评述

模型蛋白Trp-cage残基互作及外在因素对其结构折叠稳定性影响的研究

吴晓敏^{1†*}, 张孝春^{1†}, 汪萌^{1†}, 黄卓然¹, 张兴旺¹, 张强¹, 徐成振^{2*}

1. 淮北师范大学生命科学学院, 污染物敏感材料与环境修复安徽省重点实验室, 淮北 235000;

2. 淮北师范大学计算机科学与技术学院, 淮北 235000

† 同等贡献

* 联系人, E-mail: dolphin1009@sina.com; xu5115@163.com

收稿日期: 2020-08-08; 接受日期: 2020-08-18; 网络版发表日期: 2020-12-03

国家自然科学基金(批准号: 31500594)、安徽省自然科学基金(批准号: 1908085QF286, 1608085MC49)、安徽省高校省级自然科学研究重点项目(批准号: KJ2020A0037, KJ2020A0031, KJ2017A373)、安徽省高校优秀青年人才支持计划(批准号: gxyq2018163)和安徽省高校优秀青年骨干人才国外访学研修项目

摘要 蛋白质折叠驱动力主要取决于结构上残基间的相互作用, 尤其是其侧链间的作用. 迷你蛋白Trp-cage结构典型且可快速折叠, 是理论和实验研究蛋白质折叠的理想模型. 尽管该蛋白仅由20个氨基酸组成, 但其结构中残基间的相互作用甚为复杂且侧链类型多样. 由于现有力场参数对其相互作用描述的精度问题, 其折叠模型仍有分歧. 本文以模型蛋白Trp-cage为例, 首先概述了该蛋白自设计提出以来所报道的所有结构体系及其相应突变体的研究情况, 并分析了其TC5b和TC10b两个经典体系的内在残基作用网络和侧链特性; 然后详细讨论了其折叠机制模型类型的研究情况, 并探讨了不同折叠模型之间的相互转换及其与模拟温度和力场之间的关系; 最后从内在残基侧链特性和外在环境因素两个角度讨论了其对结构稳定和折叠动力学行为的影响. 各影响因素及其协同效应的阐明不仅有助于揭示蛋白质结构平衡稳定和快速折叠的驱动力成因, 也为其进一步结构调控和优化设计提供了理论指导.

关键词 模型蛋白Trp-cage, 残基相互作用, 折叠模型, 残基突变, 侧链位置与特性

分子模拟计算作为一种研究手段已经在生命科学、药物设计、材料科学和石油化工等诸多领域得到广泛的应用^[1]. 在蛋白质研究领域, 分子动力学(molecular dynamics, MD)模拟不仅可以描述蛋白质在原子水平上的详细参数, 而且能够准确提供其结构位置涨落和构象态间转变的精细过程, 并将原子的不可直接观测层面与实验可观测层面紧密联系起来, 这为相

关重要蛋白结构转变机制和生物功能发挥的理论和实验研究提供了有力的平台^[2,3]. 近10年来, 随着计算机软硬件的迅速发展, 分子模拟计算的精度和时间尺度都得以大幅提升^[4-6]. 当前MD模拟已取得一定进展, 但仍存在众多挑战^[7], 如MD模拟中力场无法完全再现蛋白质的真实势能面; 许多蛋白质折叠模拟的时间尺度需达到毫秒、秒级甚至更长; 如何获取足够多的采

引用格式: 吴晓敏, 张孝春, 汪萌, 等. 模型蛋白Trp-cage残基互作及外在因素对其结构折叠稳定性影响的研究. 中国科学: 生命科学, 2021, 51: 135-150
Wu X M, Zhang X C, Wang M, et al. Computational and experimental studies on the influence of the Trp-cage inter-residue interactions and other external factors on its structural and folding stability (in Chinese). Sci Sin Vitae, 2021, 51: 135-150, doi: [10.1360/SSV-2019-0304](https://doi.org/10.1360/SSV-2019-0304)

样以正确表征多样的(heterogeneous)折叠过程,并有效地分析这些数据.为了应对这些挑战,科研工作者做了诸多努力,如对当前现有经典力场进行不断优化^[8],开发提高采样效率的新算法^[9]和轨迹分析的新方法^[10]等.模型蛋白Trp-cage因具有快速折叠的疏水核和典型二级结构等特征得到众多科研人员的青睐,被视为研究和验证各种力场精确度和新算法有效性等的理想模型.

在该模型蛋白被设计提出的近18年里,人们已通过理论计算和实验对其结构的确切折叠机制进行了广泛研究,但关于其详细折叠事件和路径仍然存在分歧,尤其是结构中一些残基间相互作用以及各种外在因素对其结构折叠稳定性的影响^[11],而这些问题的阐明可为优化蛋白结构和多肽设计提供重要的理论依据.尤其是在使用Exenatide(艾塞那肽, Trp-cage的前体Exendin-4具有与胰高血糖素样肽-1(glucagon-like peptide-1, GLP-1)相似的降血糖活性,目前已被研制成为首个“模仿肠降血糖素”治疗II型糖尿病的药物)注射液治疗的患者发生急性胰腺炎的情况下^[12],对模型蛋白Trp-cage的研究,特别是残基间相互作用的探索,可为Exenatide的优化提供重要的理论依据.另外,模型蛋白Trp-cage的设计发展历程是众多中小型蛋白质十几年不断优化改进的缩影,对其结构体系和折叠特性的研究概述也可以为理解其他蛋白结构和自我折叠规律提供重要的理论参考^[13,14].对模型蛋白残基间相互作用特性及其外在因素在其结构稳定与折叠转变中作用的系统概述,有助于人们理解该蛋白结构快速折叠和稳定形成的驱动力成因,同时也为新方法、新力场、新算法的验证和优化提供有效参考.本文以中小型蛋白质的代表——迷你蛋白Trp-cage体系为模型,对其结构残基间相互作用及外在因素对其结构稳定性和折叠转变机制等方面所取得的进展进行了系统分析,并对其结构优化和进一步设计进行了展望.

1 模型蛋白Trp-cage及其突变体结构体系

模型蛋白Trp-cage(TC5b, PDB ID: 1L2Y, 序列: NLYI QWLKDG GPSSG RPPPS)是Andersen课题组^[15]基于Exendin-4(PDB ID: 1JRJ^[16])经过肽链截取和突变设计的一个由20个氨基酸残基组成的人工蛋白.如图1A所示,其拥有3个二级结构单元,分别为 α -螺旋

(L2-D9)、 3_{10} -螺旋(G10-S14)和C-端 P_{II} 螺旋(P17-P19), N-端 α -螺旋上的残基Y3, W6, L7和4个脯氨酸残基(P12, P17, P18和P19)形成了一个以W6为中心的三级疏水核结构.该模型蛋白结构典型且可快速折叠(在pH为7,温度为22.7℃的水溶液中约4 μ s)^[17],自被设计以来就成为实验和模拟研究蛋白质折叠相关问题的理想模型^[1,8].尽管该模型蛋白仅由20个氨基酸组成,但其结构中残基间相互作用极为复杂且侧链类型多样.一方面,结构中存在盐桥(D9-R16)、疏水作用(Y3, W6, L7, P12, P18和P19)、静电作用(D1-K8(TC10b))、阳离子- π 作用(W6-R16)和密集的氢键网络等多种非键相互作用(图2).其中氢键网络由四部分组成:(i)残基骨架间氢键网络(I4 $\rightarrow$ N1, Q5 $\rightarrow$ N1, Q5 $\rightarrow$ L2, W6 $\rightarrow$ L2, L7 $\rightarrow$ Y3, L7 $\rightarrow$ I4, K8 $\rightarrow$ I4, D9 $\rightarrow$ W6, G11 $\rightarrow$ W6, G10 $\rightarrow$ L7, G11 $\rightarrow$ L7, S13 $\rightarrow$ G10和S14 $\rightarrow$ G11);(ii)残基骨架与侧链间氢键网络(W6 $\rightarrow$ R16和S14 $\rightarrow$ D9);(iii)残基侧链与侧链间氢键网络(K8 $\rightarrow$ Q5, R16 $\rightarrow$ D9和R16 $\rightarrow$ S14);(iv)结构表面极性残基与溶剂水分子间氢键网络(N1, S13和S20).另一方面,该结构侧链位置类型多样,可分为四种:外置摆动侧链、内置摆动侧链、外置稳定侧链和内置稳定侧链^[11].

正是由于TC5b的典型结构及其残基间相互作用与侧链位置关系的复杂性,很多研究者通过残基点突变的方法对一些关键残基侧链的功能和修饰优化进行了探讨.如Andersen课题组^[18]以TC5b为基础进行了一系列突变实验发现,由L2A和I4A双突变构建的TC9b(序列: NAYAQ WLKDG GPSSG RPPPS)折叠态结构高度稳定,并且使用D1作为N-cap的结构比使用N1作为N-cap的结构更稳定.因此他们在TC9b的基础上又进行了N1D突变最终获得TC10b(PDB ID: 2JOF^[18], 序列: DAYAQ WLKDG GPSSG RPPPS).TC10b和TC5b拥有相同的二级结构和三级结构(图1B),但是TC10b的残基D1、A2和A4均有效提升了其结构稳定性.因此,TC10b作为一种更稳定的Trp-cage同样被用于许多实验和模拟研究^[19,20],但与TC5b相比,TC10b的研究仍相对较少,这可能与模型蛋白TC5b和TC10b的相似性非常高但TC5b更为典型有关.

