均值（对应于实验可观测量）为

$$\langle A \rangle = \sum_j A_j \exp(-\beta E_j) / Z \quad (8.4)$$

时间平均：分子动力学模拟（Molecular Dynamics, MD）基于对时间演化的模拟，从而宏观平均值

$$\bar{A} = \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t A(t') dt' \approx \frac{1}{M} \sum_{i=1}^M A(t_i) \quad (8.5)$$

分子模拟中常用的热力学系综有：

- 1) 微正则系综 (Microcanonical Ensemble): NVE 皆为常数。
- 2) 正则系综 (Canonical Ensemble): NVT 皆为常数。
- 3) 巨正则系综 (Grandcanonical Ensemble): μVT 皆为常数，粒子数不固定。
- 4) 等压-等温系综 (Isobaric-Isothermal Ensemble): NPT 皆为常数。
- 5) 等张力-等温系综 (Isotension-Isothermal Ensemble): 模拟盒子的形状可变。

其中 N 为粒子数， V 为体积， T 为温度， P 为压强， μ 为化学势（即单个粒子的平均自由能）。

8.2. 热力学量

一个 NVT 系综下的体系总的内能有两种划分方式：1) 分为动能和势能两部分

$$E = E_k + E_p \quad (8.6)$$

动能 E_k 反映粒子的微观运动具备的能量，势能 E_p 反映粒子间的相互作用。2) 分为自由能和熵对应的能量两部分。有限温度下，体系的总内能不可能都用来对外做功。因为温度效应导致的体系内部的混乱对应的能量是无法用于对外做功的，刻画体系混乱度的热力学量定义为熵 S ，其对应的能量为 TS 。而体系可以用来对外做功的那部分能量定义为自由能

$$F = E - TS \quad (8.7)$$

分子模拟中常用的热力学量有：

- 1) **动能：**每个粒子的动能的叠加

$$E_k = \left\langle \sum_{i=1}^N \frac{1}{2} m_i \mathbf{v}_i^2 \right\rangle \quad (8.8)$$

2) **温度**: 系统动能的平均

$$T = \frac{1}{dNk_B} \left\langle \sum_{i=1}^N m_i \mathbf{v}_i^2 \right\rangle \quad (8.9)$$

其中 d 是空间维数。

3) **势能**: 每个粒子上势能的叠加

$$E_p = \left\langle \sum_{i=1}^N E_{pi} \right\rangle \quad (8.10)$$

4) **压强**: 由维里力计算得到

$$p = \frac{k_B T N}{V} + \frac{1}{dV} \left\langle \sum_{i < j} \mathbf{f}_{ij} \cdot \mathbf{r}_{ij} \right\rangle \quad (8.11)$$

5) **焓**: 可以理解为 NPT 系综下系统的有效总内能

$$H = E + pV \quad (8.12)$$

6) **熵**:

$$S = k_B \ln \Omega(N, V, E) \quad (8.13)$$

其中 Ω 是系统的总微观状态数。

7) **Helmholtz 自由能**:

$$F = E - TS = -k_B T \ln Z \quad (8.14)$$

8) **Gibbs 自由能**: NPT 系综下系统的自由能

$$G = F + pV = E - TS + pV \quad (8.15)$$

9) **化学势**: 平均每个粒子的自由能

$$\mu = \left. \frac{\partial G}{\partial N} \right|_{T,p} = \left. \frac{\partial F}{\partial N} \right|_{T,V} \quad (8.16)$$

8.3. 模拟与采样

8.3.1. 短程作用与长程作用

三维空间中, 积分

$$\int_{r_c}^{\infty} V(r) 4\pi r^2 dr \quad (8.17)$$

收敛的称为短程作用，发散的称为长程作用。对于短程作用可以作截断，即认为 $r > r_c$ 处的 $V(r)$ 都为零。LJ 势是短程作用，而静电势是长程作用。

8.3.2. 约化单位

数值模拟时使用的内部单位，使得数值与 1 不会有太大的数量级的差别，从而保证计算精度尽量少地因为计算机的截断误差而损失，同时使得计算结果可以适用于相同模型的一类问题，不依赖于实际的数值。约化单位的计算结果需要乘上常数才能对应于实际体系的真实物理单位（国际单位制，SI units）。

只要给定四个基本物理量的单位换算：长度 L ，质量 M ，时间 t ，电荷电量 Q ，就可以换算得到所有其它物理量的单位：

1) 能量 $E = M * L^2 / t^2$

2) 温度 $T = E / k_B$

3) 压力 $P = E / L^3$

4) 质量密度 $\rho = M / L^3$

5) 数量密度 $n = 1 / L^3$

6) 介电常数 $\epsilon = \frac{N_A \cdot Q^2}{L \cdot E}$

其中 k_B 是 Boltzmann 常数， N_A 是 Avogadro 常数。

8.3.3. 模拟与采样

1) **空间的连续性**：离散模型，如伊辛（Ising）模型，和连续模型。

2) **边界条件**：周期性边界条件（Periodic Boundary Condition, PBC），自由（free）边界条件，刚性（hard）边界条件。

周期性边界条件：模拟的盒子中的粒子与无穷多的镜像中的粒子有相互作用，从而可以用 $\sim 10^3$ - 10^6 个粒子模拟 $\sim 10^{23}$ 个粒子的体系。体系的总能量表示为

$$E_{\text{tot}} = \frac{1}{2} \sum'_{i,j,\vec{n}} E(\vec{r}_{ij} + \vec{n}L) \quad (8.18)$$

