

## 基于元动力学的增强采样方法应用实例

对于大多数生物大分子体系例如蛋白，无论是经历构象变化、与配体结合，抑或是发生化学反应，通常我们所感兴趣的系统状态在自由能面上表现为低势能构象区域，而这些区域通常会被高自由能势垒( $\gg k_B T$ ) 隔开。

由于算力的限制，即使使用“长时间的”常规分子动力学模拟 (Molecular dynamics simulation, MD)，系统也很容易陷在局部能量极小的空间里无法逃逸出来，因此难以遍历最稳定、最关键的构象，更无法遍历整个自由能面。因此，为了克服时间尺度和采样不足的问题，增强采样通常是必不可少的<sup>1</sup>。

根据玻尔兹曼定律，样本  $R$  被访问的概率与玻尔兹曼因子  $\exp[-\beta F(R)]$  成正比 (其中  $\beta = 1/k_B T$  为温度因子， $k_B$  是玻尔兹曼常数， $T$  是温度)。可见，决定构象跃迁速率的基本量是  $\beta \Delta F^\ddagger$ ，即温度因子  $\beta$  乘以构象跃迁的自由能垒  $\Delta F^\ddagger$ 。

因此为了加速相空间采样，可以通过提高温度或降低能垒的方式实现。其中，元动力学方法 (Metadynamics) 作为一种功能强大的增强采样方法，是基于降低能垒的思路所提出来的方案<sup>2</sup>。该算法的一个重要的分支 Well-Tempered Metadynamics 已经广泛应用于加速分子动力学模拟中。这种偏置模拟能够非常有效地探索整个自由能面，比无偏 (即常规的) 分子动力学模拟快 4 个数量级甚至以上，采样效率非常高<sup>3,4</sup>。

本文介绍由德锐特搭建的分子动力学模拟计算平台 (SP MD)，结合元动力学增强采样方法在蛋白构象变化问题中进行了探究。与常规分子动力学模拟结果进行对比，直观呈现了元动力学在相空间的快速采样，及其收敛迅速的优势表现。

### 1. 元动力学基本思想

元动力学的基本思想是通过一系列激活-弛豫步骤来促使系统逃离所处的势阱。该方法对预先定义的集体变量 (Collective Variables, CVs) 重写运动方程，通过非马尔科夫动力学遍历自由能面<sup>5</sup>。通过计算 CVs 的概率分布，可以得到系统在该 CVs 下的自由能变化情况。元动力学能够在宏观时间尺度上，同时保留与慢速动力学相关的原子构型细节，通过对体系添加额外的历史依赖的偏置势，降低再次访问已经访问过的相空间区域的概率，从而将系统从局部最小值推向相空间的新区域<sup>6-8</sup>。

