

未折叠蛋白质的普遍初始热力学亚稳态*

杨霖^{1,2)**} 郭帅¹⁾ 马晓亮¹⁾ 侯成宇³⁾ 史丽萍¹⁾ 李佳成¹⁾ 赫晓东¹⁾⁽¹⁾ 哈尔滨工业大学特种环境复合材料技术国家级重点实验室, 哈尔滨 150080;⁽²⁾ School of Aerospace, Mechanical and Mechatronic Engineering, The University of Sydney, NSW 2006, Australia;⁽³⁾ 哈尔滨工业大学电子与信息工程学院, 哈尔滨 150080)

摘要 探索和理解蛋白质折叠问题一直是分子生物学、结构生物学和生物物理学的终极挑战. 未折叠的蛋白质应该存在一种普遍初始热力学亚稳态, 否则无法解释蛋白质是如何在剧烈的热振动干扰下完成快速精确折叠的. 本文通过分析水溶液环境和蛋白质折叠的相关性, 揭示了一种由水分子屏蔽效应引起的未折叠蛋白质的普遍初始热力学亚稳态, 该亚稳态的存在是水溶液环境中水分子的物理性质决定, 并赋予未折叠蛋白质抵抗热扰动和避免错误折叠的能力. 我们通过研究已发表的实验数据和建立分子模型, 找到了该初始热力学亚稳态存在的相关证据, 并推测了该亚稳态导致蛋白质精确折叠的相关物理学机制.

关键词 未折叠蛋白质, 蛋白质二级结构, 水分子屏蔽效应, 热力学亚稳态

中图分类号 Q71

DOI: 10.16476/j.pibb.2019.0111

蛋白质及其产物是地球生命的基础, 几乎所有已知的生物化学反应和生命现象都是由蛋白质参与完成的^[1]. 地球生物中可能存在数千万种不同的蛋白质, 每种蛋白质的生物学功能和活性都是通过其特有的三维形状表达的. 蛋白质形成天然三维结构的过程被称为蛋白质折叠现象, 蛋白质折叠因此被认为是分子生物学和结构生物学的基础^[2-3]. 未折叠的蛋白质也被称为多肽链, 其往往不具有复杂的生物学功能, 原因是多肽链的本质只是类似树脂分子的长链聚合物分子. 蛋白质折叠实现了多肽链的功能化, 例如蛋白质折叠赋予了蛋白质作为配体和受体相互对接和作用的能力. 蛋白质折叠可以被看作是一切生物存在、进化、功能化和多样性的最重要的机制和原动力^[4]. 自从 Anfinsen^[5] 揭示了蛋白质的三维结构是由蛋白质的一级结构决定的这一重要关系, 并因此获得了 1972 年的诺贝尔化学奖以来, 探索和研究蛋白质氨基酸序列和蛋白质三维结构之间的物理关系, 成为了分子生物学和生物物理学的重大前沿课题. 蛋白质折叠问题是《科学》(Science) 杂志公布的 125 个最具挑战性的科学问题之一, 该问题可以被表述为: “未折叠的蛋白质

是如何以快速和可重复的方式折叠成其天然三维结构的?”. 蛋白质折叠问题在 60 年前被提出并被认为是分子生物学和生物物理学问题, 对蛋白质折叠机理和折叠规律的探索被广泛开展, 相关研究多次获得诺贝尔化学奖^[5-13]. 但是蛋白质折叠问题在生物物理学层面的研究进展依然有限, 至今仍无法解释蛋白质是如何在极短时间内完成精确折叠的, 这限制了蛋白质研究的发展和应用^[1]. 蛋白质折叠问题被认为是基础科学的遗产, 蛋白质折叠问题的解决将会推进分子生物学、结构生物学、生物物理学、病理学、遗传学和药理学的发展, 其影响将会广泛而深入^[1]. Dill 等^[1] 将蛋白质折叠问题细化为不同层面的三个经典问题: a. 存在于蛋白质氨基酸序列中的, 决定蛋白质三维结构的物理折叠密码是什么? b. 导致蛋白质快速折叠的物理驱动力和机制是什么? c. 是否能够通过计算机算法准确预测

* 中央高校基本科研专项资金和深圳市科技计划资助项目.

** 通讯联系人.

Tel: 0451-86402739, E-mail: linyang@hit.edu.cn

收稿日期: 2019-05-18, 接受日期: 2019-08-30

蛋白质的三维结构? 此外, 为什么蛋白质折叠现象只发生在水溶液环境中并需要特定的温度范围也是蛋白质折叠的一个核心问题^[14-16].

蛋白质的折叠过程主要由如下的多种物理力所引导: 氢键的形成、范德华力、静电力、疏水作用、熵和温度^[1]. 这些物理作用力可以被分子“力场函数”描述, 这使得用分子动力学模拟方法成为了研究蛋白质折叠问题的重要手段之一. 然而, 即使使用最高性能的计算机进行蛋白质折叠的模拟研究, 蛋白质折叠问题至今依然没有被解答. 尤其是形成蛋白质天然折叠路径的机理和导致蛋白质快速折叠的驱动机制没有被确定^[17].

