

蛋白质折叠的计算机模拟研究进展

苏计国¹, 王伟卜¹, 陈慰祖², 王存新²

(1. 燕山大学理学院河北省微结构材料物理重点实验室, 秦皇岛 066004;

2. 北京工业大学生命科学与生物工程学院, 北京 100024)

摘要: 蛋白质折叠是结构生物学领域有待解决的重要科学问题, 计算机模拟方法是研究蛋白质折叠的主要手段之一. 对蛋白质折叠的计算机模拟研究进展进行了简要介绍, 主要内容包括: 全原子分子动力学模拟、Gō 模型以及弹性网络模型的主要思路及其在蛋白质折叠研究中的应用.

关键词: 蛋白质折叠; 分子动力学模拟; Gō 模型; 弹性网络模型

中图分类号: U 461; TP 308

文献标志码: A

文章编号: 0254-0037(2017)12-1818-10

doi: 10.11936/bjtxb2017060036

Research Progress in Computer Simulations of Protein Folding

SU Jiguo¹, WANG Weibu¹, CHEN Weizu², WANG Cunxin²

(1. Key Laboratory for Microstructural Material Physics of Hebei Province, School of Science,

Yanshan University, Qinhuangdao 066004, China;

2. College of Life Science and Bioengineering, Beijing University of Technology, Beijing 100124, China)

Abstract: Protein folding is an unsolved important scientific problem in the field of structural biology. Computer simulation is one of the main methods for the study of protein folding. The research progress in the computer simulations of protein folding was briefly introduced. The main contents in this paper include: the primarily methods of all-atomic molecular dynamics simulation, Gō model and elastic network model, as well as their implications in the study of protein folding.

Key words: protein folding; molecular dynamics simulation; Gō model; elastic network model

蛋白质是生物体内最为重要的生物大分子之一, 是生物体各种生命活动的具体执行者, 发挥着重要的生物学功能. 蛋白质各种生物学功能的发挥有赖于其精巧复杂的空间结构, 研究蛋白质结构的形成机制是揭示蛋白质功能的前提和基础. 储存在 DNA 序列中的遗传信息可以转录到 RNA, 然后在核糖体翻译成一定的氨基酸序列, 三联码的破译使得人们已经了解了遗传信息从 DNA 到氨基酸序列的传递过程. 然而, 翻译生成的氨基酸序列如何能够快速准确地折叠到其天然结构, 目前尚不清楚. 这一问题称为蛋白质折叠问题, 是结构生物学领域有

待解决的重要科学问题之一. 蛋白质折叠问题包含 2 个层面的内容: 特定蛋白质序列会折叠到什么样的天然结构, 以及这一结构形成过程是如何进行的, 这 2 个方面相辅相成. 关于折叠如何进行是蛋白质物理研究的重要内容, 是衔接蛋白质序列和结构的物理内容. 20 世纪 90 年代, Wolynes 等提出了“能量漏斗”模型, 该模型认为蛋白质折叠的能量曲面具有漏斗形状, 漏斗底部对应其天然结构^[1-6]. 能量漏斗模型保证了蛋白质的折叠态在热力学上的稳定性, 同时又确保折叠具有快速的动力学特性. 蛋白质的能量曲面之所以具有漏斗形状是由于长时间生

收稿日期: 2017-06-20

基金项目: 首批河北省青年拔尖人才支持计划资助项目(201319)

作者简介: 苏计国(1978—), 男, 教授, 主要从事软凝聚态物理方面的研究, E-mail: jiguo@ysu.edu.cn

物进化的选择,使得蛋白质折叠过程的能量阻挫降到最低,避免折叠过程中残基间非天然相互作用的形成,使得体系能快速折叠到它的天然构象,这一观点称为“最小阻挫”原理^[7-9]。最新的实验研究直接证实了最小能量阻挫原理的正确性^[10],能量漏斗是实现能量“最小阻挫”的一种方式。能量漏斗模型为蛋白质折叠研究提供了普适的理论框架,但是对于一个具体的蛋白质体系,其折叠路径如何还需要相应的实验或计算方法进行研究。由于现有的实验手段还无法在微观水平上精确观测蛋白质的整个折叠过程,计算机模拟方法已成为研究蛋白质微观折叠过程的重要工具。本文对蛋白质折叠研究常用的计算机模拟方法进行简要介绍。

1 基于全原子分子动力学模拟的蛋白质折叠研究

全原子尺度的分子动力学(molecular dynamics, MD)模拟是在原子水平上获得蛋白质微观运动过程细节的有效方法。MD模拟的主要思路是:首先,给定体系(包括蛋白质和溶剂)的初始坐标和初始速度,对于蛋白质折叠研究,初始坐标一般是将蛋白质天然结构通过高温去折叠得到,初始速度则采用麦克斯韦速度分布律获得;然后,基于一定的分子力场计算体系所有原子之间的相互作用力,常用的力场包括 CHARMM、AMBER、GROMOS 等;随后,采用一定的积分算法通过求解牛顿方程获得体系的折叠运动轨迹,常用的积分算法包括 Verlet 算法、速度 Verlet 算法、蛙跳算法等;最后,通过对模拟轨迹的统计分析获得体系的折叠过程以及相关的动力学和热力学性质。

20 世纪 70 年代, Karplus 研究组^[11]将 MD 模拟方法引入蛋白质动力学性质的研究,他们基于经验力场在全原子水平上模拟了牛胰蛋白酶抑制剂在平衡位置附近的构象运动。但是,在早期研究中,全原子 MD 模拟并没有成功用于蛋白质折叠过程研究的报道,这是由于蛋白质折叠过程非常复杂,折叠的时间尺度很长,利用全原子 MD 模拟蛋白质的折叠过程是一个巨大的挑战。直到 20 世纪 90 年代,国际上一些研究组成功模拟了一些多肽和小蛋白的折叠过程。Case 小组^[12]利用 AMBER/OPLS 力场模拟了 5 个氨基酸所构成的多肽 $\text{NH}_3^+ \text{-Ala-Tyr-cisPro-Tyr-Asp-NMA}$ (*cis*-AYPYD)从伸展的链到转角构象的折叠过程,模拟得到的折叠构象与 NMR 实验能很好地吻合。van Gunsteren 研究组^[13-14]基于 GROMOS96

