

蛋白质结构从头预测多级个体筛选进化算法

李章维 郝小虎 张贵军

(浙江工业大学信息工程学院 杭州 310023)

摘要 针对蛋白质高维构象空间采样多样性问题,文中提出了一种蛋白质结构从头预测多级个体筛选进化算法(MIISEA)。基于进化算法框架,首先采用基于知识的 Rosetta 粗粒度能量模型作为优化目标函数,以降低构象空间优化变量维数;其次以基于 9 片段和 3 片段的片段组装技术为不同的变异策略,增加同代种群的多样性;同时,设计多级个体筛选方法,进一步增加不同代种群间的多样性;然后利用 Monte Carlo 算法较强的局部搜索能力对每个个体做局部增强,以得到当前的局部最优解;最后,得到全局最优解以及不同的局部最优解。10 个目标蛋白的测试结果表明,所提方法能够有效提高采样多样性,得到 TMscore 大于 0.5 的预测构象,为进一步做结构修饰提供便利。

关键词 从头预测,进化算法, Monte Carlo, 片段组装, TMscore

中图分类号 TP301.6 **文献标识码** A

Multi-layer Screening Based Evolution Algorithm for De Novo Protein Structure Prediction

LI Zhang-wei HAO Xiao-hu ZHANG Gui-jun

(College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023, China)

Abstract Aiming at the diversity of sampling in high-dimensional protein conformational space, a multi-layer screening based evolution algorithm for de novo protein structure prediction (MIISEA), was proposed. On the basis of the evolution algorithm framework, the knowledge-based Rosetta coarse-grained energy model is employed as the objective function, to reduce the optimal variable dimension of protein conformational space. Taking 9-mer and 3-mer fragment assembly technique as two different kinds of mutation strategies, the diversity of the individuals in the same generation can be increased. In conjunction, multi-layer individual screening method is designed for further improving the diversity of the individuals in different generations. Then, Monte Carlo algorithm is adopted to enhance the performance for each individual to get the local optimal solution. Finally, the global resolution and different local solutions can be obtained. Test results of 10 target proteins show that the proposed method can effectively improve the diversity of sampling, the prediction conformations with TMscore greater than 0.5 can be obtained for further refinement.

Keywords De novo, Evolution algorithm, Monte Carlo, Fragment assembly, TMscore

蛋白质分子在生物的生命过程中扮演着极其重要的角色,对蛋白质分子三级结构的预测是计算生物学领域中一个非常重要的研究课题。蛋白质三级结构预测主要依据 Anfinsen 原则^[1],以计算机为工具,设计有效的算法,在广阔的构象空间进行搜索,执行能量极小化过程获得蛋白质三维预测结构。

分子动力学模拟(MD)^[2-4]、蒙特卡洛方法(MC)^[5-7]、构象空间退火(CSA)^[8-10]、副本交换方法(REM)^[11-14]等,是比较经典的蛋白质结构预测方法。近年来,遗传算法(GA)^[15-18]、差分进化算法(DE)^[19-21]等进化类算法在蛋白质结构预测中也有较为广泛的应用。

Zhang 团队在差分进化算法框架下,将抽象凸理论结合到蛋白质结构预测问题中,提出了基于抽象凸下界估计和利普希茨下界估计的蛋白质结构从头预测方法 ACUE^[22]、

LUE^[23],其对构象空间进行隐式分割得到分段线性的抽象凸下界估计,以快速排除无效搜索区域,提高构象空间搜索效率;其同样基于差分进化算法框架,将距离谱作为约束条件引入算法选择环节,提出了基于距离谱的蛋白质结构差分进化算法 DPDE^[24],能够有效地平衡能量函数不精确的问题;由于蛋白质结构预测问题本身的多模特性,其在差分进化算法的框架下,提出了多模态蛋白质差分进化算法 MDE^[25],通过控制搜索过程中各个模态的分布达到增加种群多样性的目的。董辉等在差分进化算法的框架下,引入 Monte Carlo 方法做局部增强,提出了蛋白质构象空间局部增强差分进化搜索方法 LEDE^[26],并用 5 个测试蛋白验证了其有效性;李章维等将副本交换的思想引入差分进化算法框架,提出了基于副本交换的局部增强差分进化蛋白质结构从头预测方法 RL-DE^[27]。Shehu 研究小组在进化算法框架下,采用粗粒度蛋白