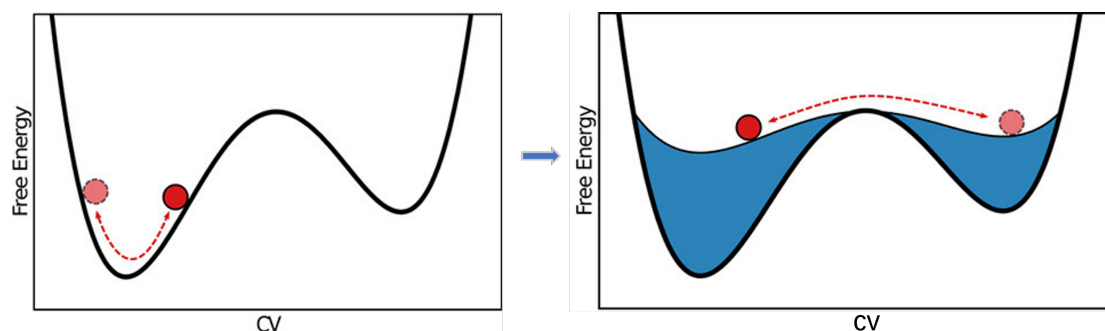


图 1. 元动力学示意图：自由能势阱逐渐被高斯型势能函数填满，从而降低能垒，促进不同亚稳态之间的跃迁<sup>9</sup>

## 2. 动力学增强采样实例：RfaH 蛋白 CTD 结构域的 $\alpha$ 和 $\beta$ 两种不同构象之间的转变

RfaH 蛋白是大肠杆菌的毒力调节因子，该蛋白质有两个结构域：C 端结构域 (CTD) 可以采用  $\alpha$ -螺旋或  $\beta$ -桶状两种折叠方式，并连接到 N 端结构域 (NTD)<sup>10</sup>。完整蛋白中的 CTD 以  $\alpha$  螺旋折叠通过与 N 末端结构域 (NTD) 的相互作用紧密结合，从而掩盖 RNA 聚合酶结合位点发挥自我抑制作用。当与 DNA 结合后，会触发结构域解离，使 NTD 与 RNA 聚合酶相互作用以促进转录，同时 CTD 折叠成  $\beta$  桶构象，与核糖体相互作用以激活翻译。而孤立状态的 CTD 会折叠成一个  $\beta$  桶形式。触发蛋白从  $\alpha$  向  $\beta$  转换需要破坏域间相互作用，两种状态都代表着同一序列可以折叠成的低自由能构象<sup>11</sup>。如果试图对蛋白的两种不同折叠的转换过程进行探究，需要对相空间进行充分采样，确定蛋白分子低能量稳定构型以及过渡态的构象分布。为了深入理解该转换机制，我们选取了 RfaH 孤立的 CTD 结构进行了分子动力学模拟，探索了 RfaH 蛋白 CTD 结构域的  $\alpha$  和  $\beta$  两种不同构象之间变化的动力学过程，计算了两种构象之间的自由能变化。

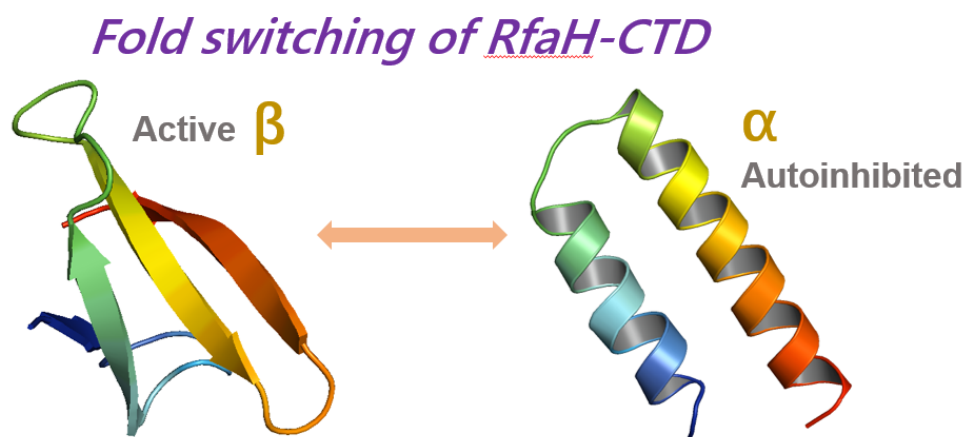


图 2. RfaH 蛋白中的 CTD 结构域折叠变化

### ➤ 常规分子动力学模拟

我们使用分子动力学模拟软件分别从  $\beta$  和  $\alpha$  构象模型出发进行了 500 ns 的常规分子动力学模拟计算。通过对体系在模拟期间的轨迹结构相对于初始结构的均方根偏差 (RMSD) 计算分析，得到了以下的结果：