力场,利用全原子 MD 模拟研究了含有 6 个氨基酸的多肽的折叠过程,模拟结果验证了 NMR 实验数据。Duan 等^[15]利用 Cray T3E 大规模并行超级计算机,采用显含溶剂的全原子 MD 对绒毛蛋白的顶部结构域(villin headpiece subdomain)进行了长达 1 μs 的模拟,模拟体系含有 36 个氨基酸,模拟结果显示该蛋白质从去折叠态出发,在折叠的早期阶段发生疏水塌缩并形成了螺旋结构,随后,构象进行了重新调整,最后到达了一个类似于天然结构的稳定的中间态,通过模拟轨迹他们发现了 2 条折叠路径。Lindorff-Larsen 等^[16]建立了一个专用的超级计算机系统 Anton,对原子水平的 MD 模拟进行了大幅度加速,此外,他们对 CHARMM 力场进行了优化,提高了力场的可移植性。在此基础上,他们对 12 个具有不同结构类型的快速折叠蛋白进行了长达 100 μs ~ 1 ms 的全原子 MD 模拟,模拟结果表明,这些蛋白质都能在模拟时间内折叠到它们的天然结构。根据这些模拟轨迹,他们发现了一些普遍规律:在折叠过程的早期,蛋白质主链具有类天然拓扑结构,一些二级结构单元和小部分的长程接触已经形成。大部分体系的折叠过程沿着一个唯一的特定路径进行折叠,并且二级结构的折叠次序与其在去折叠态下的形成倾向性高度关联。Prigozhin 等^[17]的研究工作表明长时间的全原子 MD 模拟所得到的蛋白质折叠/去折叠过程能与光谱实验数据很好地吻合。

为了提高 MD 模拟采样的效率,Sugita 等^[18]及 Garcia 等^[19]将多副本交换技术引入 MD 模拟,提出了副本交换分子动力学(replica exchange molecular dynamics, REMD)模拟方法。在 REMD 方法中,多个副本在不同的温度下并行模拟,并且根据一定的规则对不同副本中的构象进行交换。通过 REMD 方法,陷入局部极小的模拟轨迹可以与高温副本中的构象进行交换,跳出局部极小,从而提高 MD 模拟的采样效率。Sugita 和 Okamoto 利用所提出的 REMD 方法模拟了 5 个氨基酸构成的小肽——甲硫啡肽的折叠过程,研究表明该方法能有效克服体系能量曲面的多个极小点问题,通过一次模拟可以较为精确地得到体系的自由能函数(或状态密度),继而通过重赋权方法得到不同条件下的热力学性质。Garcia 和 Sanbonmatsu 利用 REMD 研究了 G 蛋白 C-端 β 发卡结构的折叠过程,得到了体系的自由能曲面,分析了体系在不同中间态之间的转变,获得了体系的熔球态结构。Peter 等^[20]基于动力学蒙特卡洛方法,提出了一种粗粒化的多副本交换动力学技术,

该方法将多副本空间进行了简化,降低了副本交换时所需要的副本数量并提高了副本交换的效率. 与传统的 REMD 方法相比,该方法能大幅降低计算量,并且显著提高 MD 模拟的采样效率. Lee 等^[21]将传统的温度副本交换方法(T-REM)与哈密顿副本交换方法(q-REM)进行结合,提出了一种新的多副本交换方法即 Tq-REM 方法,该方法可以有效提高采样的效率,同时避免了上述 2 种方法的缺点. 他们利用 Tq-REM 方法研究了甲硫啡肽、(AAQAA)₃和 Trpzip2 的折叠过程,结果表明相对于 T-REM 方法,Tq-REM 方法计算得到的自由能曲面的收敛性有了大幅度提高,进而基于模拟轨迹他们还发现了体系的多个折叠中间态. 目前,REMD 方法已经成为蛋白质折叠研究中加快采样的常用技术. 另外一个常用的增强采样的方法是 metadynamics,该方法通过构造集合变量,在集合变量空间中引入偏好势,将该偏好势加入到体系的哈密顿量来加快采样过程^[22]. Bonomi 等^[23]利用 metadynamics 模拟了 GB1 的 C 端 β 发卡结构的折叠过程,根据模拟轨迹计算获得了折叠的自由能曲面,根据该自由能曲面发现了 2 条不同的折叠路径,模拟结果与实验能很好地吻合. Han 等^[24]将 metadynamics 与 REMD 技术相结合,模拟了天然无序蛋白 α -MoRE-MeV 与 phosphoprotein P 的 X 结构域结合过程中的边折叠边结合行为. 观察到了多个折叠和结合现象,计算得到了相应的自由能曲面,揭示了该蛋白质体系折叠和结合过程的耦合机制.

在对采样轨迹的分析方面,Pande 小组发展了马尔可夫状态模型,该方法将长时间的大量模拟轨迹首先分割为离散的亚稳态,并计算亚稳态之间的转移概率,进而构造马尔可夫模型来对模拟轨迹进行分析. 该方法能有效揭示统计学上重要的折叠路径,效果优于主成分分析、构象投影等传统方法^[25-26]. 除了采样算法的改进和提高以外,硬件技术的发展也在很大程度上促进了蛋白质折叠的 MD 模拟研究^[27]. Stone 等^[28]在 2007 年首次将 GPU 技术成功引入 MD 模拟研究,他们的模拟结果表明,利用 GPU 技术,MD 模拟中长程静电和非键相互作用力的计算可以提高 10 ~ 100 倍. 随后,Nguyen 等^[29]基于 GPU 技术,采用 AMBER 模拟软件对 17 个蛋白质的折叠过程进行了长达 2.5 ms 的 MD 模拟,这 17 个蛋白质分别含有 10 ~ 92 个氨基酸,并且具有不同的拓扑特征和二级结构含量,研究表明这些蛋白都可以折叠到正确的构象. 目前,很多 MD 模拟软件

都部分或者全部支持 GPU 并行^[30],包括 GROMACS^[31]、AMBER^[32]、ACEMD^[33]、HOOMD-blue^[34]等.

2000 年斯坦福大学的 Pande 课题组发起了一个 Folding@home 分布式计算工程,该工程利用全世界各个角落的成千上万的个人电脑,在系统不忙的时候调用中央处理器进行蛋白质折叠的模拟工作,并将模拟结果传回斯坦福大学的服务器^[35]. Chong 等^[36]利用 Folding@home 在全原子水平上模拟了 p53 的折叠及二聚体形成过程,发现了 p53 二聚体的折叠核位置,与实验结果一致. Shi 等^[37]利用 Folding@home 对 CheY 蛋白的折叠过程进行了长达 42 ms 的全原子分子动力学模拟,基于模拟轨迹获得了 CheY 折叠中间态的构象特征,模拟结果与 FRET 实验结果很好地吻合.

