

doi:10.3969/j.issn.1672-5565.2016.02.09

# 距离约束的 HMC 采样算法在蛋白质结构预测中的运用

罗升<sup>1</sup>, 吕强<sup>2\*</sup>

(1. 苏州大学计算机科学与技术学院, 苏州 215006

2. 苏州大学江苏省计算机信息处理技术重点实验室, 苏州 215006)

**摘要:**蛋白质结构预测中,采样是指在构象空间中生成具有最小自由能的状态。传统的采样方法是对自由度直接赋值。这种方法在处理较少的残基时能取得好的效果。但是对于包含100个残基以上的蛋白质结构,由于构象空间的急剧增长,难以得到理想的结构。本文引入深度学习中的HMC(Hybrid Monte Carlo)采样方法,以概率分布为依据对蛋白质的自由度进行采样,能够对包含100、200甚至更多个残基的蛋白质结构进行采样。并且,在采样的过程中加入残基间的距离约束,使得一个结构中,相对于Rosetta的ab initio最多有75%(平均40%)的残基对得到优化,满足距离约束。

**关键词:**距离约束;HMC;采样;结构预测;蛋白质

**中图分类号:**TP391 **文献标志码:**A **文章编号:**1672-5565(2016)02-117-06

## Distance constrains model based hybrid monte carlo sampling algorithm in protein structure prediction

LUO Sheng<sup>1</sup>, LÜ Qiang<sup>2\*</sup>

(1. School of Computer Science & Technology, Soochow University, Suzhou 215006, China;

2. Provincial Key Laboratory for Computer Information Processing Technology of Jiangsu, Suzhou 215006, China)

**Abstract:** Sampling is defined as searching the conformational space for the status with the minimum free energy in protein structure prediction. In this paper, the Hybrid Monte Carlo(HMC) method from deep learning algorithms is introduced to better sample the conformational space of protein structures with 100, 200, or even more residues according to the probability distributions, while traditional sampling methods succeed in cases that proteins usually have less residues by assigning each value of free degrees directly. But they often fail the situation in which proteins have more than 100 residues, because of the large conformational space. In addition, residue distance constrains are added to the sampling algorithm to optimize a maximum 75 percent (40 percent on average) of residue pairs in each structure compare with ab initio in Rosetta.

**Keywords:** Distance constrain; HMC; Sampling; Structure prediction; Protein

近年来,从头预测的方法在蛋白质结构预测领域取得了不错的成绩。但是,就目前的技术手段而言,从头预测的方法仍然面临两大基本难点<sup>[1]</sup>:第一,由于蛋白质内部原子相互作用的复杂性,人们难以找到足够准确的能量函数来描述一个蛋白质构象;第二,蛋白质结构的构象空间相当大,尤其在残基序列较长的时候,如果没有合适的、高效的采样算法,蛋白质构象采样将是一个“灾难性”的计算问

题。由此可见,采样算法在从头预测的方法中有着举足轻重的地位<sup>[2]</sup>,因此,许多研究者针对这一问题开发了多种适用于蛋白质三维结构采样的算法,比如改进的遗传算法<sup>[3]</sup>、构象空间退火<sup>[4-5]</sup>和分子动力学模拟采样<sup>[6]</sup>。

在蛋白质结构预测的语境下,采样是指产生一个满足要求的构象,而根据这个要求的不同,我们可以将蛋白质构象采样分为狭义的采样和广义的采样:

收稿日期:2016-03-04;修回日期:2016-04-25.

基金项目:国家自然科学基金项目(No.61170125)。

作者简介:罗升,男,硕士生,苏州大学。研究方向:生物信息计算;E-mail: great\_luo@hotmail.com.

\*通信作者:吕强,男,教授,苏州大学。研究方向:生物信息计算、元启发搜索、并行计算;E-mail: qiang@suda.edu.cn.

