

评述

蛋白质折叠、动力学以及蛋白质-配体结合的 物理化学基础

李慧敏^{①③†}, 谢月辉^{②③†}, 刘次全^{③④}, 柳树群^{③④⑤*}

① 云南民族大学数学与计算机科学学院, 昆明 650500;

② 昆明医学院计算机教研室, 昆明 650031;

③ 云南大学生物资源保护与利用重点实验室, 教育部微生物资源重点实验室, 昆明 650091;

④ 中国科学院西南生物多样性实验室, 昆明 650223;

⑤ 云南省高校动物遗传多样性与进化重点实验室, 昆明 650091

† 同等贡献

* 联系人, E-mail: shuqunliu@gmail.com

收稿日期: 2013-08-18; 接受日期: 2013-09-17; 网络版发表日期: 2014-04-24

国家自然科学基金(批准号: 31370715, 31160181, 31360277, 30860011)、国家重点基础研究发展计划(批准号: 2013CB127500)、云南省创新团队项目(批准号: 2011CI123)和云南大学骨干教师培养计划(批准号: XT412003)资助

摘要 蛋白质是生物体的重要组成部分并参与细胞内几乎所有的生物学过程. 随着越来越多物种基因组序列的测定, 准确理解基因产物的功能并探索蛋白质功能多样性的原因, 已经成为当前的研究热点. 为了研究蛋白质的功能, 已有大量蛋白质的静态三维结构被测定. 但是, 蛋白功能最终受其动力学行为所控制, 这包括折叠过程、构象波动、分子运动以及蛋白质-配体相互作用等. 基于自由能图谱理论, 本文深入讨论了蛋白质动力学的底层物理化学机制, 并回答了以下问题: 蛋白质为什么能够折叠、以及如何折叠成其天然三维结构? 为什么蛋白质的动力学特征是固有的? 其动力学行为如何控制蛋白质的功能? 讨论结果将有助于后基因组时代生命科学研究中蛋白质结构-功能关系的理解.

关键词

自由能图谱
熵-焓
非互补性
粗糙性
驱动力
热力学
能力学

蛋白质是由一条或多条氨基酸链组成的生物大分子. 它们在生物体内执行各种各样的功能并参与几乎所有的生物学过程, 包括催化反应、免疫应答、DNA 复制、信号转导、分子转运、细胞黏附、细胞周期、基因表达调控、以及在肌肉组织和细胞骨架中所行使的结构或力学功能. 因此, 研究、了解蛋白质的功能, 对理解生物学过程以及生命的本质至关重要.

结构生物学的中心法则是: 蛋白质的氨基酸序

列包含了决定其三维结构的所有信息, 且蛋白质的结构决定蛋白质的功能^[1-3]. 高分辨率 X-射线晶体衍射方法是结构生物学的重要研究手段, 该方法通过测定大量的蛋白质结构, 极大促进了蛋白质结构-功能关系研究^[4]. 然而, 蛋白质本质上是一种“柔软的”动态物质, 因此, 依靠静态的 X-射线晶体衍射结构还无法全面解释、了解其功能. 物理学家认为蛋白质的生物学功能根植于其物理运动当中; 或者说, 蛋白

引用格式: 李慧敏, 谢月辉, 刘次全, 等. 蛋白质折叠、动力学以及蛋白质-配体结合的物理化学基础. 中国科学: 生命科学, 2014, 44: 433-449

英文版见: Li H M, Xie Y H, Liu C Q, et al. Physicochemical bases for protein folding, dynamics, and protein-ligand binding. Sci China Life Sci, 2014, 57: 287-302, doi: 10.1007/s11427-014-4617-2

结构的动力学行为决定了它们的功能特征^[5,6]。广义上讲, 这些物理运动包括: (i) 蛋白质折叠过程中的构象变化; (ii) 折叠构象状态下的动力学行为和分子运动; (iii) 蛋白质-配体识别结合过程中的动力学行为。

目前, 已有诸多方法可用于研究不同时间尺度和分辨率下蛋白质的动力学行为^[4]。实验方法如 X-射线晶体衍射、核磁共振、光谱法、Laue X-射线衍射、低温电子显微镜和小角 X-射线散射法等, 能够提供原子分辨率的结构快照和结构集合(structural ensemble), 因此可用于推测慢时间尺度上的动力学信息。其他经典的生物物理方法, 如圆二色谱、荧光共振能量转移、红外光谱、拉曼光谱和电子顺磁共振等, 尽管只能提供低分辨率的局部结构信息, 但是可用于推测不同时间尺度范围内的能力学(即构象转换)信息。计算方法如分子动力学(molecular dynamics, MD)模拟^[7,8]及其衍生方法(如模拟退火和复制-交换蒙特卡罗技术^[9])、力场简化模型(如高斯网络模型^[10]和 Gō 模型^[11])、以及优化采样技术(如元动力学(metadynamics)^[12-14]、伞状采样(umbrella sampling)^[15-18]和靶操纵模拟(targeted and steered simulations)^[19-22])等, 不仅能够提供原子水平上的完整动力学信息(即单个原子的三维坐标值随时间的变化情况), 同时还能提供构象变化相关的力和能量信息。值得指出的是, 优化采样技术还能计算蛋白-溶剂系统的自由能变化并重构自由能图谱(free energy landscape)。随着计算机硬件水平的提高以及算法的改进^[23-25], 计算机模拟方法已极大促进了蛋白质动力学相关的研究, 这包括蛋白质的折叠机制、构象空间采样、分子运动以及构象转换等, 研究结果极大提升了人们对蛋白质动力学-功能关系的理解。

针对研究蛋白质折叠和动力学的相关实验和计算方法、以及动力学-功能关系研究的例子, 目前已有大量的综述性文章^[4,6,26-29], 因此本文对此不再赘述。本文目的在于, 通过讨论蛋白质物理运动底层的物理化学原理和机制, 回答下列问题: (i) 为什么蛋白质能够快速折叠成天然构象状态? 其中的驱动力是什么? 驱动力又如何指导和加速折叠? (ii) 为什么蛋白质具有固有的动力学性质(即动态天性)? 决定其动态天性的因素有哪些? (iii) 蛋白质-配体结合的驱动力是什么? 它们又如何驱动蛋白和配体间的识别和结合过程? 为了解决上述问题, 下文将先介绍热

力学系统(thermodynamic system)的几个基本概念, 即焓(enthalpy)、熵(entropy)、自由能(free energy)和自由能图谱(free energy landscape)。

1 蛋白质-溶剂热力学系统的基本概念

蛋白质-溶剂系统本质上为热力学系统, 由溶质(即蛋白分子)、液态水、缓冲离子以及其他物质如小分子化合物、辅因子和配体组成。由于系统中的各组分间存在着复杂的相互作用和能量(热)交换, 因此系统的整体属性, 如蛋白质的宏观构象状态, 是这些微观点分平均行为的表现。热力学定律决定了系统中各组分间的关系和能量迁移规律^[30], 其中所涉及的主要概念有熵、焓和自由能。

蛋白质-溶剂系统中的熵可被看作是衡量热能(thermal energy)(物理学中也称作热(heat)或动能(kinetic energy))在系统中分布均匀程度的一个量。热力学第二定律认为, 热能总是由系统的高温区域向低温区域扩散, 因此熵在该定律中处于核心地位。热能将尽可能地均匀分布, 这一趋势会不可避免地导致系统有序性的丧失, 因此熵通常也被看作是衡量系统无序或混乱程度的一个指标^[6]。当系统处于恒温、恒压条件时, 熵起源于原子的热能, 这种能量使得原子相互推挤和碰撞, 进而导致了键的振动、基团(如氨基酸残基侧链)和残基的旋转和波动、以及分子(如水、蛋白质和配体分子)的运动。总之, 熵是热能在系统各组分间扩散、迁移的过程, 最终使得热能尽可能均匀地散布于整个系统并造成系统的无序性。尽管熵是一种无法被直接测量的量, 但是可以用系统自由度(degrees of freedom)的变化来定量熵的变化。进一步, 熵变又被划分为溶质分子(蛋白质和配体)的构象熵变、旋转熵变和平移熵变, 以及溶剂(水和离子)分子的熵变^[5]。

焓是衡量蛋白质-溶剂系统总能量的一个量。它是系统内部能量(即系统形成所需的能量)和建立系统体积和压力所需能量(即体积和压力的乘积)的总和。尽管焓是一种可计量的状态函数, 系统的总焓却无法直接测定。目前, 人们用系统能量的变化来衡量焓变, 这一作法简化了对能量转移过程的描述。例如, 蛋白内或蛋白-溶剂间原子非共价键的形成是系统的放热过程, 因此这一过程会造成系统总能量的减少而被认为是负焓变(或焓减)。相反, 原子间非共价键

的断裂是需要能量的吸热过程, 因此会造成系统能量的增加而被认为是正焓变(或焓增). 需要指出的是, 焓变为系统的全局属性, 它是系统内(包括蛋白质内、溶剂内以及蛋白质-溶剂间)各种非共价键(如范德华接触、氢键、盐桥以及其他极性和非极性相互作用)断裂和形成的综合体现^[6].

自由能是物理系统中可转化为功的能量. 在蛋白质-溶剂热力学系统中, 该能量通常被称作吉布斯(Gibbs)自由能^[31]. 也可将吉布斯自由能看作是一种能衡量系统能量变化的化学势(chemical potential). 自由能是系统的状态函数, 因此, 只有当自由能变化为负值的时候, 系统才能自发地从一种状态转换到另一种状态. 自由能的变化(ΔG)可表达为:

$$\Delta G = \Delta H - T\Delta S, \quad (1)$$

ΔH 和 ΔS 分别为系统的焓变和熵变, T 为开尔文(Kelvin)温度. 热力学的一个重要原则是: 当系统处于恒温、恒压条件并达到平衡状态时, ΔG 为负值且系统具有最小的自由能值, 也就是说, 热力学系统的平衡状态为其自由能最小化状态. 需要指出的是, 如果 2 个状态间的自由能差异 ΔG 为正值, 系统则无法进行自发的状态转换, 除非系统组分或条件发生了变化, 或者系统被做功.

蛋白质-溶剂系统的自由能值为系统的状态函数, 这一函数通常被形象地描述为自由能图谱^[32,33], 即用系统的自由能变化来描绘系统的状态变化情况. Dill^[34]将蛋白质-溶剂系统的自由能图谱定义为,

$$F(x) = F(x_1, x_2, \dots, x_n), \quad (2)$$

$F(x)$ 为系统处于“ x ”状态时的自由能值. 系统状态“ x ”可被一系列变量, 即 x_1, x_2, \dots, x_n 所描述. 系统状态是指系统内各成分(包括溶质如蛋白质、配体和辅因子、以及溶剂如水分子和离子)间复杂相互作用所造成的整体状态. 然而在研究蛋白质折叠和蛋白质-配体相互作用等具体问题时, 研究者通常只关注蛋白质的构象状态(conformational state). 简便起见, 这里的 x_1, x_2, \dots, x_n 仅用于描述蛋白质的微观状态^[34], 如蛋白肽链的二面角、本征向量投射(由本质动力学分析 MD 模拟轨迹而获得)所代表的构象变化^[6,35,36]、原子间天然接触(native contacts)数量、肽链 N-, C-末端距离、序参量(order parameter, 用于描述折叠中间体与天然构象状态的相似性)^[37]、或任何其他可描述蛋白质自由度的变量^[34].