本文受国家自然科学基金(61075062,61379020)资助。

李章维(1967—),男,博士,副教授,CCF 会员,主要研究方向为智能信息处理, E-mail: lzw@zjut.edu.cn;郝小虎(1990—),男,博士生,主要研究方向为智能信息处理、生物信息学;张贵军(1974—),男,博士,教授,主要研究方向为智能信息处理、全局优化理论及算法设计、生物信息学, E-mail: zgj@zjut.edu.cn(通信作者)。

质表达模型和片段组装技术,并结合局部搜索策略,提出了混合进化算法 HEA^[28];将局部-全局搜索策略结合到基于群体的进化算法中,提出了多目标进化算法 MOEA^[29];同样在进化算法框架下,将 Pareto 排名和 Pareto 计数引入 HEA,又提出了 MOEA-PC^[30]。

由此可见,采样问题仍然是蛋白质结构预测中一个需要解决的问题,而增加采样的多样性是其中一个行而有效的途径。因此,针对构象空间采样多样性问题,本文提出一种蛋白质结构从头预测多级个体筛选进化算法(MIIEA)。该算法基于进化算法框架,首先采用基于知识的 Rosetta 粗粒度能量模型作为优化目标函数,降低构象空间优化变量维数;其次,以基于 9 片段和 3 片段的片段组装技术为不同的变异策略,增加同代种群多样性;同时,设计多级个体筛选方法,进一步增加不同代种群间的多样性;然后,利用 Monte Carlo 算法较强的局部搜索能力对每个个体做局部增强,以得到当前局部最优解;最终,通过多样性保持进化,得到全局最优解以及不同的局部最优解。针对 10 个小目标蛋白的测试结果表明,所提方法能够有效提高采样多样性,得到 TMscore 大于 0.5 的预测构象,为进一步做结构修饰提供便利。

1 理论分析

1.1 进化算法

进化算法(Evolution Algorithm)借鉴了大自然中生物的进化操作,包括基因编码、种群初始化、交叉变异算子、精英保留机制等基本操作。进化计算是一种成熟的具有高鲁棒性和广泛适用性的全局优化方法,具有自组织、自适应、自学习的特性,能够不受问题性质的限制,有效地处理传统优化算法难以解决的复杂问题。进化规则把编码结构抽象为种群之间的相似,而将进化策略抽象为个体之间的相似。进化策略和进化规则已应用于连续优化、模式识别、机器学习、神经网络训练、系统辨识和智能控制等众多领域^[31]。

进化算法在蛋白质结构预测领域也有广泛的应用,如 Zhang 研究团队提出的 ACUE^[22], LUE^[23], DPDE^[24], MDE^[25],以及 LEDE^[26], RLDE^[27]等均以差分进化算法为计算框架;Shehu 研究小组的 HEA^[28], MOEA^[29], MOEA-PC^[30]也以进化算法为计算框架;Garza-Fabre 则以莫因算法为计算框架,提出了基于多样性采样的蛋白质结构预测方法^[32]。

1.2 粗粒度能量模型

蛋白质折叠问题本质上是典型的多尺度问题^[33]。自从 Levitt 及 Warshel 在 1975 年建立蛋白质分子粗粒度能量模型以来,陆续有研究者建立了一系列从粗粒度到全原子尺度的多尺度能量模型^[34-35],其利用粗粒度能量模型快速优化得到“诱饵构象”,然后进一步基于精度更高的全原子能量模型对其进行修正,得到精度更高的近天然态构象。