(1) 狭义的采样: 一个蛋白质构象可以通过不同的自由度来表示, 常见的有二面角、原子坐标等。如果使用二面角表示蛋白质构象, 那么这里的采样就是指对二面角的一次赋值; 如果使用原子坐标来表示蛋白质构象, 那么这里的采样就是指对原子坐标的一次赋值。Rosetta<sup>[7]</sup>中的每一个 Mover 都是这种意义下的采样。

(2) 广义的采样: 对于表示蛋白质构象的自由度, 如果我们认为它们满足某一特定的概率分布, 比如正态分布  $N(d, u)$ , 那么就可以对这个特定的概率分布进行一次采样, 同样可以得到该自由度表示下的蛋白质的构象, 而这样的采样就是广义上的采样。

对于狭义的采样, 由于是在自由度上进行不断的赋值尝试, 再通过能量函数进行打分评估。不可避免的问题就是随着残基数量的增加, 自由度的数量会成指数级增长, 即使在现有的计算能力下, 这样的计算量增长对采样算法来说, 仍然是不可行的。而且, 能量函数本身就是一个值得研究的问题。本文提出的基于距离约束模型的 HMC (Hybrid Monte Carlo) 采样算法则是对概率模型的采样, 即使残基数量增加, 蛋白质序列变长, 但是在参数数量上面临的问题会比直接对自由度赋值的方法好许多。

HMC 采样<sup>[8-9]</sup>是一种对概率模型的采样算法, 其采样机制通过引入 Hamiltonian 动力学系统<sup>[10]</sup>, 避免在巨大的采样空间中随机采样, 并且可以在采样的过程中保留概率分布所具有的统计规律。例如, 在本文的案例中, 如果我们使用原子坐标作为自由度来表示一个蛋白质结构, 每个残基选取 4 个原子, 即 N、C $\alpha$ 、C 和 O; 每个原子有 3 个笛卡尔坐标, 即 x、y 和 z 坐标。因此, 表示一个残基的就是 12 个自由度。对于一个长度为 10 残基的蛋白质构象, 用来表示它的自由度就有 120 个。进一步认为这 120 个自由度满足一个 120 个随机变量的概率分布, 那么就可以用 HMC 算法对这个概率分布进行一次采样, 可以认为这是对该结构的一次“理性”的采样。因为采样的过程是模拟的 Hamiltonian 动力学系统, 即粒子在空间中运行的状态, 那么如何保证蛋白质结构在采样的过程中不会因为随机的走动而散架呢? 本文引入了距离约束模型。

蛋白质的空间结构并不仅仅是由一个个的残基的独立位置决定的。事实上, 残基间的相互作用对蛋白质结构有着巨大的影响<sup>[11]</sup>。在蛋白质结构中, 按照残基的间隔, 一般可以分为三类距离约束<sup>[12]</sup>: 1) 短距离约束 (Short-range), 此类约束的残基间隔相对较短, 范围通常在 6-12 个残基; 2) 中

等距离约束 (Medium-range), 此类约束的残基间隔一般在 12-24 个残基; 3) 长距离约束 (Long-range), 此类约束的残基间隔一般在 24 个残基以上。同时, 在蛋白质结构中, 一般认为残基空间欧氏距离在 8 Å 以内的残基间是有相互作用的, 因此本次实验中使用的约束模型就是 8 Å 以内的长距离约束。

## 1 材料与方法

### 1.1 数据准备

蛋白质结构预测评估大会<sup>[13]</sup> (CASP: Critical Assessment of techniques for protein Structure Prediction) 是世界范围内对蛋白质结构预测技术进行评估, 评价当前蛋白质结构建模方法的能力及局限性的竞赛, 被称为蛋白质结构预测领域的“奥运会”。本文从 CASP8、CASP9 和 CASP10 当中共选取了 16 个案例, 它们的平均残基长度为 100~200, 最短的包含 101 个残基, 最长的包含 195 个残基, 每个蛋白质都是提取的单链。

如上文所述, 本文采用的自由度是原子三维坐标, 每个残基取其包括骨架原子 C $\alpha$  在内的 4 个原子的三维坐标, 即每个残基由 12 个值表示。每条蛋白质为一行, 以 300 行为一个 batch。因为本文的主要目的是验证采样算法的可行性, 所以在数据准备的时候是从蛋白质的天然结构里获取其距离约束的, 而在实际应用的时候可以考虑使用同源蛋白质的距离约束。