2 蛋白质折叠问题

蛋白质折叠问题可分为 2 个具体的问题^[38,39]: (i) 折叠的能力学问题, 即为什么蛋白质能够快速折叠. 该问题也被称作列文托悖论(Levinthal's paradox)^[40]; (ii) 折叠的热力学问题, 即作用于氨基酸残基序列上的原子间相互作用力如何指导蛋白质的折叠. 该问题也被称作折叠代码. 第(i)个问题最早由 Levinthal^[40]于 1968 年提出, 他发现尽管蛋白质具有巨大的构象空间(假如一个蛋白质由 100 个氨基酸组成, 每个氨基酸具有 3 个自由度, 那么该蛋白的折叠可能性(折叠空间)为 3^{100}), 它们仍能在毫秒到秒一级的时间尺度内快速搜索到(或折叠成)其天然构象. 为了解决蛋白质折叠的能力学问题, 首先要查明指导蛋白正确折叠的驱动力, 这就引出了折叠的热力学问题, 即折叠代码及其作用机理. 因此, 以上 2 个问题紧密关联, 只有同时解决了这 2 个问题, 才能完全理解蛋白质的折叠过程和机制^[6,41].

2.1 蛋白质折叠能力学

Zwanzig 等人^[42]通过建立数学模型解决了蛋白质折叠的能力学问题. 该模型显示, 仅需对折叠过程中的局部不利构象施加一个较小的偏倚势(或偏倚能量), 蛋白质便能在合理的时间范围内完成折叠. 也就是说, 只有当系统存在合理的自由能梯度时, 蛋白质才能快速折叠成天然构象状态.

变性剂, 如强酸、强碱、高浓度无机盐或有机溶剂(如乙醇、氯仿或尿素), 会使蛋白-溶剂系统的自由能图谱变得相对平坦, 这促使处于折叠状态的蛋白质结构失去其三级和二级结构并最终导致蛋白质变性^[43]. 如果维持变性条件不变, 解折叠的蛋白分子将在平坦的自由能表面(free energy surface)上随意“游荡”(wander randomly), 最终造成大量的随机卷曲构象状态. 因此, 在变性条件下, 系统自由能表面的平坦特性使得已变性的蛋白分子无法找到折叠出口. Anfinsen^[3]和 Anfinsen 等人^[44]实验显示, 一旦复性(或折叠)条件被恢复, 变性蛋白分子便能迅速折叠回天然构象, Anfinsen 也因此项工作获得了诺贝尔奖. 恢复蛋白-溶剂系统的折叠条件, 其本质是通过建立自由能梯度而改变系统自由能图谱的形状. 这种自由能的改变显著大于变性条件下系统自由能的波动. 小的自由能波动仅能产生小且浅的自由能陷阱(free

energy traps)或井(wells),这使得“受困于”自由能井内的蛋白分子很容易逃脱而掉进周围的自由能井,从而发生随机卷曲状态间的构象转换.相反,只有大幅度的自由能变化才能建立足够陡峭的自由能斜坡(slope),进而触发蛋白分子的折叠,即蛋白构象快速、不可逆地向全局自由能最小化区域运动.

(1) 漏斗样(funnel-like)蛋白质折叠自由能图谱. 折叠条件的恢复导致蛋白质-溶剂系统的自由能图谱形成足够大的自由能梯度或斜坡,这时,如同滑雪者跃下雪坡滑行一样,变性肽链从自由能最高处跌落并沿着斜坡向自由能最低处滚落.本质上,蛋白质的复性过程是蛋白质-溶剂系统的自由能下降过程,这一过程主要包含快速疏水坍塌(rapid hydrophobic collapse)^[29,45,46]和随后的慢瓶颈(slow bottleneck)^[47]2个阶段.随着系统自由能的下降,蛋白分子会因为溶剂的压缩效应(由溶剂熵最大化而引起)以及蛋白内和蛋白-溶剂间非共价键的形成而逐渐丧失其构象自由度(也称为构象熵).系统的自由能越低,蛋白分子所失去的构象熵就越多,因而其构象状态会显得越发有序.因此,这种伴随着系统自由能下降的蛋白构象熵减过程使得蛋白质折叠自由能图谱呈现为漏斗形^[47,48],将其称作折叠漏斗.

需要指出的是,由于蛋白质拥有数量庞大的原子且每个原子具有3个自由度,因此其自由能图谱具有相当高的维度.尽管如此,研究者通常将自由能图谱表示成一个三维表面,即以2个指定自由度为函数的自由能变化表面.在经典的三维折叠漏斗中^[47,49],漏斗深度(即折叠初始状态与最终状态间的自由能差异)代表天然折叠状态相对于变性状态的能量稳定性,漏斗的宽度代表蛋白质的构象空间或构象熵.蛋白质的天然折叠构象状态位于最窄的漏斗底部(也称为全局自由能最小化区域).漏斗的顶部通常被描述为相对平坦的、分布有很多小自由能井的表面.

(2) 折叠过程和折叠路径. 尽管蛋白质折叠问题相当复杂,但是研究者已发展出一些简化模型以描述、研究蛋白质的折叠过程,如扩散碰撞模型(diffusion collision)^[50,51]、框架模型(trajectory)^[52]、成核压缩模型(nucleation condensation)^[53,54]、拉链-装配模型(zippering-and-assembly)^[55-57]、拼图模型(jigsaw puzzle)^[58]、计量化学模型(stoichiometry)^[59,60]、疏水坍塌模型(hydrophobic collapse)^[46]、以及折叠漏斗模型(folding funnel)^[47,48]等.值得指出的是,上述模型

并非相互独立、相互排斥,相反,它们之间存在或多或少共同之处^[5],通过侧重蛋白质折叠的不同方面,极大推动了折叠机制研究.由于基于自由能图谱理论的折叠漏斗模型包含了上述其他模型的绝大部分特征,因此该模型已被广泛接受并用于描述蛋白质的折叠过程.

折叠漏斗模型认为,蛋白质的折叠过程是蛋白分子从各种变性状态沿多条路径向天然构象状态滚落的自由能下山过程(图 1A)^[47].由于自由能图谱的粗糙性(ruggedness),在下山过程的任何阶段,蛋白分子都有可能掉入大小不一的自由能陷阱(即自由能图谱表面的局部自由能最小化区域).但是,整体上大的自由能落差(或自由能坡度)足以保证折叠中间体跳出局部自由能陷阱而继续其下山过程.在下山过程的不同阶段,蛋白质分子会形成各种折叠中间体,如熔球态(molten globule state)、转换态(transition state)、玻璃转换态(glass transition state)以及其他相对不稳定的状态或子状态(substates).图 1B 显示了小球状蛋白的具体折叠过程^[61-63].折叠的第一阶段为疏水坍塌,该过程导致变性肽链集合(特指处于变性状态的蛋白质分子集合)迅速坍塌成较为紧密的熔球态.熔球态是一种非常重要的折叠中间体,它们具有如下结构特征:疏水侧链位于熔球的内部;具有最小的溶剂可及表面积;已具备一些二级结构和非特异性三级接触.但是,与天然折叠构象状态相比,熔球态所拥有的原子(或残基)间相互作用数量要少得多;换言之,熔球态还缺乏天然构象状态中的大量相互作用或天然接触^[62].目前,已有实验证据表明^[29]:(i)在低浓度化学变性剂存在时,蛋白-溶剂体系中的确存在熔球态;(ii)疏水坍塌过程非常迅速,可在亚微秒级别的时间尺度内完成.研究者认为,折叠漏斗的上半部分非常陡峭且相对平滑,这使得变性蛋白分子能够迅速滚落,最终进入包含了“山、谷和陷阱”(hills, valleys and traps)的粗糙地带(图 1B).因此,熔球态会在较长的时间里受困于山岭间的自由能陷阱内,使得接下来的折叠进程(瓶颈阶段)变得非常缓慢^[47].在折叠的瓶颈阶段,蛋白分子会重复进行一系列的上山和下山过程以搜寻可能的下山路径,这就导致了多种折叠中间体的形成.例如, Dill 和 Chan^[47]将所有位于山顶的中间体定义为“过渡态”,他们认为,过渡态或许有着各种各样的结构特征,且大量过

渡态的存在减缓了折叠过程. 进一步, 非互补熵-焓变化所造成的自由能降低使得蛋白分子折叠成所谓的“玻璃转换态”^[32,62]. 玻璃转换态类似于液体受冷后流动性降低的状态, 形象地说明这种折叠中间体处于某种结构被固定的状态而很难发生构象重排, 这导致折叠的最后一步, 即玻璃转换态向天然构象状态的转变, 变得非常缓慢^[63]. 然而, 由于折叠漏斗具有足够大的全局自由能梯度, 玻璃转换态仍能沿着少数路径继续其自由能下山过程, 最终到达折叠漏斗的最底部, 即全局自由能最小化区域^[61]. 总之, 折叠漏斗模型强调 2 点: (i) 蛋白质的折叠过程是变性分子集合的自由能下山过程; (ii) 蛋白分子可沿诸多可能的路径下山. 值得注意的是, 尽管折叠漏斗模型是在研究小球状蛋白质折叠机理的基础上建立的, 但是同样适用于描述、解释拥有复杂拓扑结构的大蛋白质折叠过程^[64].

2.2 蛋白质折叠热力学

目前普遍认为, 蛋白质折叠的驱动力(折叠代码)存在于其氨基酸序列当中. 这一观点基于以下事实: (i) 不同或相似的氨基酸序列能够折叠成不同或相似的蛋白质结构; (ii) 在所有可能的序列(即氨基酸

序列空间)当中, 仅有少量序列(即自然界当中存在的氨基酸序列)能够折叠成具有生物学功能的结构状态. 传统的观点认为, 折叠代码是由许多个单独的、小的相互作用(包括氢键、盐桥、范德华接触以及其他极性或非极性相互作用)叠加而成^[39,65], 这使得一级序列能够编码二级结构, 进而由二级结构编码三级结构. 换言之, 距离较近的原子间相互作用很容易导致形成简单的局部结构, 通过这些局部结构间的碰撞和相互作用, 便能形成更为复杂的蛋白质三维结构. 基于这一观点, 研究者建立了扩散碰撞模型和框架模型^[50-52]以描述上述分级折叠过程(hierarchical folding process). 然而, 将不同原子间相互作用的叠加看作蛋白质折叠的驱动力, 这种观点存在片面性, 原因有: (i) 该观点忽略了溶剂熵作用, 而溶剂熵作用恰恰是折叠驱动力的一个重要组成^[6]; (ii) 统计力学模型表明折叠代码不仅分布于局部序列, 同时也分布在氨基酸序列当中^[38]; (iii) 局部相互作用的形成在很大程度上是主折叠驱动力(即溶剂熵作用)所导致的结果, 而非非折叠驱动力的成因^[39].