依据粗粒度蛋白质表达模型,MIIEA 采用了基于知识的 Rosetta Score3 粗粒度能量模型,在不丢失氨基酸序列重要结构信息的前提下只保留 N, C, C α , O 这些主链骨干原子及侧链替代原子,用一系列设置为 -120° 到 120° 的二面角 $\psi(N-$

$C\alpha)$, $\psi(C\alpha-C)$ 表示氨基酸链,有效减小了计算空间的复杂度^[36],降低了构象空间优化变量维数。Rosetta Score3 粗粒度能量模型是 10 种不同能量项独立加权的线性和^[37-38],如式(1)所示,它是一种基于知识的能量函数,它隐式地体现了形成蛋白质天然结构的内在作用,计算成本较低,而且非常有效^[39]。

$$E_{\text{protein}} = W_{\text{repulsion}} E_{\text{repulsion}} + W_{\text{attraction}} E_{\text{attraction}} + W_{\text{solvation}} E_{\text{solvation}} + W_{\text{bb-sc hb}} E_{\text{bb-sc hb}} + W_{\text{bb-bb hb}} E_{\text{bb-bb hb}} + W_{\text{sc-sc hb}} E_{\text{sc-sc hb}} + W_{\text{pair}} E_{\text{pair}} + W_{\text{dunbrack}} E_{\text{dunbrack}} + W_{\text{rama}} E_{\text{rama}} + W_{\text{reference}} E_{\text{reference}} \quad (1)$$

1.3 片段组装技术

由于蛋白质构象空间的高维特性、模型的多尺度特性及不精确性,片段组装技术的应用成为蛋白质结构从头预测的一种重要手段。片段组装技术首先将预测序列划分成若干连续区段,然后通过序列比对寻找局部拟合已知蛋白质结构的片段,与优化目标蛋白质的指定片段进行替换,即 3 种二面角 ϕ, ψ, ω 的替换,组装成完整的目标结构。片段组装主要包括以下步骤:1)选择开始插入的位置 i ;2)从片段库中选择相应位置的片段进行组装;3)从选定的片段中选择插入长度 L ,替换目标序列中的相应片段。

通过片段组装可以极大地减小搜索空间,能够显著提高计算速度;另一方面,由于蛋白质三维结构具有一定的层次性和规律性,许多序列同源性很低的远亲蛋白质也存在相似的结构片段折叠模式,因此,通过片段组装可以构建更为合理的蛋白质三维结构。在现有技术水平的条件下,采用单纯的优化技术,目前只能得到长度为 5~20 左右的多肽链公认稳定构象;而引入片段组装,则可以得到序列长度为 150 的高精度预测结构^[7]。

2 MIIEA 算法

MIIEA 基于进化算法框架,采用基于知识的 Rosetta 粗粒度能量模型作为优化目标函数,降低了构象空间优化变量维数;以基于 9 片段和 3 片段的片段组装技术为不同的变异策略,增加同代种群多样性;多级个体筛选方法的应用,可以进一步增加不同代种群间的多样性;Monte Carlo 用于对每个个体做局部增强,可以得到当前局部最优解。

2.1 多级个体筛选 EA

种群个体多样性是保证进化算法有效性的重要因素。因此,为了提高进化算法中种群个体间的多样性以及不同进化代数种群间的多样性,本文设计了多级个体筛选方法。

第一级筛选在种群进化的每一代进行:对每个目标个体分别以 9 片段和 3 片段进行片段组装,产生两个不同的变异个体;这两个变异个体与当前目标个体进行两两交叉,生成至多 6 个新的测试个体,与目标个体、变异个体组成种群规模至多为 9 的子种群;在当前子种群中挑选出 3 个代表个体,进入候选种群,它们分别是当前能量最低个体、能量次最低个体以及能量值居中的个体。以此,在保证最低能量个体能够保留的前提下,同时保留能量略高的个体,用于平衡能量函数不精确的问题。