从天然结构中获取距离约束的过程如下: 首先, 遍历天然结构的残基, 计算出每对残基间隔大于 24 个残基的残基对的空间欧氏距离, 然后找出小于 8 Å 的残基对, 记录下这些残基对以及距离值, 作为约束。采样的输入数据是用 Rosetta 平台 ab initio 得到的。考虑到 ab initio 得到的结构具有较大的 RMSD, 可以更加显著的看到采样算法的效果, 而且, 采样是为了得到更加符合距离约束的 decoys, 因此选用这样的数据集作为采样的输入。

### 1.2 HMC 采样

HMC 采样算法的核心在于通过引入 Hamiltonian 动力学系统, 模拟粒子在空间的运动, 从而对一个能量模型进行采样。事实上, 蛋白质的空间结构就是由空间中一个个粒子组成的, 对蛋白质结构的采样可以想象为让这些粒子在空间中不断的运动, 以找到和天然结构最接近的构象。可见, 如果不添加任何约束, 那么采样空间将接近于无穷大, 因为粒子在空间里可以向任意方向以任意速度运

动。面对如此巨大的采样空间,任何采样方法都是无能为力的。因此必须添加一定的约束从而一定程度上缩小采样空间。

### 1.2.1 Hamiltonian 动力学系统

粒子在空间中的任何时刻都具有两种属性,即速度和位置,这两种属性对应于两种能量:动能和势能,我们把这样的一个动力系统称作 Hamiltonian 动力学系统。HMC 采样正是通过模拟这样的一个系统,对一个高维空间进行采样的。其能量函数的定义如公式(1)、(2)、(3)所示:

$$E(s) = s \in R^D \quad (1)$$

$$K(\varphi) = \varphi \in R^D \quad (2)$$

$$H(s, \varphi) = E(s) + K(\varphi) = E(s) + \frac{1}{2} \sum_i \varphi_i^2 \quad (3)$$

其中,  $E(s)$  为势能,  $K(\varphi)$  为动能。在本文的案例中,  $E(s)$  即为距离约束模型的值,在采样过程中,会沿着这个能量和的最小方向进行采样。这样的过程就一定程度上缩小了采样空间,避免了直接采样的不可行性。而在 Hamiltonian 系统中,我们并没有直接对  $p(s)$  进行采样,所用的方法是通过对于一个正则分布:

$$p(s, \varphi) = \frac{1}{Z} \exp(-H(s, \varphi)) = p(s)p(\varphi)$$

进行采样,这是因为两个概率分布  $p(s)$  和  $p(\varphi)$  是相互独立的,边缘化  $\varphi$  并不会产生很大的影响,并且可以重现原有的概率分布。

### 1.2.2 蛙跳算法

因为模拟的时间是离散的,而粒子的运动是连续的,所以在实现过程中并不能很精确的模拟这样一个系统。对于这个问题,有多种方案可以解决,其中比较好的是蛙跳算法<sup>[14-15]</sup>,可以通过三步操作解决这个问题,如公式(4)、(5)、(6)所示:

$$\varphi_i \left( t + \frac{\varepsilon}{2} \right) = \varphi_i(t) - \frac{1}{2} \frac{\partial}{\partial s_i} E(s(t)) \quad (4)$$

$$s_i(t + \varepsilon) = s_i(t) + \varepsilon \varphi_i \left( t + \frac{\varepsilon}{2} \right) \quad (5)$$

$$\varphi_i(t + \varepsilon) = \varphi_i \left( t + \frac{\varepsilon}{2} \right) - \frac{\varepsilon}{2} \frac{\partial}{\partial s_i} E(s(t + \varepsilon)) \quad (6)$$

这里首先计算了速度在时间为  $t + \varepsilon/2$  时候的势能,然后通过这个值计算出  $s_i(t + \varepsilon)$  和  $\varphi_i(t + \varepsilon)$ 。

### 1.2.3 Metropolis 接受/拒绝

由于使用有限的步长  $\varepsilon$  并不能准确的表达能量  $H(s, \varphi)$ ,因此可能在模拟的过程中产生误差。而且使用单精度的数值也可能导致舍入误差,所以对于一次新的采样,会使用 Metropolis 准则<sup>[16-17]</sup>进行