虽然全原子 MD 模拟成功得到了一些小肽和蛋白质的折叠路径,但是,目前的 MD 模拟手段仅仅能模拟一些小肽和小的蛋白质体系. 对于比较大的体系,随着蛋白质体系的增大,MD 模拟的时间尺度会大幅度增加. 实验研究表明,蛋白质折叠的时间随体系氨基酸个数平方根的增大而指数增加^[38]. Perez 等^[39]指出,在可预见的将来,全原子 MD 方法还无法模拟超过 140 个氨基酸的蛋白质体系的折叠过程. 为了解决这一难题,人们提出了各种粗粒化方法来研究蛋白质折叠/去折叠问题. 粗粒化方法对蛋白质体系进行适当的简化,从而减小计算量,加快模拟速度. 在蛋白质折叠研究中,人们建立了各种不同简化程度的粗粒化模型,比如将一个氨基酸简化为 4 ~ 6 个点、2 个点甚至 1 个点^[40]. Cornell 大学 Scheraga 小组^[41]和 Gdansk 大学 Liwo 小组^[42]发展的 UNRES 模型是非常成功的粗粒化模型之一,在蛋白质折叠研究中得到了广泛的应用. 该模型将氨基酸简化为 3 个点,首先保留每个氨基酸的 C α 原子作为 1 个简化点,每个氨基酸的侧链作为一个整体简化为 1 个点,每个氨基酸主链上的其他重原子也作为一个整体简化为 1 个点. C α 原子主要用来维持蛋白质的几何结构,不参与氨基酸之间的相互作用. 氨基酸之间的相互作用主要作用在侧链简化点和主链其他重原子简化点上. 在 MD 模拟中,相对于全原子模型,UNRES 模型的模拟速度可以提高 1000 ~ 4000 倍,可以在数小时内模拟蛋白质的从头折叠过程.

在粗粒化模型构建中,如何对蛋白质体系进行简化,并优化获得适用于相应简化模型的势能函数

是粗粒化方法的核心问题,关于势能函数中力常数的优化问题可以参见文献[43]的相关章节. 在蛋白质折叠研究的所有简化模型中,G ϕ 模型和弹性网络模型基于蛋白质天然结构来构造体系的势能函数,简化了力常数优化这一难题,在蛋白质折叠/去折叠研究中得到了广泛应用. 下面对 G ϕ 模型和弹性网络模型在蛋白质折叠研究中的应用进行简要介绍.

2 基于 G ϕ 模型的蛋白质折叠研究

能量漏斗和最小阻挫原理表明蛋白质的天然结构很大程度上决定了蛋白质的折叠过程,G ϕ 模型就是基于上述 2 个原理,从蛋白质的天然结构出发来构造体系折叠的势能函数. 在 G ϕ 模型的势能函数中,天然结构中存在相互作用的残基之间用一个吸引势来表示,在模拟过程中将这些残基拉到一起. 而天然结构中不存在相互作用的残基之间用一个排斥势来表示,阻止这些残基在模拟过程中相互靠近. G ϕ 模型的这种势能函数将蛋白质天然结构中的相互作用作为折叠的主要驱动力,使得体系的势能函数具有光滑的漏斗形状,蛋白质的天然结构位于漏斗的最底端,保证了体系折叠过程中具有最小的能量阻挫. 20 世纪 70 年代, G ϕ 模型首先用于离散的格点模型来研究蛋白质折叠问题^[44-46]. 随着计算机计算能力的提高,格点模型已经很少使用了,现在 G ϕ 模型主要结合非格点模型来模拟蛋白质的折叠过程. 由于 G ϕ 模型对体系的能量曲面做了最大程度的简化,完全忽略了残基间的能量阻挫,因此,利用常规的 MD 模拟方法就可以对蛋白质折叠的能量曲面进行详尽的研究,尤其是过渡态和一些高能的折叠中间态等这些热力学上不稳定的构象状态也可以较容易地获得^[47]. G ϕ 模型虽然异常简化,但是大量的理论模拟研究表明,该模型与大量的实验数据能定量吻合. 目前,人们对于蛋白质折叠机制的理解很大一部分都是来自于 G ϕ 模型的模拟研究. G ϕ 模型有效性的一个重要证据是 G ϕ 模型模拟计算得到的不同残基的 ϕ -值能与实验观测数据很好地吻合. ϕ -值的表达式为 $\phi = \Delta\Delta G_{\ddagger-D} / \Delta\Delta G_{N-D}$, 其中, $\Delta\Delta G_{\ddagger-D}$ 和 $\Delta\Delta G_{N-D}$ 分别表示氨基酸突变所导致的蛋白质激活自由能和折叠自由能的差异^[48]. ϕ -值分析是确定蛋白质折叠过渡态以及折叠核的重要方法,实验中通过丙氨酸突变扫描实验可以对不同残基的 ϕ -值进行测量,基于 G ϕ 模型的 MD 模拟方法也可以对 ϕ -值进行定量计算. 研究表明,G ϕ 模型模拟计算得到的 ϕ -值与实验测量值能较好地吻合,验

证了 G ϕ 模型在蛋白质折叠研究中的有效性^[49-51]. G ϕ 模型成功的另外一个重要证据是基于 G ϕ 模型的 MD 模拟计算得到的蛋白质的折叠速率能与实验测量结果很好地吻合^[52-53]. 大量成功的研究案例表明 G ϕ 模型能反映蛋白质折叠背后的一些普遍机制,是人们认识蛋白质折叠机理的一个简单而重要的研究工具. 近几十年来,G ϕ 模型已经成为研究蛋白质折叠路径、折叠核、折叠速率、折叠过渡态、折叠中间态、蛋白质稳定性、蛋白质力学性能等相关动力学和热力学性质的重要手段,得到了广泛的应用.