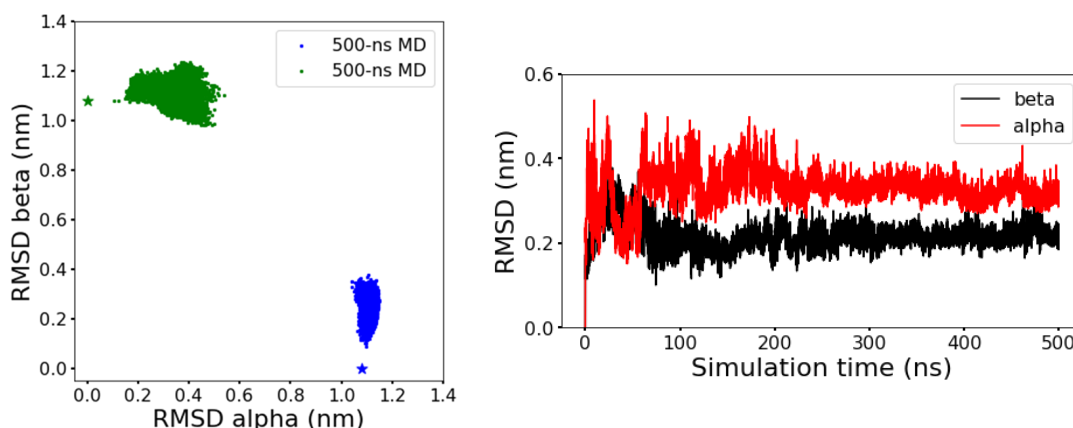


图 3. RfaH-CTD 不同构象下的 500 ns 常规分子动力学模拟的 RMSD 分析结果

其中，左图展示了从初始结构（蓝色和绿色五角星所示）出发，对相空间进行探索所能到达的范围。绿色部分为对  $\alpha$ -螺旋结构的探索区域，蓝色部分为对  $\beta$ -桶状结构的探索区域，可以看出分别从两种结构出发所探索的范围都非常有限，离初始点的距离最远不超过 0.6 nm。右图显示了从两种结构出发的模拟结构随着模拟时间的变化， $\alpha$ （红色）和  $\beta$ （黑色）RMSD 变化幅度都较小同时都能够趋于稳定。可以看出，在常规分子动力学模拟下，难以实现对该体系在自由能面上的全局性采样。

### ➤ 元动力学模拟

我们使用元动力学方法进行增强采样，选用相对于  $\alpha$  和  $\beta$  两种构象的 RMSD 作为系统的 CVs 进行模拟。我们将放置高斯函数的间隔步数设置为 500，初始高度设置为 1.2 kJ/mol，宽度设置为 0.05 nm，偏置因子设置为 60。总的模拟时间为 1  $\mu$ s。根据模拟采样结果，我们采取了两种方法来计算得到自由能面。

一种是通过 sum\_hills 对模拟产生的 HILLS 文件信息直接进行分析：

```
plumed sum_hills --hills ./HILLS --outfile fes.1000ns.dat -mintozero
```

得到自由能面：

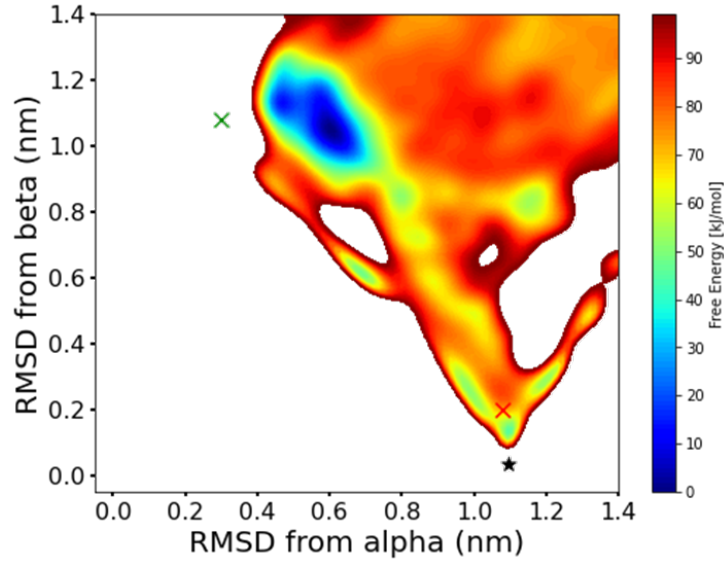


图 4. 通过 sum\_hills 方式重现 RfaH-CTD 域以  $\alpha$  和  $\beta$  折叠的 RMSD 为 CVs 的自由能面。图中红色  $\times$  标识和绿色  $\times$  标识分别为 CTD 域  $\beta$ -桶状结构和  $\alpha$ -螺旋结构在此自由能面上所处的相对位置；黑色五角星  $\star$  标识为模拟所选取的初始 CTD 结构在自由能面上所处的相对位置。

从图 4 可以看出，从初始  $\beta$  结构出发，其在模拟期间对不同结构方向的探索，同时可以看出  $\beta$  相对于  $\alpha$  的折叠方式的自由能值更低，佐证了实验所发现的孤立的 CTD 结构域以  $\beta$  折叠形式单一存在的结论。