蛋白质分子折叠主要导致蛋白质分子内部形成大量的氢键和疏水键, 这些氢键和疏水键的形成过程释放了大量的自由能并降低了蛋白质分子结构的势能, 一些研究认为可以将蛋白质折叠的过程看作是分子结构的自由能释放过程或结构松弛过程^[18]. 这些研究认为蛋白质的快速折叠可以被归因于分子随机热运动导致的未折叠蛋白质分子的构象变化, 分子随机热运动会使蛋白质分子的自由能下降并使分子结构向着势能最小的天然结构变化^[1, 19]. 但值得指出的是, 通常在温度40℃时哺乳动物体内蛋白质可以正确折叠, 但当温度上升到42℃时, 哺乳动物体内的许多蛋白质的三维结构会被热运动破坏, 使蛋白质发生不可逆转的变性, 变性后的蛋白质结构可能也是一个稳定的折叠构象^[20-21]. 也就是说, 蛋白质正确折叠所需的环境温度上限(40℃)和导致蛋白质发生热变性的温度下限(42℃)是十分接近的. 这两个相近的温度下的分子热运动的剧烈程度差别非常小, 因此仅用分子结构的自由能释放或分子结构松弛来完整解释蛋白质折叠现象是不充分的.

氢键是蛋白质结构稳定性的最主要的贡献者. 一个未折叠的蛋白质拥有巨量潜在可能的折叠构象, 由于这些可能的折叠构象往往都形成了大量的氢键和疏水键, 所以每一种可能折叠构象的自由能都远小于未折叠蛋白质的自由能. 如果把蛋白质折叠过程看成是简单的自由能释放过程, 那么蛋白质如何能快速地折叠成其唯一的天然结构而不发生错误折叠^[22]? 考虑到一些蛋白质的折叠过程只需要几微秒^[1, 23], 蛋白质是如何能不通过搜索折叠路径就避免了错误折叠的发生? 蛋白质分子存在无数种可能的折叠构象和蛋白质快速准确折叠的事实之间的矛盾被称为经典的Levinthal佯谬^[17].

科学家通过激光温度跳跃技术和单分子实验技术等实验手段揭示了蛋白质折叠的一些重要动力学规律. 蛋白质折叠首先导致蛋白质二级结构的形成, 蛋白质的三级结构是在蛋白质的二级结构基础上通过进一步折叠形成的^[2]. 在蛋白质二级结构形成的过程中, 导致未折叠蛋白质不同位置的多肽链片段形成不同二级结构的机理可以被认为是重要的蛋白质折叠密码. 激光温度跳跃技术实验还表明蛋白质二级结构的折叠速度非常快, 例如二级结构 β 折叠和 α 螺旋的折叠成型过程只需要几微秒就可以完成^[24-25]. 在 α 螺旋和 β 折叠等二级结构的形成过程中, 蛋白质分子形成局部稳定结构和释放自由能的主要方式是主链上相邻的羰基氧原子(C=O)和酰胺氢原子(N-H)彼此生成了大量的氢键. 也就是说在 α 螺旋和 β 折叠形成的数微秒时间内, 未折叠蛋白质的自由能因大量氢键连续生成而快速下降, 大量氢键的连续生成主导了蛋白质二级结构的折叠机制. 蛋白质二级结构形成的折叠机制很可能与主导蛋白质的三级结构形成的折叠机制不同. 因为引导蛋白质三级结构形成的主要是疏水作用和疏水键, 疏水作用和疏水键引导蛋白质二级结构中较大的疏水基团相互吸引形成蛋白质三级结构的疏水核心.

一些理论认为蛋白质分子的天然结构是最稳定的分子构型, 并具有最小的自由能和最小的分子结构势能^[26]. 如果这些理论成立, 蛋白质分子的二级结构可以被推测为分子结构势能最小的局部最稳定结构. 因为在蛋白质二级结构形成过程中, 蛋白质折叠如果没有确保每个局部都生成最稳定的局部二级结构, 那么很难说明在这个阶段蛋白质分子的自由能释放是充分的. 经典的蛋白质二级结构 α 螺旋和 β 折叠确实有较高的结构稳定性, 主链上的羰基氧原子(C=O)和酰胺氢原子(N-H)几乎都彼此生成了氢键^[27-28], 这些大量氢键的生成确保了 α 螺旋和 β 折叠的形成过程可以有效地释放自由能和降低分子结构的势能. 但是许多蛋白质二级结构并不是 α 螺旋或 β 折叠^[27-28], 也就是说这些蛋白质的二级结构并没有形成足够数量的主链之间的氢键来大幅减小其自由能, 所以说这些二级结构并不是局部分子结构势能最小的稳定结构^[27-28]. 也就意味着这些二级结构的稳定性并不高, 因为维持这些结构稳定的氢键数量相对较小, 其结构稳定性明显小于 α 螺旋和 β 折叠结构的稳定性. 既然蛋白质折叠的过程无法保证氨基酸序列中的每个片段都形成势

