

## 蛋白质构象空间的多模态优化算法

李章维<sup>1</sup> 肖璐倩<sup>1</sup> 郝小虎<sup>1</sup> 周晓根<sup>2</sup> 张贵军<sup>1</sup><sup>1</sup> 浙江工业大学信息工程学院 杭州 310023<sup>2</sup> 密歇根大学计算医学和生物信息学系 安娜堡 48109-2218

(lzw@zjut.edu.cn)

**摘要** 蛋白质能量模型的不精确性导致数学上的最优解并不一定对应其稳定的天然态结构,同时其巨大的构象空间使得现有方法也极易收敛到局部最优解。针对蛋白质结构能量模型不精确和高维构象空间采样可靠性低的问题,在进化算法的基础上,提出了一种基于二面角相似度的蛋白质构象多模态优化方法。首先,执行模态探测,将 Rosetta 粗粒度能量模型作为筛选高质量新个体的标准,进行种群更新,增加种群构象的多样性;然后,建立二面角相似度模型,用于评价不同构象间的相似程度,以满足多模态优化算法中相似个体快速判定的要求,并基于排挤更新策略实现模态增强,获得结构更为合理的构象。10 个测试蛋白质的实验结果表明:所提算法能够达到较高的预测精度,并且可以使种群具有良好的模态分布,得到尽可能多的高质量局部极值解,从而获得一些较好的蛋白质亚稳态结构。

**关键词:** 蛋白质结构预测;多模态优化;二面角相似度模型;Rosetta;粗粒度能量模型

**中图法分类号** TP301.6

## Multimodal Optimization Algorithm for Protein Conformation Space

LI Zhang-wei<sup>1</sup>, XIAO Lu-qian<sup>1</sup>, HAO Xiao-hu<sup>1</sup>, ZHOU Xiao-gen<sup>2</sup> and ZHANG Gui-jun<sup>1</sup><sup>1</sup> College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023, China<sup>2</sup> Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI 48109-2218, USA

**Abstract** Due to the inaccuracy of protein energy model, the optimal solution in mathematics does not consistently correspond to the stable natural structure of a given protein. The existing methods tend to converge to local optimal solutions because of the huge conformational space. To address the problem of the inaccuracy of energy model in the field of protein structure prediction and the reliability of high-dimensional conformational space sampling, a dihedral angle similarity based multimodal conformation optimization method (DASOM) for ab-initio protein structure prediction is proposed in the framework of population based algorithm. Firstly, the modal exploration is conducted, knowledge-based Rosetta coarse-grained energy model is used as the standard to select new individuals with high quality, thus the diversity of the population can be increased. Then, a dihedral angular similarity model is established to meet the requirements of similar individual determination in the multi-modal optimization algorithm. Crowding update strategy is used for optimizing the existing modal to achieve the modal enhancement and more reasonable conformation is obtained. Experimental results on 10 test proteins show that the proposed method not only achieves high prediction accuracy, but also obtains many metastable protein conformations as well.

**Keywords** Protein structure prediction, Multimodal optimization, Dihedral angular similarity model, Rosetta, Coarse-grained energy model

## 1 引言

从蛋白质氨基酸序列出发来预测蛋白质三维结构(即蛋白质结构预测)是目前生物信息学领域的热点研究问题<sup>[1-2]</sup>。构象优化是蛋白质结构预测的关键环节<sup>[3]</sup>,十几年来许多算法被提出用于解决该问题。分子动力学模拟(Molecular Dynamics, MD)<sup>[4-6]</sup>、蒙特卡罗(Monte Carlo, MC)<sup>[7-9]</sup>、遗传算法

(Genetic Algorithm, GA)<sup>[10-13]</sup>、构象空间退火(Conformational Space Annealing, CSA)<sup>[14]</sup>以及构象树指导搜索(Conformation Tree Guiding Exploration, CTGE)<sup>[15-19]</sup>等算法在蛋白质结构预测领域得到了广泛的应用。然而,在蛋白质结构预测领域,评判一种方法的唯一标准是蛋白质结构预测的精度,目前尚未出现一种能完整地表达蛋白质分子间各种作用力的能量模型。蛋白质能量模型的不精确性,导致数学上的最优解

收稿日期:2019-06-19 返修日期:2019-11-08 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:国家自然科学基金(61773346,61573317)

This work was supported by the National Natural Science Foundation of China (61773346,61573317).