从全局的角度来看, 可以将折叠驱动力理解为蛋白-溶剂系统总自由能的降低, 这就涉及到 2 个重要的驱动因子, 即熵变(包含溶剂熵和蛋白质构象熵

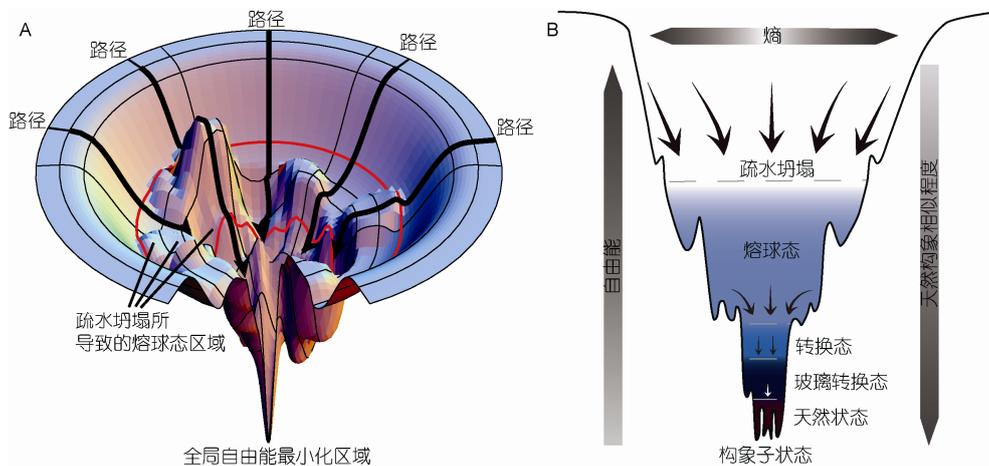


图 1 漏斗状蛋白质折叠自由能图谱(网络版彩图)

A: 包含山、陷阱/井以及自由能能垒的三维自由能图谱示意图. 图中 5 条带箭头的黑线仅用于显示可能的折叠路径, 即处于变性状态的蛋白分子从漏斗的外表面沿箭头所示的下山路径滚落至全局自由能最小化区域. 2 条红线间的粗糙区域表示一个重要的折叠中间体, 即疏水坍塌过程所导致的熔球态所在的区域. 修改自 Dill 和 Chan^[47]; B: 描述详细折叠过程以及各种折叠中间体的二维折叠漏斗示意图. 漏斗的宽度代表蛋白质构象熵的大小; 漏斗深度代表变性状态相对于天然构象状态的自由能稳定性或构象相似程度. 箭头表示自由能的下山方向以及折叠过程的不同阶段. 由淡蓝色逐渐过渡到蓝黑色的区域代表熔球态、转换态、玻璃转换态以及天然状态所在的区域. 修改自 Onuchic 等人^[61,62]以及 Wolynes 等人^[63]

的变化)和焓变(包含蛋白质内以及蛋白与溶剂间共价键的形成和断裂)^[5]. 由于蛋白质折叠过程的复杂性, 因此在不同的折叠阶段, 熵变和焓变对系统自由能降低的贡献程度会有所不同^[41].

(1) 为什么溶剂熵作用(疏水力)对蛋白质折叠非常重要? 在已折叠的蛋白质结构当中存在着数量庞大的氢键和范德华接触, 以及少量的盐桥, 这使得人们认为上述相互作用在驱动蛋白质折叠时发挥着主导作用. 氢键的重要性在于: 天然结构中含有几乎所有可能形成的氢键; 骨架酰胺基和羰基间的氢键是蛋白质二级结构的关键组成部分^[39]. 类似地, 天然折叠结构通常具有紧密的包装, 这意味着范德华接触也非常重要. 由于大多数蛋白质仅具有少量的带电残基且带电残基通常暴露于结构的表面, 因此, 与氢键和范德华接触相比, 盐桥的重要性要小得多. 以上非共价键的重要性仅见于已折叠的天然结构, 因此可以肯定它们在维持已折叠结构的稳定性方面发挥着重要作用. 然而, 这些相互作用在多大程度上能够驱动蛋白质的折叠? 这一问题还需慎重回答.

就直觉而言, 氢键在约束和稳定蛋白质的二级结构方面发挥着重要作用. 然而, 已有实验证据表明, 溶液环境中孤立的二级结构很难保持自身构象的稳定^[39]; 天然蛋白质中存在的“变色龙”序列既能采取 α -螺旋(α -helical), 也能采取 β -股(β -strand)构象^[66,67]. 这就意味着蛋白二级结构的稳定性强烈依赖于蛋白质的三级结构背景, 即蛋白结构的紧密包装方式. 这种紧密包装成型于蛋白质折叠的第一步, 即溶剂熵增所驱动的疏水坍塌过程.

目前, 研究者普遍认为溶剂熵作用(也称作疏水力或疏水相互作用)在蛋白质折叠中发挥着至关重要的作用. 该观点的直接证据为^[39]: (i) 所有天然蛋白的结构均具有疏水核心, 这就意味着溶剂水可将残基的疏水侧链隔离进蛋白结构内部而形成疏水核心; (ii) 将一个疏水残基侧链从水溶剂转移到非极性溶剂中, 需消耗 1~2 kcal/mol 的能量^[68]; (iii) 蛋白结构在非极性溶剂中会发生变性; (iv) 打乱的氨基酸序列, 只要它们还保留着正确的疏水和亲水排列模式, 便能折叠成预期的天然构象状态^[69-71].

以上证据表明, 溶剂熵的最大化是蛋白质折叠驱动力的一个重要组成, 它对蛋白质-溶剂系统的自由能降低作出了实质性的贡献. 例如, 为了满足系统溶剂熵最大化的需要, 水分子将尽可能减少与非极

性氨基酸侧链的接触, 进而将疏水侧链压缩至蛋白内部而仅将极性和带电残基暴露于蛋白质表面, 最终导致蛋白质具有疏水核心和最小的溶剂可及表面积. 疏水侧链-水溶剂系统的平衡状态为其全局自由能最小化状态, 在这一状态下, 系统的溶剂熵(即水分子的熵)得到了最大化. 这时, 疏水侧链会被一层水壳所包裹以保证水壳以外的水分子拥有最大限度的自由度. 因此, 如果将疏水侧链移入非极性溶剂, 则需要有额外的能量破坏水壳中的非共价键, 这便解释了为什么转移溶剂会消耗 1~2 kcal/mol 的能量. 有趣的是, 当疏水侧链被放入非极性溶剂之后, 二者的混合过程于熵于焓均非常有利. 这是因为, 疏水侧链和非极性溶剂均为非极性物质, 因此二者之间很容易产生接触或形成相互作用; 此外, 能量在同性物质的混合物中更容易分散, 这使得疏水侧链-非极性溶剂混合物比疏水侧链-水溶剂混合物更为“混乱”和“无序”. 以上的焓减和熵增过程最终使得疏水侧链-非极性溶剂系统达到平衡状态, 且该状态较疏水侧链-水溶剂系统的平衡状态具有更低的自由能. 类似地, 蛋白质在非极性溶剂中的变性过程同样由系统的自由能下降来驱动, 其中熵增和焓减均为自由能的降低作出了积极贡献^[41]. 以上分析从反面说明, 正是因为液态水具有独特的熵(如熵最大化需求)和焓作用(如与暴露的带电和极性残基侧链形成氢键), 它才能为蛋白质功能构象状态的形成和存在提供理想的溶液环境. 保留了正确疏水和亲水排列模式的乱序氨基酸序列能够折叠成预期的天然构象, 这一现象表明, 溶剂熵最大化所导致的疏水坍塌是蛋白质折叠过程中最重要和最关键的一步. 疏水坍塌所产生的熔球态为后续的构象调整和结构塑形奠定了基础.

(2) 熵、焓对系统自由能降低的差异性贡献. Ben-Naim^[65]建立的“溶剂诱导”力(“solvent-induced” force)统计力学模型显示, 发生于一个基团(如氨基酸残基侧链)上的小构象变化便能建立一个蛋白溶剂化自由能梯度, 这一自由能梯度将向该基团施加一个作用力, 最终使其沿力的方向进行移位. 实际上, 自由能是对蛋白质-溶剂系统整体化学势的一种近似, 因此自由能是个系统性的概念. 由于上述模型仅考虑了蛋白质溶剂化自由能变化所产生的力, 因此有其局限性. 更进一步讲, 不仅蛋白质的构象变化(即构象熵变)能导致系统自由能的变化, 同时溶剂熵变以及系统焓变(包括蛋白内以及蛋白质-溶剂间非共

价键的形成和断裂), 或任何有关溶剂环境的变化(例如向溶液中加入酸、碱、变性剂、辅因子、配体、底物或其他复合物), 都能引起系统自由能的变化^[41]. 自由能变化所造成的梯度越陡, 施加于蛋白质分子上的力将会越大, 这将导致蛋白分子以更高的概率向低自由能区域移动.

使用化学变性剂或改变系统温度能够导致蛋白质变性, 这一过程本质上是通过变性因子改变系统自由能图谱的形状, 即从漏斗形变为相对平坦的表面. 复性是变性的逆过程, 其本质在于, 通过恢复“正常的”溶剂环境而建立足够大的自由能梯度并最终将系统的自由能图谱恢复成漏斗形. 折叠漏斗上半部分具有平滑性(即少有自由能陷阱的存在)和陡峭性(即大的自由能梯度)2个显著特征. 这2个特征使得变性的蛋白分子能够迅速滚落, 最终导致变性的随机卷曲状态坍塌成熔球态. 传统的观点认为这一坍塌过程是由疏水相互作用(hydrophobic interaction)所驱动的^[38,39,45,46]. 然而, 将“疏水相互作用”一词用在这里并不恰当. 这是因为, 这里的疏水相互作用并不是由疏水基团间的相互吸引而形成的; 换言之, 它不是一种主动相互作用, 而是由水分子挤压所造成的被动相互作用. 依据热力学第二定律, 疏水侧链聚集于蛋白质的内部, 这一结果是为了满足系统熵最大化的需求, 即通过牺牲蛋白质的部分构象熵而获得最大的溶剂熵, 最终使得整个系统获得最大的熵(即最大熵增或熵最大化).

在蛋白质-溶剂系统中, 水分子与蛋白分子相比, 无论在数量还是质量上均占据着绝对优势^[41]. 溶剂熵最大化的要求将使水分子间保持尽可能多的动态氢键, 因此, 水分子会尽量减少与疏水侧链基团的接触, 并通过蛋白质溶剂可及表面积的最小化, 将疏水基团隔离进坍塌体(熔球态)的内部. 同时, 在疏水坍塌过程中, 系统还需要额外的能量以将变性蛋白分子表面的水分子剥离下来, 这种正焓变不利于系统自由能的降低; 构象坍塌也造成了蛋白质构象自由度的丢失, 这种构象熵减同样不利于系统自由能的降低. 尽管如此, 坍塌所获得的溶剂熵增远远大于系统焓增和蛋白构象熵减, 因此溶剂熵增对系统总自由能的下降作出了实质性贡献. 综上所述, 本文认为, 溶剂熵作用是蛋白质折叠代码的主导组分, 而疏水相互作用是疏水坍塌的结果而并非驱动蛋白质折叠的原因. 尽管如此, 熔球态中的疏水接触带来的焓减

对系统自由能的降低也是有利的.