能最小化的稳定结构, 那么蛋白质折叠过程就不应该被简单地理解为松弛过程或自由能最小化过程. 因为很难解释蛋白质折叠是如何通过放弃了大量的局部更稳定的二级结构来追求三级结构的稳定性和分子势能最小化的. 此外, 液体核磁共振 (solution NMR) 实验发现许多蛋白质存在天然未折叠部分, 实验结果表明这些蛋白质未折叠的部分并不是稳定的分子结构^[29]. 显然, 蛋白质分子的自然结构是最稳定的分子构型并具有最小自由能的假说, 对于这些天然局部未折叠的蛋白质来说并不成立. 蛋白质的疏水内核的形成应该是发生在蛋白质折叠过程的后期, 疏水键的形成是需要疏水氨基酸相互靠近到表面张力相互作用的距离, 因此疏水作用很可能没有主导蛋白质折叠的全部过程. 综上, 蛋白质折叠过程中不同氨基酸序列片段形成不同二级结构的现象是不能简单地用分子结构自由能最小化来解释的. 导致蛋白质在特定的氨基酸序列片段形成特定的二级结构的机制是未知的, 也就是蛋白质二级结构的物理折叠密码是未知的.

1 水分子对蛋白质的屏蔽效应

蛋白质只是在水溶液环境中才能正确折叠, 没有实验证明蛋白质可以在非水溶液中正确折叠^[30-31], 全面理解未折叠的蛋白质与水的相互作用可能会揭示蛋白质折叠问题的答案. 相比于其他的溶剂分子, 水分子有其独特性, 水分子的尺寸非常小, 且每个水分子的氢原子和氧原子分别带有强正电和强负电, 这使得水分子可以进入到未折叠蛋白质侧链之间的空隙中和未折叠蛋白质的每个亲水侧链形成氢键, 也就是这些亲水侧链生成氢键的能

力可以被大量的水分子饱和, 这就避免了亲水侧链之间生成氢键的可能. 在水溶液环境下, 水分子也会阻碍未折叠蛋白质的亲水侧链和主链之间形成氢键. 生物体内存在大量表面亲水的蛋白质, 水分子会在蛋白质表面形成水合层, 从而避免了蛋白质分子团聚, 这直接证明了水溶液中的水分子对亲水侧链有屏蔽效应. 一些激光温度跳跃实验表明, 未折叠的蛋白质可以在较高的温度下稳定存在 (约 10℃)^[24], 随着温度的快速上升, 未折叠的蛋白质的二级结构快速形成. 这意味着, 这些实验中的未折叠蛋白质处于没有二级结构的完全未折叠的状态. 完全未折叠的蛋白质在室温环境下稳定存在很可能依赖溶液中的水分子对亲水侧链的屏蔽效应.

有许多证据可以证明水分子屏蔽了亲水侧链之间形成氢键. 首先在经典 CHARMM 势函数描述的蛋白质分子模型中, 水分子中的氧原子带电量为 ($-0.834e$), 其电量普遍大于蛋白质亲水侧链原子的带电量 (图 1)^[32]. 所以在水环境中, 蛋白质亲水侧链中的正性原子只会和水分子中的氧原子先形成氢键. 基于同样的原因, 蛋白质亲水侧链中的带负电的氧原子只会和水分子中的氢原子先形成氢键. 另一个证据是很多的蛋白质天然结构并不是紧实的或局部未折叠的^[27-28], 这些未折叠的局部结构之间可以通过生成蛋白质分子间的氢键继续折叠, 形成更为紧密的更为稳定的蛋白质结构, 这些不完全紧实的蛋白质结构的存在也佐证了存在水分子对亲水侧链的屏蔽效应. 此外, 蛋白质二级结构是通过主链间形成氢键来稳定的, 支链没有参与蛋白质二级结构的形成, 这也证明蛋白质的亲水侧链被水分子屏蔽了.

