


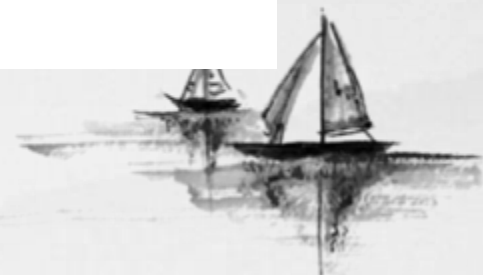
On-the-Fly Learning and Sampling of Ligand Binding by High-Throughput Molecular Simulations

S. Doerr and G. De Fabritiis*

Computational Biophysics Laboratory (GRIB-IMIM), Universitat Pompeu Fabra, Barcelona Biomedical Research Park (PRBB),
C/Doctor Aiguader 88, 08003 Barcelona, Spain

 Supporting Information

报告人：王宏伟



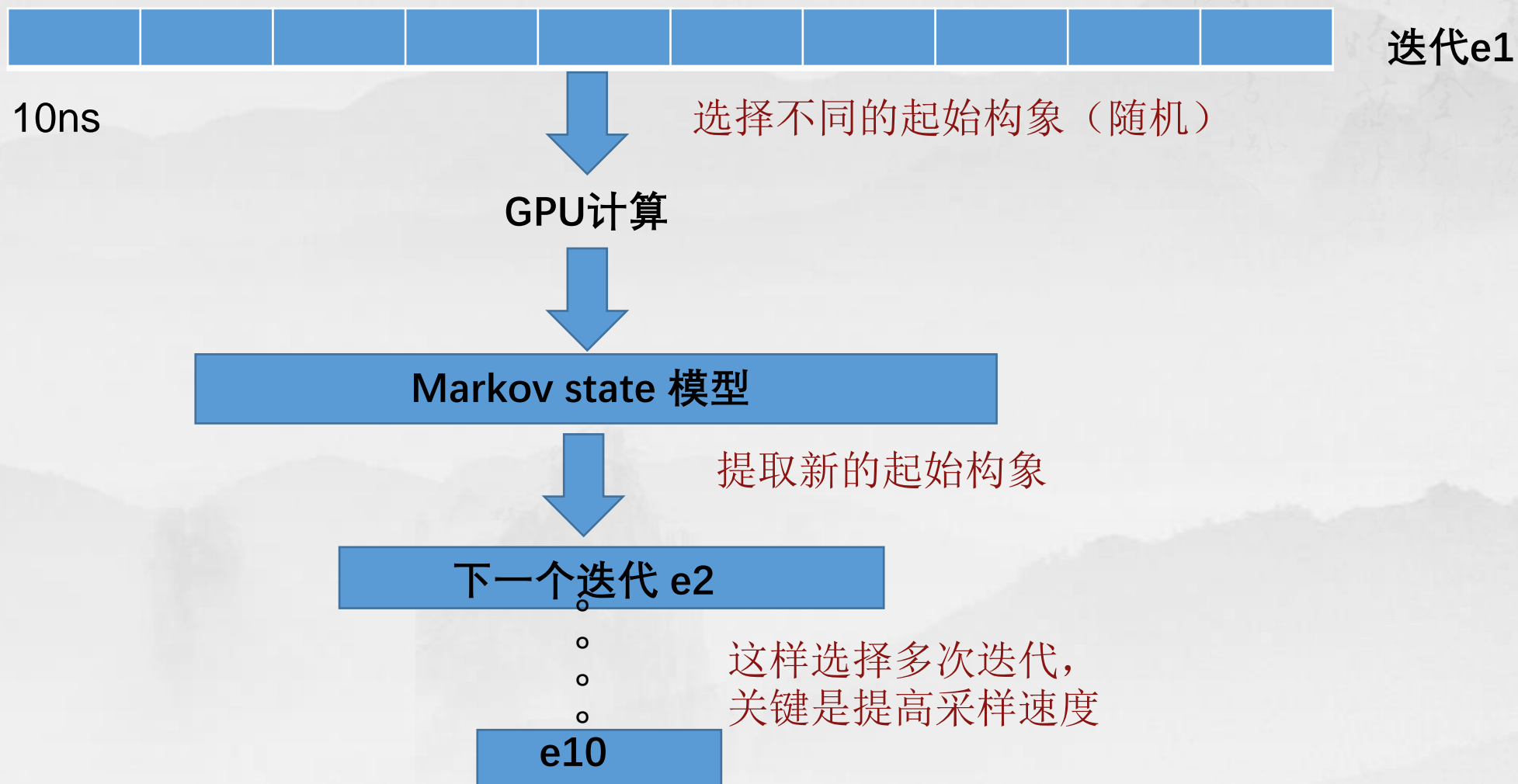
1. BACKGROUND

Adaptive sampling methods.