通信作者:张贵军(zgj@zjut.edu.cn)

并不一定对应其稳定的天然态结构;即便有完美的能量模型,其巨大的构象空间也使现有方法极易收敛到局部最优解。

设计一种多模态蛋白质构象优化方法在蛋白质结构预测领域具有十分重要的意义<sup>[20]</sup>。一方面,多模态优化可以弥补蛋白质知识能量模型不精确的缺陷;另一方面,可以提高构象优化算法本身的可靠性。1997年, Lee提出了构象空间退火(Conformation Space Annealing, CSA)<sup>[21]</sup>算法。2010年, Wong等使用CGA-mixed方法<sup>[22]</sup>得到了蛋白质多种模态的稳定结构。2017年, Zhang等提出了(Multi-subpopulation Differential Evolution, MDE)方法<sup>[23]</sup>,先使用快速模态识别方法生成给定数量的模态,然后使用差分进化算法进行模态保持,同时引入抽象凸知识进行模态增强。该算法可以得到若干个包含一系列与能量值成比例的构象特定簇,在不需要聚类操作的情况下,可以直接从生成的簇中得到具有代表性的构象。此外, Lee<sup>[13, 24-27]</sup>、Shehu<sup>[15-18, 28-30]</sup>及Kandathil<sup>[31]</sup>等研究小组也在该方向开展了一些研究工作。然而,多模态优化中极其关键的结构相似度指标、算法早熟收敛以及计算复杂等问题仍然存在<sup>[31]</sup>。

针对蛋白质能量模型不精确的问题,本文在进化算法的基础上提出了一种基于二面角相似度的蛋白质构象多模态优化方法(Dihedral Angle Similarity based multimodal conformation optimization method, DASOM)。该算法分为模态探测和模态增强两个阶段:首先执行模态探测,以Rosetta粗粒度能量模型作为筛选高质量新个体的标准,进行种群更新,增加种群构象的多样性;然后建立二面角相似度模型,以满足多模态优化算法中相似个体判定的需求,并基于排挤更新策略实现模态增强。10个测试蛋白质的实验结果表明:所提方法能够达到较高的预测精度,并得到一系列的局部最优构象。

## 2 能量模型

由于蛋白质分子体系复杂,构象空间的自由度极高,因此构建蛋白质粗粒度表达模型能够有效地降低计算复杂度和计算代价。Rosetta粗粒度能量模型在保留蛋白质重要结构信息的基础上,确定N, C, C<sub>α</sub>和O这些主链的骨干原子以及侧链替代原子的信息,有效降低了计算复杂度和代价<sup>[32]</sup>。

Rosetta从头预测算法的4个阶段中使用的能量函数不同。本文采用的能量函数是Rosetta Score3,它同时考虑了10种不同的能量项,比Rosetta中其余3种能量函数更加全面,表达式如下:

$$E_{\text{protein}} = W_{\text{repulsion}} E_{\text{repulsion}} + W_{\text{attraction}} E_{\text{attraction}} + W_{\text{solvation}} E_{\text{solvation}} + W_{\text{bb-sc hb}} E_{\text{bb-sc hb}} + W_{\text{bb-bb hb}} E_{\text{bb-bb hb}} + W_{\text{sc-sc hb}} E_{\text{sc-sc hb}} + W_{\text{pair}} E_{\text{pair}} + W_{\text{dunbrack}} E_{\text{dunbrack}} + W_{\text{rama}} E_{\text{rama}} + W_{\text{reference}} E_{\text{reference}}$$