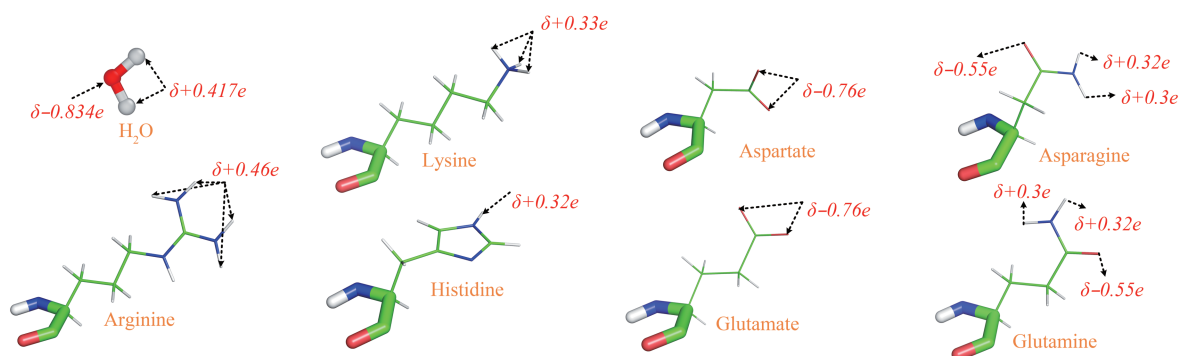


Fig. 1 Charge comparison between water molecule and charged side-chains

2 未折叠蛋白质的普遍初始热力学亚稳态

蛋白质折叠需要水溶液环境和适宜的温度,许多蛋白质折叠所需的温度高于室温,室温环境引起纳米尺度未折叠蛋白质分子的热运动是非常剧烈的,甚至可以认为未折叠的蛋白质分子是在剧烈热振动中完成准确折叠的.在热运动如此剧烈的环境中,蛋白质的快速折叠几乎不可能是完全由分子结构松弛所决定的.因为未折叠的蛋白质侧链存在大量的亲水基团,剧烈的热振动会导致这些亲水基团随机相遇形成氢键并由此导致错误的折叠.由此可以推测未折叠的蛋白质分子可能存在某种热力学亚稳态,这种亚稳态可以避免未折叠的蛋白质在热运动中进行随机的折叠.蛋白质的折叠可能是利用分子热振动来精确破坏某些氨基酸序列片段的亚稳态来实现蛋白质二级结构的精确折叠,这解释了为什么许多蛋白质折叠需要特定的温度范围.因此,寻找未折叠蛋白质的热力学亚稳态可能是解答蛋白质折叠问题的关键.如果所有未折叠的蛋白质都普遍存在同一种热力学亚稳态来保证蛋白质能按清晰的物理折叠密码折叠,那么该亚稳态很可能是在蛋白质折叠发生前就存在了,也就是说多肽分子链本身就具有某种热力学稳定状态.因为一旦蛋白质折叠开始,折叠过程中的不同种类的蛋白质很难形成类

型相同的某种热力学亚稳态.

在环境水分子的屏蔽效应作用下,一个未折叠蛋白质很可能会进入一个热力学亚稳态来避免蛋白质发生错误折叠,这个亚稳态可能也是蛋白质物理折叠密码生效的必要条件(图2a).在水溶液环境屏蔽亲水侧链的情况下,未折叠蛋白质的状态很可能是由蛋白质主链上的相邻带电原子决定.每个肽平面的羰基氧原子和相邻肽平面内的酰胺氢原子分别带有负电和正电,水分子无法屏蔽相邻如此近的两个原子间的静电吸引.这种酰胺氢原子和羰基氧原子间的正负电吸引会导致相应的N—H键和C—O键趋于平行(图2a).

这同时会导致N—H键和C—O键所在的两个相邻的肽平面趋于平行,考虑到肽平面是刚性平面结构,我们会发现这种状态下的未折叠的蛋白质是无法折叠的,因为主链上所有的单键的旋转都受到了限制.如果没有热运动破坏这种羰基氧和酰胺氢原子间的静电吸引,那么不可能发生蛋白质的折叠.只有相邻的羰基氧原子和酰胺氢原子远离彼此并摆脱了彼此的静电吸引,主链上肽平面和单键才能自由扭转,肽平面旋转才会导致主链之间生成氢键和二级结构的折叠.这种相邻的肽平面保持平行的状态可能是未折叠的蛋白质的一种普遍的热力学亚稳态.

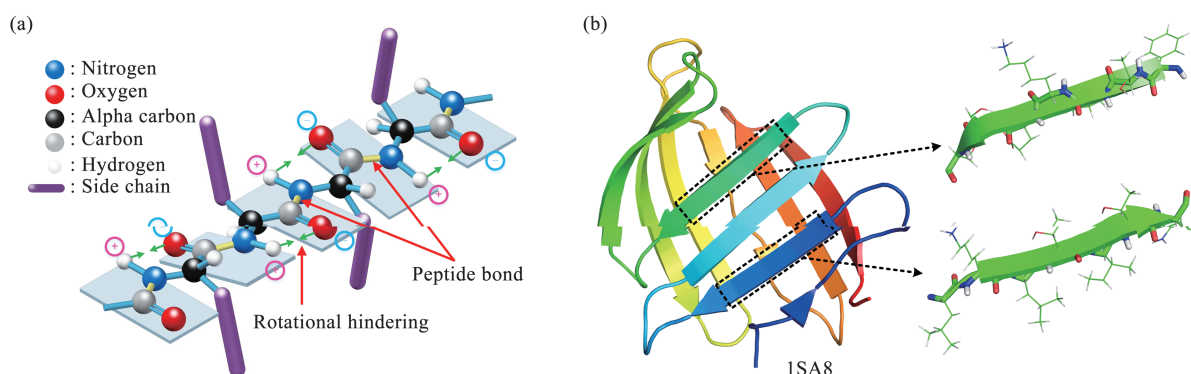


Fig. 2 Initial thermodynamic metastable state of unfolded proteins

(a) Adjacent peptide planes are parallel distributed due to the electrostatic attraction between each carbonyl oxygen atom and adjacent amide hydrogen atom. (b) Parallel distributed state of peptide planes in a β -sheet of the protein 1SA8.

存在这种相邻肽平面保持平行亚稳态的最直接证据是, β 片层结构的分子结构状态就是相邻的肽平面保持平行,这意味着该亚稳态大量的存在于蛋白质天然结构中(图2b).在其他的蛋白质二级结

构中,这种相邻的肽平面保持平行的状态也被发现普遍存在(图3),这证明了该状态对于所有多肽链来说是一种重要的热力学稳定状态.