一次接受/拒绝判定。其判定过程如公式(7)所示:

$$P_{acc}(x, x') = \min \left( 1, \frac{\exp(-H(s', \varphi'))}{\exp(-H(s, \varphi))} \right) \quad (7)$$

Metropolis 准则是模拟退火算法中最常用的接受准则,其基本思想是以一定的概率接受新状态。它的重点在于如果  $H(s', \varphi')$  小于  $H(s, \varphi)$  那么,新状态被接受,反之,则以  $P_{acc}(x, x')$  的概率接受,而不是拒绝。

### 1.2.4 HMC 采样流程

在本文的实验中,势能是根据蛋白质残基的距离约束构建的模型,动能是一个随机的均匀分布,整个采样的流程如下:

- (1) 从均匀分布中产生一个随机的速度(动能);
- (2) 执行  $n$  次蛙跳算法,获取新的状态  $x'$  (势能 + 动能),即对残基的三维坐标进行  $n$  次调整,模拟粒子的运动,每次赋值都是通过计算势能函数的梯度得到的,并不是随机的赋值;
- (3) 对新的状态  $x'$  和旧的状态  $x$  执行一次 Metropolis 接受判断,是否可以接受;
- (4) 根据当前接受率和目标接受率对步长进行一次调整,调整采样效率;
- (5) 重复以上步骤,直到达到特定的步数;

HMC 采样的过程中,有几个重要的参数,比如步长、蛙跳次数、步长增益幅度等等,这些参数直接涉及到采样的效果和速度(见表1)。

表1 采样参数

Table 1 The parameters of sampling

参数名	取值	参数意义
initial_stepsize	0.000 001 0	初始步长
stepsize_dec	0.980 000 0	步长减小比例
stepsize_min	0.000 000 1	最小步长
stepsize_max	0.000 015 0	最大步长
stepsize_inc	1.020 000 0	步长增加比例
n_steps	20.000 000 0	蛙跳次数

实验中,步长的初始值和变化范围设置的比较小,这是可以理解的,因为蛋白质原子空间的变化本身就不是很大,太大幅度的变化可能导致空间严重变形,或者直接越过最佳位置。而且本文仅仅加了距离约束,没有添加其它约束,如果设置过大,那么将会出现不可预期的结构。

### 1.3 模型构建和实验流程

实验中使用的是残基间隔大于 24 并且残基空间欧氏距离小于 8 Å 的残基对作为约束。在实验的过程中,对于每一次的采样结果,都会与天然结构进

行计算。主要分为三步:

- (1) 计算标记的每个残基对的空间欧氏距离;
- (2) 将这些距离与天然结构中对应残基对的空间欧氏距离作差;
- (3) 计算这些差值的最小平方和。

用这个值作为 HMC 采样的势能函数。这里,使用最小平方和作为模型输出,是因为该方法本身具有的优点:它可以找出与数据最佳的匹配函数,并且对于曲线有很好的拟合能力。在本实验中,该模型如公式(8)~(11)所示:

$$\Gamma_D = \sqrt{((D_{lx} - D_{rx})^2 + (D_{ly} - D_{ry})^2 + (D_{lz} - D_{rz})^2)} \quad (8)$$

$$\Gamma_N = \sqrt{((N_{lx} - N_{rx})^2 + (N_{ly} - N_{ry})^2 + (N_{lz} - N_{rz})^2)} \quad (9)$$

其中,  $\Gamma_D$  和  $\Gamma_N$  分别表示 decoys 和天然结构中,两个标记的残基对之间的空间欧氏距离。然后对它们求差值:

$$\Delta_i = \Gamma_{D_i} - \Gamma_{N_i} \quad (10)$$

最后,再将这些差值求平方和:

$$\varphi = \sum_{i=0}^n \Delta_i \quad (11)$$

$\varphi$  的大小反映了采样得到的构象是否满足从天然结构中获取的距离约束。其值越小,说明标记的每对残基对的空间欧氏距离和天然结构中对应的残基对的空间欧氏距离越接近,因此我们也就认为在采样的过程中,距离约束起到了作用,得到了更加满足距离约束的构象。