传统的 G ϕ 模型仅仅考虑了蛋白质天然结构对蛋白质折叠行为的决定性作用,忽略了蛋白质具体的氨基酸序列对于折叠过程的影响. 而实际上,正是蛋白质不同的序列决定了蛋白质特定的空间结构,序列的差异会导致折叠行为的不同. 比如,单个残基的突变可能会导致折叠动力学以及折叠态稳定性的改变. 一些关键残基的突变可能会造成蛋白质的部分结构的去折叠以及淀粉样聚集体的形成^[54]. 因此,一些研究工作将蛋白质序列信息加入到 G ϕ 模型中,引入一些能量阻挫,探讨蛋白质序列对于折叠过程的影响. Karanicolas 等^[55]利用 M-J 统计势对天然结构中残基间的相互作用进行加权,发现考虑到不同类型氨基酸之间相互作用的差异后,G ϕ 模型能很好地解释 L 蛋白和 G 蛋白这 2 个结构相似而折叠路径不同的行为差异. Das 等^[56]在 G ϕ 模型考虑了蛋白质序列信息的影响,通过迭代方法对体系中氨基酸之间的天然接触能量和非天然接触能量进行了优化,使得天然结构与非天然结构之间的能隙最大. 利用该模型,他们研究了 src-SH3 的折叠过程,与已知的所有实验结果和理论预测能很好地吻合. 该研究表明残基间的非天然相互作用和能量阻挫对折叠过程也有一定的影响. 南京大学王伟教授领导的课题组^[57]利用不同氨基酸物理化学性质的不同来对残基能量进行加权,在此基础上研究了核糖体蛋白 S6 以及其突变体的折叠路径、不同氨基酸的 ϕ -值以及折叠核的位置,与实验结果吻合得很好. Chan 课题组^[58]将溶剂化-去溶剂化能垒加入到传统 G ϕ 模型中,增加折叠能量曲面的受挫性质,该模型能再现一些重要的去溶剂化中间态. 王存新教授领导的课题组^[59]以及苏计国教授课题组^[60]把带电残基之间的静电相互作用加入到传统 G ϕ 模型中,研究了常温蛋白、嗜热蛋白以及超嗜热蛋白的热稳定性、折叠速率以及折叠路径等,与实验结果一致. 图 1 显示了带电残基突变对于冷休克蛋白热容

曲线的影响. 热容曲线的峰值对应于体系的折叠温度, 可以看出 E46A、R3E 和 R3E/L66E 突变使得体系的折叠温度降低, 表明这些带点残基突变降低了体系的热稳定性. 图中温度和热容均为约化温度. 图 2 显示了体系处于去折叠态概率的对数与模拟步数之间的关系. 图中直线斜率的绝对值表示折叠速率, 可以看出上述残基突变能显著降低体系的折叠速率. 这些研究揭示了静电相互作用对蛋白质热稳定性和折叠行为的重要影响, 为蛋白质的功能改造提供了有价值的信息. 由于 G ϕ 模型对体系的能量曲面进行了光滑处理, 加快了折叠模拟速度和采样效率. 近年来的研究趋势是利用 G ϕ 模型对大的具有复杂拓扑结构的蛋白质体系以及复杂的折叠现象进行研究. Das 等^[61] 基于 G ϕ 模型研究了具有多结构域的乳糖阻遏蛋白的折叠过程, 与实验结果能较好地吻合. Soler 等^[62] 利用 G ϕ 模型研究了打结蛋白质在分子伴侣 GroEL-GroES 空腔中的折叠过程, 研究表明分子伴侣空腔在空间上的限制以及蛋白质的局部柔性对打结蛋白质的折叠具有重要影响. 南京大学王伟课题组^[63] 利用 G ϕ 模型研究了在细胞拥挤环境中蛋白质的边折叠边对接过程, 能解释一些重要的实验现象. 王伟课题组和日本京都大学 Takada 课题组基于 G ϕ 模型, 将蛋白质折叠、辅助离子结合以及别构运动整合到一起, 发展一个统一的理论模型, 成功模拟了钙调蛋白的折叠、别构运动以及辅助离子的结合过程之间的耦合机制和协同运作过程^[64]. 他们还在 G ϕ 模型的基础上, 发展了基于原子层次相互作用的粗粒化(AICG)模型, 并将蛋白质分子的局部柔性考虑到模型中, 研究了多结构域蛋白 adenylate kinase 以及打结蛋白 2ouf-knot 的折叠过程, 发现了多个亚稳态以及这些亚稳态之间的转变过程^[65].

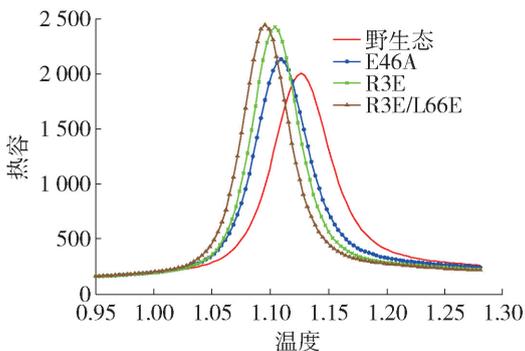


图 1 冷休克蛋白野生型及突变体的热容曲线^[59]

Fig. 1 Temperature dependence of the specific heat for the wild type cold shock protein and its mutants^[59]

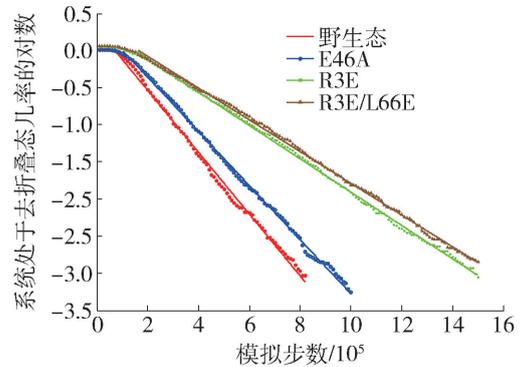


图 2 冷休克蛋白野生型以及 E46A、R3E 和 R3E/L66E 突变体的折叠速率^[59]

Fig. 2 Folding rates for the wild-type cold shock protein, and its three mutants E46A, R3E and R3E/L66E, respectively^[59]

虽然, G ϕ 模型对蛋白质折叠的能量曲面进行了极大的简化, 可以加快折叠模拟的速度, 提高构象空间的采样效率, 但是, 要获得蛋白质折叠的热力学性质, 需要对体系的折叠过程进行长时间的模拟, 并对模拟轨迹进行大量的采样, 基于这些采样, 通过系综平均计算得到体系的宏观热力学量. 相对于 G ϕ 模型, 弹性网络模型对体系的势能函数做了进一步简化, 将残基之间的相互作用用谐振势来代替, 把蛋白质简化为一个由大量弹簧所构成的弹性网络. 弹性网络模型的优势在于体系的配分函数以及相关的热力学量可以解析地求出, 避免了构象空间的采样问题, 大大降低了计算量. 下面对弹性网络模型在蛋白质折叠/去折叠研究中的应用进行简要介绍.