其中各能量项的能量函数表达式的具体参数配置请见文献<sup>[33-34]</sup>。

## 3 DASOM 算法描述

### 3.1 二面角相似度模型

蛋白质构象空间的复杂度极高,因此能够快速执行蛋白质结构比对的方法在蛋白质结构多模态优化过程中极为重要。本文基于蛋白质结构粗粒度表达模型,根据蛋白质结构

的二面角角度信息设计了二面角相似度模型,不需要对两个蛋白质进行空间旋转平移,有效地降低了计算代价。

粗粒度蛋白质模型只保留了主链骨干原子N, C<sub>α</sub>, C, O以及侧链替代原子。如图1所示,用二面角 $\phi$ ,  $\psi$ 和 $\omega$ 来表示氨基酸链, $\phi$ 表示C-N-C<sub>α</sub>所在平面与N-C<sub>α</sub>-C所在平面构成的二面角, $\psi$ 表示N-C<sub>α</sub>-C所在平面与C<sub>α</sub>-C-N所在平面构成的二面角, $\omega$ 表示C<sub>α</sub>-C-N所在平面与C-N-C<sub>α</sub>所在平面构成的二面角。

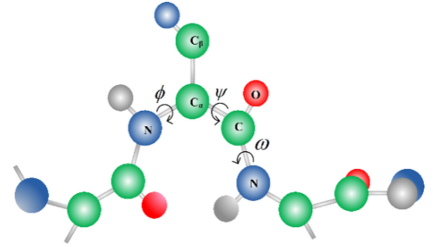


图1 粗粒度蛋白质表达模型

Fig. 1 Coarse-grained protein representation

文献<sup>[35]</sup>给出了蛋白质结构二面角的定义与计算方法。本文设计的二面角相似度模型从蛋白质二面角出发,根据式(1)计算两个蛋白质的二面角相似度,以此来衡量两个蛋白质结构的相似程度。

$$S(p, q) = \sum_{m=1}^l \sqrt{\delta^2(\phi_m^p, \phi_m^q) + \delta^2(\psi_m^p, \psi_m^q) + \delta^2(\omega_m^p, \omega_m^q)} \quad (1)$$

其中, $p$ 和 $q$ 表示两个蛋白质构象; $m$ 为蛋白质残基序号; $l$ 是蛋白质序列的长度; $\phi$ ,  $\psi$ 和 $\omega$ 表示蛋白质结构的3个二面角; $\delta$ 表示两个蛋白质结构中相同位置对应二面角的差异,如式(2)所示:

$$\delta(x, y) = \begin{cases} |x - y|, & \text{if } |x - y| < 180 \\ 360 - |x - y|, & \text{else} \end{cases} \quad (2)$$

其中, $x$ 和 $y$ 分别表示构象 $p$ 和 $q$ 的3个二面角。由式(1)可知, $S(p, q)$ 的值越小,两个蛋白质结构越相似。

### 3.2 DASOM 算法

DASOM基于进化算法框架,以Rosetta粗粒度能量函数为优化目标函数,首先对构象空间进行充分的模态探测,执行算法初步搜索过程,然后基于建立的二面角相似度模型,进一步使用排挤更新策略进行模态增强,完成算法对构象空间的深度搜索。

#### 3.2.1 模态探测阶段

1)初始化种群:输入目标蛋白的序列与片段库(通过Rosetta在线服务器获得),产生初始种群 $pop_{t_1} = \{x_1, x_2, \dots, x_i, \dots, x_{\text{popsize}}\}$ ,并设置种群大小 $\text{popsize}$ 、模态探测阶段算法的迭代次数 $G_1$ 、模态探测阶段的当前迭代次数 $t_1 = 0$ 。

2)对种群 $pop_{t_1}$ 中的个体 $x_i$ 执行变异操作,其中 $i \in \{1, 2, \dots, \text{size}\}$ ,得到变异个体 $x_i'$ 。

3)计算个体 $x_i$ 的能量 $E(x_i)$ 和变异个体 $x_i'$ 的能量 $E(x_i')$ 。

4)若 $E(x_i) \geq E(x_i')$ ,则用变异个体 $x_i'$ 替换个体 $x_i$ ,否则不替换。

5)以种群中的每个个体为目标个体,执行步骤2)一步骤4),得到新一代种群, $t_1 = t_1 + 1$ 。

6)重复执行步骤5),直至达到最大迭代次数 $G_1$ ,模态探测阶段结束。

### 3.2.2 模态增强阶段

1)将模态探测阶段的最后一代种群作为模态增强阶段的初始种群  $pop_{t_2}$ ,模态增强阶段的最大迭代次数为  $G_2$ ,当前迭代次数为  $t_2=0$ 。

