第 41 卷 第 4 期 力 学 进 展 Vol. 41 No. 4
2011 年 7 月 25 日 ADVANCES IN MECHANICS July 25, 2011

蛋白质折叠的统计热力学基础
—— 非共价作用的统计热力学原理*

孙卫涛 †

清华大学周培源应用数学研究中心, 北京 100084

摘 要 蛋白质折叠是一个复杂的物理、化学和生物过程, 涉及热力学、统计力学、高分子动力学等多个领域的相关知识, 单独从宏观或微观角度出发, 试图建立蛋白折叠物理模型, 都面临着巨大的困难. 统计热力学方法是研究蛋白质分子运动规律的重要途径, 本文从溶液环境下蛋白质折叠的非共价作用出发, 介绍了疏水作用、氢键、静电力和范德华力的统计热力学机理, 并探讨了这些作用力对蛋白质结构稳定性的影响. 通过深入理解微观结构相互作用与系统熵、焓、自由能之间关系, 有助于我们发展蛋白质折叠的统计热力学理论, 进而为蛋白质结构宏观热力学研究提供重要的微观统计力学机理.

关键词 蛋白质折叠, 自由能, 统计热力学

1 前 言

蛋白质是由细胞内核糖体 (ribosome) 合成的链状氨基酸高分子聚合物 (也称为多肽链 (polypeptide chain)), 是生命活动必不可少的物质基础. 蛋白质分子包含从几十个到上千个数量不等的氨基酸单元 (称为残基 (residue)), 普通蛋白质分子由数百个氨基酸串联组成. 自然界存在 20 种氨基酸, 不同种类氨基酸在蛋白质链上的排列顺序称为氨基酸序列 (sequence), 氨基酸序列决定了蛋白质之间的差别.

蛋白质通过多肽链折叠形成特殊的三维空间分子结构, 并由此实现对应的生物功能, 这种空间结构称为天然态结构 (native structure). 蛋白质天然态结构特征可以通过 X 射线散射 [1]、核磁共振 (NMR)[2-6] 和低温电镜 (cryo-electron microscopy)[7-11] 等实验技术加以观测. Pauling 等 [12-13] 指出蛋白质内部存在规则的螺旋结构和片层结构, 正是这些结构导致了 X 射线中的 α 和 β 散射图样. 此后不久, Linderstrom 和 Schellman[14] 发现蛋白质内部的结构组织具有层次性, 并首次引入了一级结构、二级结构、三级结构的名称. 一般认为, 蛋白质结构可以分为 4 种组织层次 [15]: 一级结构、二级结构、三级结构和四级结构. 一级结构指蛋白质多肽链的氨基酸序列. 二级结构是多肽链局部折叠形成的稳定规则结构, 包括 α 螺旋 (helix)、β 片层 (sheet) 和转角结构 (turn) 等. 二级结构的生成与维持需要依靠不同氨基酸的 C=O 和 N—H 基团间的氢键作用. 三级结构指多个二级结构在空间中紧密排列所形成的三维结构. 蛋白质的三级结构通常依靠疏水核心、双硫键、氢键等来稳定结构. 四级结构指不同蛋白质多肽链间相互作用形成具有不同功能的复合物分子形态.

与天然态结构相对应的是去折叠结构 (unfolded structure). 理想状态的去折叠态蛋白质结构

收稿日期 : 2010-09-17, 修回日期 : 2011-06-03
* 清华大学自主科研计划 (20101081751) 和林家翘专项捐款资助
† E-mail: sunwt@tsinghua.edu.cn

呈现无规卷曲 (random coil) 状态, 没有形成规则的二级结构. 研究表明, 在强变性剂环境下 (如 8M 尿素), 去折叠态蛋白质的流体力学属性与无规卷曲多肽链相符 [16-17]. 但是在没有变性剂的情况下, 即使是极端的 pH 值和温度环境, 去折叠蛋白质也不是真正的无规卷曲结构 [18-20]. 在无规卷曲状态下, 多肽链不同部分之间的相互作用与结构–溶液之间相互作用完全达到平衡. 然而, 蛋白质由 20 种具有不同物理化学属性的氨基酸组成, 这种异质多肽链 (heterogeneous polypeptide chain) 结构内部相互作用不易与结构–溶液间作用力完全平衡, 多肽链本身不同部分之间的相互作用往往更占优势, 因此去折叠态的蛋白质结构比无规卷曲结构更加紧密, 更加有规则.