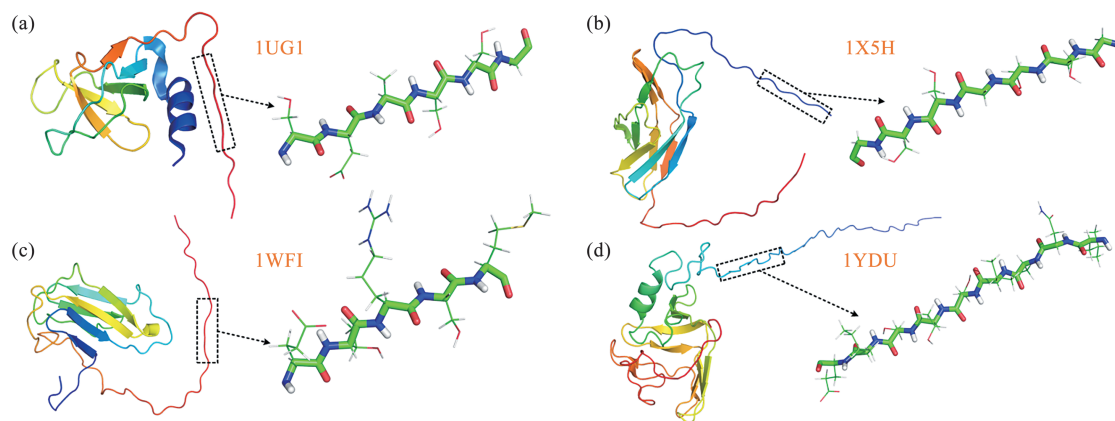


Fig. 3 Parallel distribution of adjacent peptide planes is common in partially unfolded proteins

(a) 1UG1; (b) 1X5H; (c) 1WFI; (d) 1YDU.

3 蛋白质折叠的启动

从PDB分子模型中可以发现 β 折叠中相邻的羰基氧原子和酰胺氢原子的距离约为2.2 Å, 羰基氧原子和酰胺氢原子在这个距离上的相互作用已属于弱氢键连接, 这说明该亚稳态是很稳定的. 该亚稳态中相邻的肽平面间的这种弱氢键连接可以让未折叠的蛋白质链承载一定的扭矩和弯矩, 当一个肽平面的旋转会带动相邻的肽平面也发生旋转, 也就是说扭转波会沿着未折叠的蛋白质分子链传播. 只有当两个相邻的肽平面摆脱了羰基氧原子和酰胺氢原子相互之间的静电吸引, 这两个相邻的肽平面才会发生相对扭转, 从而导致启动折叠和主链间氢键的生成.

由于蛋白质折叠是需要一定的温度范围的, 可以推测是温度引起的热运动破坏了未折叠蛋白质序列中特定氨基酸上的亚稳态, 也就是热运动使该氨基酸的羰基氧原子和酰胺氢原子摆脱了两者之间的静电吸引状态, 导致了该氨基酸所在的位置发生了折叠. 那么这种热力学亚稳态下未折叠的蛋白质是如何准确折叠的? 这要看什么原因会导致羰基氧原子和酰胺氢原子间的静电吸引被破坏. 考虑到热环境必然会引起未折叠蛋白质的肽平面和支链振动, 最有可能的原因就是相邻肽平面和支链发生了较大的振动或转动, 破坏了相应的羰基氧原子和酰胺氢原子间的静电吸引.

值得注意的是, 脯氨酸 (proline) 并不能形成热力学亚稳态, 因为脯氨酸并不包含酰胺氢原子

(N—H), 也就是说脯氨酸没有羰基氧原子和酰胺氢原子间的静电吸引来阻碍蛋白质折叠. 由于热力学亚稳态的存在, 未折叠的蛋白质分子可以在低温环境下保持未折叠状态一段时间, 但是脯氨酸不存在热力学亚稳态, 脯氨酸残基抵抗扭转的能力相比于其他的氨基酸残基来说是微不足道的. 因此, 当其他的氨基酸残基保持亚稳态时, 分子热运动会导致脯氨酸残基先发生折叠. 所以基于未折叠蛋白质的热力学亚稳态理论, 脯氨酸一定会导致二级结构折叠发生. 如果实验测定的蛋白质结构当中脯氨酸都存在于二级结构的转角位置, 那么就可以证明未折叠蛋白质的热力学亚稳态存在. 我们在PDB数据库中随机抽取的107个蛋白质天然结构中 (表1) 找到了418个脯氨酸残基, 经过统计发现99%的脯氨酸残基都位于二级结构的折叠转角处. 这也证明了未折叠蛋白质热力学亚稳态很有可能是存在的.

Table 1 Proteins used for validating the folding codes.