第二级筛选在每一代进化结束时进行:第一级筛选后组成的候选种群中,个体数目为初始种群的3倍,为保持种群数目恒定,并且在能量选择压力下尽可能增加种群个体多样性,从候选种群中选择能量最低的前2/3的个体,以及1/3能量值居中的个体进入下一代种群,执行进化过程。

通过两级个体筛选,在以能量为主的选择压力下,保留大多数能量值最低的个体,同时为了平衡能量函数不精确的问题,一部分能量值较高的个体也被保留了下来,它们作为潜在的优秀个体,有利于在构象空间中搜索得到结构更优的近天然态构象。

2.2 MIIEA 算法

MIIEA 算法采用伪代码形式描述,如算法1所示。其中 $popSize$ 为种群大小, $generation$ 为算法迭代次数, $seqLen$ 为序列长度, P 表示种群, p 表示个体。

算法1 MIIEA 算法

输入:蛋白质序列信息

输出:预测蛋白质三维结构 PDB

```

1. 初始化,对伸展氨基酸链进行 seqLen 次 9 片段组装和 3 片段组装,
   生成 popSize 个个体,组成初始种群  $P_{init}$ 
2. for g in range(0, generation):
3.   for i in range(0, popSize):
4.      $P_{target} \leftarrow P_i$ 
5.      $P_{mutation9} \leftarrow P_{target}$  (9 片段组装)
6.      $P_{mutation3} \leftarrow P_{target}$  (3 片段组装)
7.      $P_{cross_{t9}} \leftarrow P_{target} \cdot P_{mutation9}$  交叉
8.      $P_{cross_{t3}} \leftarrow P_{target} \cdot P_{mutation3}$  交叉
9.      $P_{cross_{9t}} \leftarrow P_{mutation9} \cdot P_{target}$  交叉
10.     $P_{cross_{3t}} \leftarrow P_{mutation3} \cdot P_{target}$  交叉
11.     $P_{sub} \leftarrow P_{target} \cdot P_{mutation9} \cdot P_{mutation3} \cdot P_{cross_{t9}} \cdot P_{cross_{t3}} \cdot P_{cross_{9t}} \cdot P_{cross_{93}} \cdot P_{cross_{39}} \cdot P_{cross_{3t}}$ 
12.     $P_{lowest1}, P_{lowest2}, P_{mid} \leftarrow P_{sub}$ 
13.     $P_{lowest1\_MC} \leftarrow P_{lowest1}$  (Monte Carlo)
14.     $P_{lowest2\_MC} \leftarrow P_{lowest2}$  (Monte Carlo)
15.     $P_{mid\_MC} \leftarrow P_{mid}$  (Monte Carlo)
16.     $P_{candidate} \leftarrow P_{lowest1\_MC} \cdot P_{lowest2\_MC} \cdot P_{mid\_MC}$ 
17.   end for
18.  $P_{lowest_{2/3}}, P_{mid_{1/3}} \leftarrow P_{candidate}$ 
19.  $P_{init} \leftarrow P_{lowest_{2/3}}, P_{mid_{1/3}}$ 
20. end for
21. end

```

3 实验和结果

3.1 实验环境及测试材料准备

MIIEA 基于 Rosetta 平台,采用 Python 语言实现,选取折叠类型为 α 、 α/β ,序列长度从 33 到 115 的 10 种小蛋白进行测试,并与 Rosetta Ab-initio Protocol (Rosetta) 进行比较。测试蛋白分别为:2MRF,1VII,1GYZ,2MU2,2EZK,这 5 个测

试蛋白选自参考文献,其余 5 个测试蛋白 T0773-D1, T0792-D1, T0816-D1, T0759-D2, T0855-D1 是 CASP11 竞赛的目标蛋白,其中 T0773-D1, T0792-D1 分类为 Easy Target, T0816-D1, T0759-D2, T0855-D1 分类为 Hard Target。它们的实验测定结构可以从 PDB 库¹⁾下载得到。