除了天然折叠态和去折叠结构, 许多蛋白质还具有稳定的中间态结构. 具有如下特征的中间态结构一般称为熔球态 (molten globule)[21]: (1) 整体大小十分接近折叠态结构; (2) 其二级结构的含量与天然态蛋白质相仿; (3) 结构内部氨基酸侧链的空间分布更均匀; (4) 与折叠态相比, 结构内部氨基 (amide) 与溶液之间氢原子的交换更加迅速, 但是相比去折叠态更缓慢; (5) 中间态结构的焓与去折叠态结构十分接近, 与折叠态结构相去甚远; (6) 中间态/去折叠态结构之间转换迅速且具有非协作性 (nonco-operative), 中间态/折叠态结构之间的转换缓慢且具有协作性 (co-operative)[22-24].

自然状态下, 蛋白质在水溶液中从去折叠态结构转变为天然态结构, 以实现相应的生物功能, 这个过程称为蛋白质折叠. 去折叠结构和天然态结构是蛋白质折叠过程起点和终点. 较大蛋白质分子 (如多域蛋白 (multi-domain protein)) 的折叠过程通常会出现中间态结构 (见图 1), 而小蛋白的折叠过程是去折叠态和天然态之间的结构转变.

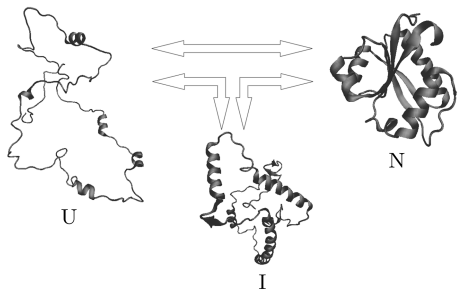


图 1 蛋白质折叠过程中去折叠态 (U)、中间态 (I) 和天然态 (N) 结构转变示意图. 较大蛋白质分子具有相对稳定的折叠中间态 (I), 小蛋白分子折叠过程是去折叠态 (U) 和折叠态 (N) 之间的二态转变

蛋白质折叠是个可逆过程, 在添加变性剂、改变温度和 pH 值、施加压力等情况下, 处于天然态结构的蛋白质分子通常会发生去折叠 (unfold) 现象. 同时, 如果溶液条件允许, 即使在生物体外 (如实验室条件下), 不同初始结构的蛋白质分子 (这里主要指水溶性球蛋白) 也能够实现快速折叠, 并完全恢复到相同的天然态结构 [25-26]. 早在 19 世纪 30 年代, 中国科学家吴宪就提出球蛋白的变性 (denaturation) 是从紧密有序结构向松散无序结构的转变 [27]. 同时, 人们还发现蛋白质结构在发生变性后, 能够自发地重新折叠到天然态结构 [28-29]. 20 世纪 70 年代, Anfinsen 等 [25-26] 通过核糖核酸酶的去折叠–折叠实验表明, 蛋白质的天然折叠态结构处于热力学平衡状态, 天然态结构只依赖于氨基酸序列和溶液环境. 因此, 蛋白质天然态结构特征仅依赖于氨基酸序列, 与折叠是否发生在生物体内无关. 特殊的氨基酸序列是如何决定蛋白质三维空间结构, 这其中的奥秘依然不为人所知, 这也就是我们现在要研究的主题.

在蛋白质折叠和结构预测研究中, 可以分别从热力学、量子力学和统计力学多个角度研究. 表 1 给出了 3 种理论的对比关系 [30].

表 1 热力学、量子力学和统计力学的比较 [30]

	热力学	量子力学	统计力学
研究原理	涉及温度、热量、功、能量、可逆性等, 全部是宏观系统物理量	通过波函数表示物理系统	涉及物理系统的平均值和概率
研究对象	处于平衡态的宏观真实物体	电子和原子核, 或者是其他粒子的集合	真实粒子或者模型粒子的集合, 这些粒子可以是复合物
研究结论	建立可测量的宏观平衡态属性之间的关系	得到量子态的能级, 或者不同量子态之间跃迁的概率	真实或者模型系统的宏观属性, 热力学、输运和弛豫 (relaxation) 系数

统计热力学 (statistical thermodynamics) 将热力学与统计力学联系起来, 从物质的分子 (原子) 结构出发直接导出热力学函数, 使对热力学定理的理解更为清晰. 本文将重点介绍蛋白质折叠过程中疏水, 氢键、静电力等主要物理作用的统计热力学基础, 并探讨非共价作用对结构折叠过程中自由能的影响.