Retro Trp-cage(PDB ID: 2LUF, 序列: SPPPR GSSPG GDKLW QIYLN)^[21]是将Trp-cage的N端和C端序列完全逆转所构建的一种特殊的Trp-cage,它保留了Trp-cage原有的序列组成和顺序. Retro Trp-cage在

图 1 模型蛋白Trp-cage体系结构示意图。所示意的结构分别从蛋白质结构数据库中下载得到, 且这些结构均是核磁共振实验测定; 体系中的 α -螺旋, 3_{10} -螺旋和 P_{II} -螺旋分别用红色、黄色和蓝色表示; 疏水残基侧链采用棒状模型表示

Figure 1 Ribbon diagram of the Trp-cage structures. These illustrated structures are downloaded from the PDB database, which are obtained by the nuclear magnetic resonance (NMR) experiment. The secondary units including α -helix, 3_{10} -helix and P_{II} -helix are shown in red, yellow and blue, respectively. The hydrophobic side chains are shown as stick model

图 2 模型蛋白Trp-cage(TC5b)的氢键网络示意图

Figure 2 Ribbon diagram of hydrogen-bonding interaction network of Trp-cage mini-protein (TC5b)

水溶液中并不存在稳定的构象, 但在30%的2,2,2-三氟乙醇(2,2,2-trifluoroethanol, TFE)溶液中形成了与Trp-cage(TC5b)类似的相对稳定的结构, 其拥有一个 α -螺旋(S8-I17)(图1C)。模型蛋白Trp-cage结构体系除上述提到的TC5b, TC10b和Retro Trp-cage三个具有代表性结构外, 还有很多体系结构, 如TC12b和TC16b等, 其详细的结构序列见表1。除上述体系结构外, 该模型蛋白的很多突变体结构也都被进行了深入研究, 本文将所

有的突变体及其参考文献汇总于表2中。

2 模型蛋白Trp-cage折叠模型的研究

2.1 两态折叠模型

研究人员通过近18年的理论计算和实验对Trp-cage的确切折叠机制进行了广泛且深入的研究, 但关于其详细折叠事件和路径仍然存在一定的分歧。通过模拟和实验研究获得的大量Trp-cage折叠轨迹可以简单地表征为两态折叠和非两态折叠模型。简单的两态折叠模型是指在折叠过程中仅存在未折叠态和折叠

表 1 模型蛋白Trp-cage结构体系^{a)}

Table 1 The system of Trp-cage mini-protein model^{a)}

结构体系	序列
EX-4 ^[16]	-SKQMEEEAVRLFIEWLKNNGPSSGAPPPS
TC1a ^[15]	SEDEAVRLFIEWLKNNGPSSGAPPPS
TC1c ^[15]	SEDEAVRLFIEWLKNNGPSSGADPPS
TC2 ^[15]	NEVELFIRWLKNNGPSSGAPPPS
TC3b ^[15]	NLFIEWLKNNGPSSGAPPPS
TC4a ^[15]	DLFIEWLKNNGPSSGRPPPS
TC4c ^[15]	KGLFIEWLKNNGPSSGRPPPS
TC5a ^[15]	NLFIQWLKDGGPSSGRPPPS
TC5b ^[15]	NLYIQWLKDGGPSSGRPPPS
TC6a ^[22-25]	NLYIQWLKNNGPSSGRPPPS
TC6b ^[22-24,26]	NLYIQWLKEGGPSSGRPPPS
TC8a ^[18]	NLYAQWLKDGGPSSGRPPPS
TC8b ^[18]	DLYAQWLKDGGPSSGRPPPS
TC9b ^[18]	NAYAQLKDGGPSSGRPPPS
TC10b ^[18]	DAYAQLKDGGPSSGRPPPS
TC11b1 ^[18]	GAYAQLKDGGPSSGRPPPS
TC12a ^[18]	AYAQLKDGGPSSGRPPPS
TC12b ^[18]	NYAQLKDGGPSSGRPPPS
TC12b ^[18]	NYAQLKDGGPSSGRPP-NH ₂
TC13b ^[27]	DAYAQLADGGPASGRPPPS
TC15b ^[27]	DAYAQLADaGPASGRPPPS
T ² C15b ^[27]	DAYAQLADaGWASGRPPPS
TC16b ^[27,28]	DAYAQLADaGPASaRPPPS
trTC16b ^[27]	Ac-AYAQLADaGPASaRPPPS-NH ₂
Retro Trp-cage ^[21,29,30]	SPPPRGSSPGDKLWQIYLN

a) 符号“-”表示蛋白序列的其他省略部分; 残基D-Ala突变在序列中显示为“-a”; TC5b的D9N突变在文献中为TC6a^[24]或TC5c体系^[25], 在此统称为TC6a体系; TC5b的D9E突变体称为TC6b体系^[24]

表 2 模型蛋白Trp-cage突变体^{a)}

Table 2 The mutations of Trp-cage mini-protein model^{a)}

结构体系	突变体
TC5b	W6F ^[15,31-33] , W6H ^[33] , W6G ^[34] , L7W ^[35] , K8A ^[36] , D9A ^[36] , D9S ^[23,26] , G10A ^[37] , G10Q ^[37] , G10 ^D A ^[37,38] , G10 ^D Q ^[37,38] , G10 ^D N ^[37,38] , GLR10 ^[39] , GLR11 ^[39] , GLR15 ^[39] , P12W ^[34,35,40] , R16A ^[23,36] , R16hR ^[23] , D9N&R16A ^[23] , D9Aad&R16K ^[23] , R16W ^[40] , P18W ^[35] ,
TC9b	Q5A ^[18] , K8A ^[18] , D9N ^[18] , S14T ^[18] , R16Nva ^[18]
TC10b	Y3F ^[18] , Y3L ^[18] , Y3A ^[18,41] , L7A ^[18,41] , K8A ^[18,28,41] , D9E ^[42] , G10 ^D A ^[28] , P12A ^[18] , S13A ^[18,28] , S14A ^[18] , G15 ^D A ^[28] , R16K ^[18,42] , R16Om ^[42] , D9E&R16K ^[42] , D9R&R16E ^[42] , D9E&R16Om ^[42] , P17A ^[18] , P17G ^[18] , P12A&P17A ^[18] , P18L ^[18] , P18A ^[18] , P19A ^[18]
TC11b1	Y3F ^[18]
trTC16b	Q5A ^[27] , D9N ^[27] , Q5A&D9N ^[27] , Q5A&R16Nva ^[27] , R16Nva ^[27] , R16Cit ^[27] , Q5A&D9N&R16Nva ^[27]
系列突变	甘氨酸残基点突变(除G10、G11和G15) ^[43] , 丙氨酸残基虚拟筛选点突变(除P12和P17-19) ^[44] , 丙氨酸残基点突变和连续4个丙氨酸突变 ^[45]

a) Aad为 α -氨基己二酸, hR为高精氨酸, Orn为鸟氨酸, Nva为正缬氨酸, Cit为瓜氨酸, D-氨基酸突变在命名中显示为D, GLR代表甘氨酸自由基, TC5b的P12W突变也称为Trp²-cage

态两种结构态, 不形成中间态结构, 显示为单指数弛豫动力学^[17]。Qiu和Streicher等人^[17,46]使用差示扫描量热法(differential scanning calorimetry, DSC)、圆二色谱(circular dichroism, CD)和激光温度跃迁(laser temperature-jump, LT-J)光谱等实验方法, 以及Snow和Day等人^[47,48]分别通过Folding@Home项目和副本交换分子动力学(replica exchange molecular dynamics, REMD)模拟均表明, Trp-cage的折叠符合两态折叠模型; Jungwirth课题组^[49]通过实验(CD、DSC和核磁共振(nuclear magnetic resonance, NMR))和MD模拟相结合的方法也发现, Trp-cage虽然在尿素和胍的变性过程中各个方面具有不同的敏感性, 但其始终指向简单的双态展开过程; 吴云东课题组使用基于经典力场优化后获得的RSFF(residue-specific force field)系列力场^[50,51]对众多小型蛋白进行REMD模拟也发现, Trp-cage(TC5b和TC10b)整体上符合两态折叠模型^[41,52]。同样是研究TC10b, Pang^[53]使用FF12MC方法分别在280 K和300 K两种温度下对TC10b进行模拟, 发现其遵循两态折叠模型^[4], 这与Andersen课题组^[54]的NMR实验结果相一致。

2.2 非两态的折叠模型

与两态折叠相反, 非两态折叠模型是指在折叠过程中除未折叠态和折叠态结构外还额外形成中间态结构, 通过更复杂的分级折叠来完成整个折叠过程。Zhou^[55,56]利用REMD模拟发现了一种包含两个疏水核的中间态结构, 分别由Y3, W6, L7和四个脯氨酸残基