选取这些蛋白质做测试是因为它们在生物体内有着重要的作用,并且它们的三级结构已经由实验测定,被广泛用作算法测试。相关测试蛋白信息如表 1 所列。

表 1 测试蛋白信息及测试结果

No.	PDB ID	Len	Fold	Max TMscore	
				MIIEA	Rosetta
1	2MRF	33	α/β	0.54	0.53
2	1VII	36	α	0.69	0.68
3	1GYZ	60	α	0.74	0.72
4	2MU2	71	α	0.69	0.70
5	2EZK	99	α	0.70	0.69
6	T0773-D1	67	α/β	0.73	0.71
7	T0792-D1	78	α/β	0.51	0.50
8	T0816-D1	68	α	0.66	0.59
9	T0759-D2	62	α/β	0.70	0.69
10	T0855-D1	115	α/β	0.42	0.34

执行片段组装的片段库通过序列比对的方法获得:基于 PISCES 服务器²⁾,以分辨率小于 2\AA ,同源性小于 30% 为参数,从筛选得到的序列中进行比对搜索,在查询序列的每个片段位置产生 200 个得分最高的片段,将这些片段的二面角等信息记录下来构成片段库文件。在执行片段组装时,片段库中的片段将被随机选择,其 ϕ, ψ, ω 3 个二面角将替换目标序列中相应位置的二面角,从而得到新的构象。

MIIEA 运行环境为搭载 Inter Core i7 处理器,16GB 运行内存的 64 位 Ubuntu 操作系统。算法参数设置: $popSize = 30$, $generation = 10000$,算法独立运行 30 次。

3.2 实验结果

MIIEA 与 Rosetta 在所有测试蛋白上的测试结果如图 1 所示,横坐标为 MIIEA 算法得到的 TMscore 值,纵坐标表示 Rosetta 得到的 TMscore 值,图中点在对角线下方则表示 MIIEA 得到的结果优于 Rosetta,在对角线上方则表示 Rosetta 优于 MIIEA。

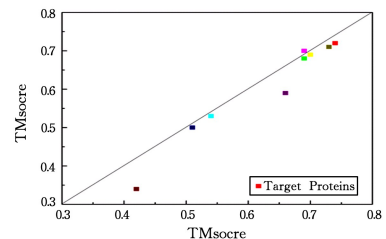


图 1 测试结果对比图

部分蛋白的测试结果如图 2 所示,图中横坐标为预测结构和实验方法测定结构的相似度指标 TMscore;纵坐标为计算得到构象的能量分值,红色的点表示 RLDE 搜索过程中产生的所有构象,蓝色的点表示采用 Rosetta 协议计算产生的构象, Rosetta 根据文献[40]实现。部分蛋白质预测结构和实验方法测定结构的三维相似性对比如图 3

¹⁾ <http://www.rcsb.org>

²⁾ <http://dunbrack.fccc.edu/PISCES.php>

所示,图中黄色、红色和蓝色的结构分别表示实验室测定结构、采用 MIIEA 算法所得到的预测结构、采用 Rosetta 得到的预测结构。测试蛋白信息及由不同算法计算得到的

预测结构和实验方法测定的结构最大相似度指标 TMscore 值如表 1 所列。其中结果较好(TMscore 值大)的用加粗字体表示。

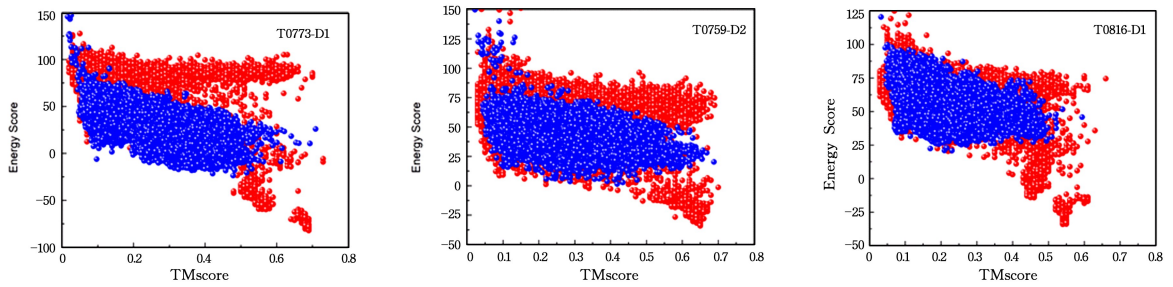


图 2 算法搜索过程示意图(电子版为彩色)