如前所述, 疏水坍塌所形成的熔球态是一组相对稳定的蛋白质折叠中间体状态, 该状态具有较为紧密的包装, 同时还包含少量的二级结构和三级接触. 由于熔球态为局部相互作用和全局三级相互作用的进一步形成营造了结构环境, 因此, 本文认为这一中间体的形成对蛋白质的折叠至关重要. 由熔球态向天然构象状态的进一步折叠(即瓶颈过程)会涉及转换态和玻璃转换态2种主要中间体. 尽管瓶颈过程相比疏水坍塌要慢得多, 这一过程仍由系统整体自由能的降低来驱动. 具体地, 熔球态首先要脱离困住它们的自由能陷阱才有可能找到继续下山的路径. 这一自由能爬坡过程由熔球态的构象熵增天性(蛋白质结构在任何状态下都具有构象熵增加的趋势, 本文将其定义为构象熵增天性)所触发. 熔球态中的局部和三级相互作用数量有限, 因此不足以将其构象稳定维持在某个特定状态. 构象熵增天性首先会打断最弱的非共价键并瓦解这些键所束缚的结构区域, 从而触发构象调整过程, 即原子间非共价键的断裂以及不同原子间新非共价键的形成. 这种竞争性相互作用通常会重复多次, 只有当键的强度足以抵消构象熵增作用时才会停止. 本文认为, 蛋白质的构象熵增驱动了自由能上山(或爬坡)过程, 而系统焓减驱动了自由能下山过程. 瓶颈阶段的折叠表现为蛋白分子持续上山和下山以搜索更低自由能状态的过程, 随着越来越多稳定的非共价键的形成, 系统熵减(或负焓变)将足以弥补蛋白质的构象熵减所带来的负效应, 最终使得系统的整体自由能进一步下降. 需要指出的是, 这里的系统焓减不仅来自蛋白内原子间非共价键的形成, 同时还来自蛋白质-溶剂间非共价键的形成. 蛋白质表面水分子网络的形成(水分子的焓减)也对系统自由能的降低作出了贡献. 总之, 由熔球态向天然状态转变的瓶颈阶段仍由系统自由能的整体下降所驱动. 在这一阶段, 负焓变是折叠代码的主导组分, 它通过抵消蛋白质构象熵和溶剂熵损失所带来的负作用, 最终将蛋白质分子引导至全局自由能最小化区域.

3 蛋白质动力学的物理化学基础

折叠完成后, 蛋白质分子位于漏斗状蛋白-溶剂系统自由能图谱的底部, 即全局自由能最小化区域.

然而, 这一区域并不平坦, 同样包含自由能井以及位于自由能井之间的山或能垒. 本文认为, 自由能图谱底部的粗糙性不仅允许蛋白质分子以不同构象状态或子状态集合的形式共存于这一区域, 同时也允许不同构象状态(或子状态)间的构象转换. 具体地, 不同的构象状态集合位于不同的自由能井内, 自由能井的深度和宽度(即自由能井尺寸的大小)决定着某种构象状态的群体大小、发生概率或持续时间. 蛋白质构象状态的群体分布是蛋白质动力学的一个重要性质, 即热力学性质. 相邻自由能井之间的能垒高度决定着2个状态之间的构象转换速率, 这是蛋白质动力学的另一个重要性质, 即能力学性质.

位于某个大尺寸自由能井内的蛋白结构并不是静止的, 相反, 该结构能够围绕其平均/平衡(average/equilibrium)状态进行波动, 进而产生一组紧密关联的构象子状态. 由于这种波动持续的时间较长(毫秒到秒一级时间尺度), 因此被认为是蛋白质的平衡波动(equilibrium fluctuations)或平衡动力学(equilibrium dynamics). 一般认为, 蛋白质的平衡波动控制、支配蛋白质的生物学功能^[4,6], 而发生于构象转换期间的非平衡波动(non-equilibrium fluctuations)似乎对蛋白功能以及生物学过程的整体速率影响不大^[4]. 由于构象转换需要蛋白质分子翻越自由能能垒且转换中间态只有很短的存在时间, 因此目前的实验手段很难捕捉到转换中间态和非平衡波动^[5,6].

3.1 蛋白质动力学的层次性

蛋白质波动的层次性是蛋白质动力学的重要特征, 具体指蛋白结构的不同组分具有不同的波动幅度和波动方向, 且不同的波动幅度发生的时间尺度也有所不同^[4,72]. 如图2所示, 蛋白质的构象状态或子状态位于漏斗底部的各个自由能井内. 井的大小各有不同, 它们的深度和宽度决定着蛋白波动的幅度和时间尺度. 转换中间态位于井间的自由能能垒上, 且能垒的高度决定着波动的方向性以及相邻2个状态间的构象互换速率. 下文将根据图2对蛋白质的动力学幅度及其时间尺度做详细描述.

图2所示的自由能图谱含有2个大尺寸的自由能井(tier-0井), 同时, tier-0井的底部包含小的(tier-1井)或者更小的(tier-2井)自由能井, 例如, 左边的 tier-0井底包含2个 tier-1井, 且这2个 tier-1井的底部还各包含4个和3个更小的 tier-2井. 这种嵌套的自由能

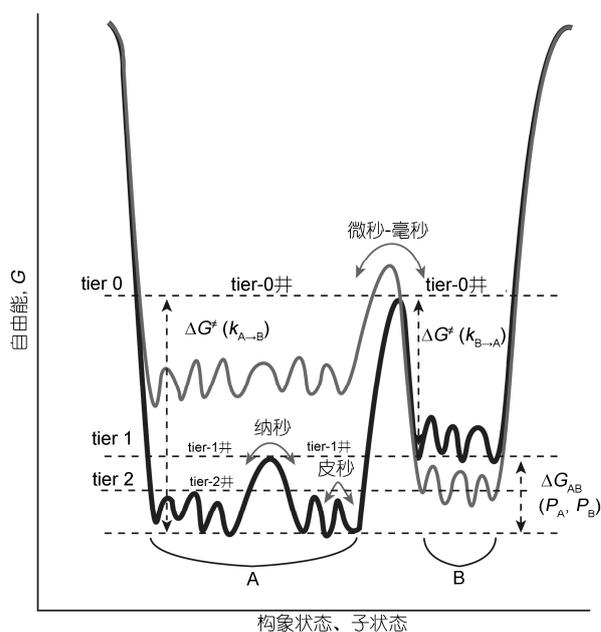


图2 漏斗状蛋白质-溶剂系统自由能图谱底部区域示意图

图中显示了自由能图谱底部的粗糙性、可变性以及分层组织等特征, 这些特征不仅决定了蛋白质的波动幅度和方向性, 同时还决定蛋白质的动力学等级及其时间尺度; 因此, 该图可用于表征蛋白质的热力学和能力学性质. 蛋白质的构象状态或子状态位于不同的自由能井中, 例如, 2个 tier-0 状态, A 和 B, 分别位于 2 个大而深的 tier-0 自由能井内, 且二者间的自由能差异 ΔG_{AB} 决定了状态 A 和 B 的群体尺寸(例如具有更低自由能的状态 A 较具有较高自由能的状态 B 有着更大的群体尺寸, 即 $P_A > P_B$). 状态 A 和 B 之间的自由能能垒(ΔG^\ddagger)高度差异决定了二者间构象互换速率的差异(如 $k_{A \rightarrow B} < k_{B \rightarrow A}$). tier-0 动力学为整个蛋白质分子在微秒到毫秒时间尺度范围内的波动. 尽管 A 和 B 状态平衡共存且具有不同的群体尺寸, 系统环境的改变却会引起自由能图谱的变化(从黑线到灰线), 并最终导致状态 A 和状态 B 的重新分布. tier-1 和 tier-2 子状态分别位于 tier-1 和 tier-2 自由能井内. 围绕 tier-1 状态的构象波动(tier-1 动力学)为纳秒时间尺度上的环和转角运动; tier-2 动力学为皮秒时间尺度上的侧链旋转运动. 修改自 Henzler-Wildman 和 Kern^[4]以及 Ansari 等人^[72]

井组织方式决定了蛋白质波动的多层次性以及波动的的时间尺度. 例如, 2 个主要的构象状态, 即 tier-0 状态 A 和 B, 分别存在于 2 个最大的 tier-0 井内, 因此这 2 个构象状态拥有大的群体尺寸、长的存在时间或高发生概率(P_A, P_B). 围绕 tier-0 状态的构象波动(tier-0 动力学)为波及整个蛋白结构的大幅度波动, 其发生的时间尺度为微秒到毫秒范围, 文献中将这种平衡波动称作大尺度协同运动(large-scale concerted motions)或集体运动(collective motions). 尽管状态 A 和 B 以平衡方式共存, 但是二者间的自由能差异(ΔG_{AB})使得它们在群体大小上也存在差异.

例如, 状态 A 向状态 B 的转换比 B 向 A 的转换需克服更高的自由能垒($\Delta G^\ddagger(k_{A \rightarrow B}) > \Delta G^\ddagger(k_{B \rightarrow A})$), 这就导致 B \rightarrow A 较 A \rightarrow B 具有更快的转换速率($k_{B \rightarrow A} > k_{A \rightarrow B}$), 最终使得状态 A 比状态 B 具有更大的群体尺寸或更高的分布概率. tier-1 井内的构象波动为快动力学波动(tier-1 动力学), 这种波动发生的时间尺度为纳秒级且仅涉及二级结构单元间的环区和转角, 最终致使在 tier-0 井的底部形成一组紧密相关的构象子状态. 更快的构象波动(tier-2 动力学)发生于 tier-1 井底部的 tier-2 井内, 这种波动是发生在皮秒时间尺度上的氨基酸侧链旋转运动, 而侧链旋转则源自最快的波动(tier-3 动力学), 即发生在飞秒尺度上的共价键振动.

3.2 晶体结构与位于层次化自由能井内构象状态的关系

X-射线晶体衍射方法所测定的蛋白结构本质上是结晶条件下的平均或平衡构象状态, 也可以认为是系统自由能图谱的 tier-0 井所“捕捉”到的构象状态. 晶体 B-因子(crystallographic B-factors)能够提供有关原子围绕 tier-0 状态的空间分布信息或平衡波动信息, 这说明晶体结构同样包含 tier-1 和 tier-2 构象状态信息. 因此, 晶体结构快照, 尤其是它们的 B-因子, 能用于蛋白质结构-功能关系研究.

对于同一个蛋白质, 可能会在 PDB(protein data bank)蛋白结构数据库(<http://www.pdb.org>)中搜索到多个晶体结构. 但是, 这些结构在整体构象上的差异并不明显, 它们的结构核心几乎完全一致, 仅能在表面环区上观察到构象差异, 这意味着, (i) 不同实验室采取的结晶条件非常相似, 因此可将测定的结构理解为存在于 tier-0 井底不同 tier-1 井中的构象状态; (ii) 绝大多数蛋白质或许只具有一个 tier-0 井, 且该井中发生的平衡波动控制着蛋白质的生物学功能.