(P12, P17, P18和P19)组成, 并通过D9-R16盐桥分隔和稳定该中间态结构, 而Juraszek和Han等人^[57-60]也通过过渡路径采样(transition path sampling, TPS)和混合分辨率模型(hybrid-resolution model, PACE)证实了这一中间态结构的存在。但是2005年, Ahmed等人^[61]通过紫外共振拉曼光谱(UV resonance Raman spectroscopy, UVRS)却发现了另一种紧凑的折叠中间态结构, 其拥有部分折叠的 α -螺旋和比天然态更紧凑的疏水核。这种与天然态十分相似的中间态结构也在之后的荧光相关光谱(fluorescence correlation spectroscopy, FCS)、时间分辨红外光谱(time-resolved infrared spectroscopy, TRIS)实验^[62,63]和MD模拟^[64,65]中被发现。无论是拥有两个疏水核的折叠中间态结构还是与天然态非常相似的熔球状中间态结构都表明, 模型蛋白Trp-cage的折叠并不是简单的两态折叠模型。尽管当前关于模型蛋白Trp-cage是两态折叠还是非两态折叠仍未有定论, 但通过文献统计发现, 大多数研究人员较为认同Trp-cage的折叠并不是简单的两态折叠模型, 而是通过更复杂的中间态结构所介导。

对于模型蛋白Trp-cage的具体折叠事件, 研究人员通过统计相关研究工作发现, 二级结构和疏水核的形成以及残基W6侧链吡咯环的埋藏是Trp-cage折叠的基本步骤, 但是二级结构和疏水核形成的先后顺序仍存在分歧。因此, 以此作为区分的标志, 可将之总结归纳为三种折叠模型: (i) 成核-聚集(nucleation-condensation)模型, 其主要特点为疏水核的塌缩先于 α -螺旋的形成。Chowdhury和Mou等人^[32,66]的MD模拟计算发

现, 疏水核心吡啶环的包埋是Trp-cage折叠的关键限速步骤; Mok等人^[67]使用NMR脉冲标记(NMR pulse-labelling)实验也发现, 非折叠态结构中存在疏水核结构, 且疏水核的塌缩驱动其整个折叠进程. (ii) 扩散-碰撞(diffusion-collision)模型, 其主要特点为 α -螺旋或 3_{10} -螺旋二级结构率先形成, 然后疏水核心再塌缩. Meuzelaar和Peter等课题组^[62,65,68-72]模拟结果表明, 其结构折叠的限速步骤并非疏水核心的包埋, 而是 α -螺旋或 3_{10} -螺旋结构形成于折叠过程的早期, 随后疏水残基逐渐向疏水核心(W6)靠拢塌缩, 最终疏水核心吡啶环完整包埋. (iii) 多种路径折叠(multi-path folding)模型, 即上述两种折叠模型在Trp-cage的折叠过程中同时存在^[10,73,74]. 如图3A所示, Juraszek课题组^[57-59]通过TPS模拟发现, 上述两种模型在其折叠形成路径中同时存在, 并且在所有的折叠轨迹中, 路径 I (成核-聚集模型)约占80%~90%, 路径 II (扩散-碰撞模型)约占10%~20%, 这一观点同样得到了Andryushchenko和Han等课题组^[60,75-77]模拟计算结果的支持. 而Shigeta

课题组^[78-80]的模拟结果虽然也赞同多种路径折叠模型, 但是与Juraszek等人相反, 结果显示路径 II 在折叠过程中占据大多数, 而路径 I 占据少数. 本课题组推测, 这可能与Juraszek课题组使用了OPLSAA力场, 而Shigeta课题组使用AMBER(ff14SB和ff03)力场有关. 除以上三种折叠模型外, 最近Ferguson课题组^[19]针对TC10b提出了一个九个宏观态的马尔科夫模型(图3D). 但是通过研究这九个宏观态之间的结构转化, 本课题组发现, 该马尔科夫模型仍归属于上文总结的成核-聚集模型. 如图3D所示, 所有结构均通过熔球状动力学中心S8与其他亚稳态结构之间相互转换, 并且折叠至天然态S7仅有三条选择路径: S8—S7, S8—S0—S7和S8—S1—S7. 这说明融球状结构S8的形成是TC10b折叠至天然态过程中的限速步骤.

研究人员通过对过去十几年研究工作的统计发现, 多种路径折叠模型占据着绝对优势, 并且发现, 由盐桥稳定两个疏水核的中间态结构常发生在扩散-碰撞模型中. 盐桥的早期形成不仅可以加速Trp-cage的

图3 模型蛋白Trp-cage的折叠路径^[19,57-59,64,81]. A~C: 蛋白的多种折叠路径模型; D: 表示蛋白的九态折叠模型; E: 该蛋白的折叠能量漏斗示意图

Figure 3 The folding pathways of Trp-cage mini-protein model^[19,57-59,64,81]. A~C: Multi-path folding model; D: the nine-state folding model; E: a schematic of the Trp-cage folding funnel

折叠动力学,也可以阻碍Trp-cage的折叠动力学,因此由中间态结构转换至天然态结构的过程中,体系结构需要能量来打破已经形成的盐桥.而与天然态十分相近的融球状中间态结构在扩散-碰撞和成核-聚集模型中均被发现,这可能与这种中间态结构存在于折叠末期有关.

2.3 非两态折叠模型中各类折叠路径之间的转换及其与温度的依赖性

在以上三种折叠模型中,多种路径折叠模型目前更容易被接受和认同,那么在折叠过程中共存的两种折叠路径是否存在相互转换? 2.2小节介绍了Juraszek课题组^[57,58]发现了Trp-cage折叠过程中存在两条共存的折叠路径,此外,在TPS模拟中Juraszek等人还观察到N-L和N-P_d-I途径之间的转换事件(图3B).这些转换通过N-P_d-I路径,然后快速平衡到另一个主要路径,并且从N-P_d-I到N-L的概率(0.029)比其他方式(0.007)高4倍,这些概率与在折叠整体中L-N路径是I-N路径的4倍相一致.同样Debenedetti课题组^[64]通过扩展映射(diffusion maps)的方法对Trp-cage模拟轨迹进行分析,也发现了两条共存的折叠路径之间存在相互转换的现象(图3C),并且结构转化主要发生在HLX-I(路径II中间态)和LOOP-I(路径I中间态)结构之间,这两个中间态结构均包含了完整的 α -螺旋.尽管这些转换事件的具体细节不太一致,但可确定两种折叠路径的转换多发生在折叠中间态之间,即折叠过程中局部势能最小值之间的切换,这证明了模型蛋白Trp-cage可以通过多个折叠路径模型完成自我折叠过程.

在模拟过程中适当提高模拟温度能加快构象转化速率,进而有效促进结构跨越能垒实现更为全面的构象采样,因此利用REMD模拟通过不断调温扩大构象采样空间,可快速搜索全局能量最低的稳定平衡结构^[82].尽管不同模拟温度条件不会改变蛋白结构的折叠路径,但会影响模型蛋白Trp-cage对不同折叠路径模型的选择. Andryushchenko和Chekmarev^[75,77]利用蛋白质折叠的流体动力学(hydrodynamic, 折叠过程被认为是折叠流体的运动)探讨了两种折叠路径在不同温度下的相对贡献.结果表明,随着温度的升高,在285 K和300 K时占折叠流总量90%左右的路径I在325 K时下降到约10%,即路径II占据主导地位.而在315 K时,两种路径的贡献近似相等.所有这些温度都低于计算

的熔点(420 K),这表明Trp-cage的折叠机制是由动力学因素而非热力学决定的(图4).

2.4 蛋白折叠类型与力场的关系

目前针对模型蛋白Trp-cage结构折叠模拟计算实验所采用的力场涵盖了现有的常用力场: OPLSAA/L^[57]、AMBER系列(ff03, ff99, ff99SB-ildn, ff99SB, ff96, ff94等)^[10,68,73,78,83]、CHARMM系列(22, 22CMAP, 36等)^[74,75,84]、GROMOS96系列(43a1, 53a6等)^[57,85,86]以及相关课题组优化开发的分子力场(如PACE, CVFF, GAFF, RSFF系列等)^[8,60,87,88].尽管能够很好地利用这些分子力场研究蛋白质结构,然而它们的经验势函数对原子间非键相互作用的描述精度差异显著,尤其是对相互作用庞大复杂的蛋白质结构,计算误差会随着原子相互作用数目增大而迅速积累,从而导致模拟计算结果错误.例如,在GROMOS96 43a1力场下Trp-cage的天然折叠态结构并不能稳定形成,模拟计算得到的全局能量最小构象和NMR实验结构存在较大差异^[57];在AMBER ff96和Amber99SB-ildn力场下Trp-cage甚至会错误折叠形成一个稳定的 β -发卡结构^[10,83].本课题组通过采用OPLSAA/L力场进行常规动力学模拟发现,Trp-cage也会错误折叠形成稳定的 β -发卡结构.研究还发现,目前还存在一些力场对Trp-cage折叠热力学行为的描述不够准确.例如,Zhou^[56]和Paschek等人^[89]分别发现,OPLSAA/L和AMBER ff94力场计算得到的折叠熔融温度为400 K和450 K,显著高于实验观察值315 K^[15,17,18].尽管Garcia课题组^[48]也发现,AMBER ff99SB力场在Trp-cage折叠过程中计算得到的熔融温度、自由能和比热容(specific heat)和实验值具有很好的一致性,但是熵和焓变化量与实验观察值存在显著差异.