3 基于弹性网络模型的蛋白质折叠/去折叠研究

弹性网络模型(ENM)的基本思想来自于 Floy 对随机链的统计力学研究^[66]. 随后, Tirion^[67] 将经典分子力场简化为谐振势来研究蛋白质的构象运动, 取得了令人惊奇的结果, 发现如此简单的模型可以准确地计算蛋白质的动力学性质. 1997 年, Bahar 等^[68]、Haliloglu 等^[69] 建立了残基水平的 ENM. 该模型将蛋白质简化为一个弹性网络, 蛋白质中的每个氨基酸用其 C α 原子来代替, 简化为一个点, 氨基酸之间的相互作用简化为谐振势, 用弹簧来代替. 并且, 所有弹簧的强度系数都相同. 根据上述简化方法, 蛋白质体系的势能函数等于所有弹簧的势能之和, 可以写成

$$V = \frac{\gamma}{2} \sum_{i,j}^N (\Delta R_{ij})^2 = \frac{\gamma}{2} \Delta R^T \Gamma \Delta R$$

式中: ΔR_{ij} 为残基 i 和 j 之间距离的涨落; N 为蛋白质残基的个数; $\Delta R^T = \{\Delta R_1, \Delta R_2, \dots, \Delta R_N\}$ 表示各个残基偏离平衡位置的涨落, 上标 T 表示转置; γ 为弹簧的强度系数; Γ 为 Kirchhoff 矩阵, 其矩阵元为

$$\Gamma_{ij} = \begin{cases} -1, & i \neq j, \text{ 且 } R_{ij} \leq r_c \\ 0, & i \neq j, \text{ 且 } R_{ij} > r_c \\ -\sum_{i, i \neq j}^N \Gamma_{ij}, & i = j \end{cases}$$

式中: R_{ij} 为残基 i 和 j 之间的平衡距离; r_c 为距离截断半径. 通过对 Kirchhoff 矩阵进行对角化, 可以获得体系的各个运动模式以及相应的频率. 进而, 通过统计物理学理论, 可以计算体系各个物理量的系综平均值. 目前, ENM 广泛用于蛋白质功能性运动、活性位点识别、结构域划分、加快构象采样、蛋白质柔性分析、变构机制等相关问题的研究.

近年来, 人们也开始尝试将 ENM 用于蛋白质折叠/去折叠过程的研究, 取得了较好的效果.

Micheletti 等^[70] 首先将 ENM 引入蛋白质折叠研究, 对传统的 ENM 做了改进, 在他们的模型中, 非共价相互作用残基之间的弹簧允许断裂, 并将温度引入到势能函数中, 可以计算不同温度下体系平衡态的性质, 通过调整温度来研究蛋白质的折叠/去折叠过程, 该模型所得到的结果与实验数据能很好地吻合. 随后, Dietz 等^[71] 将 ENM 用于蛋白质机械外力拉伸去折叠的研究, 基于 ENM 探讨了绿色荧光蛋白在不同拉伸方向外力作用下的力学性能和去折叠过程, 能很好地解释原子力显微镜的实验结果. Sulkowska 等^[72] 利用 ENM 来预测单分子拉伸实验中残基之间接触的断裂次序, 与实验观测结果很好地吻合. Su 等^[73-74] 基于 ENM, 提出了一种模拟蛋白质去折叠过程的方法. 该方法根据残基间均方涨落的大小, 通过依次打断残基间非共价接触的方法来模拟蛋白质去折叠过程. 该方法对应于体系的缓慢升温过程, 蛋白质去折叠的每一步都保证体系的熵增最大. 图 3 显示了利用该方法所得到的 CI2 蛋白的去折叠过

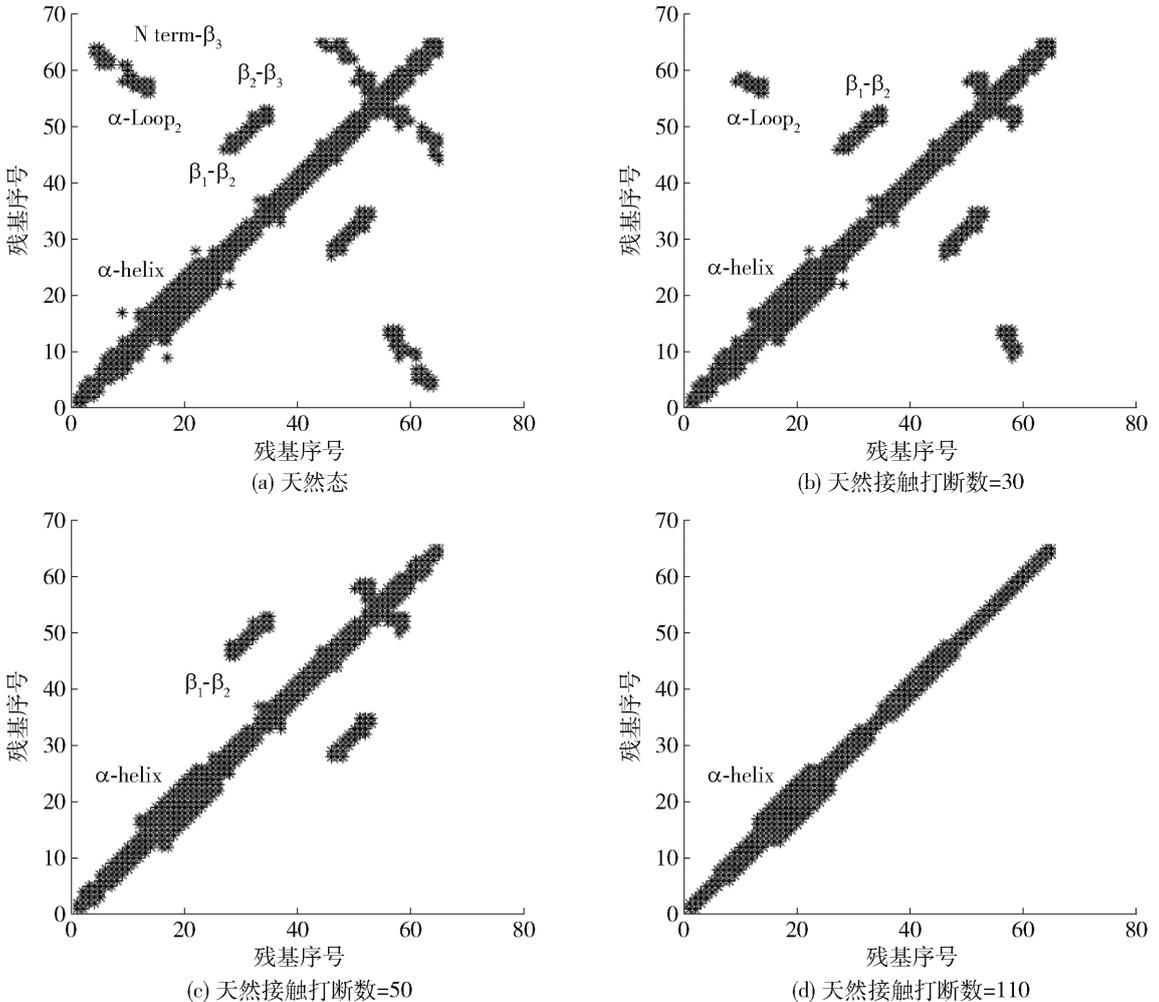


图 3 模拟得到的 CI2 蛋白的去折叠过程^[73]