- 1. 现如今的分子模拟技术通过一些动力学软件已经能分析到微妙到毫秒级别，但往往得到的长轨迹分析起来比较耗时，利用Markov模型为这些大量的模拟提供了分析手段。但是，在生命体从蛋白质与配体结合到蛋白质折叠这种大的构象转变往往需要几秒量级。
- 2. 为了到达几秒级的时间尺度，我们利用迭代的方法，从那些能量低的样品中选择有用的特性，进行下一步的模拟。但这些人工的重复采样方法受人为性干扰，只允许少量的重叠采样。
- 3. 因此作者提出一个完全自动的、快速迭代的方法“自适应抽样方法”

2. METHODS

Run Ai ($1\mu\text{s}$) 为了验证结果收敛性, 跑了A1—A1010个 adaptive run



2. METHODS

Learning the Model

- 1. The resulting data of the 10 parallel simulations is postprocessed to calculate a metric for each Conformation that are clustered using a k-center method into K clusters. $K = 50$ (胰蛋白酶-苯甲醚)
- 2. Using this conformational clustering, a Markov state model is built. (The count and transition probability matrices are estimated using a lag time $t_{lag} = 100$ ps, which is the simulation sampling rate.)

2. METHODS

Sampling Method

如何从Markov模型中选取一定比例的结果作为下一迭代的起始构象。“自适应抽样”基本原理如下：

$$p(m) \propto k_B T \log \left(\frac{k_{\text{on}}^m}{k_{\text{off}}^m} \right)$$

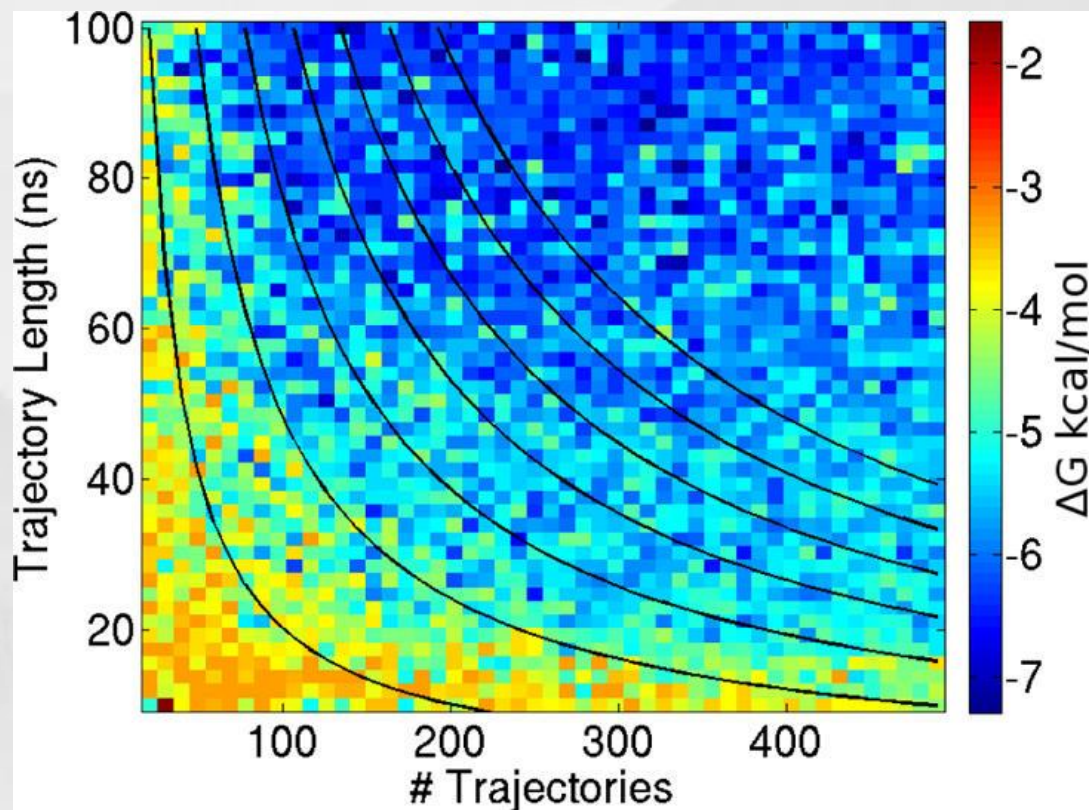
通过上面的公式，我们可以得到概率分布 $p(m)$ ，这样我们可以得到每个构象的模拟数，然后放到GPU上计算，这样进行迭代10次。