正是由于当前分子力场的精度对其不同结构的相互作用描述存在差异,所以不同力场仍然会对折叠机制产生不同的倾向.正如Shao和Paschek等人^[83,90]所指出的,Trp-cage稳定性和中间结构对蛋白质力场和水模型选择具有显著依赖关系.本课题组统计发现,采用OPLSAA/L或CHARMM(22, 27)力场模拟结果都更倾向于支持成核-聚集模型和多种路径折叠模型;而AMBER力场对上述三种折叠模型存在不同程度的支持,这可能与其力场版本较多有一定关系,但它更倾向于支持扩散-碰撞模型,这可能与AMBER(ff03, ff99SB,

图 4 模型蛋白Trp-cage在285 K, 300 K, 315 K和325 K温度下的折叠动力学网络^[75-77]. 箭头线的宽度与聚类间的转换数成正比, 这里仅表示“主要”转换, 且其转换数量大于10

Figure 4 The Trp-cage folding kinetic network at temperatures of 285 K, 300 K, 315 K and 325 K^[75-77]. The widths of the arrow lines are proportional to the net numbers of inter-cluster transitions. Only “main” transitions are displayed, for which the net number of transitions is larger than ten

ff94)力场过分强调螺旋结构稳定性有关^[65]. 此外, 现有模拟结果已证实, 使用这些力场大部分轨迹可正确折叠成天然态结构, 但也存在一些轨迹发生解折叠或折叠为假天然态结构(pseudonative structure)(图3E)^[72,81,91]. 因此, 现有力场的参数精度能否客观有效地描述残基间相互作用及其对蛋白质结构的影响仍是一个重要的问题.

3 模型蛋白Trp-cage残基互作对其结构和折叠转变过程的影响

3.1 带电残基对结构和折叠转变过程的影响

在模拟计算实验中为提高蛋白质的稳定性, 降低末端残基对结构折叠的影响, 研究人员通常对蛋白质结构进行封端处理, 即分别在蛋白质结构的N-端添加乙酰基(Ac-)、C-端添加甲基(-CH₃)以消除末端点电荷. 然而在正常生理条件下蛋白质的末端携带有两性离子, 即具有带正电荷的N-端(-NH₃⁺)和带负电荷的C-端(-COO⁻). 因此, 在模拟体系中保留蛋白质两端两性离子, 使之与生理条件下的结构保持一致具有重要意义. English和García^[92]使用REMD模拟对模型蛋白Trp-cage的末端点电荷对其折叠的影响进行了探索并发

现, 相比于封端, 带电端的加入使Trp-cage模拟平衡时间增加了约2.5倍, 并且还在折叠过程中额外形成了两个亚稳态(折叠中间态和具有 β -片层的构象态). 与提高温度不改变蛋白质的折叠路径类似, Trp-cage末端电荷态的变化仅增加了模拟平衡的时间长度, 并不改变最终折叠构象. 这同样表明, 即使Trp-cage这样的小型蛋白质, 其折叠自由能地形图也可能是极其复杂的, 并且会由于序列电荷态的简单变化而发生显著改变.

在模拟体系中, 模型蛋白Trp-cage不仅存在C-端和N-端的点电荷, 其自身结构体系中还包含了K8, D9和R16三个携带电荷的氨基酸残基, 并且残基D9-R16形成盐桥相互作用. 正如2.2节折叠机制中所述, 在折叠过程中存在由盐桥稳定的折叠中间态结构. 并且Perczel课题组^[22,23]通过CD、NMR实验和MD模拟也发现, 与拥有盐桥的结构(Tc5b或D9E)相比, 缺少盐桥的D9N和D9S突变体消除了COO⁻侧链, 使折叠中间体失稳, 而在保留COO⁻侧链的突变体(D9E)中, 却未发现其对中间态稳定性有显著影响, 并且在酸性pH条件下野生型Trp-cage中间体的数量显著高于盐桥缺失突变体. 这些研究都表明, D9-R16盐桥的形成对Trp-cage折叠中间态的形成和稳定发挥重要的作用^[15,17,87]. Kim等人^[64]通过对模拟轨迹分析发现, 尽管盐桥的早期形成

可以加速Trp-cage的折叠动力学, 但是这也成为结构后期进一步折叠转变的阻碍, 因为中间态结构折叠转换至天然态结构必须打破盐桥, 这就导致了长期存在的亚稳定中间态. Andersen课题组^[42]通过NMR和CD实验检测了D9E, D9R&R16E, R16Orn(Orn为鸟氨酸), D9E&R16Orn, R16K和D9E&R16K共6种TC10b盐桥突变体发现, D9/R16盐桥在折叠协同性和折叠稳定性方面是最优的, 并且电离的D9显著降低了螺旋结构的固有螺旋度, 并认为这是保持结构折叠稳定性的必要条件. 当然, 在对盐桥突变体的探索中也形成了一些相对稳定的体系, 例如TC6a^[24](TC5b的D9N突变体), Wu和Narsimhan^[25]称之为TC5c和TC6b(TC5b的D9E突变体).

另一个携带正电荷的残基K8并未参与疏水核的形成, 但是其形成的氢键相互作用有助于增强 α -螺旋的稳定. 吴晓敏和蒋帆等人^[41,43]分别对残基K8突变体K8G(TC5b)和K8A(TC10b)进行研究, 发现K8突变均导致了 α -螺旋的稳定性变差且形成较晚, 但是这对最终结构的影响不大(均方根偏差(root mean squared deviation, RMSD): K8G为1.4 Å; K8A为0.33 Å). 这些结果均表明, K8在Trp-cage折叠过程中并未发挥主要作用, 但是其残基氢键对螺旋结构的稳定性具有一定影响. 通过总结发现, 末端残基电荷态的改变会导致结构折叠平衡时间的延长, 但这并未影响其最终形成的折叠结构; 盐桥在模型蛋白Trp-cage的折叠转变和结构稳定中发挥重要作用, 而携带正电荷的残基K8虽然未对最终的折叠结构产生影响, 但其对局域二级结构 α -螺旋的稳定性仍具有重要贡献.

3.2 疏水残基对结构和折叠转变过程的影响

在模型蛋白Trp-cage的折叠过程中, 残基W6侧链吲哚环与残基Y3、P12和P19侧链之间的疏水相互作用是其结构快速折叠的主要驱动力, 并且残基W6的侧链吲哚环的包埋还是其结构折叠至天然态的关键步骤. Rogne等人^[34]通过构建W6G突变体消除W6侧链吲哚环发现, 关键疏水相互作用的缺失导致结构形成错误的疏水塌缩构象, 同时还阻碍了二级结构的正确形成. 而Chowdhury等人^[32]构建的W6F突变体模拟结果表明, W6F突变体折叠速度非常快, 且与野生型的折叠路径相似, 但其结构稳定性较差. Rogne与Chowdhury结果差异的原因在于残基G6的侧链仅为一个氢原

子, 而残基F6仍存在疏水性的侧链, 但是W6侧链吲哚环的缺失同样无法稳定结构. 在对Trp-cage结构稳定性的探讨中, 吴晓敏等人^[43]通过模拟发现, 构建的W6G突变体快速从天然态构象折叠转变至无规则卷曲构象, 且结构中疏水核松散并发生水合作用, 大量的天然接触与氢键作用都受到破坏. 上述这些研究从不同的角度证明了残基W6吲哚环对整个模型蛋白的折叠与稳定都发挥最为关键的作用.

其他疏水残基(Y3, L7, P12, P17, P18, P19)虽然也参与了疏水核的构建, 但是这些残基对Trp-cage的折叠和稳定的贡献存在差异. Andersen和吴晓敏课题组^[18,43]的研究发现, P19A和P19G突变体均会导致结构出现显著的不稳定, 而且Y3A突变体同P19A突变一样不稳定, 这表明T3/P19, Y3/W6和P19/W6疏水相互作用对于结构稳定性发挥着同等重要的作用^[44]. 尽管Y3A突变体具有低稳定性, 但这并未影响Trp-cage(TC10b)折叠至近天然构象(RMSD为0.42 Å)^[41]. 与残基Y3类似, 残基L7(突变体L7G/L7A^[18,43])同样被发现对结构稳定性具有重要影响, 但是对结构折叠并未产生较大的影响^[41]. 而其他三个脯氨酸的杂环吡咯烷侧链同样参与了疏水核的形成, 但通过构建突变体发现, 其对于结构稳定的影响显著低于残基P19^[43]. 在对脯氨酸的研究过程中, Gai课题组^[40]基于W6/W12叠合可产生更强的疏水效应并极大地提高折叠速率(1 μ s)设计了“Trp²-cage”(突变体P12W), 这种设计还加强了残基的侧链包埋程度, 使其 T_m 值(melting temperature)值由42°C增加至56.9°C^[34,35,54].