2khe	2l2n	2m7s	3zpz	2n76	5unk	1rbs	2l0q	2lyx	3adg
2lvn	2l6q	2m66	4hcs	2rue	5y6h	1rdu	1v95	2djj	2mgx
2klz	2l7k	2mct	4n6t	4qyw	1okd	1rzw	1whb	2fwg	1z9b
2knz	2l9r	2mew	4od6	5aiw	1onb	1s3a	1wik	2hst	2crp
2krb	2l33	2mh2	4ou0	5dfg	1ovq	1sbo	1pc2	2b5x	1rw2
2krk	2lhc	2mtl	2mh8	5jyu	1ovy	1srv	2n6e	5ub0	1trs
2kt2	2lqj	2rql	2msw	5ks5	1p6q	1t4y	2kl3	1vkr	2ftv
2kv8	2lsg	2ru9	2n2t	5npa	1pqn	1ti3	1xfl	2jnw	2dun
2kxg	2luq	2wnm	2n2u	5npg	1pux	1tmy	1xoa	1n3k	2ki8
2kyz	2ly3	2wqg	2n3z	5szw	1qzm	1tof	1tq1	2ayy	2jy9
2chf	1wlm	1xwc	1z2d	1r26	5kph	5djx			

温度是通过什么具体的物理学机制破坏了未折叠蛋白质上氨基酸上的羰基氧原子和酰胺氢原子间的静电吸引,从而启动折叠的?相邻的肽平面保持平行的热力学亚稳态会导致每个肽平面都不能单独自由旋转,所以由肽平面和支链热振动产生的扭转波会沿着未折叠蛋白质的主链传播.当蛋白质氨基酸序列中相邻氨基酸侧链的扭转阻力差异阻碍了扭转波沿蛋白质主链的传递,就有可能导致折叠的发生^[33].此外,温度引起的支链热运动中可能会启动相邻带电支链间的静电吸引或静电排斥,相邻支链因静电吸引或静电排斥导致的运动很可能会破坏这些支链所在氨基酸残基的热力学亚稳态,从而导致蛋白质折叠的发生.

考虑到蛋白质折叠需要特定的温度范围,氨基酸残基的亚稳态破坏很可能是由热振动触发.处于亚稳态的蛋白质主链上相邻羰基氧原子(C=O)和酰胺氢原子(N-H)间的氢键可以被热振动破坏,从而导致相邻的肽平面发生扭转,并启动了蛋白质的折叠过程.值得注意的是,主链上相邻羰基氧原子(C=O)和酰胺氢原子(N-H)的热振动不可避免地会引起两者之间的侧链振动,侧链的振动起到抑制热振动的阻尼器作用.然而在水环境中,较长的亲水侧链(如精氨酸)的旋转阻力应远大于较短的疏水侧链(如甘氨酸)的旋转阻力(图1).这意味着具有长亲水侧链氨基酸残基的亚稳态在热环境中应该比具有短疏水侧链氨基酸残基的亚稳态更稳定,因为具有更大旋转阻力的长亲水

侧链提供了更强的抗热振动阻尼.相邻氨基酸残基侧链的扭转阻力差异、亲水疏水性质差异和带电荷差异,会导致某一温度下的热振动选择性地破坏特定的氨基酸残基的亚稳态,并导致蛋白质精确折叠.例如,当未折叠蛋白质的某个氨基酸序列片段中存在某个容易失去热力学亚稳态的氨基酸残基(如甘氨酸和脯氨酸),当热运动破坏了该氨基酸残基的亚稳态并导致了其相邻的氨基酸残基也失去热力学亚稳态,那么很可能这几个氨基酸残基的失稳形成了 β 转角,从而导致 β 片层结构的形成.这意味着 β 片层结构的折叠过程很可能起始于 β 转角的形成,其他的氨基酸残基的亚稳态并没有被破坏.当热振动无法破坏某个氨基酸序列片段中的氨基酸残基的热力学亚稳态,这个氨基酸序列片段很可能就会成为蛋白质天然无折叠的部分.

根据未折叠蛋白质的热力学亚稳态理论, α 螺旋的产生过程可以看作是处于亚稳态的4个或5个相邻的氨基酸残基发生压缩和扭转造成的.当未折叠蛋白质的氨基酸序列中存在两个亲水氨基酸介于两个疏水氨基酸片段中间时,在熵的作用下,这两个疏水区域存在使其塌缩合并的疏水作用力,疏水作用力和侧链热振动配合也很可能破坏这几个氨基酸的亚稳态.当这个片段的氨基酸失去了亚稳态后,其疏水区域的塌缩很可能将这几个氨基酸的片段形成了螺旋结构,这样也就形成了连续贯通的疏水区域(图4).

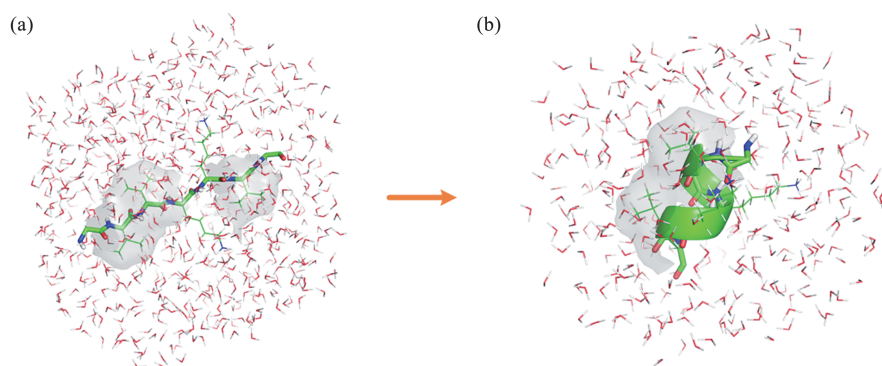


Fig. 4 Hydrophobic-interaction-induced α -helix formation

丙氨酸(alanine)、缬氨酸(valine)、亮氨酸(leucine)、异亮氨酸(isoleucine)的支链有很强的疏水性,可以被归类为疏水氨基酸(疏).精氨酸(arginine)、赖氨酸(lysine)、谷氨酰胺