其中 L 是立方体模拟盒子的边长, \vec{n} 是三维整数矢量, Σ' 表示当 $i = j$ 时排除 $\vec{n} = 0$ 项相加。

运动到模拟区域外的粒子通过周期性边界条件从另一端放入模拟区域。模拟时 L 的取值必须大于最长的作用力截断距离的两倍, 从而模拟区域内的粒子只与邻近镜像内的粒子有短程相互作用。长程相互作用必须用 Ewald Summation 的方法计算。

3) 特征长度: 某一特定物理量在空间的相关性的长度。原则上, 模拟盒子的边长应该大于所关心的物理量的特征长度。具体操作上, 可以通过变化模拟尺寸来了解有限尺度效应 (finite size effect) 的影响。特征长度可以通过计算空间关联函数估算得到

$$c(\vec{r}) = \langle \delta A(\vec{r}_0 + \vec{r}) \delta A(\vec{r}_0) \rangle \quad (8.19)$$

其中 $\delta A(\vec{r}) = A(\vec{r}) - \langle A(\vec{r}) \rangle$ 。例如粒子运动关联性的特征长度在固态和液态都很短, 而在熔化相变点附近变长, 在相变点处发散。

原则上, 为了使周期性边界条件所引入的有限尺寸效应 (finite size effect) 尽量小, 模拟盒子的边长应该大于所关心的物理量的特征尺度。实际应用时, 可以通过变化模拟尺寸来定性理解有限尺寸效应的大小, 甚至可以定量计算并外推得到热力学极限下的性质。

4) 作用势的截断距离: 如前所述, 需要使得作用势的截断距离 $r_c < \frac{L}{2}$, 从而避免与同一粒子的两个镜像相互作用。截断作用势的处理方法有:

a) 简单截断:

$$V(r) = \begin{cases} V^{\text{real}}(r) & r \leq r_c \\ 0 & r > r_c \end{cases} \quad (8.20)$$

截断点处势能不连续。简单截断的问题有: 1) 势能的绝对值有偏差; 2) 做分子动力学模拟时断点处的力发散; 3) 不连续点对压强有一个“排斥”的贡献。

b) 截断平移:

$$V(r) = \begin{cases} V^{\text{real}}(r) - V^{\text{real}}(r_c) & r \leq r_c \\ 0 & r > r_c \end{cases} \quad (8.21)$$

截断点处势能连续, 为最常用的方法。

c) 最小镜像法：模拟盒子中的所有粒子的最邻近镜像的相互作用。不适用于 MD。

5) 采样：在有限时间内进行重要性采样 (importance sampling)。常规的分子模拟算法中一般采用均匀时间间隔进行采样。如果采样间隔太近，样本之间会有关联，从而影响统计误差的估计。原则上采样间隔应该足够长，从而使得样本不相关。

6) 初始构型：如果采样足够充分 (样本数趋于无穷大)，模拟得到的系统平衡态的性质应该是与初始构象无关的。实际模拟采样得到的样本数总是有限的，所以严格说来多少会相关。如果相关度大到不可忽略，有两种可能：

a) 系统不满足各态历经假设。如玻璃态，不同的势能面区域被无穷大势垒分割开。

b) 初始构型远离平衡态，即系统尚未平衡。更严格地说，是出现几率极小的初始构型及其邻近构型以很大的权重出现在时间平均中。所以实际模拟时，若不能确定初始构型为平衡态，需要先进行预平衡，确信系统平衡后再采样。需要某些参数来量化观察系统是否平衡 (如液体的体积很容易平衡，势能其次，扩散系数则较难)。

7) 采样样本的相关度：不影响平均值，影响误差范围。

a) 弛豫时间 (relaxation time)：随机运动的系统经过弛豫时间 τ_{relax} 后，“忘记”了其初始构型，即时间相关度趋于零。可以通过计算时间关联函数来确定 τ_{relax} ：

$$\begin{aligned} c(t) &= \langle \delta A(t_0 + t) \delta A(t_0) \rangle = \langle A(t_0 + t) A(t_0) \rangle - \langle A \rangle^2 \\ \delta A(t) &= A(t) - \langle A \rangle \end{aligned} \quad (8.22)$$

原则上采样间隔要大于 τ_{relax} 才真正在统计上是正确的。

b) 块平均 (block average)：把包含 N 个数据的数据集均匀分成 $M = \frac{N}{B}$ 份，每份包含 B 个数据，则平均值为：

$$\bar{A} = \frac{1}{N} \sum_{i=1}^N A_i = \frac{1}{M} \sum_{j=1}^M \bar{A}_j \quad (8.23)$$

方差为

$$S = \frac{\sum_{i=1}^M \left(\frac{1}{B} \sum_{j=1}^B A_{iB+j}^2 - \bar{A}^2 \right)}{(M-1)^{\frac{1}{2}}} \quad (8.24)$$