2)对种群  $pop$  中的个体  $x_i$  进行变异操作,其中  $i \in \{1, 2, \dots, size\}$ ,得到变异个体  $x_i'$ 。

3)根据式(1)分别计算变异个体  $x_i'$  与种群  $pop_{t_2}$  中每个个体  $x_j$  的二面角相似度  $S(x_i', x_j)$ ,其中  $j \in \{1, 2, 3, \dots, popsize\}$ ,得到  $popsize$  个二面角相似度值。

4)选取二面角相似度值最小的前 10 个构象作为与变异个体  $x_i'$  最相似的前十名个体。最相似的前十名个体按二面角相似度值从小到大依次是  $x_{top1}, x_{top2}, \dots, x_{topn}, \dots, x_{top10}$ ,其中  $b$  为最相似的前十名个体的索引序号,  $n \in \{1, 2, 3, \dots, 10\}$ 。

5)分别计算变异个体  $x_i'$  的能量  $E(x_i')$  和与变异个体  $x_i'$  最相似的前十名个体的能量  $E(x_{top1}), E(x_{top2}), E(x_{topn}), \dots, E(x_{top10})$ ,同时找到种群中能量最高的个体  $x_{worst}$ ,并计算其能量值  $E(x_{worst})$ ,设置  $n=1$ 。

6)若  $E(x_i') \leq E(x_{topn})$ ,则用变异个体  $x_i'$  替换种群中的个体  $x_{topn}$ 。

7)若  $E(x_i') > E(x_{topn})$ ,则产生一个随机数  $rand1$ ,且  $rand1 \in (0, 1)$ ,如果  $rand1 \leq \exp(-(E(x_i') - E(x_{top3}))/KT)$ ,则用变异个体  $x_i'$  替换种群中的个体  $x_{topn}$ ,否则不替换。

8)判断是否替换成功,若替换成功,则转至步骤 10);若替换失败,且  $n < 10$ ,则  $n = n + 1$ ,转至步骤 6)迭代执行,直至替换成功。当  $t_3 \geq 10$  时,转至步骤 9)。

9)若  $E(x_i') \leq E(x_{worst})$ ,则用变异个体  $x_i'$  替换种群中能量最高即最差的个体  $x_{worst}$ ,否则不替换。

10)以种群中每个个体为目标个体,执行步骤 2)一步骤 9),得到新一代种群,  $t_2 = t_2 + 1$ 。

11)重复执行步骤 10),直至达到最大迭代次数  $G_2$ ,模态增强阶段结束。

## 4 算法测试及结果分析

### 4.1 实验环境及测试材料

本文基于 Rosetta 平台,采用 C++ 语言实现 DASOM 算法,选取折叠类型为  $\alpha, \beta, \alpha/\beta$  且序列长度为 64~131 的 10 种蛋白质对 DASOM 算法进行测试,并将其与 Rosetta 算法进行比较。这些测试蛋白的 PDB ID 分别是:1ctf,1opd,1c8cA,1cc5,1kpe,4icb,4uex,1eyv,1lisu,1wit。上述测试蛋白的三维结构已经由实验方法测定,被广泛用于蛋白测试。

DASOM 算法的运行环境为搭载 Inter Core i5 处理器,8GB 运行内存的 64 位 Linux 操作系统。根据经验,算法参数设置如下: $popsize=300, G_1=3L$ ,其中  $L$  为蛋白质序列长度,  $G_2=20, KT=2$ 。DASOM 算法和 Rosetta 算法分别独立运行 20 次,取最优结果为该蛋白质的预测结构。

### 4.2 实验结果与分析

图 2 是蛋白质相似度(Root Mean Square Deviation, RMSD)<sup>[36]</sup>和二面角相似度的计算时间对比图,横坐标为不同测试蛋白,纵坐标为进行 18000000 次结构比对实验所花费的时间。图 2 所示结果表明,采用二面角相似度进行蛋白