2　蛋白质折叠问题研究状况

生物体通过 DNA/RNA 遗传物质的基因表达, 仅仅确定了蛋白质一级结构的氨基酸序列, 而非氨基酸多肽链的三维空间结构. 然而, 特殊的氨基酸排布顺序决定了残基在结构中相互接触 (contact) 和相互作用的模式 (pattern), 这种非局部 (nonlocal) 的相互作用影响了蛋白质的折叠过程. 氨基酸序列包含蛋白质天然态结构特征的 "密码", 但两者之间并非严格的对应关系, 当氨基酸序列上的某些氨基酸发生突变时, 蛋白质仍能够保持结构不被破坏, 并继续执行生物功能, 这对物种的进化过程是至关重要的 [31]. 氨基酸序列的差异导致了蛋白质结构的不同, 因此蛋白质折叠的 "隐秘规则" 一定包含在代表氨基酸差别的残基侧链以及残基的排列顺序中. 对氨基酸序列如何决定蛋白质空间三维结构的研究, 被称为 "蛋白质折叠问题".

蛋白质折叠问题包含 3 个不同的方面 [21,32]: (1) 蛋白质折叠的热力学问题, 即在原子相互作用下氨基酸序列怎样形成稳定存在的天然态结构; (2) 蛋白质折叠的动力学问题, 即蛋白质通过何种折叠路径快速达到天然态结构; (3) 蛋白质结构预测的计算方法问题, 即怎样从蛋白质氨基酸序列出发预测出天然态蛋白质空间结构. 完全理解这 3 个问题是蛋白质折叠问题的最终目标.

达尔文的进化论观点认为, 自然界存在 "物竞天择, 适者生存" 的竞争和淘汰规律. 自然界存在数以百万计的蛋白质, 这些蛋白质具有不同的氨基酸序列, 折叠成不同的三维空间结构. 在自然选择和进化压力的驱动下, 为了满足结构快速折叠和正确执行生物功能, 蛋白质氨基酸序列得到不断优化 [33-34]. 在形成特定的氨基酸序列之后, 结构折叠的动力学过程和热力学平衡态则是物理、化学、生物等研究领域的问题. 随着晶体结构和去折叠实验的不断发展, 支配蛋白质结构发生折叠的物理原理日渐明显. 蛋白质分子在溶液中处于布朗运动状态, 存在各种各样的空间结构. 氨基酸多肽链的不同结构状态构成了分子结构的系综 (ensemble), 在溶液中每一种结构出现概率与对应 Gibbs 自由能状态有关. 折叠过程中, 疏水基团在蛋白质分子内部聚集, 带电基团之间形成氢键, 同时球蛋白分子与溶液接触面积减少, 这些现象都表明蛋白质折叠伴随着 Gibbs 自由能降低. 因此, 热力学基本定律认为, 在恒定的生理环境下, 蛋白质折叠是一个自由能下降过程 [35].

如果把去折叠过程中天然态结构三级结构的打开、二级结构的消失看作 "溶解" 过程 [36-37], 天然态结构与去折叠结构之间的自由能差别看作 "溶解自由能"(solvation free energy), 则蛋白质分子 "溶解自由能" 决定了蛋白质的折叠–去折叠过程. 蛋白质分子在溶液中的溶解自由能与溶液分子的性质有关 (包括 pH 值、极化率等). 这些性质对疏水作用、静电作用、氢键作用等物理作用都有重要影响, 因而也决定了折叠过程中不同分子结构的自由能高低. 溶液条件影响着蛋白质折叠效率, 在溶液环境中, 扩散速度、溶解效应等因素决定了蛋白质分子不同折叠状态之间的自由能差别. 对于给定氨基酸序列, 蛋白质分子是处于天然态还是去折叠态、是否存在中间态结构, 都由对应结构的稳定性 (即自由能高低) 决定, 从根本上说, 也就是由溶液的物理、化学性质决定.