按照上述模型,在本地机器上(CPU: Intel(R) Xeon(R) E5-2620 v2 @ 2.10Hz 内存:16GB)跑完 HMC 算法用时 3.5 h,每个案例在 6~26 min 不等。整个采样实验流程如下:

- (1) 使用 Rosetta 平台,对每个案例用 ab initio 的方法得到 300 个 decoys 作为采样的输入;
- (2) 从天然结构中计算得到距离约束,即包括残基对和对应的空间欧氏距离;
- (3) 对于 ab initio 得到的 decoys,提取每个残基骨架原子的三维坐标,构成一个行数为 300,列数为序列长度  $\times$  12 的输入矩阵;
- (4) 将做好的矩阵输入到 HMC 当中,进行多次的迭代采样,并输出最终得到的构象。

之所以选用 Rosetta 平台制作输入集,也是为了作为对比结果。Rosetta 的 ab initio 就是一种不带距离约束的采样,本文使用这样的 decoys 作为输入,经过带距离约束的 HMC 采样,然后比较输出 decoys 的距离约束,分析 HMC 采样的效果。

## 2 结果与分析

实验一共使用了 16 个案例,每个案例用 300 个 decoys 作为采样的输入,每个案例采样 1 001 次,其中有 1 000 次作为 HMC 的 burn-in 阶段,丢弃。保留最后一次的结果。对比采样之前的 input-decoys 和采样之后得到的 output-decoys,分别对每个 decoys 中标记的距离小于 8Å 的残基对的个数变化做了统计并比较,变化比例的计算如公式(12)所示:

$$\text{变化比例} = \frac{\text{采样前大于 } 8\text{\AA} \text{ 的残基数}}{\text{采样后小于 } 8\text{\AA} \text{ 的残基数}} \times \frac{\text{总距离约束数量}}{\text{总距离约束数量}} \quad (12)$$

实验统计结果如表 2 所示,其中第一列是所选案例的 PDB ID,二、三列分别是残基和约束对的个数,最后三列则是变化的百分比。

表 2 采样后距离小于 8 Å 的残基对个数变化比例  
Table 2 Proportion changes of quantities of residue pairs within 8 Å after sampling

PDB ID	残基数 (个)	约束数 (个)	变化比例(%)		
			最大	前 1% 平均	平均
1WWV	101	127	22.83	21.78	6.18
1WWW	101	122	31.15	28.69	6.32
4EFV	101	123	30.89	24.93	5.77
1IKN	104	131	44.27	34.35	6.28
1U3Z	104	136	56.62	45.35	7.09
3JV5	104	132	71.79	51.01	7.84
3JV6	104	135	75.56	66.42	8.07
1G4U	175	205	34.15	33.50	8.49
1HH4	175	227	40.97	35.53	8.23
2G0N	175	219	48.40	45.05	11.05
2WKP	175	226	55.75	49.11	9.37
1ZHI	195	160	18.12	16.46	2.52
3TU4	195	139	16.55	15.59	2.41
4JJN	195	152	13.16	11.84	2.34
4KUD	195	150	20.67	19.78	4.52
4KUI	195	143	13.99	13.52	2.15

从表 2 中可以看出,在所有的案例中,采样后的 300 个 decoys 里,小于 8Å 的残基对个数有了明显的增长。如案例 3JV6 里,最大的一个构象有 75.56% 的残基对在采样之后空间欧氏距离小于 8Å。即 3JV6 的天然结构中,一共有 104 个残基,长距离残基对个数为:  $1 + 2 + 3 + \dots + 79 + 80 = 3\ 240$ 。其中,空间欧氏距离小于 8Å 的有 132 个。变化最大的一个 decoy,在采样的过程中,这 132 个残基对里