3.3 氢键网络对结构和折叠转变过程的影响

密集且复杂的氢键网络同样对Trp-cage的折叠转变和结构稳定发挥着重要作用. Barua和Andersen^[33]发现, W6H突变会破坏埋藏的W6侧链吲哚环上的氨基与R16的羰基氧(W6-indole-NH ϵ 1 $\rightarrow$ O=C-R16)原子之间的长程氢键, 导致部分暴露在水溶剂中. 而在野生型中, 这种埋藏的氢键对W6吲哚环的稳定性和折叠形成都有显著贡献, 而末端残基与溶剂水分子之间的氢键作用却并未对Trp-cage的稳定性产生显著的影响. 通过对N1G、L2G和S20G三个突变体的模拟发现, 这些突变体的优势构象中仍具有完整紧凑的疏水核. 这表明Trp-cage结构中末端残基突变不会引起结构稳定性的降低, 但会引起结构两端局域构象自由度的增加, 这

是由于末端残基与溶剂水发生的氢键作用以及侧链摆动作用远低于邻近残基Y3/P19之间的疏水叠合稳定作用。但是通过对末端残基引入额外的电荷却会对结构折叠和稳定产生显著的影响。如Taricska等人^[93]构建的N1R突变体并未与C-端形成盐桥相互作用, 反而由于 α -螺旋偶极子与R1的正电荷冲突导致了88%构象总体呈U型未折叠态。而TC9b的N1D突变体(TC10b)^[18]具有更大的螺旋N-cap倾向(特别是在其电离状态下), 稳定性获得显著提高。

此外, 3_{10} -螺旋结构上的残基对结构的折叠和稳定也具有一定的作用, 如Andersen课题组^[15]在设计Trp-cage时发现, G11在不同的Trp-cage突变体中都是保守的, 因此G11的氨基与W6、L7的羰基氧(G11-NH $\rightarrow$ O=C-W6和/或G11-NH $\rightarrow$ L7-C=O)之间不断交换形成包埋在结构内部的长程氢键, 与前文提及的W6侧链吡啶环上的氨基与R16的羰基氧形成的氢键对Trp-cage的折叠形成发挥重要的协同作用。Nanda和Andersen等课题组^[28,37,38,54]通过构建残基G10突变体G10^DA和G10^DQ发现, L-型氨基酸被D-型氨基酸替换降低构象熵值来增加结构的热力学稳定性。另外, 具有相同侧链的氨基酸残基(S13和S14)由于其侧链所处位置的不同导致其对结构稳定性的贡献程度也不同。S14侧链上的羟基处于模型蛋白Trp-cage天然态结构的内部, 与其他氨基酸侧链形成了较强的内部氢键作用(S14-OH $\rightarrow$ O=C-D9, R16-NH $\rightarrow$ OH-S14)和天然接触(S14-D9, S14-G10, S14-G11, S14-P12, S14-R16)。在突变体S14G中, 由于被侧链仅有氢原子的甘氨酸所替换, 导致这些特殊的稳定结构的作用力受到破坏, 因此S14的点突变导致了结构中N-端 α -螺旋局部结构解旋松散和S14所涉及形成氢键的断裂。残基突变于S14引起结构不稳定的程度远超过突变于S13上。尽管氢键作用对整个结构折叠稳定的贡献是相当小的, 每个氢键作用形成的能量仅在5 kJ级或更少, 这也反映出结构在去折叠态中与溶剂水发生氢键作用的补偿效应^[94], 而蛋白结构从去折叠态到折叠态转变形成中, 这种极性基团与溶剂水形成氢键作用被结构内部极性基团间形成的氢键所取代需要的成本则会大于20 kJ^[95]。因此, 氢键作用对于蛋白结构稳定与折叠转变影响的程度一方面取决于残基侧链特性本身, 另一方面与其所处的位置和朝向以及侧链本身稳定度都存在紧密联系。

4 外在环境因素对其折叠态结构稳定和形成过程的影响(温度、压力和pH)

当外部环境因素(温度、压力和pH等)改变时, 蛋白质的稳定性将不可避免的受到影响, 其结构的变化必定会导致生物学功能的损失甚至丧失。因此外部环境因素对蛋白质折叠和稳定的影响的研究有助于人们理解在极端环境(如低温、高温、高压和低pH等)下生物的生长代谢机制, 同时为蛋白质类食品、生物制品的制备、储存和运输提供合理的设计改进方向(特别是治疗糖尿病所需的胰岛素和艾塞那肽(Trp-cage的前体)等蛋白类药物仍需在低温下储存使用)。模型蛋白Trp-cage的实验 T_m 值为315 K, 这意味着在温度高于315 K时, Trp-cage会出现变性展开^[96,97], 而在更高的温度下, Trp-cage结构完全解折叠变性且二级和三级结构完全丧失, 结构内部的疏水残基均暴露于溶液中^[98]。

与高温吸热变性不同, Kim和Taricska等人^[93,98]发现, 在210 K的低温下, Trp-cage变性是由于肽链结构局部发生解折叠, 进而导致结构内分子间接触减少, 残基间距离增大(W6-S14), 蛋白-水氢键数量增加, 并且D9-R16盐桥缺失; 但其中二级结构 α -螺旋在低温中能够继续保留, 且结构稳定性有所增加, 虽然在低温条件下也可观察到 3_{10} -螺旋, 但是其位置向C-端位移(3_{10} -螺旋中心向S14移动), 并且这种位移伴随着水侵入Trp-cage核心和内部埋藏的疏水残基发生水合作用。另外, 通过计算显示, 低温展开Trp-cage能量的主要贡献来自于亲水性残基的水化, 但这种低温引起的结构解折叠变性既没有导致溶剂可及表面积(solvent accessible surface areas, SASA)的显著增加, 也没有导致回旋半径(radius of gyration, R_g)的显著增加。这也进一步支持了冷变性导致分子内相互作用放松而不是蛋白质骨架大量展开的观点^[93]。在冷变性的过程中, Trp-cage可保留紧凑的天然结构, 形成一种渗透溶剂的状态, 而不是展开。但是在高压下, 冷变性导致二级和三级结构产生更大的破坏, 此时冷变性和热变性表现出渐进的相似性^[99]。如图5所示, 在压力-温度(P-T)平面内, 蛋白质的稳定性呈椭圆区域。在室温及以上的温度下, 结构的稳定性随着压力的增加反而降低, 但在较低的温度下(250 K), 结构对压力呈现非单调依赖的关系。

除温度和压力外, 另一个对蛋白质具有重要影响

的因素是pH值. Taricska等人^[93]通过NMR实验发现, 当pH<3时, 模型蛋白Trp-cage的主导构象与其冷变性时的构象非常相似, 并且结构中的D9-R16盐桥消失, 同时结构中还存在1%的无规卷曲构象. 与Taricska等人的研究结果相一致, Kitazawa等人^[100]同样通过NMR实验发现Trp-cage在pH<3时光谱色散值较低, 表明结构中存在无规卷曲构象. Andersen课题组^[27]的最新文献表明, 在Trp-cage折叠状态下, N-cap的D1残基(TC10b-TC16b)、D9-R16盐桥、肽链两端残基之间的相互作用以及残基D9侧链参与的相互作用等众多稳定结构的相互作用都具有很强的pH依赖性. 因此, 他们对Trp-cage结构体系进行优化并最终获得了TC13b, TC15b, T²C15b和TC16b等众多在pH=2.5下稳定的Trp-cage体系.

除本文讨论概述的温度、压力和pH三种外在因素对Trp-cage结构和折叠的影响外, 还有很多外在环境因素对其结构产生影响. 例如, 模型蛋白Trp-cage在无机和有机材料表面^[101~105]、拥挤环境^[106~108]和溶剂中含有变性剂、稳定剂^[49,84,88,109~118]、分子伴侣^[119,120]等一系列环境因素下的折叠动力学研究, 将为人们全面深入理解结构内因与外因协同作用影响蛋白结构与功能提供重要参考.

5 总结与展望

蛋白质残基间相互作用、结构折叠转变及其二者关系一直是结构生物学的研究重点. 本文以模型蛋白Trp-cage为例首先详细介绍了已报道的该蛋白所有结构体系类型和各自典型特征, 重点描述了TC5b, TC10b和Retro Trp-cage三个重要结构体系, 并以经典模型TC5b和TC10b为例详细分析阐明了其结构组成、残基间相互作用类型及其侧链位置属性. 针对该模型蛋

图5 在压力-温度(P-T)曲线下模型蛋白Trp-cage的稳定性变化示意图^[99]

Figure 5 Schematic diagram showing the change of Trp-cage stability under the pressure-temperature (P-T) curve^[99]

白结构折叠模型类型的研究予以归纳, 可分为两态和非两态折叠模型两类, 并将后者非两态折叠模型按照折叠形成事件的差异进一步分为成核-聚集模型、扩散-碰撞模型和多种路径折叠模型. 同时讨论了不同折叠模型之间的相互转换及其与温度、力场之间的关系. 另外, 本文进一步分析了影响模型蛋白结构的内在因素和外在环境因素: 带电残基、疏水残基、氢键网络、温度、压力、pH值. 阐明这些因素对蛋白结构稳定性和折叠动力学行为的影响不仅是各个因素的独立分析探讨, 更重要的是搞清楚这些因素之间是如何协同影响调控蛋白结构及其折叠转变过程. 同时, 细胞内蛋白结构稳定平衡和折叠动力学行为的调控不仅受到上述内外因素的影响, 还会受到拥挤环境、分子伴侣等各种潜在内外因素的共同影响, 这些影响因素及其协同效应的阐明不仅有助于人们理解蛋白结构平衡稳定和折叠转变的驱动力成因, 也为进一步结构调控和优化设计提供了理论指导.