(glutamine)、谷氨酸(glutamate)、组氨酸(histidine)、天冬酰胺(asparagine)、天冬氨酸(aspartate)、丝氨酸(serine)、苏氨酸(threonine)的支链有很强的亲水性,可以被归类为亲水氨基酸

(亲). 如果发现“疏-亲-亲-疏-疏”5个基酸所组成的折叠码广泛存在蛋白质结构的 α 螺旋中, 就可以确定其为蛋白质的物理折叠密码. 我们共查找了200个这样的码, 发现84%的该码都存在于 α 螺旋中(用来确定折叠密码的蛋白质列于表1). 这也证明了未折叠蛋白质的普遍初始热力学亚稳态的存在.

4 结 论

未折叠蛋白质主链上可旋转的单键使得氨基酸序列中每个氨基酸残基都有两个折叠的自由度, 这导致了蛋白质有无数种可能的折叠构象. 考虑到蛋白质折叠所需的环境温度会使未折叠蛋白质产生剧烈的热振动和错误折叠的发生, 可以推测存在一种普遍初始热力学亚稳态来避免蛋白质在折叠过程中受到热振动的干扰. 通过发现水分子对蛋白质侧链亲水性的屏蔽效应, 我们揭示了以相邻肽平面保持平行特征的一种未折叠蛋白质的普遍初始热力学亚稳态. 该亚稳态赋予了未折叠的蛋白质抵抗一定扭矩和弯矩的能力, 这避免了蛋白质发生错误折叠. 处于该亚稳态未折叠蛋白质的相邻肽平面并不是完全平行的, 未折叠蛋白质在整体上可以是弯曲的或团聚的. 三种可以破坏未折叠蛋白质氨基酸序列中特定氨基酸残基亚稳态的物理机制被指出, 沿未折叠蛋白质主链传播的扭转波是启动这些物理学机制的关键, 通过研究这些物理机制可以破解蛋白质折叠的物理学密码. 这里我们揭示的未折叠蛋白质的普遍初始热力学亚稳态预示着蛋白质折叠过程不完全由随机热运动主导, 蛋白质折叠问题可能存在以牛顿时力学为基础的答案, 蛋白质折叠问题可能是基础科学留给牛顿时力学的遗产.

参 考 文 献

- [1] Dill K A, MacCallum J L. The protein-folding problem, 50 years on. *Science*, 2012, **338**(6110): 1042-1046
- [2] Dill K A, Ozkan S B, Shell M S, *et al.* The protein folding problem. *Annual Review of Biophysics*, 2008, **37**: 289-316
- [3] Lednev Igork. Amyloid fibrils: the eighth wonder of the world in protein folding and aggregation. *Biophysical Journal*, 2014, **106**(7): 1433-1435
- [4] Grishin N V. Fold change in evolution of protein structures. *Journal of Structural Biology*, 2001, **134**(2-3): 167-185
- [5] Anfinsen C B. Principles that govern the folding of protein chains. *Science*, 1973, **181**(4096): 223-230
- [6] Kendrew J C, Dickerson R E, Strandberg B E, *et al.* Structure of myoglobin: a three-dimensional fourier synthesis at 2 Å resolution. *Nature*, 1960, **185**(4711): 422-427
- [7] Wakabayashi K. Accomplishment of Dr. Aaron Klug, winner of Nobel prize in chemistry, 1982. *Tanpakushitsu Kakusan Koso Protein Nucleic Acid Enzyme*, 1983, **28**(2): 156-157
- [8] Rodnina M V, Wintermeyer W. The ribosome goes Nobel. *Trends in Biochemical Sciences*, 2010, **35**(1): 1-5
- [9] Brown H, Sanger F, Kitai R. The structure of pig and sheep insulins. *Biochemical Journal*, 1955, **60**(4): 556-565
- [10] Rasmussen S G F, Devree B T, Zou Y, *et al.* Crystal structure of the β_2 adrenergic receptor - Gs protein complex. *Nature*, 2011, **477**(7366): 549-555
- [11] Levitt M. The birth of computational structural biology. *Nature Structural Biology*, 2001, **8**(5): 392-393
- [12] Dubochet J. Cryo-EM—the first thirty years. *Journal of Microscopy*, 2012, **245**(3): 221-224
- [13] Prusiner S B. Prions. *Proc Natl Acad Sci USA*, 1998, **95**(23): 13363-13383
- [14] Levy Y, Onuchic J N. Water and proteins: a love-hate relationship. *Proc Natl Acad Sci USA*, 2004, **101**(10): 3325-3326
- [15] Snow C D, Nguyen H, Pande V S, *et al.* Absolute comparison of simulated and experimental protein-folding dynamics. *Nature*, 2002, **420**(6911): 102-105
- [16] Alfano C, Sanfelice D, Martin S R, *et al.* An optimized strategy to measure protein stability highlights differences between cold and hot unfolded states. *Nature Communications*, 2017, **8**: 15428
- [17] Zwanzig R, Szabo A, Bagchi B. Levinthal's paradox. *Proc Natl Acad Sci USA*, 1992, **89**(1): 20-22
- [18] G D Rose A, Wolfenden R. Hydrogen bonding, hydrophobicity, packing, and protein folding. *Annual Review of Biophysics and Biomolecular Structure*, 1993, **22**(1): 381-415
- [19] Dill K A, Chan H S. From Levinthal to pathways to funnels. *Nature Structural Biology*, 1997, **4**(1): 10-19
- [20] Lepock J R, Frey H E, Ritchie K P. Protein denaturation in intact hepatocytes and isolated cellular organelles during heat shock. *The Journal of Cell Biology*, 1993, **122**(6): 1267-1276
- [21] Lepock J R. Protein denaturation during heat shock. *Advances in Molecular and Cell Biology*, 1997, **19**(8): 223-259
- [22] Clark A C. Protein folding: are we there yet? *Archives of Biochemistry and Biophysics*, 2008, **469**(1): 1-3
- [23] Mayor U, Johnson C M, Daggett V, *et al.* Protein folding and unfolding in microseconds to nanoseconds by experiment and simulation. *Proc Natl Acad Sci USA*, 2000, **97**(25): 13518-13522
- [24] Clarke D T, Doig A J, Stapley B J, *et al.* The α -helix folds on the millisecond time scale. *Proc Natl Acad Sci USA*, 1999, **96**(13): 7232-7237
- [25] Chen E H L, Lu T T Y, Hsu J C C, *et al.* Directly monitor protein rearrangement on a nanosecond-to-millisecond time-scale. *Scientific Reports*, 2017, **7**(1): 8691-8691
- [26] Alm E, Baker D. Prediction of protein-folding mechanisms from free-energy landscapes derived from native structures. *Proc Natl Acad Sci USA*, 1999, **96**(20): 11305-11310
- [27] Berman H, Henrick K, Nakamura H. Announcing the worldwide