有  $132 \times 0.7556 \approx 100$  个残基对,由原来大于  $8\text{\AA}$  的距离变成小于  $8\text{\AA}$  的距离,由不满足距离约束变为了满足距离约束。这就可以说明,本次实验中,HMC 采样成功的加入了距离约束,使一些原本大于  $8\text{\AA}$  的残基对的距离小于  $8\text{\AA}$ 。这样的意义在于使原来没有满足距离约束的残基对变得满足距离约束,如此一来,这样的构象就是比较好的构象,对下一步的三级结构预测有很大的意义。

另外,表 2 是根据残基数从少到多的顺序排列的,通过观察数据还可以看出,残基数对采样效果的影响很少。残基数较少的案例其采样的效果并不一定好,例如 1WWV、1WWW 等;而残基数较多的案例其采样的效果也不一定会差,例如 1HH4、2GON、2WKP。这也印证了之前所说的,HMC 采样避免了因序列增长,自由度数量急剧增长而导致的采样空间巨大,采样算法不可行的问题。

同时,本文也计算了采样前后蛋白质结构 RMSD 的变化,发现 RMSD 的变化很小,分析可能的原因有两点:

第一,采样的步长取值很小,也就意味着每一次采样对三维坐标的变化很小。模拟到 Hamiltonian 动力学系统中就是指粒子在空间中每次移动的速度很慢,因此原子移动距离很小,而且,实验中仅仅取了  $C_{\alpha}$ ,并没有全原子采样。事实上,如果步长较大,那么反而会出现不可预期的结果。虽然 RMSD 变化不大,但是在经过多次移动之后,仍然可以将大于  $8\text{\AA}$  的残基距离缩小到  $8\text{\AA}$  以内。可见采样的过程还是符合预期的结果;

第二,在采样的过程中,标记出的约束对里,虽然有部分残基对空间欧氏距离从大于  $8\text{\AA}$  降到  $8\text{\AA}$  以内,但也有部分从  $8\text{\AA}$  以内拉长到  $8\text{\AA}$  以外。但是,从整体相互作用的残基对数变化的情况来看,采样还是可以获得比较好的构象的。事实上,采样得到的 decoys 并不是都要的,我们只需要保留比较好的构象。本文对这部分数据也做了统计,如表格 3 所示。

从表 3 中可以看出,对于好的采样案例,空间欧氏距离小于  $8\text{\AA}$  的残基对个数是增加的,而且增加的不少。在统计的 16 个案例中,有 3 个案例出现了超过 50% 的增幅(3JV5、3JV6、2WKP)。表中还统计了前 1% 平均和前 10% 平均,可以说,这些构象是比较好的构象,也正是我们需要的。结合表 2 和表 3 可见,基于距离约束的 HMC 采样算法在蛋白质结构预测方面是可行的。并且,相对于目前常用的 Rosetta 方法,能够得到更加满足距离约束的采样结果,取得更好效果,对于三级结构预测有很大的帮助。

表 3 标记的残基对中距离小于  $8\text{\AA}$  的残基对增加个数及每个案例用时

Table 3 Increased Number of Marked Residue Pairs within  $8\text{\AA}$  and Time Use for Each Case

PDB ID	约束数 (个)	最大 (个)	前 1% 平均 (个)	前 10% 平均 (个)	用时 (分钟)
1WWV	127	28	27.00	7.40	6
1WWW	122	36	32.00	8.00	7
4EFV	123	38	29.00	6.50	7
1IKN	131	55	41.33	7.70	8
1U3Z	136	66	53.33	10.03	8
3JV5	132	90	63.00	12.13	7
3JV6	135	102	85.33	16.80	8
1G4U	205	70	68.67	7.60	9
1HH4	227	75	71.34	17.30	12
2GON	219	97	94.33	25.33	11
2WKP	226	123	107.67	24.97	12
1ZHI	160	29	25.67	5.70	18
3TU4	139	23	21.67	4.37	23
4JJN	152	20	15.33	3.40	24
4KUD	150	29	27.33	7.03	26
4KUI	143	20	19.00	4.07	24