2. METHODS

预测标准结合自由能

$$\Delta G^\circ = \Delta G + \Delta G_V = -k_B T \log \left(\frac{p_{\text{eq}}^{\text{bound}}}{p_{\text{eq}}^{\text{bulk}}} \frac{V_u}{V_0} \right)$$

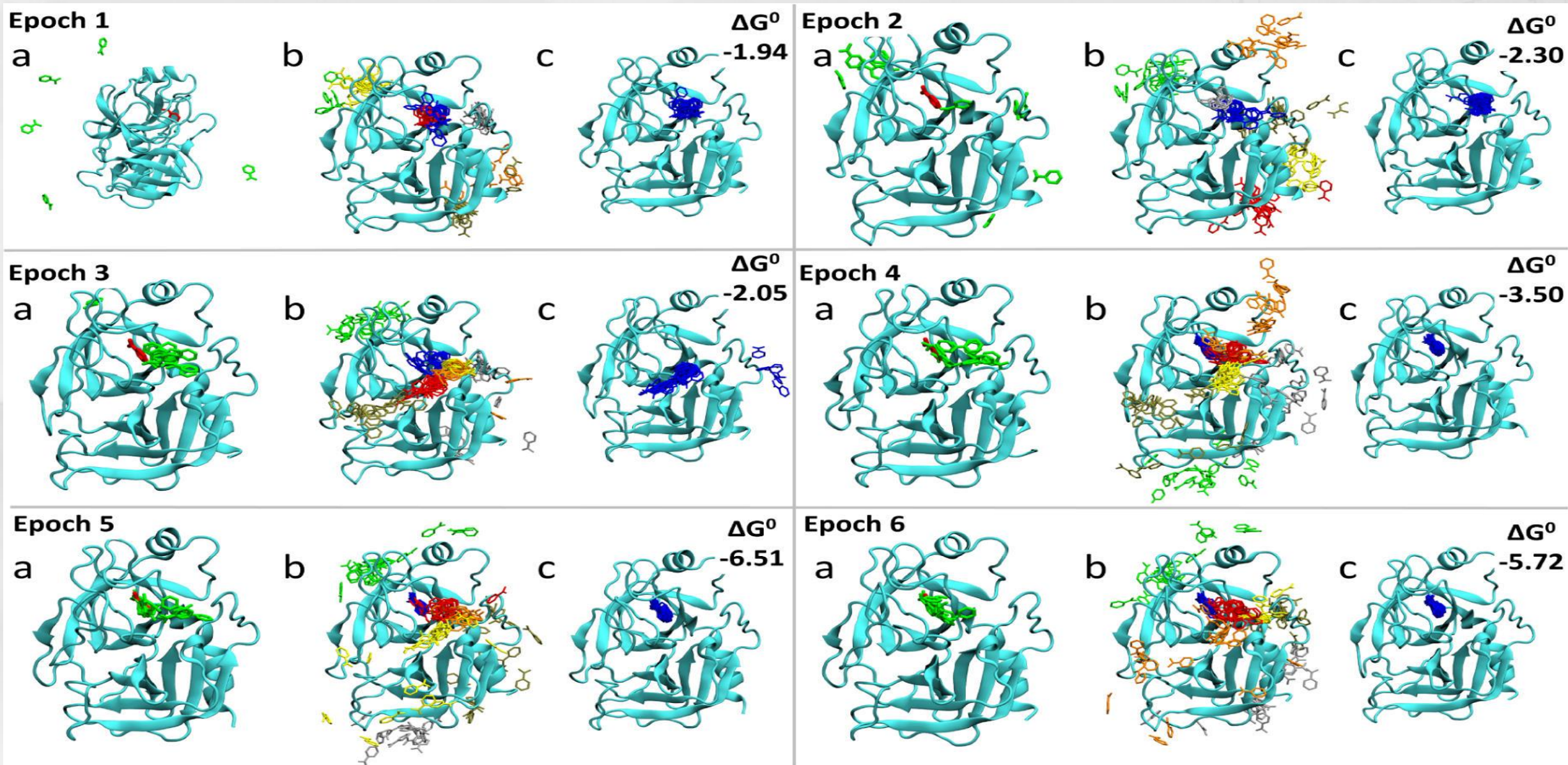
3. Result



1. 我们计算了每个轨迹的长度和轨迹数组合的标准自由能的状态
2. 黑色的曲线是等值线
3. 优点：即使是几个短的模拟、没有足够的时间来采样，也可以正确估计的结合自由能