另一种通过 reweighting 方法来进行分析：

*plumed driver --noatoms --plumed plumed\_reweight.dat*

得到相应的自由能面：

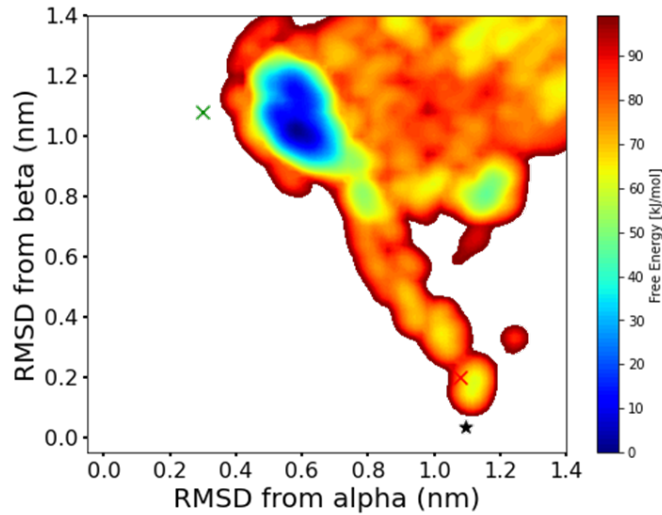


图 5. 通过 reweighting 方式重现 RfaH-CTD 以  $\alpha$  和  $\beta$  折叠的 RMSD 为 CVs 的自由能面。图中红色  $\times$  标识和绿色  $\times$  标识分别为 CTD 域  $\beta$ -桶状结构和  $\alpha$ -螺旋结构在此自由能面上所处的相对位置；黑色五角星  $\star$  标识为模拟所选取的初始 CTD 结构在自由能面上所处的相对位置。

图 5 中展现的自由能面同样表现了从初始  $\beta$  结构出发可以达到的采样区域较为广泛，模拟期间对不同构象所进行的不断采集， $\beta$  结构相对于  $\alpha$  折叠的自由能值更低的数据结果与图 4 中给出的数值结果一致。

上述两种方式得出的自由能面高度一致，很好地表明了该模拟具有较好的收敛性。而且模拟过程中所添加的高斯函数高度逐渐下降至趋近于 0（如图 6 所示），也进一步印证了这一点。

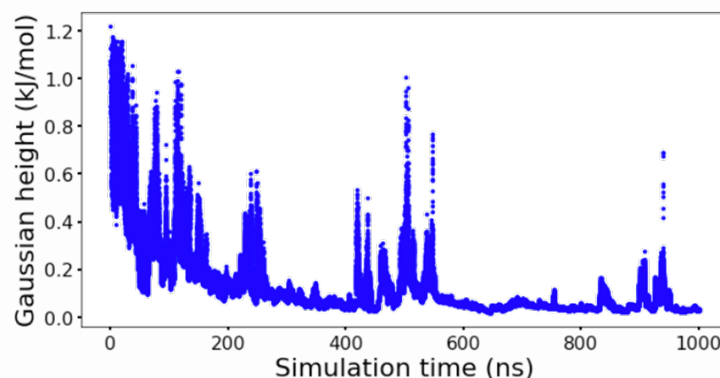


图 6. 随着计算模拟的演进，所添加的高斯函数其高度的变化情况

同时，从图 7 可以看出，随着模拟时间的演进，CTD 域在  $\alpha$ -螺旋和  $\beta$ -桶状两种结构之间反复变化。这也证明了在元动力学的偏置势推动下，该体系在整个 CVs 相空间能够较为充分地扩散；体系在 1 微秒的元动力学模拟不仅完全覆盖了常规模拟所能采样到的范围，并且探索到两种构象相较于初始结构更远的相空间区域（如图 8 所示），证明了 WT-MetaD 模拟在增强采样上的优异表现。