质结构比对所花费的时间远远短于采用 RMSD 进行比对所用的时间。因此,二面角相似度模型能够实现蛋白质结构的快速比对,达到降低计算代价、提高搜索效率的目的。

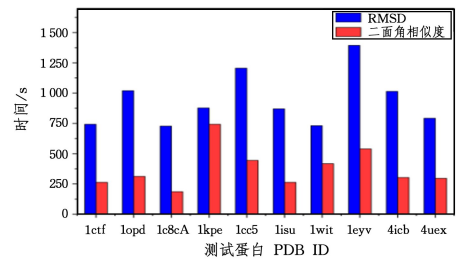


图 2 时间对比(电子版为彩色)

Fig. 2 Comparison of running time

图 3 以 1c8cA 为例给出了算法搜索过程中的种群平均能量趋势。在种群模态探测阶段初期,种群构象的平均能量快速下降,形成了大致的折叠拓扑;在随后的探测过程中,种群平均能量的下降趋势明显放缓,算法在更为广阔的构象空间中进行充分的采样,实现了更多模态的探测。模态探测过程完成后,算法进入模态增强阶段,执行贪婪的搜索过程,种群平均能量急剧下降,算法向构象空间中能量更低的区域进行采样,以找到结构更为合理的构象。从图中可以看出,算法搜索成果与预期一致,能够实现多模态优化过程中的模态探测和增强,采样得到蛋白质近天然态的构象。

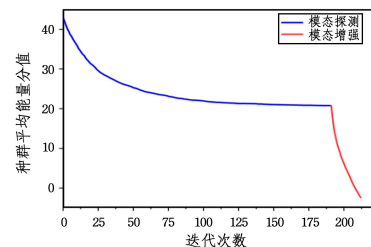


图 3 种群的平均能量趋势(电子版为彩色)

Fig. 3 Mean energy of population

部分蛋白的测试结果如图 4 所示。部分测试蛋白质结构与实验方法测定结构的三维对比效果如图 5 所示,图中绿色表示实验测定结构,蓝色表示采用 DASOM 算法所得到的预测结构。

图 4 清晰地展示了模态探测和模态增强的过程。经过模态探测阶段对构象空间进行充分的搜索之后,种群在构象空间中形成了不同的模态;进入模态增强阶段后,排挤更新策略使得种群中潜在的合理构象被保留下来,种群构象质量得到提高;在最后一代种群中,不仅得到了能量最低的全局最优解,也可以得到一系列局部极值解。在测试蛋白 1c8cA, 1cc5, 4uex, 1opd, 1eyv 和 1kpe 上,算法展示出了多模优势,即各个模态近似均匀地分布在构象空间中能量较低的区域,这也说明蛋白质能量模型的不精确性使数学上能量最低的构象并不一定对应稳定的天然态结构。因此多模态蛋白质构象空间优化算法在蛋白质结构预测领域有着非常重要的意义,它不仅可以得到数学上的最优解,而且可以得到一系列局部极值解,而这些局部极值解很有可能是与天然结构非常相似的构象,从而提高了构象优化算法的可靠性。表 1 列出了所有测试蛋白信息、采用 Rosetta 预测得