Fig. 3 Unfolding process of CI2 predicted by our method^[73]

程,可以发现其去折叠路径为 $N\text{-term-}\beta_3 \rightarrow \beta_2-\beta_3 \rightarrow \alpha\text{-Loop}_2 \rightarrow \beta_1-\beta_2 \rightarrow \alpha\text{-helix}$,与全原子 MD 模拟结果和实验结果能很好地吻合^[73]. 该方法还被成功用于 RNA 去折叠过程的研究,取得了良好的预测效果^[75]. 基于类似的思想,Arkun 等^[76]成功预测了蛋白质的折叠过程,在他们的方法中,蛋白质折叠的每一步都保证体系的熵减最小,熵的变化通过 ENM 计算获得. 随后,与上述方法类似,Srivastava 等^[77-78]利用 ENM,通过依次打断残基间非共价接触的方法来模拟蛋白质的受力去折叠和高温去折叠过程,并且发现了去折叠过程中的一些协同效应. Williams 等^[79]将 ENM 与 MD 模拟相结合来模拟蛋白质的折叠路径,与实验结果较好地吻合.

4 总结与展望

计算机模拟方法是研究蛋白质折叠物理机制的重要手段,本文简要介绍了全原子 MD 模拟和粗粒化模型在蛋白质折叠研究中的进展. 对于全原子 MD 模拟,如何加快构象空间采样是一个热点和难点问题. 本文介绍了增强采样的 REMD 方法和 metadynamics 方法,以及在并行和硬件技术方面的发展,包括 Folding@home 和 GPU 技术. 此外,还介绍了对采样轨迹进行有效分析的马尔可夫状态模型. 对于粗粒化模型,介绍了 G ϕ 模型和 ENM 模型在蛋白质折叠研究中的应用.

由于经典分子力场还无法做到完全精确,并且受到计算机计算能力的限制,全原子 MD 模拟方法模拟大的蛋白质体系的折叠过程还存在很大的困难. 发展粗粒化模型仍然是蛋白质折叠研究的主要方向之一. 鉴于 G ϕ 模型和弹性网络模型对蛋白质折叠的能量曲面进行了极度简化,可能会导致一些影响蛋白质折叠的重要信息的丢失. 如何发展有效的粗粒化模型用于蛋白质折叠研究仍然是一个有待解决的重要问题.

参考文献:

[1] KEN A. Polymer principles and protein folding [J]. Protein Science, 1999, 8(8): 1166-1180.

[2] BALDWIN R L. The nature of protein folding pathways: the classical versus the new view [J]. Journal of Biomolecular NMR, 1995, 5(2): 103-109.

[3] BRYNGELSON J D, ONUCHIC J N, SOCCI N D, et al. Funnels, pathways, and the energy landscape of protein folding: a synthesis [J]. Proteins Structure Function & Bioinformatics, 1995, 21(3): 167-95.

[4] DOBSON C M, SALI A, KARPLUS M. Protein folding: a perspective from theory and experiment [J]. Angewandte Chemie International Edition, 1998, 37(7): 868-893.

[5] ONUCHIC J N, WOLYNES P G, LUTHEYSCHULTEN Z, et al. Toward an outline of the topography of a realistic protein-folding funnel [J]. Proceedings of the National Academy of Sciences of the United States of America, 1995, 92(8): 3626-3630.

[6] DILL K A, CHAN H S. From Levinthal to pathways to funnels [J]. Nature Structural Biology, 1997, 4(1): 10-19.

[7] BRYNGELSON J D, WOLYNES P G. Spin glasses and the statistical mechanics of protein folding [J]. Proceedings of the National Academy of Sciences of the United States of America, 1987, 84(21): 7524-7528.

[8] WOLYNES P G. Recent successes of the energy landscape theory of protein folding and function [J]. Quarterly Reviews of Biophysics, 2005, 38(38): 405-410.

[9] FERREIRO D U, HEGLER J A, KOMIVES E A, et al. Localizing frustration in native proteins and protein assemblies [J]. Proceedings of the National Academy of Sciences, 2007, 104(50): 19819-19824.

[10] TZUL F O, VASILCHUK D, MAKHATADZE G I. Evidence for the principle of minimal frustration in the evolution of protein folding landscapes [J]. Proceedings of the National Academy of Sciences, 2017, 114(9): E1627-E1632.

[11] MCCAMMON J A, GELIN B R, KARPLUS M. Dynamics of folded proteins [J]. Nature, 1977, 267(5612): 585-590.

[12] DEMCHUK E, BASHFORD D, CASE D A. Dynamics of a type VI reverse turn in a linear peptide in aqueous solution [J]. Folding & Design, 1997, 2(1): 35-46.

[13] XAVIER D, van GUNSTEREN W F, DARIO R, et al. Studying the stability of a helical β -heptapeptide by molecular dynamics simulations [J]. Chemistry - A European Journal, 1997, 3(9): 1410-1417.

[14] DAURA X, JAUN B, SEEBACH D, et al. Reversible peptide folding in solution by molecular dynamics simulation I [J]. Journal of Molecular Biology, 1998, 280(5): 925-932.

[15] 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.

[16] LINDORFF-LARSEN K, PIANA S, DROR R O, et al. How fast-folding proteins fold. [J]. Science, 2011, 334(6055): 517-520.