参考文献

- 1 Dill K A, MacCallum J L. The protein-folding problem, 50 Years On. *Science*, 2012, 338: 1042-1046
- 2 Li C D, Junaaid M, Chen H, et al. Helix-switch enables C99 dimer transition between the multiple conformations. *J Chem Inf Model*, 2019, 59: 339-350
- 3 Junaaid M, Khan M T, Malik S I, et al. Insights into the mechanisms of the pyrazinamide resistance of three pyrazinamidase mutants N11K, P69T, and D126N. *J Chem Inf Model*, 2019, 59: 498-508
- 4 Pang Y P. How fast fast-folding proteins fold in silico. *Biochem Biophys Res Commun*, 2017, 492: 135-139

- 5 Wong-Ekkabut J, Karttunen M. The good, the bad and the user in soft matter simulations. *Biochim Biophys Acta*, 2016, 1858: 2529–2538
- 6 Abraham M J, Murtola T, Schulz R, et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 2015, 1: 19–25
- 7 Freddolino P L, Harrison C B, Liu Y, et al. Challenges in protein-folding simulations. *Nat Phys*, 2010, 6: 751–758
- 8 Jiang F, Wu H N, Kang W, et al. Developments and applications of coil-library-based residue-specific force fields for molecular dynamics simulations of peptides and proteins. *J Chem Theor Comput*, 2019, 15: 2761–2773
- 9 Peter E K. Adaptive enhanced sampling with a path-variable for the simulation of protein folding and aggregation. *J Chem Phys*, 2017, 147: 214902
- 10 Lemke T, Peter C. EncoderMap: dimensionality reduction and generation of molecule conformations. *J Chem Theor Comput*, 2018, 15: 1209–1215
- 11 Wu X M, Yuan X H, Xue S L, et al. Research progress of the Trp-cage formation and its folding mechanism (in Chinese). *Acta Phys-Chim Sin*, 2013, 29, 1842–1850 [吴晓敏, 袁晓辉, 薛书蕾, 等. 模型多肽Trp-cage折叠形成机制的研究进展. *物理化学学报*, 2013, 29: 1842–1850]
- 12 Singh S, Chang H Y, Richards T M, et al. Glucagonlike peptide 1-based therapies and risk of hospitalization for acute pancreatitis in type 2 diabetes mellitus. *JAMA Intern Med*, 2013, 173: 534–539
- 13 Lv C, Qiang S M, Sun T T, et al. Computational design of heterotrimeric collagen helices (in Chinese). *Sci Sin Vitae*, 2019, 49: 615–624 [吕成, 强书敏, 孙添添, 等. 胶原蛋白异源三聚体的理性设计. *中国科学: 生命科学*, 2019, 49: 615–624]
- 14 Wang Y J, Wu C. Oligomerization of yeast prion Sup35NM (in Chinese). *Sci Sin Vitae*, 2018, 48: 662–670 [王艳景, 吴奇. 酵母朊蛋白Sup35NM的寡聚化过程研究. *中国科学: 生命科学*, 2018, 48: 662–670]
- 15 Neidigh J W, Fesinmeyer R M, Andersen N H. Designing a 20-residue protein. *Nat Struct Biol*, 2002, 9: 425–430
- 16 Neidigh J W, Fesinmeyer R M, Prickett K S, et al. Exendin-4 and glucagon-like-peptide-1: NMR structural comparisons in the solution and micelle-associated states. *Biochemistry*, 2001, 40: 13188–13200
- 17 Qiu L, Pabit S A, Roitberg A E, et al. Smaller and faster: The 20-residue Trp-cage protein folds in 4 μ s. *J Am Chem Soc*, 2002, 124: 12952–12953
- 18 Barua B, Lin J C, Williams V D, et al. The Trp-cage: optimizing the stability of a globular miniprotein. *Protein Eng Des Sel*, 2008, 21: 171–185
- 19 Sidky H, Chen W, Ferguson A L. High-resolution markov state models for the dynamics of Trp-cage miniprotein constructed over slow folding modes identified by state-free reversible VAMPnets. *J Phys Chem B*, 2019, 123: 7999–8009
- 20 Miao Y, Feixas F, Eun C, et al. Accelerated molecular dynamics simulations of protein folding. *J Comput Chem*, 2015, 36: 1536–1549
- 21 Vymětal J, Bathula S R, Cerný J, et al. Retro operation on the Trp-cage miniprotein sequence produces an unstructured molecule capable of folding similar to the original only upon 2,2,2-trifluoroethanol addition. *Protein Eng Des Sel*, 2014, 27: 463–472
- 22 Rovó P, Stráner P, Láng A, et al. Structural insights into the Trp-cage folding intermediate formation. *Chem Eur J*, 2013, 19: 2628–2640
- 23 Rovó P, Farkas V, Hegyi O, et al. Cooperativity network of Trp-cage miniproteins: probing salt-bridges. *J Pept Sci*, 2011, 17: 610–619
- 24 Hudáky P, Stráner P, Farkas V, et al. Cooperation between a salt bridge and the hydrophobic core triggers fold stabilization in a Trp-cage miniprotein. *Biochemistry*, 2008, 47: 1007–1016
- 25 Wu X, Narsimhan G. Characterisation of the effect of electrostatic interaction on the structure of Trp-cage using molecular dynamics simulation. *Mol Simul*, 2010, 36: 1086–1095
- 26 Kardos J, Kiss B, Micsonai A, et al. Phosphorylation as conformational switch from the native to amyloid state: Trp-cage as a protein aggregation model. *J Phys Chem B*, 2015, 119: 2946–2955
- 27 Graham K A, Byrne A, Son R, et al. Reversing the typical pH stability profile of the Trp-cage. *Biopolymers*, 2019, 110: e23260
- 28 Williams D V, Barua B, Andersen N H. Hyperstable miniproteins: additive effects of D- and L-Ala mutations. *Org Biomol Chem*, 2008, 6: 4287–4289
- 29 Zerze G H, Stillinger F H, Debenedetti P G. Computational investigation of retro-isomer equilibrium structures: Intrinsically disordered, foldable, and cyclic peptides. *FEBS Lett*, 2020, 594: 104–113
- 30 Bathula S R. Structural comparison between Trp-cage and Retro Trp-cage peptide structures. *MOJ Proteom Bioinform*, 2016, 3
- 31 Aldeghi M, de Groot B L, Gapsys V. Accurate calculation of free energy changes upon amino acid mutation. *Methods Mol Biol*, 2019, 1851: 19–47
- 32 Chowdhury S, Lee M C, Duan Y. Characterizing the rate-limiting step of Trp-cage folding by all-atom molecular dynamics simulations. *J Phys*