一般认为, 在一定的溶液条件下, 蛋白质天然态结构具有最小自由能 (free energy)[25-26,38]. 然而, 蛋白质分子可能的结构数量是十分庞大的 (例如, 100 个氨基酸组成的分子链可以有 10^{18} 种不同空间结构), 在如此广袤的结构空间中遍历所有结构并搜索最小自由能状态是艰巨的任务 (保守估计需要 10^{30} 年)[39], 如此长的折叠时间显然是不现实的, 绝大多数蛋白质分子都能在数秒内完成折叠过程, 这就是所谓的 Levithal 佯谬 (paradox). Levithal 佯谬表明, 蛋白质折叠不是随机遍历所有可能的结构, 而是沿一条包含有限数目构象的特殊折叠途径达到天然态结构.

目前, 对蛋白质结构空间的研究普遍认为, 不同能量状态下结构数量是不同的, 随着结构自由能降低, 可取的蛋白质分子结构数量逐渐减少. 结构自由能曲面整体呈漏斗 (funnel) 状 (见图 2), 天然态蛋白质结构对应于能量最低点 [40-44]. 这个模型允许蛋白质从大量各自不同的去折叠态结构出发, 在不断降低的自由能引导下, 经过不同折叠路径和中间态结构迅速折叠到天然态结构. 该模型得到了蛋白质结构模拟计算和实验数据的证实 [45].

蛋白质折叠路径沿漏斗状能量面下降, 这个猜想与疏水塌缩假设紧密相关. 球形蛋白质天然态结构内部通常包含由非极性氨基酸组成的疏水核心, 而极性和带电残基则分布在蛋白质表面, 暴

露在溶液环境中. 这样的结构特征将非极性残基与水分子隔离开来, 结构能量达到最稳定状态. 在蛋白质折叠初始阶段, 氨基酸多肽链上的疏水残基快速聚集, 形成紧密疏水核心并引起蛋白质结构迅速缩小, 这个过程称为疏水塌缩. 疏水塌缩发生在大量二级结构和三级结构出现之前, 并引导蛋白质进入熔球态结构. 熔球态对应于能量漏斗中相对较低的部分, 但是距离能量最低点的天然态结构仍有一段距离, 需要较长时间才能逐渐进入天然态结构.

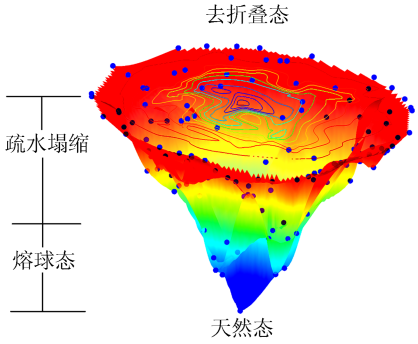


图 2 蛋白质结构漏斗形能量曲面示意图. 图中的点代表具有不同能量状态的蛋白质分子结构状态

近年来, 蛋白质折叠结构预测计算方法取得了很大的进展 [32,46]. 随着越来越多的蛋白质晶体结构被解析出来, 基于蛋白质数据库 (PDB) 的结构预测方法突飞猛进. 较早运用数据库信息是二级结构预测算法 [47-49], 该算法根据数据库统计信息, 从氨基酸序列出发预测蛋白质局部二级结构, 对球形蛋白来说能达到 80% 的正确性 [50]. 二级结构预测方法已被应用于蛋白质折叠识别 (fold recognition)[51-52]、从头开始 (ab initio) 蛋白质结构预测 [53-54]、结构基序 (structural motif) 分类 [55] 和序列比对 (sequence alignment)[56-57] 等研究中.