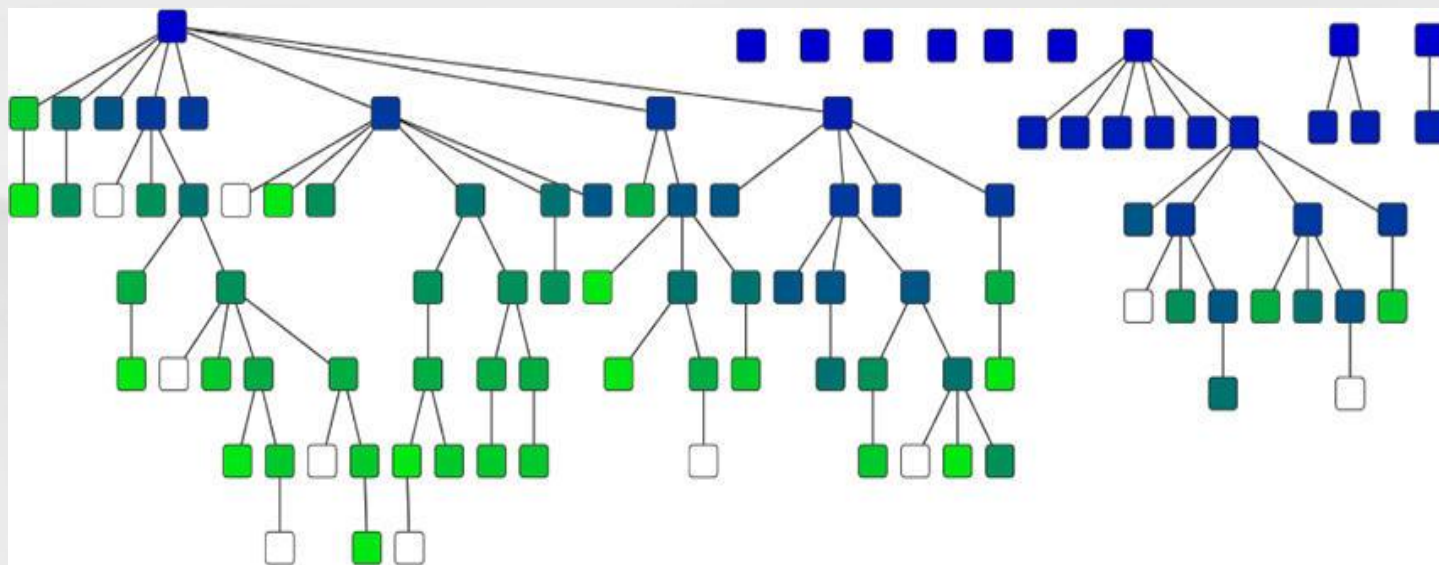
3. Result

自适应采样在每个迭代过程的情况



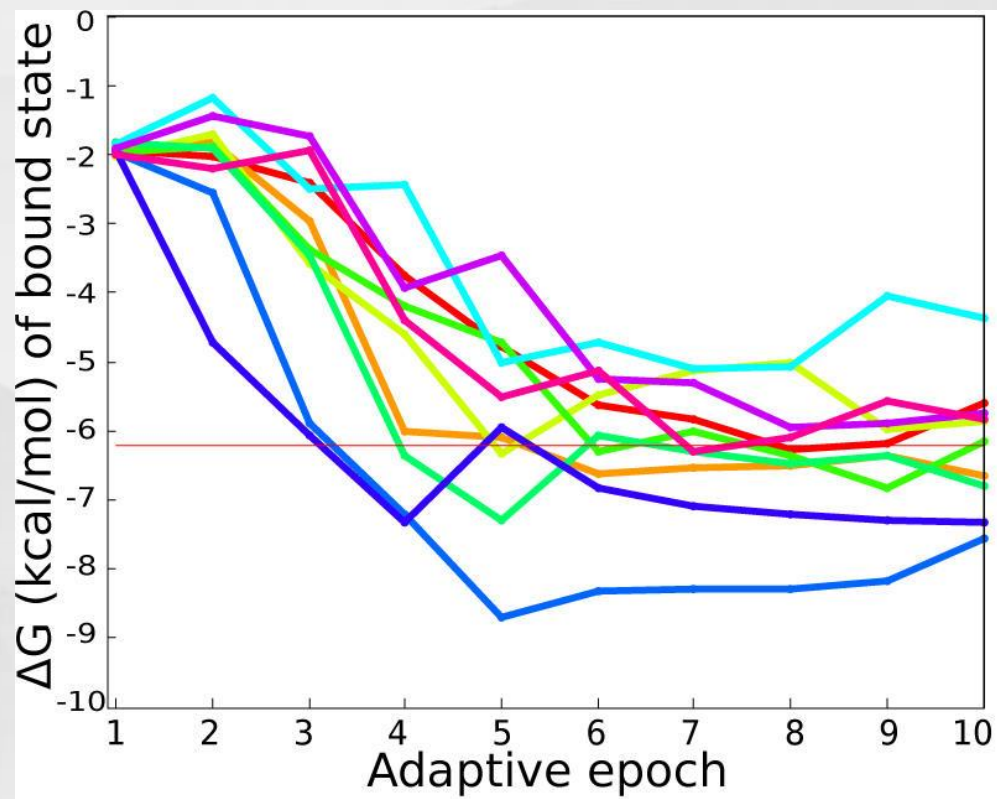
3. Result

Genealogical tree of a single adaptive run

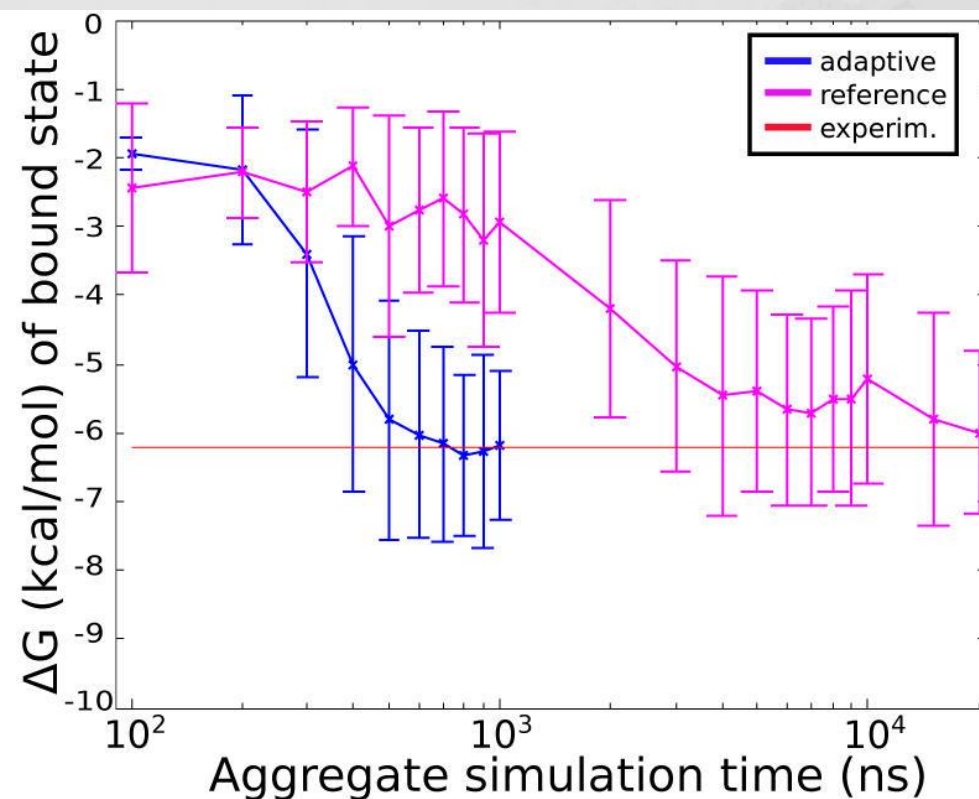


从这10个初始构象的模拟中发现只有2个产生显著的采样。因为在迭代1模拟中只有少数的稳定构象或者这些构象与胰蛋白酶结合时，停留时间远超过其他构象

3. Result



(a)



(b)

4. CONCLUSION

优点：该方法使用马尔可夫模型学习的简化模型的模拟，采用适应性抽样的方法，即便是从很少的样本中采样，也能达到很高的精确度，而且比经典的采样快1个数量级

缺点：该方法选取的样本还是相对比较少，有待于选取更多的样本进行模拟，将简化模型中的参数进行略微的改进，达到更高的精确度

收获：将马尔科夫模型巧妙的融合进动力学分析中，不仅能达到实验的效果，还有利于加快动力学模拟效率

山水秀雅
峰峦叠嶂
烟波浩渺
舟楫往来
渔歌唱晚
牧笛声声
白鹭飞飞
红蓼青青

欣 赏
谢 谢