具有至少 2 个 tier-0 状态的著名例子是人类免疫缺陷病毒类型 1(human immunodeficiency virus, HIV-1)的包膜糖蛋白 gp120. 该蛋白质可以以 2 种构象形式存在, 即 unliganded 状态(结合受体 CD4(cluster of differentiation 4)之前的构象状态)和 CD4-bound 状态(结合 CD4 之后的状态), 且这 2 种状态间存在着巨大的构象差异, 这意味着它们各自栖居于 2 个不同的 tier-0 井内. 在结晶环境中, 无论有无 CD4 分子的存在, X-射线晶体衍射方法所测定的 gp120 结构均呈现为 CD4-bound 状态^[73-78], 这便暗示

CD4-bound 状态稳定栖居于一个大而深的 tier-0 井内. 相反, 由于 unliganded 状态无法结晶而很难被测定结构, 目前 PDB 数据中唯一一个 unliganded 状态的晶体结构来自猿类免疫缺陷病毒(simian immunodeficiency virus, SIV)的 gp120^[79]. 本文认为, unliganded 状态之所以无法结晶是因为该状态非常不稳定且栖居于另一个具有更高自由能水平的 tier-0 井内. 进一步, unliganded 状态比 CD4-bound 状态具有更高的构象柔性^[80,81], 这也意味着前者的 tier-0 井较后者具有更粗糙的井底和更多样化的构象子状态. 然而, 与体外实验条件不同的是, 在 HIV-1 功能病毒轴中, gp120 是以 gp120₃/gp41₃ 三聚复合物形式存在的. 因此, gp120 可与 gp41 或相邻的结构亚单位发生相互作用而将其构象稳定在 unliganded 状态. 由于针对 unliganded 构象的抗体很少, 因此该状态的稳定存在有利于 HIV-1 逃避人体的免疫监控^[78]. gp120 通过与其他分子的相互作用而提高其非优势状态的构象稳定性, 这说明蛋白质-溶剂系统的自由能图谱并非静止的, 相反, 溶剂条件的改变会引起系统自由能的变化(图 2).

3.3 自由能图谱的动态天性

图 2 所示的自由能井和能垒仅为蛋白质-溶剂系统动态自由能图谱的一种静态表示方式. 当系统环境, 如温度、压力、pH、离子强度以及系统组成(包括溶质和溶剂组分)保持恒定时, 自由能图谱底部的粗糙形态将得以维持, 具体表现为自由能井以及井间能垒的稳定分布. 然而, 这种稳定性是相对的, 其本质是自由能动态平衡波动的表现形式. 在蛋白质-溶剂系统中, 分子内的原子热运动、键振动以及分子间的布朗碰撞(Brownian collisions)会不可避免地导致系统自由能在一定范围内波动(即平衡波动), 这类类似于微风拂过湖面时引起的相对“静止”的涟漪. 稳定存在的自由能井和井间能垒将允许: (i) 蛋白分子构象状态的稳定存在(即不同的构象状态长时间栖居于各自的自由能井内); (ii) 平衡波动的发生(即发生于自由能井内的围绕平衡/平均状态的构象波动); (iii) 不同构象状态间较为恒定的转换速率. 这就解释了为什么蛋白质分子能以构象状态(或子状态)集合的形式稳定分布(或共存)于折叠漏斗的全局自由能最小化区域.

任何能引起蛋白质-溶剂系统自由能变化的因素

都会打破上述平衡并建立新的平衡, 最终导致蛋白质构象状态的重新分布^[4,41]. 引起系统自由能改变的因素有外因和内因, 前者包括温度、压力、pH、离子强度、变性剂, 以及配体、辅因子、底物、小分子化合物或任何其他分子的加入; 内因主要包括蛋白的氨基酸突变、蛋白构象熵和溶剂熵作用、以及蛋白残基间和残基-溶剂间的竞争性相互作用. 前文已经讨论过, 高温或变性剂导致蛋白变性的原因在于它们能够引起系统自由能图谱形状的实质性改变. 同理, 这些因子也能引起自由能图谱形状的适度变化(图 2, 灰色线条), 并最终导致蛋白质构象状态的重新分布. 例如, 配体和蛋白质分子间的随机碰撞会破坏二者表面的水分子网络, 这是一个溶剂熵增并伴随着溶剂焓增的过程; 进一步的相互作用(即蛋白和配体间非共价键的形成和累积)又会引起系统的焓减以及二者旋转和平移熵的丧失. 以上非互补的熵-焓变化最终导致了系统自由能的波动, 并引起自由能井尺寸(即井的宽度和深度)和井间能垒高度的变化. 本文将在后文详细讨论这些变化在蛋白质-配体结合过程中的作用.

本质上, 非互补的熵-焓变化起源于系统在熵方面的最大化需求, 即尽可能均匀地将原子热能分布于整个系统. 原子热能会引起成键原子沿共价键的谐振荡并造成共价键的振动(即上文所描述的飞秒级 tier-3 动力学). 键振动的累积会产生大量的动能, 而动能的释放又会引起水分子的布朗运动和氨基酸残基侧链的旋转运动(即皮秒级的 tier-2 动力学). 布朗碰撞能够打断相邻水分子间的氢键, 进而导致另外 2 个水分子间氢键的形成, 这种重复的氢键断裂和形成能够将水分子动能尽可能均匀地释放到整个系统当中, 并最终导致溶剂熵的最大化. 蛋白残基侧链旋转运动的累积会打断最弱结构束缚区域(一般为暴露于蛋白表面的环区和转角区域)的非共价键, 而这种局部的非共价键断裂又为形成新的非共价键(包含蛋白残基间以及残基与水分子间的非共价键)提供了条件. 如果新形成的非共价键强度不够, 蛋白质的构象熵增作用仍能将其断裂, 进而会触发新一轮的竞争性相互作用, 这一过程最终导致蛋白表面的环区和转角发生较大的构象波动(即纳秒级的 tier-1 动力学). 环和转角的运动能够通过蛋白表面的水分子网络以及特异的结构组分(如铰链结构)传播、扩散至整个蛋白结构, 最终导致了蛋白质分子的大尺度协同运动

或集体运动(即最慢的 tier-0 动力学).

综上所述, 蛋白质的宏观动力学行为是系统中原子和原子基团微观运动级联放大的结果, 其根本为起源于原子热能的熵. 在恒定的溶剂条件下, 系统的熵增趋势在很大程度上能够被系统的焓增作用所弥补, 因此只可能发生较小的自由能波动, 反过来, 稳定的自由能波动决定了相对稳定的构象状态分布以及不同构象状态间稳定的转换速率. 溶剂环境的剧烈变化, 如变性剂的加入, 会将巨大的正熵引入系统. 由于系统的正焓无法抵消如此巨大的熵增, 这就造成系统自由能图谱的实质变化并最终导致蛋白质变性. 相反, 配体和小分子化合物的加入仅能引起适度的非互补熵-焓变化, 这种变化通过改变自由能垒的高度以及自由能井的尺寸, 最终导致蛋白质构象状态在全局自由能最小化区域的重新分布.

4 蛋白质-配体结合的物理化学基础

蛋白质是通过与其他分子如蛋白质和小肽、核酸、辅因子、配体、底物、小分子化合物、氧分子以及金属离子等发生相互作用而实现其功能并参与生物学过程的^[5,6,82]. 因此, 要详细了解蛋白质的生物学功能就必须深入理解蛋白质-配体的识别和结合机制, 反过来, 这也将极大推动药物发现、设计等相关领域的发展.

4.1 蛋白质-配体的结合过程及驱动力

和蛋白质折叠过程一样, 蛋白质-配体的识别和结合过程同样由蛋白质-配体-溶剂系统的自由能下降来驱动, 且自由能的降低同样涉及到熵和焓的共同贡献^[5,41,83]. 所不同的是, 蛋白质-配体-溶剂系统的自由能降低过程由更为精细、微秒的非互补熵-焓变化所控制.

描述蛋白质-配体结合过程的传统模型有“锁钥”(“lock-and-key”)模型^[84]和“诱导契合”(“induced-fit”)模型^[85]. 前者假设蛋白质(锁)和配体(钥)均为刚性分子且二者的结合表面能够完美匹配, 因此配体对蛋白质分子的结合类似于钥匙插入锁孔的过程(图 3A). 后者认为, 蛋白质和配体的结合表面并没有达到理想的匹配程度, 且蛋白分子的结合位点区域具有构象柔性. 因此, 配体对蛋白质的结合会诱导蛋白结合位点的构象变化, 最终使得配体与蛋白质完全

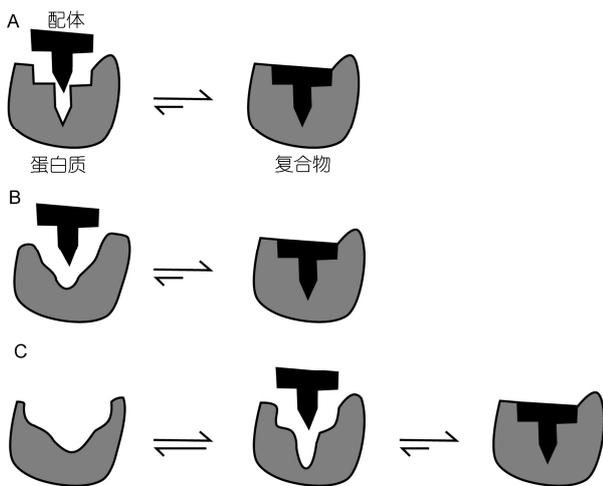


图3 蛋白质-配体结合机制模型示意图

A: 锁钥模型; B: 诱导契合模型; C: 构象选择模型. 修改自 Tobi 和 Bahar^[86]

结合在一起(图 3B). 目前, 上述 2 个模型仍被广泛用于解释酶-底物、靶蛋白-药物以及受体-配体的识别和相互作用机制. 下文将详细讨论这 2 个模型中熵、焓对结合驱动力的不同贡献.

蛋白质-配体结合的第一步为二者在溶剂环境中的扩散以及随后的碰撞过程. 尽管这一过程常常被讨论结合机制的文献所忽视, 但是, 由于分子间的初始接触和碰撞是发生进一步相互作用的先决条件, 因此扩散-碰撞过程在蛋白质-配体结合中显得非常重要. 前文已经讨论过, 水分子动能的释放会引起它们的布朗运动. 这些布朗运动一方面能够导致水溶剂自身熵的最大化; 另一方面, 还能引起蛋白质和配体分子的旋转、平移和扩散运动. 应该牢记的是, 在蛋白质-配体-溶剂系统中, 溶质分子周围存在着数量巨大的水分子, 它们会对溶质进行强烈的布朗轰炸而最终导致溶质分子的扩散和碰撞. 无论对锁钥模型还是诱导契合模型, 系统中溶质的浓度越高, 蛋白质和配体碰撞、接触的概率就越大, 因此二者之间建立进一步相互作用关系的概率也越大.