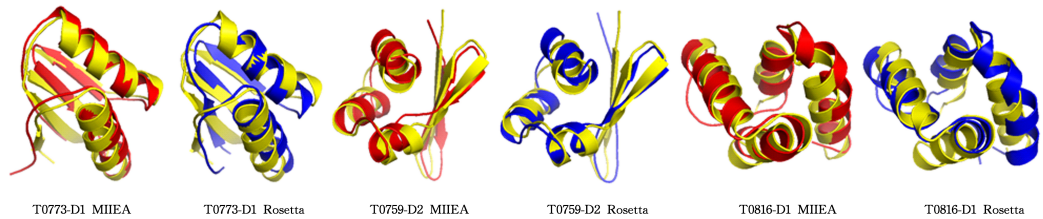


图 3 预测结构和实验方法测定结构的三维相似性对比图(电子版为彩色)

4 分析与讨论

实验结果中,MIIEA 算法预测得到最高精度的蛋白质是 1GYZ, TMscore 值为 0.74, 在所有测试蛋白上平均达到了 0.64, 且其中 9 个测试蛋白的 TMscore 值均在 0.5 以上。图 1 直观展示了 MIIEA 算法与 Rosetta Ab-initio Protocol 方法的预测结果, 其中 9 个测试蛋白均的预测结果均表明 MIIEA 优于 Rosetta, 其余一个测试结果表明二者相当, 并没有明显的优势。图 2 展示了两种算法在构象空间中采样的过程, 其中红色的点表示 MIIEA 算法采样得到的构象, 蓝色的点表示 Rosetta 采样得到的构象。显然, MIIEA 算法较 Rosetta 算法有更好的全局和局部搜索能力: MIIEA 算法能够在更为广阔的构象空间中采样并且逐步趋向能量更低的区域, 而 Rosetta 算法的搜索区域明显较小且对低能量区域的采样明显不足。图 3 清晰地展示了采用 MIIEA 算法和 Rosetta 算法得到的蛋白质三维结构与实验室测定的结构之间的相似程度, 两者重叠部分越多, 则表明越相似。从图 3 可以看出, 采用 MIIEA 算法预测得到的三维结果(红色)与实验测定的三维结构(黄色)之间的重叠更为紧密, 相似程度更高。

MIIEA 算法是针对采样多样性提出的一种蛋白质结构预测方法, 因此, 本文针对多样性问题进行分析: MIIEA 采用了不同以往的子代选择策略, 较好地控制了进化算法的贪婪特性, 通过双层排序筛选, 使得进入子代的个体之间保持了一定的差异性, 这也是采样多样性的保证和来源。从图 2 搜索过程示意图中可以清晰地看到, MIIEA 算法从起始构象开始, 逐渐向能量较低的区域搜索, 同时在能量相对较高的区域仍然保留搜索个体, 对构象空间进行充分的采样。以 T0773-D1 为例, 在 TMscore 值为 0.6 到 0.7 之间的区域中, 构象在能量约为 -80 到 100 之间均有分布, 且构象数目较多, 与之相反, Rosetta 在 TMscore 值为 0.6 到 0.7 之间的区域中, 构象能量分布仅仅从 0 到 40。其他的测试蛋白均有相似的搜索过程, 这充分说明成功达到了 MIIEA 算法设计的目的, 有

效地增加了采样的多样性, 能够预测得到较好的结构。但是, 这也同时说明了一个问题, 能量函数仍然不够精确, 以至于能量高的区域仍然存在较好的结构, 而能量最低的地方, 结构却并不一定对应天然态构象。