到的最优结构 RMSD 值、采用 DASOM 方法预测得到的

最优构象 RMSD 值。

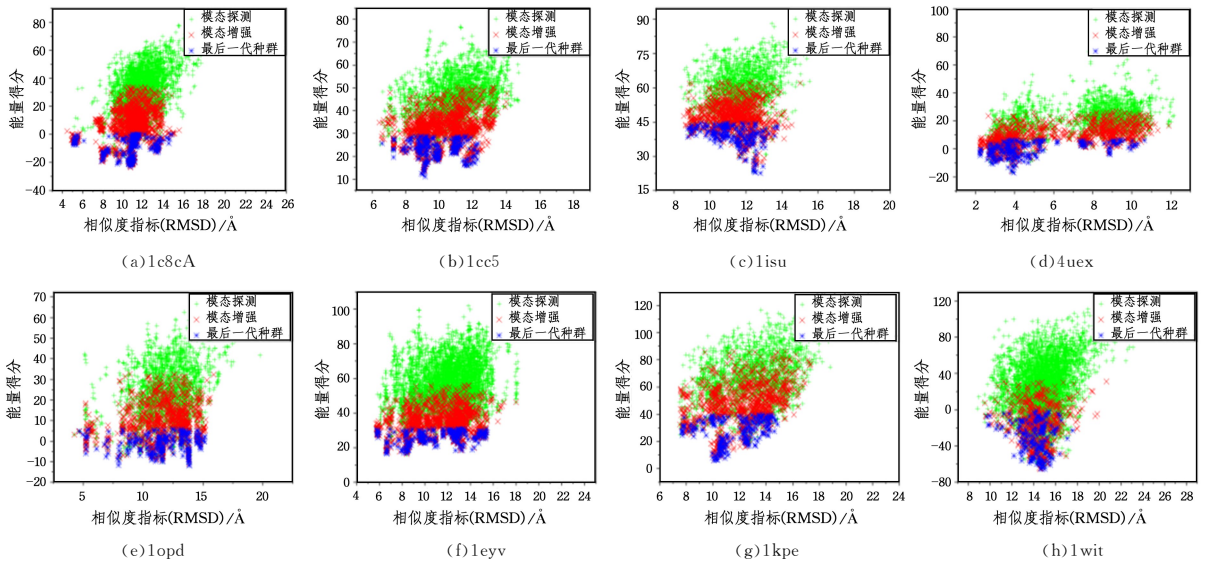


图 4 DASOM 算法的搜索过程图(电子版为彩色)

Fig. 4 Sampling process of DASOM

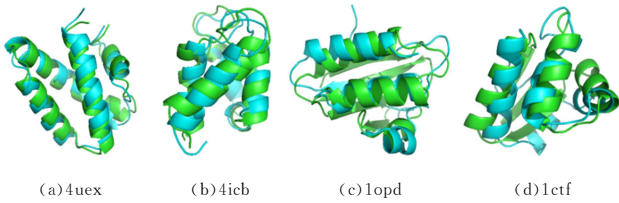


图 5 预测结构与实验方法测定结构的三维对比(电子版为彩色)

Fig. 5 3D similarity comparison between predicted structures and experimental structures

表 1 测试蛋白信息及测试结果

Table 1 Test protein information and test results

PDB	Len	Fold	RMSD <sub>min</sub> /Å	
			DASOM	Rosetta
1isu	62	α	7.66	7.45
4uex	82	α	1.72	2.60
1cc5	83	α	6.58	7.55
4icb	76	α	2.81	3.71
1eyv	131	α	5.76	5.92
1c8cA	64	α/β	5.09	6.12
1ctf	68	α/β	4.30	3.75
1opd	80	α/β	4.69	5.29
1kpe	113	α/β	7.66	8.56
1wit	93	β	9.29	9.90

表 1 的测试结果中, DASOM 对大部分蛋白的实验结果都优于 Rosetta 预测得到的结果, 表明算法能够有效地得到精度更高的预测结果; 而对蛋白 1isu 的预测结果略逊于 Rosetta 算法得到的结果。从图 4 所示的蛋白 1isu 的测试结果图可以发现, 能量高的构象 RMSD 反而高, 违反了热力学原理<sup>[37]</sup>, 即蛋白质的天然结构对应于自由能最小的结构, 导致实验过程中保留的能量低的构象反而与真正的天然态结构差异较大, 使得最终的结果并不理想。