- [17] PRIGOZHIN M B, CHAO S H, SUKENIK S, et al. Mapping fast protein folding with multiple-site fluorescent probes [J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2015, 112(26): 7966-7971.
- [18] SUGITA Y, OKAMOTO Y. Replica-exchange molecular dynamics method for protein folding [J]. *Chemical Physics Letters*, 1999, 314(1/2): 141-151.
- [19] GARCIA A E, SANBONMATSU K Y. Exploring the energy landscape of a beta hairpin in explicit solvent[J]. *Proteins*, 2001, 42(3): 345-354.
- [20] PETER E K, SHEA J E, PIVKIN I V. Coarse kMC-based replica exchange algorithms for the accelerated simulation of protein folding in explicit solvent [J]. *Physical Chemistry Chemical Physics Pccp*, 2016, 18(18): 13052-13065.
- [21] LEE M, YOON J, JANG S, et al. Conformational sampling of metastable states: Tq-REM as a novel replica exchange method. [J]. *Physical Chemistry Chemical Physics Pccp*, 2017, 19(7): 5454.
- [22] BARDUCCI A, BONOMI M, PARRINELLO M. Metadynamics [J]. *WIREs Computational Molecular Science*, 2011, 1(5): 826-843.
- [23] BONOMI M, BRANDUARDI D, GERVASIO F L, et al. The unfolded ensemble and folding mechanism of the C-terminal GB1 β -hairpin [J]. *Journal of the American Chemical Society*, 2008, 130(42): 13938-13944.
- [24] HAN M, XU J, REN Y, et al. Simulation of coupled folding and binding of an intrinsically disordered protein in explicit solvent with metadynamics [J]. *Journal of Molecular Graphics and Modelling*, 2016, 68: 114-127.
- [25] BOWMAN G R, VOELZ V A, PANDE V S. Atomistic folding simulations of the fivehelix bundle protein λ_{6-85} [J]. *Journal of the American Chemical Society*, 2010, 133: 664-667.
- [26] LANE T J, BOWMAN G R, BEAUCHAMP K, et al. Markov state model reveals folding and functional dynamics in ultra-long MD trajectories[J]. *Journal of the American Chemical Society*, 2011, 133: 18413-18419.
- [27] LLANES A, MUÑOZ A, BUENO-CRESPO A, et al. Soft computing techniques for the protein folding problem on high performance computing architectures [J]. *Current Drug Targets*, 2016, 17(14): 1626-1648.
- [28] STONE J E, PHILLIPS J C, FREDDOLINO P L, et al. Accelerating molecular modeling applications with graphics processors [J]. *Journal of Computational Chemistry*, 2007, 28(16): 2618-2640.
- [29] NGUYEN H, MAIER J, HUANG H, et al. Folding simulations for proteins with diverse topologies are accessible in days with a physics-based force field and implicit solvent [J]. *Journal of the American Chemical Society*, 2014, 136(40): 13959-13962.
- [30] ALEKSEENKO A, KONONOVA O, KHOLODOV Y, et al. SOP-GPU: influence of solvent-induced hydrodynamic interactions on dynamic structural transitions in protein assemblies [J]. *Journal of Computational Chemistry*, 2016, 37(17): 1537-1551.
- [31] ABRAHAM M J, MURTOLA T, SCHULZ R, et al. GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers[J]. *Softwarex*, 2015, 1: 19-25.
- [32] SALOMON-FERRER R, GOÖTZ A W, POOLE D, et al. Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh Ewald[J]. *Journal of Chemical Theory and Computation*, 2013, 9(9): 3878-3888.
- [33] HARVEY M J, GIUPPONI G, FABRITHS G D. ACEMD: accelerating biomolecular dynamics in the microsecond time scale [J]. *Journal of Chemical Theory and Computation*, 2009, 5(6): 1632-1639.
- [34] ANDERSON J A, LORENZ C D, TRAVESSET A. General purpose molecular dynamics simulations fully implemented on graphics processing units [J]. *Journal of Computational Physics*, 2008, 227(10): 5342-5359.
- [35] BEBERG A L, ENSIGN D L, JAYACHANDRAN G, et al. Folding@Home: lessons from eight years of volunteer distributed computing from eight years of volunteer distributed computing [C] // *Proc IEEE International Parallel and Distributed Processing Symposium (IPDPS)*. Atlanta: IEEE Computer Society, 2009: 1-8.
- [36] CHONG L T, SNOW C D, RHEE Y M, et al. Dimerization of the p53 oligomerization domain; identification of a folding nucleus by molecular dynamics simulations [J]. *Journal of Molecular Biology*, 2005, 345: 869-878.
- [37] SHI J, NOBREGA R P, SCHWANTES C, et al. Atomistic structural ensemble refinement reveals non-native structure stabilizes a sub-millisecond folding intermediate of CheY [J]. *Scientific Reports*, 2017, 7: 44116.
- [38] NAGANATHAN A N, MUÑOZ V. Scaling of folding times with protein size [J]. *Journal of the American Chemical Society*, 2005, 127(2): 480-481.
- [39] PEREZ A, MORRONE J A, SIMMERLING C, et al. Advances in free-energy-based simulations of protein folding and ligand binding [J]. *Current Opinion in*