综上所述, MIIEA 算法在所有测试蛋白上整体预测结果的精度较高。在大多数测试蛋白上都可以预测得到结构较好的个体(TMscore > 0.5)。但是对于较大的蛋白, MIIEA 和 Rosetta 均未成功预测出较好的结构, 如测试蛋白 T0855-D1, 只得到了 TMscore 为 0.4 左右的结构。可能造成这种结果的因素有很多: 首先, 随着蛋白序列长度的增加, 原子间相互作用更为复杂, 且折叠类型为 β 和 loop 的蛋白质原子间作用力多为远程作用力, 这可能导致力场模型不够精确, 使得能量函数不够精确; 其次, 随着蛋白序列长度的增加, 构象空间能量曲面将变得更为粗糙和复杂, 这将使得算法在搜索过程中陷入局部极小值而不能进一步在更为广阔的空间搜索; 最后, 片段库的质量问题也可能导致不能得到结构较好的个体。

结束语 本文针对构象空间采样多样性的问题, 提出了一种蛋白质结构从头预测多级个体筛选进化算法(MIIEA)。基于进化算法框架, 首先采用基于知识的 Rosetta 粗粒度能量模型作为优化目标函数, 降低构象空间优化变量维数; 其次, 以基于 9 片段和 3 片段的片段组装技术为不同的变异策略, 增加同代种群多样性; 同时, 设计多级个体筛选方法, 进一步增加不同代种群间的多样性; 然后, 利用 Monte Carlo 算法较强的局部搜索能力对每个个体做局部增强, 以得到当前局部最优解; 最终, 通过多样性保持进化, MIIEA 将得到全局最优解以及不同的局部最优解。针对 10 个小目标蛋白测试结果表明, 所提方法能够有效提高采样多样性, 得到 TMscore 大于 0.5 的预测构象, 为进一步做结构修饰提供便利。

参考文献

- [1] ANFINSEN C B. Principles that govern the folding of protein chains [J]. Science, 1973, 181(96): 223-230.