**结束语** 本文提出一种基于二面角相似度的蛋白质构象多模态优化方法, 设计了可以快速进行蛋白质结构比对的二面角相似度模型, 降低了算法的计算代价, 提高了算法的计算效率。提出的基于二面角相似度的蛋白质结构预测多模态优

化算法不仅能得到精度较高的结构, 还能得到一系列亚稳态结构。实验结果表明, DASOM 算法具有良好的搜索性能和预测精度, 是一种有效的蛋白质结构预测方法。

## 参考文献

- [1] 许忠能. 生物信息学[M]. 北京: 清华出版社, 2008.
- [2] LU Z P, HUANG W Q. Local Search Algorithm for Solving Protein Structure Prediction Problem [J]. Computer Science, 2005, 32(11): 148-149.
- [3] KANDATHIL S M, LOVELL S C, HANDL J. Towards a detailed understanding of search trajectories in fragment assembly approaches to protein structure prediction [J]. Proteins-structure Function & Bioinformatics, 2016, 84(4): 411-426.
- [4] DUAN Y, KOLLMAN P A. Pathways to a Protein Folding Intermediate Observed in a 1-Microsecond Simulation in Aqueous Solution [J]. Science, 282(5389): 740-744.
- [5] SCHERAGA H A, KHALILI M, LIWO A. Protein-Folding Dynamics; Overview of Molecular Simulation Techniques [J]. Annual Review of Physical Chemistry, 2007, 58(1): 57-83.
- [6] LINDORFF-LARSEN K, TRBOVIC N, MARAGAKIS P, et al. Structure and Dynamics of an Unfolded Protein Examined by Molecular Dynamics Simulation [J]. Journal of the American Chemical Society, 2012, 134(8): 3787-3791.
- [7] ZHANG Y, KIHARA D, SKOLNICK J. Local energy landscape flattening: Parallel hyperbolic Monte Carlo sampling of protein folding [J]. Proteins, 2002, 48(2): 192-201.
- [8] SHEN Y M, PICORD G, GUYON F, et al. Detecting Protein Candidate Fragments Using a Structural Alphabet Profile Comparison Approach [J]. PLoSONE, 2013, 8(11): e80493.
- [9] XU D, ZHANG Y. Toward optimal fragment generations for ab-initio protein structure assembly [J]. Proteins: Structure, Function, and Bioinformatics, 2013, 81(2): 229-239.
- [10] TANTAR A A, MELAB N, TALBI E G, et al. A parallel hybrid genetic algorithm for protein structure prediction on the computational grid [J]. Future Generation Computer Systems, 2007,

23(3):398-409.