对于这 2 个模型, 接下来的步骤会有所不同. 就锁钥模型而言, 如果最初的碰撞发生在蛋白质和配体之间的互补界面, 那么这一碰撞就会造成大量水分子的移除. 碰撞之前, 蛋白质和配体的表面被一层水分子网络所包裹. 水网络的形成是为了满足溶剂熵最大化的需要(即通过牺牲水网络的小部分熵而获得其他水分子的最大熵), 同时, 水网络中以及水分

子与蛋白质间氢键的形成也为系统自由能的降低作出了贡献. 蛋白质和配体的最初碰撞会打断部分氢键, 在这一过程中, 源自分子动能的正熵补偿了存储于水分子网络中的正焓(能量), 因而会造成系统自由能的略微下降. 随着水网络中越来越多水分子的移除, 溶剂所获得熵便越来越多. 由于锁钥模型中的蛋白质和配体具有完美匹配的结合表面, 因此碰撞会触发并造成大量的氢键断裂和水分子移除, 最终导致溶剂获得巨大的正熵. 这一正熵变不仅过度补偿了水网络中氢键断链所带来的正焓变, 同时还补偿了蛋白质和配体分子的旋转和平移熵损失(负熵变), 最终会导致系统自由能的大幅下降. 综上所述, 锁钥模型所描述结合过程主要是一个熵驱动的过程, 它类似于蛋白质折叠过程中的疏水坍塌, 其中溶剂熵增(或溶剂熵的最大化)为系统自由能的下降作出了实质性贡献. 蛋白质和配体结合表面的水分子被移除后, 二者间会进一步形成非共价键(即负焓变), 这也将进一步降低系统的自由能. 这一过程类似于蛋白质折叠的瓶颈过程, 即通过竞争性相互作用将初始复合物(或熔球态)“塑形”成紧密结合的蛋白质-配体复合物(或蛋白质天然构象状态). 由于蛋白质和配体结合表面的完美匹配是获取最大可能溶剂熵的先决条件, 锁钥模型能够用于解释配体结合的特异性.

就诱导契合模型而言, 蛋白质和配体在结合之前可能会发生多次尝试性碰撞, 直到找到它们之间合适的匹配位点为止^[87]. 在这里, “合适的匹配位点”是指那些不一定具有完全互补的接触界面、但是却具有足够好的匹配程度的位点. 这种匹配位点能为初始复合物提供足够大的强度和足够长的存在时间以允许进一步相互作用的发生, 而进一步的相互作用又会诱导结合位点在构象上发生变化. 诱导契合过程同样会移除蛋白质和配体接触表面间的水分子, 但是由于接触表面的匹配程度有限, 因此与锁钥模型相比, 其所移除的水分子数量要少得多, 这使得溶剂熵增对系统自由能下降的贡献也要小得多. 然而, 结合位点的高构象柔性会允许初始复合物进行构象调整, 以适应配体的占位并最终形成蛋白质-配体间的完全接触. 这一过程中, 源自非共价键形成的负焓变不仅能够弥补蛋白在构象熵上的损失, 同时还能弥补蛋白质和配体在旋转和平移熵上的损失, 因而为系统自由能的降低作出了实质性贡献. 综上所述, 诱导契合模型所描述的结合过程主要是一个焓驱动

的过程. 由于诱导契合所导致的大量非共价键能够将配体长时间稳定在蛋白的结合位点之内, 本文认为该模型适用于解释配体的结合亲和性.

4.2 蛋白质-配体结合的构象选择机制

尽管锁钥和诱导契合模型可分别用于描述配体与刚性和柔性蛋白受体的结合, 但是二者均为理想化模型, 且仅在单分子而非分子群体水平上描述结合过程. 前文已经讨论过, 蛋白质分子以不同构象状态集合的形式存在于系统自由能图谱的底部, 构象状态(子状态)间能够进行相互转换(能力学性质), 且转换速率的大小决定了构象状态的分布概率和群体大小(热力学性质). 基于蛋白质的热力学和能力学性质, 研究者们提出了一个更为“现实”的模型, 即构象选择(conformational selection)模型^[86,88,89], 它能够在分子群体水平上解释蛋白质-配体的结合机制.

在构象选择模型中, 配体能够选择性地结合仅具有最佳互补表面的蛋白质构象集合, 选择性结合完成后, 蛋白质可进行进一步构象调整, 最终将构象状态间的平衡迁移至复合状态(也称为配体结合状态)(图 3C). 以上 2 个连续步骤分别类似于锁钥和诱导契合模型所描述的结合过程, 因此构象选择结合过程为溶剂熵增和系统焓减所共同驱动. 构象选择模型强调蛋白质的非结合状态(unliganded state)和复合(或结合)状态(complexed or bound state)平衡分布于系统自由能图谱的底部, 而配体对复合状态的选择性结合会打破这一平衡, 最终使得更多的蛋白分子采取复合构象状态. 换言之, 配体结合通过改变自由能垒的高度(或构象转换速率)使得构象状态在群体尺寸上发生变化. 在构象选择模型下, 配体的选择性结合不仅能有效降低蛋白的构象熵惩罚(例如诱导契合所无法避免的蛋白构象熵损失), 同时还能获得尽可能多的溶剂熵(类似于锁钥结合过程的溶剂熵增). 需要指出的是, 自然界中的绝大多数蛋白质都具有构象柔性(尤其是在配体结合区域), 这不仅允许蛋白质分子以多状态集合的形式存在, 同时还有利于蛋白通过构象调整而建立配体和蛋白间的完全接触.

综上所述, 构象选择模型的 2 个步骤, 即配体的选择性结合和随后的蛋白构象调整过程, 分别为溶剂熵增和系统焓减所驱动. 这一模型具有 3 个特点: (i) 它囊括了锁钥和诱导契合模型所描述的结合方

式; (ii) 它是基于自由能图谱理论而提出的, 因此能够在分子群体水平上描述蛋白质-配体的结合; (iii) 它能同时解释配体结合的特异性和亲和性. 因此, 本文认为, 构象选择模型在解释、描述蛋白质-配体结合机制以及结合过程方面, 显得更为现实和全面.

4.3 “错误折叠疾病”(“misfolding disease”)实质上为“蛋白-蛋白结合”(“protein-protein binding”)疾病

阿尔茨海默氏症(Alzheimer's disease)和朊疾病(prion disease), 如“疯牛病”(“mad cow”)、羊瘙痒症(scrapie)和克罗伊茨费尔特-雅各布病(Creutzfeldt-Jakob diseases)等^[90]被统称为“错误折叠疾病”或“构象病”. 之所以有这样的命名, 是因为上述疾病的症状被认为是由相关蛋白质的错误折叠以及错误折叠蛋白在脑组织或其他组织中的聚合所引起的.

已有实验证据表明, 淀粉样 β (amyloid beta, A β)蛋白和朊蛋白(prion protein)的某些结构区域能够由它们的“天然”(“native”) α -螺旋构象转变为易于形成斑块或导致瘙痒的“天然解折叠”(“natively unfolded”) β -股构象^[91,92]. 这意味着 A β 蛋白和朊蛋白自由能图谱的底部具有粗糙特征, 以允许天然状态和天然解折叠状态的同时存在^[93]. 本文推测, 天然解折叠状态很可能栖居于一个具有更高自由能水平的小自由能井内, 而天然状态则栖居于另一个具有较低自由能水平的大自由能井内, 因此天然解折叠状态较天然状态有着更小群体尺寸和更短的存在时间. 尽管如此, 天然解折叠构象的柔性表面为蛋白质间的识别与结合提供了机会. 一般将最早结合在一起的蛋白-蛋白复合物称为种子(seed), 种子一旦形成, 将会有更多的蛋白分子加入进来, 通过它们之间的结合而形成纤维状复合物. 本文认为, 纤维的迅速增长由溶剂熵增所驱动, 换言之, 蛋白质通过排除互补表面间的大量水分子降低了系统的自由能. 随后, 结合界面间非共价键的形成(焓减)又进一步降低了系统自由能. 最终, 天然解折叠状态所在的自由能井被加深和加宽, 这将导致蛋白质分子的构象平衡由天然状态向解折叠状态迁移, 即解折叠状态以更大群体蛋白聚合物(或纤维, 也就是病人组织中所观察到的斑块)的形式稳定存在于该自由能井内. 因此, 斑块中观察到的“错误折叠”(“misfolded”)构象是天然解折叠状态蛋白分子相互结合的结果, 而非蛋白聚集

的原因.

本文认为, $\alpha\beta$ 蛋白和肌蛋白中的累积突变会通过减弱某些结构区域的构象束缚而增加其构象熵; 反过来, 构象熵的增加改变了系统自由能图谱的形状(如形成2个 tier-0 井), 最终允许2种构象状态共存于图谱的全局自由能最小化区域. 在这里, 可以将分子聚集看作是蛋白质分子间相互结合的一种表现形式; 也就是说, 尽管处于解折叠状态的蛋白分子具有较高的自由能水平和小的群体尺寸, 但是通过它们之间的相互识别和选择性结合, 最终可将这种状态稳定在具有低自由能水平和大群体尺寸的蛋白聚合物中^[94]. 细胞的蛋白质质量控制体系无法有效清除这类蛋白质聚合物, 因而致使其在组织中积累并最终引发疾病. 综上所述, 错误折叠疾病并非是由蛋白质的“错误折叠”而引起的, 其根本原因在于蛋白分子间的错误结合. 由于自由能图谱是动态可变的, 因此通过改变溶剂条件(细胞环境)以改变自由能图谱的形状, 便有可能帮助聚合蛋白质“跳出”自由能陷阱, 这为制定“错误折叠疾病”的治疗策略提供了思路^[41].

5 结论

蛋白质折叠以及蛋白质-配体结合具有相似的过程, 且二者均为系统的自由能下降所驱动. 它们的区别仅在于系统分子组分的链连接性上, 如蛋白质的氨基酸链由共价键所连接, 而氨基酸链和配体之间不存在共价连接, 因此可将蛋白质折叠和结合过程分别看做是分子内和分子间的识别结合过程^[5,95,96]. 然而, 无论这种识别和结合发生于分子内或分子间, 源自非互补熵-焓变化的系统自由能下降驱动了蛋白质天然构象状态以及蛋白质-配体复合物的形成^[97].

就蛋白质折叠而言, 伴随着系统自由能的下降, 蛋白质折叠中间体的构象自由度也逐渐减少, 这就决定了蛋白质折叠的自由能图谱必然呈现为漏斗形. 尽管蛋白质-配体的结合过程发生于折叠漏斗底部的自由能井内, 结合界面间水分子的移除以及非共价键的形成仍能进一步降低系统的自由能; 同时, 结合过程中蛋白质和配体会丧失它们的旋转和平移自由度(旋转和平移熵), 且蛋白分子的构象自由度(构象熵)也会逐渐减少. 因此, 蛋白质-配体的结合进一步加深并窄化了发生折叠的自由能井并使之也呈现为漏斗形, 因此将其称为折叠漏斗^[97].

基于自由能图谱理论, Nussinov 小组提出了一个构造块模型(building block model)以显示蛋白质折叠和蛋白质-配体结合的相似性^[95,98]. 在这一模型中, 蛋白质或配体被看作是一系列构造块, 且这些构造块位于不同的类似于小漏斗的自由能井内. 蛋白质的折叠或蛋白质-配体结合则被看作是漏斗融合所驱动的构造块间的识别和结合过程. 无论构造块之间是否存在链连接性, 诸多小漏斗最终能融合成一个更大尺寸的漏斗. 因此, 构造块模型的本质是一系列小漏斗融合事件所驱动的系统自由能下降过程, 最终导致折叠状态的蛋白分子或蛋白质-配体复合物位于全局自由能最小化区域, 即大漏斗的底部.