- Protein Data Bank. Nature Structural Biology, 2003, **10**(12): 980-980
- [28] Berman H, Henrick K, Nakamura H, *et al.* The worldwide protein data bank (wwPDB): ensuring a single, uniform archive of PDB data. Nucleic Acids Research, 2007, **35**(suppl_1): 301-303
- [29] Uversky V N. Natively unfolded proteins: a point where biology waits for physics. Protein Science, 2002, **11**(4): 739-756
- [30] Collet O. How does the first water shell fold proteins so fast? The Journal of Chemical Physics, 2011, **134**(8): 02B633
- [31] Soares C M, Teixeira V H, Baptista A M. Protein structure and dynamics in nonaqueous solvents: insights from molecular dynamics simulation studies. Biophysical Journal, 2003, **84**(3): 1628-1641
- [32] Brooks B R, Brooks C L 3rd, Mackerell A D Jr. , *et al.* CHARMM: the biomolecular simulation program. Journal of Computational Chemistry, 2009, **30**(10): 1545-1614
- [33] Eisenberg D. Three-dimensional structure of membrane and surface proteins. Annual Review of Biochemistry, 1984, **53**(1): 595-623

Universal Initial Thermodynamic Metastable State of Unfolded Proteins*

YANG Lin^{1,2)**}, GUO Shuai¹⁾, MA Xiao-Liang¹⁾, HOU Cheng-Yu³⁾,
SHI Li-Ping¹⁾, LI Jia-Cheng¹⁾, HE Xiao-Dong¹⁾

¹⁾National Key Laboratory of Science and Technology on Advanced Composites in Special Environments,
Harbin Institute of Technology, Harbin 150080, China;

²⁾School of Aerospace, Mechanical and Mechatronic Engineering, The University of Sydney, NSW 2006, Australia;

³⁾School of Electronics and Information Engineering, Harbin Institute of Technology, Harbin 150080, China)

Abstract Understanding and explaining how proteins fold have been the ultimate challenges in molecular and structure biology. Protein folding should initiate from a thermodynamic metastable state of unfolded proteins, otherwise, it is difficult to explain how an unfolded protein chain folds exactly into its native 3D structure in an expeditious and reproducible manner under severe thermal disturbance from temperature. Considering dependency of protein folding on aqueous environments and unique physical properties of water molecules, this study uncovers an initial thermodynamic metastable state of unfolded proteins in aqueous environment that enables them to that resists thermal-motion and avoids misfolding. The existence of the thermodynamic metastable state of unfolded proteins are verified by analyzing the results available from experiments in the literature. The principles of physics are applied to estimate the breakage of the thermodynamic metastable state of specific amino acids, that enable the leading to the inferred as physical folding mechanisms and codes for proteins.

Key words unfolded proteins, secondary structure of protein, shielding effect of water molecules, thermodynamic metastable state

DOI: 10.16476/j.pibb.2019.0111

* This work was supported by grants from Fundamental Research Funds for the Central Universities of China and Shenzhen Science and Technology Program.

** Corresponding author.

Tel: 86-451-86402739, E-mail: linyang@hit.edu.cn

Received: May 18, 2019 Accepted: August 30, 2019