- [2] DUAN Y, KOLLMAN P A. Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution [J]. *Science*, 1998, 282(5389): 740-744.
- [3] SCHERAGA H A, KHALILI, LIWO A. Protein-folding dynamics; overview of molecular simulation techniques [J]. *Annual Review of Physical Chemistry*, 2007, 58(1): 57-83.
- [4] 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.
- [5] ZHANG Y, KIHARA D, SKOLNICK J. Local energy landscape flattening: Parallel hyperbolic Monte Carlo sampling of protein folding [J]. *Proteins: Structure, Function and Bioinformatics*, 2002, 48(2): 192-201.
- [6] SHEN Y, PICORD G, GUYON F, et al. Detecting protein candidate fragments using a structural alphabet profile comparison approach [J]. *PLoS One*, 2013, 8(11): e80493.
- [7] 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.
- [8] DOTU I, CEBRIA M, VAN H P, et al. On Lattice Protein Structure Prediction Revisited [J]. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2011, 8(6): 1620-1632.
- [9] TYKA M D, JUNG K, BAKER D. Efficient sampling of protein conformational space using fast loop building and batch minimization on highly parallel computers [J]. *Journal of Computational Chemistry*, 2012, 33(31): 2483-2491.
- [10] JOO K, LEE J, SIM S, et al. Protein structure modeling for CASP10 by multiple layers of global optimization [J]. *Proteins: Structure Function and Bioinformatics*, 2014, 82(S2): 188-195.
- [11] SUGITA Y, OKAMOTO Y. Replica-exchange molecular dynamics method for protein folding [J]. *Chemical Physics Letters*, 1999, 314(1-2): 141-151.
- [12] SUGITA Y, OKAMOTO Y. Replica-exchange multicanonical algorithm and multicanonical replica-exchange method for simulating systems with rough energy landscape [J]. *Chemical Physics Letters*, 2000, 329(3): 261-270.
- [13] CZAPLEWSKI C, KALINOWSKI S, LIWO A, et al. Application of Multiplexed Replica Exchange Molecular Dynamics to the UNRES Force Field: Tests with alpha and alpha+beta Proteins [J]. *Journal of chemical theory and computation*, 2009, 3(5): 627-640.
- [14] HANSMANN U H E. Parallel tempering algorithm for conformational studies of biological molecules [J]. *Chemical Physics Letters*, 1997, 281(1): 140-150.
- [15] TANTA 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.
- [16] 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.
- [17] 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): 58-576.
- [18] CUSTÓDIO F L, BARBOSA H J C, DARDENNE L E. A multiple minima genetic algorithm for protein structure prediction [J]. *Applied Soft Computing*, 2014, 15(2): 88-99.
- [19] STORN R, PRICE K. Differential Evolution-A Simple and Efficient Heuristic for global Optimization over Continuous Spaces [J]. *Journal of global optimization*, 1997, 11(4): 341-359.
- [20] ZOU D X, WU J H, GAO L Q, et al. A modified differential evolution algorithm for unconstrained optimization problems [J]. *Neurocomputing*, 2013, 120(11): 469-481.
- [21] CASCIATI, SARA. Differential evolution approach to reliability-oriented optimal design [J]. *Probabilistic Engineering Mechanics*, 2014, 36(4): 72-80.
- [22] HAO X H, ZHANG G J, ZHOU X G, et al. A novel method using abstract convex underestimation in ab-initio protein structure prediction for guiding search in conformational feature space [J]. *IEEE/ACM Transaction Computer Biol. Bioinf.*, 2016, 13(5): 887-900.
- [23] HAO X H, ZHANG G J, ZHOU X G. Guiding exploration in conformational feature space with Lipschitz underestimation for ab-initio protein structure prediction [J]. *Computational Biology and Chemistry*, 2018, 73: 105-119.
- [24] ZHANG G J, ZHOU X G, YU X F, et al. Enhancing protein conformational space sampling using distance profile-guided differential evolution [J]. *IEEE/ACM Transaction Computer Biol. Bioinf.*, 2017, 14(6): 1288-1301.
- [25] HAO X H, ZHANG G J, ZHOU X G. Conformational Space Sampling Method Using Multi-Subpopulation Differential Evolution for De novo Protein Structure Prediction [J]. *IEEE Transaction Nano Bioscience*, 2017, 16(7): 618-633.
- [26] 董辉, 郝小虎, 张贵军. 蛋白质构象空间局部增强差分进化搜索方法 [J]. *计算机科学*, 2015, 42(11A): 22-26.
- [27] 李章维, 郝小虎, 张贵军. 基于副本交换的局部增强差分进化蛋白质结构从头预测方法 [J]. *计算机科学*, 2017, 44(5): 211-217.
- [28] CLAUSEN R, SHEHU A. A multiscale hybrid evolutionary algorithm to obtain sample-based representations of multi-basin protein energy landscapes [C] // *ACM-BCB*, Newport Beach, CA, USA, 2014: 269-278.
- [29] OLSON B, DE JONG K, SHEHU A. Off-lattice protein structure prediction with homologous crossover [C] // *15th Annu. Conference Genet. Evol. Computer*. 2013: 287-294.
- [30] OLSON B, SHEHU A. Multi-objective optimization techniques for conformational sampling in template-free protein structure prediction [C] // *6th Int. Conf. Bioinf. Computer. Biol.* 2014: 143-148.
- [31] DAS S, SUGANTHAN P N. Differential evolution: a survey of the state-of-the-art [J]. *IEEE Transactions on Evolutionary Computation*, 2011, 15: 4-31.
- [32] 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.
- [33] KMIECIEK S, JAMROZ M, KOLINSKI A. Multiscale Approaches to Protein Modeling [M]. Berlin: Springer Science, 2011: 281-293.