### 3 结束语

本文以蛋白质结构中的距离约束为模型,使用 HMC 采样对蛋白质空间三维坐标进行采样。实验已可以证明 HMC 采样可以用在蛋白质结构预测方面,并且相对于目前常用的 Rosetta 有不错的效果。因为模拟了 Hamiltonian 动力学系统,避免了其它采样的随机性和盲目性,所以即使同时加入了上百对距离约束,仍然能在计算能力可及的范围内取得结果。然而,实验本身并不是完美的,仍然有需要改进的地方。比如,在加入距离约束的同时,如何能降低蛋白质结构的 RMSD。因为结构预测的最终目的是获得近天然结构,距离约束只是为了更好的预测结构。因此,如何优化模型,加入其它信息,是本文的下一个工作重点。

### 参考文献

- [1] HARDIN C, POGORELOV T V, Luthey-Schulten Z. Ab initio protein structure prediction [J]. Current Opinion in Structural Biology, 2002, 12(2): 176-181.
- [2] KIM D E, BLUM B, BRADLEY P, et al. Sampling bottlenecks in de novo protein structure prediction [J]. Journal of

- Molecular Biology, 2009, 393(1): 249–260.
- [3] JUDSON R S, COLVIN M E, MEZA J C, et al. Do intelligent configuration search techniques outperform random search for large molecules? [J]. International Journal of Quantum Chemistry, 1992, 44(2): 277–290.
- [4] LEE J, SCHERAGA H A, RACKOVSKY S. New optimization method for conformational energy calculations on polypeptides: conformational space annealing [J]. Journal of Computational Chemistry, 1997, 18(9): 1222–1232.
- [5] LEE J, SCHERAGA H A. Conformational space annealing by parallel computations: extensive conformational search of Met-enkephalin and of the 20-residue membrane-bound portion of melittin [J]. International Journal of Quantum Chemistry, 1999, 75(3): 255–265.
- [6] LEE M R, TSAI J, BAKER D, et al. Molecular dynamics in the endgame of protein structure prediction [J]. Journal of Molecular Biology, 2001, 313(2): 417–430.
- [7] ROHL C A, STRAUSS C E M, MISURA K M S, et al. Protein structure prediction using Rosetta [J]. Methods in Enzymology, 2004, 383: 66–93.
- [8] DUANE S, KENNEDY A D, PENDLETON B J, et al. Hybrid monte carlo [J]. Physics Letters B, 1987, 195(2): 216–222.
- [9] SEXTON J C, WEINGARTEN D H. Hamiltonian evolution for the hybrid monte carlo algorithm [J]. Nuclear Physics B, 1992, 380(3): 665–677.
- [10] DIRAC P A M. Generalized hamiltonian dynamics [C] // Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences. The Royal Society, 1958, 246(1246): 326–332.
- [11] WANG Z, XU J. Predicting protein contact map using evolutionary and physical constraints by integer programming [J]. Bioinformatics, 2013, 29(13): i266–i273.
- [12] MONASTYRSKY B, D'ANDREA D, FIDELIS K, et al. Evaluation of residue-residue contact prediction in CASP10 [J]. Proteins: Structure, Function, and Bioinformatics, 2014, 82(S2): 138–153.
- [13] KRYSHTAFOVYCH A, MONASTYRSKY B, FIDELIS K. CASP prediction center infrastructure and evaluation measures in CASP10 and CASP ROLL [J]. Proteins, 2014, 82(S2): 7–13.
- [14] VAN GUNSTEREN W F, BERENDSEN H J C. A leap-frog algorithm for stochastic dynamics [J]. Molecular Simulation, 1988, 1(3): 173–185.
- [15] SNYMAN J A. The LFOPC leap-frog algorithm for constrained optimization [J]. Computers & Mathematics with Applications, 2000, 40(8): 1085–1096.
- [16] NEAL R M, EDU E R T. Probabilistic inference using markov chain monte carlo methods [J]. Physics and Chemistry, Nato Science Series C: Mathematical and Physical Sciences, 1993, 92(440): 497–537.
- [17] ULYANOV N B, SCHMITZ U, JAMES T L. Metropolis monte carlo calculations of DNA structure using internal coordinates and NMR distance restraints: an alternative method for generating a high-resolution solution structure [J]. Journal of Biomolecular NMR, 1993, 3(5): 547–568.