- Chem B, 2004, 108: 13855–13865
- 33 Barua B, Andersen N H. Determinants of miniprotein stability: can anything replace a buried H-bonded Trp sidechain? *Lett Pept Sci*, 2001, 8: 221–226
- 34 Rogne P, Ozdowdy P, Richter C, et al. Atomic-level structure characterization of an ultrafast folding mini-protein denatured state. *PLoS ONE*, 2012, 7: e41301
- 35 Anderson J M, Kier B L, Jurban B, et al. Aryl-aryl interactions in designed peptide folds: Spectroscopic characteristics and optimal placement for structure stabilization. *Biopolymers*, 2016, 105: 337–356
- 36 Kim S B, Sparano E M, Singh R S, et al. Microscopic origin of hysteresis in water sorption on protein matrices. *J Phys Chem Lett*, 2017, 8: 1185–1190
- 37 Rodriguez-Granillo A, Annavarapu S, Zhang L, et al. Computational design of thermostabilizing D-amino acid substitutions. *J Am Chem Soc*, 2011, 133: 18750–18759
- 38 Culik R M, Annavarapu S, Nanda V, et al. Using D-amino acids to delineate the mechanism of protein folding: Application to Trp-cage. *Chem Phys*, 2013, 422: 131–134
- 39 Owen, M C, Csizmadia I G, Viskolcz B, et al. Protein stability and unfolding following glycine radical formation. *Molecules*, 2017, 22: 655
- 40 Bunagan M R, Yang X, Saven J G, et al. Ultrafast folding of a computationally designed Trp-cage mutant: Trp²-cage. *J Phys Chem B*, 2006, 110: 3759–3763
- 41 Zhou C Y, Jiang F, Wu Y D. Folding thermodynamics and mechanism of five Trp-cage variants from replica-exchange MD simulations with RSFF2 force field. *J Chem Theor Comput*, 2015, 11: 5473–5480
- 42 Williams D V, Byrne A, Stewart J, et al. Optimal salt bridge for Trp-cage stabilization. *Biochemistry*, 2011, 50: 1143–1152
- 43 Wu X, Yang G, Zu Y, et al. The Trp-cage miniprotein with single-site mutations: Studies of stability and dynamics using molecular dynamics. *Comput Theor Chem*, 2011, 973: 1–8
- 44 Gapsys V, de Groot B L. pmx Webserver: A user friendly interface for alchemistry. *J Chem Inf Model*, 2017, 57: 109–114
- 45 Wang J, Ferguson A L. Recovery of protein folding funnels from single-molecule time series by delay embeddings and manifold learning. *J Phys Chem B*, 2018, 122: 11931–11952
- 46 Streicher W W, Makhatadze G I. Unfolding thermodynamics of Trp-cage, a 20 residue miniprotein, studied by differential scanning calorimetry and circular dichroism spectroscopy. *Biochemistry*, 2007, 46: 2876–2880
- 47 Snow C D, Zagrovic B, Pande V S. The Trp-cage: folding kinetics and unfolded state topology via molecular dynamics simulations. *J Am Chem Soc*, 2002, 124: 14548–14549
- 48 Day R, Paschek D, Garcia A E. Microsecond simulations of the folding/unfolding thermodynamics of the Trp-cage miniprotein. *Proteins*, 2010, 8: 1889–1899
- 49 Heyda J, Kozisek M, Bednarova L, et al. Urea and guanidinium induced denaturation of a Trp-cage miniprotein. *J Phys Chem B*, 2011, 115: 8910–8924
- 50 Jiang F, Zhou C Y, Wu Y D. Residue-specific force field based on the protein coil library. RSFF1: Modification of OPLS-AA/L. *J Phys Chem B*, 2014, 118: 6983–6998
- 51 Zhou C Y, Jiang F, Wu Y D. Residue-specific force field based on protein coil library. RSFF2: Modification of AMBER ff99SB. *J Phys Chem B*, 2014, 119: 1035–1047
- 52 Jiang F, Wu Y D. Folding of fourteen small proteins with a residue-specific force field and replica-exchange molecular dynamics. *J Am Chem Soc*, 2014, 136: 9536–9539
- 53 Pang Y P. FF12MC: A revised AMBER forcefield and new protein simulation protocol. *Proteins*, 2016, 84: 1490–1516
- 54 Byrne A, Williams D V, Barua B, et al. Folding dynamics and pathways of the Trp-cage miniproteins. *Biochemistry*, 2014, 53: 6011–6021
- 55 Zhou R. Exploring the protein folding free energy landscape: coupling replica exchange method with P3ME/RESPA algorithm. *J Mol Graph Model*, 2004, 22: 451–463
- 56 Zhou R. Trp-cage: Folding free energy landscape in explicit water. *Proc Natl Acad Sci USA*, 2003, 100: 13280–13285
- 57 Juraszek J, Bolhuis P G. Sampling the multiple folding mechanisms of Trp-cage in explicit solvent. *Proc Natl Acad Sci USA*, 2006, 103: 15859–15864
- 58 Juraszek J, Bolhuis P G. Rate constant and reaction coordinate of Trp-cage folding in explicit water. *Biophys J*, 2008, 95: 4246–4257

- 59 Juraszek J, Vreede J, Bolhuis P G. Transition path sampling of protein conformational changes. *Chem Phys*, 2012, 396: 30–44
- 60 Han W, Schulten K. Characterization of folding mechanisms of Trp-cage and WW-domain by network analysis of simulations with a hybrid-resolution model. *J Phys Chem B*, 2013, 117: 13367–13377
- 61 Ahmed Z, Beta I A, Mikhonin A V, et al. UV-resonance raman thermal unfolding study of Trp-cage shows that it is not a simple two-state miniprotein. *J Am Chem Soc*, 2005, 127: 10943–10950
- 62 Meuzelaar H, Marino K A, Huerta-Viga A, et al. Folding dynamics of the Trp-cage miniprotein: Evidence for a native-like intermediate from combined time-resolved vibrational spectroscopy and molecular dynamics simulations. *J Phys Chem B*, 2013, 117: 11490–11501
- 63 Neuweiler H, Doose S, Sauer M. A Microscopic view of miniprotein folding: Enhanced folding efficiency through formation of an intermediate. *Proc Natl Acad Sci USA*, 2005, 102: 16650–16655
- 64 Kim S B, Dsilva C J, Kevrekidis I G, et al. Systematic characterization of protein folding pathways using diffusion maps: Application to Trp-cage miniprotein. *J Chem Phys*, 2015, 142: 085101
- 65 Chen W, Ferguson A L. Molecular enhanced sampling with autoencoders: On-the-fly collective variable discovery and accelerated free energy landscape exploration. *J Comput Chem*, 2018, 39: 2079–2102
- 66 Mou L, Jia X, Gao Y, et al. Folding simulation of Trp-cage utilizing a new AMBER compatible force field with coupled main chain torsions. *J Theor Comput Chem*, 2014, 13: 1450026
- 67 Mok K H, Kuhn L T, Goetz M, et al. A pre-existing hydrophobic collapse in the unfolded state of an ultrafast folding protein. *Nature*, 2007, 447: 106–109
- 68 Peter E K, Cerny J. Enriched conformational sampling of DNA and proteins with a hybrid hamiltonian derived from the protein data bank. *Int J Mol Sci*, 2018, 19: 3405
- 69 Shao Q, Shi J, Zhu W. Determining protein folding pathway and associated energetics through partitioned integrated-tempering-sampling simulation. *J Chem Theor Comput*, 2017, 13: 1229–1243
- 70 Shao Q, Zhu W. Effective conformational sampling in explicit solvent with gaussian biased accelerated molecular dynamics. *J Chem Theor Comput*, 2017, 13: 4240–4252
- 71 Lee I H, Kim S Y. Dynamic folding pathway models of the Trp-cage protein. *Biomed Res Int*, 2013, 2013: 1–9
- 72 Marinelli F. Following easy slope paths on a free energy landscape: The case study of the Trp-cage folding mechanism. *Biophys J*, 2013, 105: 1236–1247
- 73 Chen W, Tan A R, Ferguson A L. Collective variable discovery and enhanced sampling using autoencoders: Innovations in network architecture and error function design. *J Chem Phys*, 2018, 149: 072312
- 74 Deng N, Dai W, Levy R M. How kinetics within the unfolded state affects protein folding: An analysis based on markov state models and an ultra-long MD trajectory. *J Phys Chem B*, 2013, 117: 12787–12799
- 75 Andryushchenko V A, Chekmarev S F. Temperature evolution of Trp-cage folding pathways: An analysis by dividing the probability flux field into stream tubes. *J Biol Phys*, 2017, 43: 565–583
- 76 Andryushchenko V A, Chekmarev S F. Turbulence in protein folding: Vorticity, scaling and diffusion of probability flows. *PLoS ONE*, 2017, 12: e0188659
- 77 Andryushchenko V A, Chekmarev S F. A hydrodynamic view of the first-passage folding of Trp-cage miniprotein. *Eur Biophys J*, 2016, 45: 229–243
- 78 Harada R, Yoshino R, Nishizawa H, et al. Temperature-pressure shuffling outlier flooding method enhances the conformational sampling of proteins. *J Comput Chem*, 2019, 40: 1530–1537
- 79 Harada R, Shigeta Y. Hybrid Cascade-type molecular dynamics with a markov state model for efficient free energy calculations. *J Chem Theor Comput*, 2018, 15: 680–687
- 80 Harada R, Nakamura T, Shigeta Y. Sparsity-weighted outlier FLOODing (OFLOOD) method: Efficient rare event sampling method using sparsity of distribution. *J Comput Chem*, 2016, 37: 724–738
- 81 Ota M, Ikeguchi M, Kidera A. Itinerary profiling to analyze a large number of protein-folding trajectories. *Biophysics*, 2016, 13: 295–304
- 82 Sugita Y, Okamoto Y. Replica-exchange molecular dynamics method for protein folding. *Chem Phys Lett*, 1991, 314: 141–151
- 83 Shao Q, Shi J, Zhu W. Enhanced sampling molecular dynamics simulation captures experimentally suggested intermediate and unfolded states in the folding pathway of Trp-cage miniprotein. *J Chem Phys*, 2012, 137: 125103