相比二级结构预测方法, 三级结构预测的难度更大. 目前, 三级结构预测方法主要包括对比蛋白质建模方法 (comparative protein modelling) 和从头开始预测方法两类. 前者以结构数据库中已经解析的蛋白质片段为模板 (template), 通过同源建模 (homology modeling) 和序列比对等手段, 拼接数据库中已知的蛋白质结构片段 [51-52,58]. 该方法的基本依据是蛋白质三级结构样式的有限性, 尽管自然界有数百万种不同的蛋白质, 但是自然界中独特的三级结构折叠样式却有 2 000 种左右, 因此蛋白质结构可以表示为这些独特样式的不同组合. 对于较小的球形蛋白分子, 预测结构与实验结果已经十分接近 (误差在 2~6 Å)[59-62]. 从头开始的预测方法试图直接从物理原理出发预测三维空间结构, 而不需要使用已经解析的蛋白质结构信息. 这类方法包括基于牛顿运动方程的分子动力学方法 [63-74]、基于 Langevin 方程 [75-82] 和 Monte Carlo 算法 [83-86] 的随机动力学方法 [87-89] 等. 基于物理原理的方法已经取得了巨大进展 [90-92], 但是该方法需要很大的计算量, 往往局限于较小蛋白质或者简化蛋白质模型的结构预测. 这种方法的优势非常明显, 如可以帮助人们理解蛋白质在生物制药中的结构变化机理, 理解蛋白质折叠、结构转变等动力学过程的机理, 帮助设计合成蛋白质或者可以折叠的多聚物等等.

3 蛋白质折叠的统计热力学方法

热力学 (thermodynamics) 和动力学 (kinetics) 理论是蛋白质折叠机理研究中两种主要思路. 热力学通过温度、体积和粒子数目刻画包含大量粒子的系统, 用熵 (entropy)、焓 (enthalpy) 和总能量等参数描述系统整体行为, 而非研究系统中每个单独的粒子. 热力学理论认为蛋白质的自然态结构具有最小自由能, 蛋白质折叠过程具有漏斗状的能量面. 动力学理论认为自然态蛋白质具有动力学意义上最容易达到结构, 蛋白质折叠过程中存在特殊路径 [93-96](如生成某些中间态结构), 氨基酸多肽链可以沿折叠路径达到天然态结构, 而不需要搜索所有结构. 以扩散碰撞 (diffusion-collision) 模型为例 [97-98], 该理论认为蛋白质折叠过程中存在微结构域 (microdomain)(如二级结构的初始形态、疏水残基团簇等) 的涨落 (fluctuation), 微结构域的扩散运动和碰撞导致相互结合并形成中间态结构, 中间态结构的存在加速了蛋白质的折叠过程. 此外, 分阶段的折叠观点认为蛋白质折叠过程可以分为多个阶段 [99]. 在折叠初始阶段, 溶液与残基之间的疏水作用驱使蛋白质结构快速塌缩, 在特定的残基相互作用模式 (由氨基酸序列特殊性决定) 帮助下, 疏水塌缩促使二级结构出现; 当蛋白质接近天然态结构时, 热力学机制占据主导地位, 引导蛋白质结构向自由能最低的状态转变.

蛋白质折叠是一个随机过程 (stochastic process). 蛋白质在生物体内以大量分子的集体形式存在, 分子群体运动形成可观测的宏观物理量. 此外, 单个分子在溶液环境下作无规则布朗运动, 对

单个分子的运动轨迹进行描述是极其困难的. 因此, 统计力学的方法是研究蛋白质分子运动规律的最佳途径. 同时, 蛋白质分子并非只有单一空间结构, 而是相互关联、相互转换的微观状态 (microstate) 组成的系综. 即使是 X 射线和核磁共振 (NMR) 观测到的天然态蛋白质晶体结构, 实际上也是不同微观状态的系综平均. 从这个意义上说, 自然态蛋白质结构是一组微观结构状态的集合 [100], 这些微观状态代表了蛋白质晶体结构的随机扰动. 蛋白质的宏观物理属性和可观测生物功能是微观结构状态集合的系综平均, 因此, 自然态结构系综的随机涨落对蛋白质的物理属性和生物功能产生重要影响.

通过研究原子、分子间相互作用, 掌握蛋白质系综内部微观状态的结构特征、能量分布规律, 有助于深入了解蛋白质折叠机理, 也有助于从微观角度解释宏观可观测物质属性. 同时, 在蛋白质折叠过程中, 始终存在熵和焓之间的微妙平衡关系, 结构状态与自由能的微小变化紧密相关 [101]. 热力学状态是连接蛋白质结构与生物功能的桥梁, Gibbs 自由能刻画了蛋白质结构状态的出现概率 (即结构稳定性), 也反映了不同结构状态之间发生转变的可能性. 因此, 在统计力学基础上, 还需要研究溶液与蛋白质分子的热力学性质.