- [11] HOQUE M T, CHETTY M, LEWIS A, et al. Twin Removal in Genetic Algorithms for Protein Structure Prediction Using Low-Resolution Model [J]. *IEEE/ACM transactions on computational biology and bioinformatics*, 2011, 8(1):234-245.
- [12] ISLAM M K, CHETTY M. Clustered Memetic Algorithm With Local Heuristics for Ab Initio Protein Structure Prediction [J]. *IEEE Transactions on Evolutionary Computation*, 2013, 17(4):558-576.
- [13] CUSTÓDIO, FÁBIO L, BARBOSA H J C, et al. A multiple minima genetic algorithm for protein structure prediction [J]. *Applied Soft Computing*, 2014, 15:88-99.
- [14] LEE J, SCHERAGA H A, RACKOVSKY S. New optimization method for conformational energy calculations on polypeptides: Conformational space annealing [J]. *Journal of Computational Chemistry*, 2015, 18(9):1222-1232.
- [15] SHEHU A. An Ab-initio tree-based exploration to enhance sampling of low-energy protein conformations [C]// *Robotics: Science and Systems*. 2009:241-248.
- [16] SHEHU A, OLSON B. Guiding the search for native-like protein conformations with an Ab-initio tree-based exploration [J]. *Robotics Research*, 2010, 29(8):1106-1127.
- [17] OLSON B, MOLLOY K, SHEHU A. In search of the protein native state with a probabilistic sampling approach [J]. *Journal of Bioinformatics and Computational Biology*, 2011, 9(3):383-398.
- [18] OLSON B, SHEHU A. Evolutionary-inspired probabilistic search for enhancing sampling of local minima in the protein energy surface [J]. *Proteome Science*, 2012, 10(Suppl 1):S5.
- [19] MOLLOY K, SALEH S, SHEHU A. Probabilistic search and energy guidance for biased decoy sampling in Abinitio protein structure prediction [J]. *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)*, 2013, 10(5):1162-1175.
- [20] CHENG Z H, ZHANG G J, DENG Y Y, et al. Novel Multimodal Optimization Algorithm for Protein Structure Prediction [J]. *Computer Science*, 2012, 40(9):212-215.
- [21] JOOYOUNG L, HAROLD A S, SHALOM R. New optimization method for conformational energy calculations on polypeptides: conformational space annealing [J]. *Journal of Computational Chemistry*, 1997, 18(9):1222-1232.
- [22] WONG K C, LEUNG K S, WONG M H. Protein structure prediction on a lattice model via multimodal optimization techniques [C]// *Genetic and Evolutionary Computation Conference*. Portland, USA: ACM, 2010:155-162.
- [23] HAO X H, ZHANG G J, GEN X. Conformational Space Sampling Method Using Multi-Subpopulation Differential Evolution for De novo Protein Structure Prediction [J]. *IEEE Transactions on Nano Bioscience*, 2017, PP(99):1-1.
- [24] 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.
- [25] LEE J, KIM S Y, JOO K, et al. Prediction of protein tertiary structure using PROFESY, a novel method based on fragment assembly and conformational space annealing [J]. *Proteins: Structure Function and Bioinformatics*, 2004, 56(4):704-714.
- [26] LEE J, LEE J, SASAKI T N, et al. De novo protein structure prediction by dynamic fragment assembly and conformational space annealing [J]. *Proteins: Structure Function and Bioinformatics*, 2011, 79(8):2403-17.
- [27] JOO K, LEE J, SIM S. Protein structure modeling for CASP10 by multiple layers of global optimization [J]. *Proteins: Structure Function and Bioinformatics*, 2014, 82(S2):188-195.
- [28] Clausen R, Shehu A. A multiscale hybrid evolutionary algorithm to obtain sample-based representations of multi-basin protein energy landscapes [C]// *ACM-BCB*. 2014:269-278.
- [29] OLSON B, SHEHU A. Multi-objective optimization techniques for conformational sampling in template-free protein structure prediction [C]// *BiCoB*. 2014:143-148.
- [30] CLAUSEN R, SHEHU A. A Data-driven evolutionary algorithm for mapping multibasin protein energy landscapes [J]. *Journal of Computational Biology*, 2015, 22(9):844-860.
- [31] GARZA-FABRE M, KANDATHIL S M, HANDL J, et al. Generating, maintaining, and exploiting diversity in a memetic algorithm for protein structure prediction [J]. *Evolutionary Computation*, 2016, 24(4):577-607.
- [32] SALEH S, OLSON B, SHEHU A. A population-based evolutionary search approach to the multiple minima problem in de novo protein structure prediction [J]. *BMC Structural Biology*, 2013, 13(1):1-28.
- [33] KORTEMME T, MOROZOV A V, BAKER D. An Orientation-dependent Hydrogen Bonding Potential Improves Prediction of Specificity and Structure for Proteins and Protein-Protein Complexes [J]. *Journal of Molecular Biology*, 2003, 326(4):1239-1259.
- [34] HANDL J, KNOWLES J, BAKER D, et al. The dual role of fragments in fragment-assembly methods for de novo protein structure prediction [J]. *Proteins: Structure, Function, and Bioinformatics*, 2012, 80(2):490-504.
- [35] NEUMAIER A. Molecular Modeling of Proteins and Mathematical Prediction of Protein Structure [J]. *SIAM Review*, 1997, 39(3):407-460.
- [36] KOEHL P. Protein structure similarities [J]. *Current Opinion in Structural Biology*, 2001, 11(3):348-353.
- [37] ANFINSEN C. Principles that govern the folding of protein chains [J]. *Science*, 1973, 181(96):223-230.



**LI Zhang-wei**, born in 1976, Ph. D, associate professor, is a member of China Computer Federation. His main research interests include intelligent information processing.



**Zhang Gui-jun**, born in 1974, Ph. D, professor, is a member of China Computer Federation. His main research interests include intelligent information processing, optimization theory and algorithm design and bioinformatics.