- Structural Biology, 2016, 36: 25-31.
- [40] TOZZINI V. Coarse-grained models for proteins. [J]. Current Opinion in Structural Biology, 2005, 15(2): 144-150.
- [41] LIWO A, KHALILI M, SCHERAGA H A. Ab initio simulations of protein-folding pathways by molecular dynamics with the united-residue model of polypeptide chains [J]. Proceedings of the National Academy of Sciences of the United States of America, 2005, 102: 2362-2367.
- [42] SCHERAGA H A, KHALILI M, LIWO A. Protein-folding dynamics: overview of molecular simulation techniques[J]. Annual Review of Physical Chemistry, 2007, 58: 57-83.
- [43] KOLINSKI A. 蛋白质模拟的多尺度方法[M]. 王存新, 译. 北京: 科学出版社, 2014: 102-118.
- [44] TAKETOMI H, UEDA Y, GŌ N. Studies on protein folding, unfolding and fluctuations by computer simulation [J]. Chemical Biology & Drug Design, 1975, 7(6): 445-459.
- [45] GŌ N, TAKETOMI H. Respective roles of short-and long-range interactions in protein folding[J]. Proceedings of the National Academy of Sciences, 1978, 75(2): 559-563.
- [46] GŌ N. Theoretical studies of protein folding[J]. Annual Review of Biophysics and Bioengineering, 1983, 12(1): 183-210.
- [47] HILLS R, BROOKS C. Insights from coarse-frained Gō models for protein folding and dynamics[J]. International Journal of Molecular Sciences, 2009, 10(3): 889.
- [48] FERSHT A R, DAGGETT V. Protein folding and unfolding at atomic resolution[J]. Cell, 2002, 108(4): 573-582.
- [49] CLEMENTI C, NYMEYER H, ONUCHIC J N. Topological and energetic factors: what determines the structural details of the transition state ensemble and “en-route” intermediates for protein folding? An investigation for small globular proteins [J]. Journal of Molecular Biology, 2000, 298(5): 937-953.
- [50] KOGA N, TAKADA S. Roles of native topology and chain-length scaling in protein folding: a simulation study with a Go-like model[J]. Journal of Molecular Biology, 2001, 313(1): 171.
- [51] DOKHOLYAN N V, BULDYREV S V, STANLEY H E, et al. Identifying the protein folding nucleus using molecular dynamics. [J]. Journal of Molecular Biology, 2000, 296(5): 1183-1188.
- [52] KOGA N, TAKADA S. Roles of native topology and chain-length scaling in protein folding: a simulation study with a Go-like model[J]. Journal of Molecular Biology, 2001, 313(1): 171.
- [53] JEWETT A I, PANDE V S, PLAXCO K W. Cooperativity, smooth energy landscapes and the origins of topology-dependent protein folding rates[J]. Journal of Molecular Biology, 2003, 326(1): 247-253.
- [54] DYSON H J, WRIGHT P E. Intrinsically unstructured proteins and their functions. [J]. Nature Reviews Molecular Cell Biology, 2005, 6(3): 197-208.
- [55] KARANICOLAS J, BROOKS C L 3rd. The origins of asymmetry in the folding transition states of protein L and protein G[J]. Protein Science, 2002, 11(10): 2351-2361.
- [56] DAS P, MATYSIAK S, CLEMENTI C. Balancing energy and entropy: a minimalist model for the characterization of protein folding landscapes [J]. Proceedings of the National Academy of Sciences, 2005, 102(29): 10141-10146.
- [57] WU L, ZHANG J, WANG J, et al. Folding behavior of ribosomal protein S6 studied by modified G[o]-like model[J]. Physical Review E Statistical Nonlinear & Soft Matter Physics, 2007, 75(3): 031914.
- [58] LIU Z, CHAN H S. Desolvation is a likely origin of robust enthalpic barriers to protein folding[J]. Journal of Molecular Biology, 2005, 349(4): 872-889.
- [59] SU J G, CHEN W Z, WANG C X. Role of electrostatic interactions for the stability and folding behavior of cold shock protein. [J]. Proteins Structure Function & Bioinformatics, 2010, 78(9): 2157-2169.
- [60] SU J G, HAN X M, ZHAO S X, et al. Impacts of the charged residues mutation S48E/N62H on the thermostability and unfolding behavior of cold shock protein: insights from molecular dynamics simulation with Gō model[J]. Journal of Molecular Modeling, 2016, 22(4): 1-11.
- [61] DAS P, WILSON C J, FOSSATI G, et al. Characterization of the folding landscape of monomeric lactose repressor: quantitative comparison of theory and experiment. [J]. Proceedings of the National Academy of Sciences of the United States of America, 2005, 102(41): 14569-74.
- [62] SOLER M A, REY A, FAÍSCA P F N. Steric confinement and enhanced local flexibility assist knotting in simple models of protein folding [J]. Physical Chemistry Chemical Physics, 2016, 18: 26391-26403.
- [63] WANG W, XU W X, LEVY Y, et al. Confinement effects on the kinetics and thermodynamics of protein

- dimerization. [J]. Proceedings of the National Academy of Sciences, 2009, 106(14): 5517.
- [64] LI W, WANG W, TAKADA S. Energy landscape views for interplays among folding, binding, and allostery of calmodulin domains [J]. Proceedings of the National Academy of Sciences of the United States of America, 2014, 111(29): 10550-10555.
- [65] LI W, TERAOKAWA T, WANG W, et al. Energy landscape and multiroute folding of topologically complex proteins adenylate kinase and 2ouf-knot[J]. Proceedings of the National Academy of Sciences of the United States of America, 2012, 109(44): 17789-17794.
- [66] GORDON M, FLORY P J, MCCRUM N G. Statistical thermodynamics of random networks: discussion [J]. Royal Society of London Proceedings, 1976, 351(1666): 379-380.
- [67] TIRION M M. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis [J]. Physical Review Letters, 1996, 77(9): 1905.
- [68] BAHAR I, ATILGAN A R, ERMAN B. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential [J]. Folding & Design, 1997, 2(3): 173-181.
- [69] HALILOGLU T, BAHAR I, ERMAN B. Gaussian dynamics of folded proteins [J]. Physical Review Letters, 1997, 79(16): 3090-3093.
- [70] MICHELETTI C, BANAVAR J R, MARITAN A. Conformations of proteins in equilibrium [J]. Physical Review Letters, 2001, 87(8): 088102.
- [71] DIETZ H, RIEF M. Elastic bond network model for protein unfolding mechanics [J]. Physical Review Letters, 2008, 100(9): 098101.
- [72] SULKOWSKA J I, KLOCZKOWSKI A, SEN T Z, et al. Predicting the order in which contacts are broken during single molecule protein stretching experiments [J]. Proteins: Structure, Function, and Bioinformatics, 2008, 71(1): 45-60.
- [73] SU J G, LI C H, HAO R, et al. Protein unfolding behavior studied by elastic network model [J]. Biophysical Journal, 2008, 94(12): 4586-4896.
- [74] SU J G, ZHANG X, HAN X M, et al. The intrinsic dynamics and unfolding process of an antibody fab fragment revealed by elastic network model [J]. International Journal of Molecular Sciences, 2015, 16(12): 29720-29731.
- [75] LI C, LÜ D, ZHANG L, et al. Approach to the unfolding and folding dynamics of add A-riboswitch upon adenine dissociation using a coarse-grained elastic network model [J]. Journal of Chemical Physics, 2016, 145(1): 11.
- [76] ARKUN Y, ERMAN B. Prediction of optimal folding routes of proteins that satisfy the principle of lowest entropy loss: dynamic contact maps and optimal control [J]. PloS One, 2010, 5(10): e13275.
- [77] SRIVASTAVA A, GRANEK R. Cooperativity in thermal and force-induced protein unfolding: integration of crack propagation and network elasticity models [J]. Physical Review Letters, 2013, 110(13): 138101.
- [78] SRIVASTAVA A, GRANEK R. Protein unfolding from free-energy calculations: integration of the Gaussian network model with bond binding energies [J]. Physical Review E, 2015, 91(2): 022708.
- [79] WILLIAMS G, TOON A J. Protein folding pathways and state transitions described by classical equations of motion of an elastic network model [J]. Protein Science A Publication of the Protein Society, 2010, 19(12): 2451.

(责任编辑 张 蕾)