蛋白质折叠和蛋白质-配体结合具有相似的自由能下山过程. 蛋白质折叠的第一步是溶剂熵最大化所驱动的疏水坍塌过程. 这一过程所形成熔球态中间体已具备部分二级结构和少量的三级接触(相互作用), 但仍缺乏天然折叠构象状态中的大量近距离残基接触或天然接触. 由于熔球态为进一步的构象调整(如构象熵增所触发的蛋白残基间以及残基和水分子间的竞争性相互作用)营造了结构环境, 因此本文认为, 熔球态是蛋白质折叠过程中的一个关键中间体. 类似地, 蛋白质-配体结合的第一步也由溶剂熵增作用所驱动, 其中包含2个连续的子步骤: (i) 溶剂熵增趋势所引起的溶质分子的布朗运动及随后的碰撞过程; (ii) 溶剂熵最大化需求所引起的碰撞界面间水分子网络的移除过程. 尽管所形成的蛋白质-配体初始复合物已具备了部分分子间非共价接触, 但是二者之间的连接仍较为松散. 同样地, 初始复合物为进一步转变成紧密结合的复合物状态提供了结构环境. 因此, 本文认为, 初始复合物类似于折叠过程中的熔球态, 是蛋白质-配体结合过程中的一个关键中间体. 蛋白质折叠和蛋白质-配体结合的第二步均为竞争性相互作用所导致的系统负焓变所驱动. 总之, 无论是蛋白质折叠还是蛋白质-配体结合, 溶剂熵增所驱动的第一步和系统焓减所驱动的第二步均为系统的自由能下降作出了贡献, 最终使得折叠和结合自由能图谱表现为漏斗形, 因此将其分别称作折叠漏斗和结合漏斗.

非互补的熵-焓变化会引起系统自由能的波动, 具体表现为自由能图谱的粗糙性. 值得注意的是, 这种粗糙特征同时存在于漏斗壁和漏斗底部. 在蛋白

质折叠过程中, 尽管漏斗壁的粗糙程度会影响蛋白质折叠的效率, 系统自由能的整体下降趋势(即变性状态与折叠状态间的巨大自由能落差)仍能保证解折叠蛋白分子向漏斗底部持续“滚落”。折叠完成后, 粗糙的漏斗底部使得蛋白质分子以不同状态或子状态集合的方式平衡分布于各自的自由能井内, 且井的大小决定了构象状态的群体尺寸(即蛋白质动力学的热力学性质), 而井间自由能垒的高度决定了相邻构象状态间的互换速率(即蛋白质动力学的动力学性质)。然而, 溶剂条件的改变(如配体的加入)或蛋白分子的自身变化(如氨基酸突变)均能引起系统自由能的变化而打破这一平衡, 并最终导致构象状态集合的重新分布。

尽管系统自由能的波动/扰动源自非互补的熵-焓变化, 熵增趋势却是触发自由能变化的最根本原因^[99]。这是因为: (i) 蛋白质变性过程中, 变性因子(例如化学变性剂或系统温度的增加)最初攻击、破坏的部位是具最弱构象束缚的蛋白结构区域, 如具有大构象熵的表面暴露环区和转角区域^[100]; (ii) 蛋白质折叠和蛋白质-配体结合的第一步, 即疏水坍塌和初始复合物的形成, 均为溶剂熵最大化作用所驱动。

尽管它们的第二步为竞争性相互作用而导致的系统熵减所驱动, 但是竞争性相互作用却由蛋白质的构象熵增天性所触发; (iii) 对于已处于折叠状态的蛋白质分子, 它们的构象熵增天性使得具有较弱构象束缚的结构区域(一般为位于结构表面的环区和转角区域)发生波动而形成一组紧密相关构象子状态集合。此外, 环区和转角的构象波动, 连同蛋白质表面水网络的波动(这一波动源自溶剂熵最大化的需要)一起, 能够传播至整个蛋白质结构, 最终引起功能相关的蛋白分子大尺度协同运动^[6,36,80,81,101-103]。有趣的是, 刘赟等人^[104]将相对熵作为最小目标函数(minimization object function), 已发展出有效的蛋白质设计方法, 这表明熵在蛋白质折叠和解折叠过程中均发挥着重要作用。

综上所述, 本文认为, 在蛋白质-溶剂热力学系统中, 起源于原子热能的熵最大化趋势是蛋白质折叠、结合以及动力学的最根本驱动因子。相反, 另一个重要因子, 即系统的焓减, 则能对抗熵增趋势而维持系统的有序性。正是熵与焓的轮番博弈, 才允许蛋白质-溶剂系统在全局或局部自由能最小化区域达到平衡。

参考文献

- 1 Wright P E, Dyson H J. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J Mol Biol*, 1999, 293: 321-331
- 2 Anfinsen C B. The formation and stabilization of protein structure. *Biochem J*, 1972, 128: 737-749
- 3 Anfinsen C B. Principles that govern the folding of protein chains. *Science*, 1973, 181: 223-230
- 4 Henzler-Wildman K A, Kern D. Dynamic personalities of proteins. *Nature*, 2007, 450: 964-972
- 5 Liu S Q, Xie Y H, Ji X L, et al. Protein folding, binding and energy landscape: a synthesis. In: Kaumaya P T P, ed. *Protein Engineering*. Rijeka: Intech, 2012. 207-252
- 6 Yang L Q, Sang P, Xie Y H, et al. Protein dynamics and motions in relation to their functions: several case studies and the underlying mechanisms. *J Biomol Struct Dyn*, 2014, 32: 372-393
- 7 Karplus M, McCammon J A. Molecular dynamics simulations of biomolecules. *Nat Struct Biol*, 2002, 9: 646-652
- 8 Hansson T, Oostenbrink C, van Gunsteren W. Molecular dynamics simulations. *Curr Opin Struct Biol*, 2002, 12: 190-196
- 9 Kannan S, Zacharias M. Simulated annealing coupled replica exchange molecular dynamics—an efficient conformational sampling method. *J Struct Biol*, 2009, 166: 288-294
- 10 Haliloglu T, Bahar I, Erman B. Gaussian dynamics of folded proteins. *Phys Rev Lett*, 1997, 79: 3090-3093
- 11 Scheraga H A, Khalili M, Liwo A. Protein-folding dynamics: overview of molecular simulation techniques. *Annu Rev Phys Chem*, 2007, 58: 57-83
- 12 Laio A, Parrinello M. Escaping free-energy minima. *Proc Natl Acad Sci USA*, 2002, 99: 12562-12566
- 13 Granata D, Camilloni C, Vendruscolo M, et al. Characterization of the free-energy landscapes of proteins by NMR-guided metadynamics. *Proc Natl Acad Sci USA*, 2013, 110: 6817-6822
- 14 Laio A, Gervasio F L. Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. *Rep Prog Phys*, 2008, 71: 126601-126623
- 15 Kastner J. Umbrella sampling. *Wiley Interdiscip Rev Comput Mol Sci*, 2011, 1: 932-942
- 16 Torrie G M, Valleau J P. Non-physical sampling distributions in monte-carlo free-energy estimation: umbrella sampling. *J Comput Phys*,

- 1977, 23: 187–199
- 17 Kumar S, Rosenberg J M, Bouzida D, et al. Multidimensional free-energy calculations using the weighted histogram analysis method. *J Comput Chem*, 1995, 16: 1339–1350
- 18 Roux B. The calculation of the potential of mean force using computer simulations. *Comput Phys Comm*, 1995, 91: 275–282
- 19 Adcock S A, McCammon J A. Molecular dynamics: survey of methods for simulating the activity of proteins. *Chem Rev*, 2006, 106: 1589–1615
- 20 Paci E, Karplus M. Forced unfolding of fibronectin type 3 modules: an analysis by biased molecular dynamics simulations. *J Mol Biol*, 1999, 288: 441–459
- 21 Schlitter J, Engels M, Kruger P, et al. Targeted molecular-dynamics simulation of conformational change—application to the T \leftrightarrow R transition in insulin. *Mol Simul*, 1993, 10: 291–308
- 22 Jarzynski C. Nonequilibrium equality for free energy differences. *Phys Rev Lett*, 1997, 78: 2690–2693
- 23 Stone J E, Phillips J C, Freddolino P L, et al. Accelerating molecular modeling applications with graphics processors. *J Comput Chem*, 2007, 28: 2618–2640
- 24 Brown W M, Kohlmeyer A, Plimpton S J, et al. Implementing molecular dynamics on hybrid high performance computers—particle-particle particle-mesh. *Comput Phys Commun*, 2012, 183: 449–459
- 25 Harvey M, Giupponi G, De Fabritiis G. Acemd: accelerated molecular dynamics simulations in the microseconds timescale. *J Chem Theory Comput*, 2009, 5: 1632–1639
- 26 Liu S Q, Liang L M, Tao Y, et al. Structural and dynamic basis of serine proteases from nematophagous fungi for cuticle degradation. In: Stoytcheva M, ed. *Pesticides in the Modern World—Pests Control and Pesticides Exposure and Toxicity Assessment*. Rijeka: Intech, 2011. 333–376
- 27 Rapaport D C. *The Art of Molecular Dynamics Simulation*. Cambridge: Cambridge University Press, 1998
- 28 Saguí C, Daren T A. Molecular dynamics simulations of biomolecules: long-range electrostatic effects. *Annu Rev Biophys Biomol Struct*, 1999, 28: 155–179
- 29 Haran G. How, when and why proteins collapse: the relation to folding. *Curr Opin Struct Biol*, 2012, 22: 14–20
- 30 Haynie D T. *Biological Thermodynamics*. Cambridge: Cambridge University Press, 2001
- 31 Gibbs J W. A method of geometrical representation of the thermodynamic properties of substances by means of surfaces. *Trans Conn Acad Arts Sci*, 1873, 2: 382–404
- 32 Bryngelson J D, Onuchic J N, Socci N D, et al. Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Protein Struct Funct Genet*, 1995, 21: 167–195
- 33 Wales D J. *Energy Landscapes*. Cambridge: Cambridge University Press, 2003
- 34 Dill K A. Polymer principles and protein folding. *Protein Sci*, 1999, 8: 1166–1180
- 35 Amadei A, Linssen A B M, Berendsen H J C. Essential dynamics of proteins. *Protein Struct Funct Genet*, 1993, 17: 412–425
- 36 Tao Y, Rao Z H, Liu S Q. Insight derived from molecular dynamics simulation into substrate-induced changes in protein motions of proteinase K. *J Biomol Struct Dyn*, 2010, 28: 143–157
- 37 Kapon R, Nevo R, Reich Z. Protein energy landscape roughness. *Biochem Soc Trans*, 2008, 36: 1404–1408
- 38 Dill K A, Ozkan S B, Weikl T R, et al. The protein folding problem: when will it be solved? *Curr Opin Struct Biol*, 2007, 17: 342–346
- 39 Dill K A, Ozkan S B, Shell M S, et al. The protein folding problem. *Annu Rev Biophys Biomol Struct*, 2008, 37: 289–316
- 40 Levinthal C. Are there pathways for protein folding? *J Chim Phys*, 1968, 65: 44–45
- 41 Yang L Q, Ji X L, Liu S Q. The free energy landscape of protein folding and dynamics: a global view. *J Biomol Struct Dyn*, 2013, 31: 982–992
- 42 Zwanzig R, Szabo A, Bagchi B. Levinthal's paradox. *Proc Natl Acad Sci USA*, 1992, 89: 20–22
- 43 Li W, Qin M, Tie Z, et al. Effects of solvents on the intrinsic propensity of peptide backbone conformations. *Phys Rev E*, 2011, 84: 041933
- 44 Anfinsen C B, Haber E, Sela M, et al. The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc Natl Acad Sci USA*, 1961, 47: 1309–1314
- 45 Agashe V R, Shastry M C R, Udgaonkar J B. Initial hydrophobic collapse in the folding of barstar. *Nature*, 1995, 377: 754–757
- 46 Dill K A. Theory for the folding and stability of globular proteins. *Biochemistry*, 1985, 24: 1501–1509
- 47 Dill K A, Chan H S. From levinthal to pathways to funnels. *Nat Struct Biol*, 1997, 4: 10–19
- 48 Leopold P E, Montal M, Onuchic J N. Protein folding funnels: a kinetic approach to the sequence-structure relationship. *Proc Natl Acad Sci USA*, 1992, 89: 8721–8725
- 49 Dobson C M. The nature and significance of protein folding. In: Pain R H, ed. *Mechanisms of Protein Folding*. 2nd ed. Oxford: Oxford