- 84 Goyal S, Chattopadhyay A, Kasavajhala K, et al. Role of urea-aromatic stacking interactions in stabilizing the aromatic residues of the protein in urea-induced denatured state. *J Am Chem Soc*, 2017, 139: 14931–14946
- 85 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. *Phys Chem Chem Phys*, 2016, 18: 13052–13065
- 86 Zavadlav J, Marrink S J, Praprotnik M. Multiscale simulation of protein hydration using the SWINGER dynamical clustering algorithm. *J Chem Theor Comput*, 2018, 14: 1754–1761
- 87 Wu X, Yang G, Zu Y, et al. Molecular dynamics characterisations of the Trp-cage folding mechanisms: in the absence and presence of water solvents. *Mol Simul*, 2012, 38: 161–171
- 88 Borgohain G, Paul S. Folding/unfolding of protein Trp cage in aqueous osmolyte solutions under polar confinement. *J Mol Liq*, 2017, 233: 431–441
- 89 Paschek D, Hempel S, García A E. Computing the stability diagram of the Trp-cage miniprotein. *Proc Natl Acad Sci USA*, 2008, 105: 17754–17759
- 90 Paschek D, Day R, García A E. Influence of water-protein hydrogen bonding on the stability of Trp-cage miniprotein. A comparison between the TIP3P and TIP4P-Ew water models. *Phys Chem Chem Phys*, 2011, 13: 19840–19847
- 91 Ota M, Ikeguchi M, Kidera A. Phylogeny of protein-folding trajectories reveals a unique pathway to native structure. *Proc Natl Acad Sci USA*, 2004, 101: 17658–17663
- 92 English C A, García A E. Charged termini on the Trp-cage roughen the folding energy landscape. *J Phys Chem B*, 2015, 119: 7874–7881
- 93 Taricska N, Bokor M, Menyhárd D K, et al. Hydration shell differentiates folded and disordered states of a Trp-cage miniprotein, allowing characterization of structural heterogeneity by wide-line NMR measurements. *Sci Rep*, 2019, 9: 2947
- 94 Baldwin R L. In search of the energetic role of peptide hydrogen bonds. *J Biol Chem*, 2003, 278: 17581–17588
- 95 Fleming P J, Rose G D. Do all backbone polar groups in proteins form hydrogen bonds? *Protein Sci*, 2005, 14: 1911–1917
- 96 Palese L L. Correlation analysis of Trp-cage dynamics in folded and unfolded states. *J Phys Chem B*, 2015, 119: 15568–15573
- 97 Kamo F, Ishizuka R, Matubayasi N. Correlation analysis for heat denaturation of Trp-cage miniprotein with explicit solvent. *Protein Sci*, 2016, 25: 56–66
- 98 Kim S B, Palmer J C, Debenedetti P G. Computational investigation of cold denaturation in the Trp-cage miniprotein. *Proc Natl Acad Sci USA*, 2016, 113: 8991–8996
- 99 Uralcan B, Debenedetti P G. Computational investigation of the effect of pressure on protein stability. *J Phys Chem Lett*, 2019, 10: 1894–1899
- 100 Kitazawa S, Fossat M J, McCallum S A, et al. NMR and computation reveal a pressure-sensitive folded conformation of Trp-cage. *J Phys Chem B*, 2017, 121: 1258–1267
- 101 Cai Z, Zhang Y. Hydrophobicity-driven unfolding of Trp-cage encapsulated between graphene sheets. *Colloids Surf B Biointerfaces*, 2018, 168: 103–108
- 102 Zhang H, Lv Y, Tan T, et al. Atomistic simulation of protein encapsulation in metal-organic frameworks. *J Phys Chem B*, 2016, 120: 477–484
- 103 Levine Z A, Fischer S A, Shea J E, et al. Trp-cage folding on organic surfaces. *J Phys Chem B*, 2015, 119: 10417–10425
- 104 Marino K A, Bolhuis P G. Confinement-induced states in the folding landscape of the Trp-cage miniprotein. *J Phys Chem B*, 2012, 116: 11872–11880
- 105 Shao Q, Hall C K. Protein adsorption on nanoparticles: model development using computer simulation. *J Phys Condens Matter*, 2016, 28: 414019
- 106 Irbäck A, Mohanty S. Protein folding/unfolding in the presence of interacting macromolecular crowders. *Eur Phys J Spec Top*, 2017, 226: 627–638
- 107 Bille A, Mohanty S, Irbäck A. Peptide folding in the presence of interacting protein crowders. *J Chem Phys*, 2016, 144: 175105
- 108 Bille A, Linse B, Mohanty S, et al. Equilibrium simulation of Trp-cage in the presence of protein crowders. *J Chem Phys*, 2015, 143: 175102
- 109 Bandyopadhyay D, Kamble Y, Choudhury N. How different are the characteristics of aqueous solutions of *tert*-butyl alcohol and trimethylamine-*N*-Oxide? A molecular dynamics simulation study. *J Phys Chem B*, 2018, 122: 8220–8232
- 110 Uralcan B, Kim S B, Markwalter C E, et al. A Computational study of the ionic liquid-induced destabilization of the miniprotein Trp-cage. *J Phys Chem B*, 2018, 122: 5707–5715
- 111 Borgohain G, Paul S. Atomistic level understanding of the stabilization of protein Trp cage in denaturing and mixed osmolyte solutions. *Comput*

- [Theor Chem](#), 2018, 1131: 78–89
- 112 Adamczak B, Kogut M, Czub J. Effect of osmolytes on the thermal stability of proteins: replica exchange simulations of Trp-cage in urea and betaine solutions. [Phys Chem Chem Phys](#), 2018, 20: 11174–11182
- 113 Roussel G, Caudano Y, Matagne A, et al. Peptide-surfactant interactions: A combined spectroscopic and molecular dynamics simulation approach. [Spectrochim Acta Part A](#), 2018, 190: 464–470
- 114 Borgohain G, Paul S. Effect of nonpolar confinement on protein Trp cage conformation in aqueous osmolyte solutions. [J Mol Liquids](#), 2017, 231: 174–184
- 115 Borgohain G, Paul S. Temperature-mediated switching of protectant-denaturant behavior of trimethylamine-*N*-oxide and consequences on protein stability from a replica exchange molecular dynamics simulation study. [Mol Simul](#), 2017, 43: 52–64
- 116 Chalyavi F, Hogle D G, Tucker M J. Tyrosine as a non-perturbing site-specific vibrational reporter for protein dynamics. [J Phys Chem B](#), 2017, 121: 6380–6389
- 117 Su Z, Mahmoudinobar F, Dias C L. Effects of trimethylamine-*N*-oxide on the conformation of peptides and its implications for proteins. [Phys Rev Lett](#), 2017, 119: 108102
- 118 Chevrot G, Fileti E E, Chaban V V. Protein remains stable at unusually high temperatures when solvated in aqueous mixtures of amino acid based ionic liquids. [J Mol Model](#), 2016, 22: 258
- 119 Kmiecik S, Kolinski A. Simulation of chaperonin effect on protein folding: A shift from nucleation-condensation to framework mechanism. [J Am Chem Soc](#), 2011, 133: 10283–10289
- 120 Bechtluft P, Kedrov A, Slotboom D J, et al. Tight hydrophobic contacts with the SecB chaperone prevent folding of substrate proteins. [Biochemistry](#), 2010, 49: 2380–2388

Computational and experimental studies on the influence of the Trp-cage inter-residue interactions and other external factors on its structural and folding stability

WU XiaoMin¹, ZHANG XiaoChun¹, WANG Meng¹, HUANG ZhuoRan¹, ZHANG XingWang¹,
ZHANG Qiang¹ & XU ChengZhen²

1 Anhui Province Key Laboratory of Pollutant Sensitive Materials and Environmental Remediation, College of Life Sciences, Huaibei Normal University, Huaibei 235000, China;

2 College of Computer Science and Technology, Huaibei Normal University, Huaibei 235000, China

The major driving force of protein folding depends on their inter-residue interactions, especially for sidechain-sidechain interactions. The Trp-cage is a fast-folding mini-protein, which has already been acknowledged as an ideal model for tackling complicated folding-related problems of proteins. Although this protein model is merely composed of 20 amino acid residues, it exhibits an extremely complex interaction network between residues with different types of sidechain orientation. Because current force fields exhibit different degrees of accuracy inadequacies in the descriptions of atomic interactions, it remains a matter of debate at what stage of folding the helix, hydrophobic cluster and salt-bridge is being conclusively formed in its possible folding mechanisms. The structures and mutations of all the Trp-cage system reported since 2002 are summarized in this paper, and its two typical structures (TC5b and TC10b) are taken as examples to summarize the composition and types of their inter-residue interactions as well as the sidechain features and orientations. Meanwhile, the recent computational and experimental studies of several possible folding mechanisms are also generalized and discussed, and their relationships with simulation temperature and force field as well as the switching from one folding pathway to the other are also discussed in detail. Finally, the effects of numerous factors on structural stability and folding dynamics have been elaborately discussed, including different types of inter-residue interactions (such as salt-bridge, hydrophobic and hydrogen-bonding interactions) and environmental factors (such as temperature, pressure, pH and other potential factors). We also propose that the clarification of the above factors and their synergistic effects are beneficial not only to reveal the driving force of protein stabilization and rapid formation, but also to provide theoretical guidance for further structural regulation and optimal design.

Trp-cage mini-protein model, inter-residue interactions, folding model, residue mutation, sidechain orientation and property

doi: [10.1360/SSV-2019-0304](https://doi.org/10.1360/SSV-2019-0304)