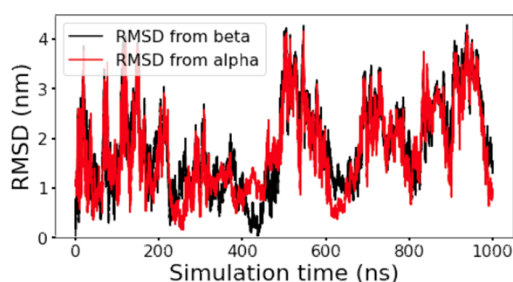


图 7. CTD 域在模拟过程中相对于  $\alpha$ -螺旋和  $\beta$ -桶状两种结构的 RMSD 变化。

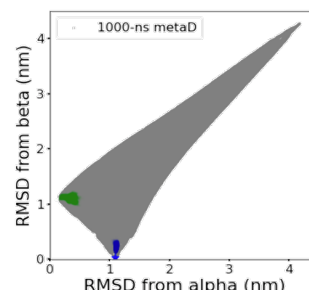


图 8. 元动力学模拟（灰色部分）相较于常规动力学模拟（蓝色和绿色部分）CTD 域的构象采样范围

与常规分子动力学模拟结果对比可以看出，元动力学增强采样方法在克服采样不足的难题方面有着不俗的表现。对体系所确立的相关自由度引入外部偏置势，能够在模拟过程中自适应地添加，能够达到减低甚至完全抹平自由能垒的目的，从而加速了我们所感兴趣的稀有事件的发生。

注：本示例模拟详细信息如下：

🔧 分子动力学模拟软件 AMBER 22:

<https://ambermd.org>

🔧 增强采样插件 Plumed 2.8.1:

<https://github.com/plumed/plumed2>

元动力学模拟的体系结构文件、输入文件和结果分析脚本：

<https://github.com/freeenergylab/articlesWeChat.git>

#### [参考文献]

1. Dama, J. F.; Parrinello, M.; Voth, G. A., Well-tempered Metadynamics Converges Asymptotically. *Physical Review Letters* **2014**, *112*(24), 240602.
2. Laio, A.; Parrinello, M., Escaping Free-energy Minima. *Proceedings of the National Academy of Sciences* **2002**, *99* (20), 12562-12566.
3. Ilott, A. J.; Palucha, S.; Hodgkinson, P.; Wilson, M. R., Well-Tempered Metadynamics as a Tool for Characterizing Multi-Component, Crystalline Molecular Machines. *The Journal of Physical Chemistry B* **2013**, *117* (40), 12286-12295.
4. Barducci, A.; Bussi, G.; Parrinello, M., Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Physical Review Letters* **2008**, *100* (2), 020603.
5. Jambrina, P. G.; Aldegunde, J, Chapter 20 – Computational Tools for the Study of Biomolecules. In *Computer-aided Chemical Engineering*; Elsevier, **2016**; Vol. 39, pp 583-648.
6. Dama, J. F.; Rotskoff, G.; Parrinello, M.; Voth, G. A., Transition-Tempered Metadynamics: Robust, Convergent Metadynamics via On-the-Fly Transition Barrier Estimation. *Journal of Chemical Theory and Computation* **2014**, *10* (9), 3626-3633.
7. Hénin, J.; Lelièvre, T.; Shirts, M.; Valsson, O.; Delemotte, L., Enhanced Sampling Methods for Molecular Dynamics Simulations. *Living Journal of Computational Molecular Science* **2022**, *4* (1), 1583.
8. Khalatur, P. G., 1.16 - Molecular Dynamics Simulations in Polymer Science: Methods and Main Results. In *Polymer Science: A Comprehensive Reference*; Elsevier, **2012**; Vol. 1, pp 417-460.
9. Pampel, B.; Valsson, O., Improving the Efficiency of Variationally Enhanced Sampling with Wavelet-Based Bias Potentials. *Journal of Chemical Theory and Computation* **2022**, *18* (7), 4127-4141.
10. Ramirez-Sarmiento, C.; Noel, J.; Valenzuela, S.; Artsimovitch, I., Interdomain Contacts Control Native State Switching of RfaH on a Dual-Funneled Landscape. *PLoS Computational Biology* **2015**, *11*, e1004379.
11. Parui, S.; Brini, E.; Dill, K. A., Computing Free Energies of Fold-Switching Proteins Using MELD x MD. *Journal of Chemical Theory and Computation* **2023**, *19* (19), 6839-6847.