- University Press, 2000. 1–33
- 50 Karplus M, Weaver D L. Protein-folding dynamics. *Nature*, 1976, 260: 404–406
- 51 Karplus M, Weaver D L. Protein folding dynamics: the diffusion-collision model and experimental data. *Protein Sci*, 1994, 3: 650–668
- 52 Kim P S, Baldwin R L. Specific intermediates in the folding reactions of small proteins and the mechanism of folding. *Annu Rev Biochem*, 1982, 51: 459–489
- 53 Nölting B, Agard D A. How general is the nucleation-condensation mechanism? *Proteins*, 2008, 73: 754–764
- 54 Yan Z, Wang J, Zhang Y, et al. Nucleation process in the folding of a domain-swapped dimer. *Phys Rev E*, 2010, 81: 021910
- 55 Dill K A, Fiebig K M, Chan H S. Cooperativity in protein-folding kinetics. *Proc Natl Acad Sci USA*, 1993, 90: 1942–1946
- 56 Dill K A, Bromberg S, Yue K, et al. Principles of protein folding—a perspective from simple exact models. *Protein Sci*, 1995, 4: 561–602
- 57 Sun L, Wang J, Wang W. Dissection of the zipping-and-assembly mechanism for folding of model proteins. *Chinese Phys Lett*, 2010, 27: 038702
- 58 Harrison S C, Durbin R. Is there a single pathway for the folding of a polypeptide chain? *Proc Natl Acad Sci USA*, 1985, 82: 4028–4030
- 59 Mittal A, Jayaram B, Shenoy S R, et al. A stoichiometry driven universal spatial organization of backbones of folded proteins: are there chargaff's rules for protein folding? *J Biomol Struct Dyn*, 2010, 28: 133–142
- 60 Ji X L, Liu S Q. Is stoichiometry-driven protein folding getting out of thermodynamic control? *J Biomol Struct Dyn*, 2011, 28: 621–623
- 61 Onuchic J N, Luthey-Schulten Z, Wolynes P G. Theory of protein folding: the energy landscape perspective. *Ann Rev Phys Chem*, 1997, 48: 545–600
- 62 Onuchic J N, Wolynes P G, Luthey-Schulten Z, et al. Towards an outline of the topography of a realistic protein folding funnel. *Proc Natl Acad Sci USA*, 1995, 92: 3626–3630
- 63 Wolynes P G, Onuchic J N, Thirumalai D. Navigating the folding routes. *Science*, 1995, 267: 1619–1620
- 64 Li W, Terakawa T, Wang W, et al. Energy landscape and multiroute folding of topologically complex proteins adenylate kinase and 2out-knot. *Proc Natl Acad Sci USA*, 2012, 109: 17789–17794
- 65 Ben-Naim A. Levinthal's question revisited, and answered. *J Biomol Struct Dyn*, 2012, 30: 113–124
- 66 Mezei M. Chameleon sequences in the PDB. *Protein Eng*, 1998, 11: 411–414
- 67 Minor D L, Kim P S. Context-dependent secondary structure formation of a designed protein sequence. *Nature*, 1996, 380: 730–734
- 68 Wolfenden R. Experimental measures of amino acid hydrophobicity and the topology of transmembrane and globular proteins. *J Gen Physiol*, 2007, 129: 357–362
- 69 Cordes M H J, Davidsons A R, Sauer R T. Sequence space, folding and protein design. *Curr Opin Struct Biol*, 1996, 6: 3–10
- 70 Kamtekar S, Schiffer J M, Xiong H, et al. Protein design by binary patterning of polar and nonpolar amino acids. *Science*, 1993, 262: 1680–1685
- 71 Hecht M H, Das A, Go A, et al. *De novo* proteins from designed combinatorial libraries. *Protein Sci*, 2004, 13: 1711–1723
- 72 Ansari A, Berendzen J, Bowne S F, et al. Protein states and protein quakes. *Proc Natl Acad Sci USA*, 1985, 82: 5000–5004
- 73 Kwong P D, Wyatt R, Robinson J, et al. Structure of an HIV-1 gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature*, 1998, 393: 648–659
- 74 Kwong P D, Wyatt R, Majeed S, et al. Structures of HIV-1 gp120 envelope glycoproteins from laboratory-adapted and primary isolates. *Structure*, 2000, 8: 1329–1339
- 75 Huang C C, Tang M, Zhang M Y, et al. Structure of a V3-containing HIV-1 gp120 core. *Science*, 2005, 310: 1025–1028
- 76 Huang C C, Lam S N, Acharya P, et al. Structures of the CCR5 N terminus and of a tyrosine-sulfated antibody with HIV-1 gp120 and CD4. *Science*, 2007, 317: 1930–1934
- 77 Pancera M, Majeed S, Ban Y A, et al. Structure of HIV-1 gp120 with gp41-interactive region reveals layered envelope architecture and basis of conformational mobility. *Proc Natl Acad Sci USA*, 2010, 107: 1166–1171
- 78 Kwon Y D, Finzi A, Wu X, et al. Unliganded HIV-1 gp120 core structures assume the CD4-bound conformation with regulation by quaternary interactions and variable loops. *Proc Natl Acad Sci USA*, 2012, 109: 5663–5668
- 79 Chen B, Vogan E M, Gong H, et al. Structure of an unliganded simian immunodeficiency virus gp120 core. *Nature*, 2005, 433: 834–841
- 80 柳树群, 符云新, 刘次全. 不同构象状态 HIV-1 gp120 分子运动特征及构象转换能力研究. *科学通报*, 2007, 52: 2494–2506
- 81 Liu S Q, Liu S X, Fu Y X. Molecular motions of human HIV-1 gp120 envelope glycoproteins. *J Mol Model*, 2008, 14: 857–870
- 82 Perozzo R, Folkers G, Scapozza L. Thermodynamics of protein-ligand interactions: history, presence, and future aspects. *J Recept Signal Transduct Res*, 2004, 24: 1–52
- 83 Ji X L, Liu S Q. Thinking into mechanism of protein folding and molecular binding. *J Biomol Struct Dyn*, 2011, 28: 995–996
- 84 Fischer E. Einfluss der configuration auf die wirkung der enzyme. *Ber Dtsch Chem Ges*, 1894, 27: 2984–2993
- 85 Koshland D E J. Application of a theory of enzyme specificity to protein synthesis. *Proc Natl Acad Sci USA*, 1958, 44: 98–104

- 86 Tobi D, Bahar I. Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state. *Proc Natl Acad Sci USA*, 2005, 102: 18908–18913
- 87 Bosshard H R. Molecular recognition by induced fit: how fit is the concept? *News Physiol Sci*, 2001, 16: 171–173
- 88 Foote J, Milstein C. Conformational isomerism and the diversity of antibodies. *Proc Natl Acad Sci USA*, 1994, 91: 10370–10374
- 89 Changeux J P, Edelman S. Conformational selection or induced fit? 50 years of debate resolved. *F1000 Biol Rep*, 2011, 3: 19
- 90 Serban D, Taraboulos A, DeArmond S J, et al. Rapid detection of Creutzfeldt-Jakob disease and scrapie prion proteins. *Neurology*, 1990, 40: 110–117
- 91 Kosik K S. Alzheimer's disease: a cell biological perspective. *Science*, 1992, 256: 780–783
- 92 Pan K M, Baldwin M, Nguyen J, et al. Conversion of α -helices into β -sheets features in the formation of the scrapie prion proteins. *Proc Natl Acad Sci USA*, 1993, 90: 10926–10966
- 93 Weinreb P H, Zhen W, Poon A W, et al. NACP, a protein implicated in Alzheimer's disease and learning, is natively unfolded. *Biochemistry*, 1996, 35: 13709–13715
- 94 Csermely P, Palotai R, Nussinov R. Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. *Trends Biochem Sci*, 2010, 35: 539–546
- 95 Ma B, Kumar S, Tsai C J, et al. Folding funnels and binding mechanisms. *Protein Eng*, 1999, 12: 713–720
- 96 Tsai C J, Xu D, Nussinov R. Protein folding via binding and vice versa. *Fold Des*, 1998, 3: R71–R80
- 97 Xie Y H, Sang P, Tao Y, et al. Protein folding and binding funnels: a common driving force and a common mechanism. *J Biomol Struct Dyn*, 2013, 31: 100–101
- 98 Tsai C J, Kumar S, Ma B, et al. Folding funnels, binding funnels, and protein function. *Protein Sci*, 1999, 8: 1181–1190
- 99 Xie Y H, Tao Y, Liu S Q. Wonderful roles of the entropy in protein dynamics, binding and folding. *J Biomol Struct Dyn*, 2013, 31: 98–100
- 100 Xu X, Su J, Chen W, et al. Thermal stability and unfolding pathways of Sso7d and its mutant F31A: insight from molecular dynamics simulation. *J Biomol Struct Dyn*, 2011, 28: 717–727
- 101 Liu S Q, Liu C Q, Fu Y X. Molecular motions in HIV-1 gp120 mutants reveal their preferences for different conformations. *J Mol Graph Model*, 2007, 26: 306–318
- 102 Liu S Q, Meng Z H, Fu Y X, et al. Insights derived from molecular dynamics simulation into the molecular motions of serine protease proteinase K. *J Mol Model*, 2010, 16: 17–28
- 103 Liu S Q, Meng Z H, Fu Y X, et al. The effect of calciums on the molecular motions of proteinase K. *J Mol Model*, 2011, 17: 289–300
- 104 刘赞, 王宝翰, 王存新, 等. 基于相对熵的蛋白质设计新方法. *中国科学 G 辑: 物理学 力学 天文学*, 2003, 33: 348–356