统计热力学认为: 在给定的温度、体积和粒子总数条件下, 蛋白质分子微观结构状态是千差万别的; 而在整个宏观系统中, 蛋白质分子处于统计概率最大的结构状态 [102]. 因此, 统计热力学给出了宏观系统平衡态性质的微观分子理论解释 [103]. 统计热力学理论包含经典热力学和统计力学的内容, 但是又具有本身的特点 [104]. 与经典热力学之间的差别在于, 经典热力学通过实验测量得到各种热力学函数, 而统计热力学则基于统计力学基本原理, 从分子模型出发经过理论推导得到各种热力学函数; 与统计力学的区别在于, 在统计热力学中, 从统计力学原理出发得到的热力学函数和法则将被广泛用于系统宏观属性研究.

蛋白质折叠过程受到多种作用力的影响, 包括疏水作用、氢键作用、静电力作用和范德华力作用等. 从能量角度看, 虽然天然态蛋白质结构比去折叠态结构更稳定, 但是两者之间的能量差别很小, 仅有 5~10 kcal/mol(蛋白质内部一个氢键强度约为 1~4 kcal/mol[105-106], 注 1 cal = 4.18 J), 因此, 显然无法忽略各种分子间作用力对折叠过程的微小影响.

在诸多作用力中, 疏水作用是驱使蛋白质主链快速塌缩并形成疏水核心的主要力量 [107], 在蛋白质折叠过程中发挥最主要作用. 这些重要作用表现在: 球形蛋白质通常具有疏水核心, 使疏水残基远离水环境使结构能量更加稳定; 如果将蛋白质放入非极性溶液中, 由于缺少极性溶液环境下的疏水作用, 蛋白质主链通常会打开, 进入去折叠状态. 研究表明, 即使缺少适当的静电和氢键作用, 如果蛋白质主链疏水/亲水残基分布特征正确, 蛋白质分子依然能够折叠成天然态结构 [108-110]. 主链氢键作用是形成二级结构的关键作用力 [12], 但是二级结构本身通常在溶液中是不稳定的, 蛋白质分子主链的紧密填埋使 α 螺旋和 β 片层结构更加稳定; 同时, 也只有 α 螺旋和 β 片层结构这样的规则结构才能使氨基酸多肽链在空间排布更紧密. 作为一种长程作用力, 静电作用通常不被认为是蛋白质折叠的重要力量, 一方面是因为多数蛋白质的带电残基含量较少, 而且分布在蛋白质表面具有较高介电常数区域, 另一方面蛋白质结构稳定性似乎与溶液离子浓度无关, 带电残基的突变通常对结构稳定性的影响很小. 但是, 随着研究的不断深入, 静电作用的影响逐渐变得不可忽视. 范德华力在折叠过程中也是至关重要的, 这是因为天然态蛋白质具有紧密的内部结构, 范德华力作为一种近程作用发挥着不可替代的作用 [111].

蛋白质内部以及蛋白质–溶液环境之间的相互作用力与结构稳定性紧密相关, 特别是作用力与结构自由能之间的关系, 决定了蛋白质天然态结构能否稳定存在. 因此, 对疏水作用、氢键作用、静电作用和范德华力等非共价作用的统计热力学原理研究至关重要. 对非共价作用力的研究要点包括: (1) 作用力的物理机制; (2) 作用力在蛋白质折叠过程中起到的作用; (3) 作用力变化对结构自由能和稳定性的影响; (4) 作用力强度和对应能量的计算方法等. 统计热力学要求蛋白质折叠是一个各态遍历 (ergodic) 过程, 这是一个重要的前提假设. 非共价作用力研究的基本方式是从物理定律 (如库仑定律) 出发, 在合理的模型简化基础上 (如点电荷模型、均匀介电常数假设), 推导出作用力的计算公式.

基于统计热力学原理的非共价作用力分析近年来取得了许多积极进展. 通过对蛋白质折叠过程中过渡态 (transition state) 结构系综的热力学平

衡态测量, 发现了使过渡结构稳定存在的作用力, 这些作用力来自于某些特殊氨基酸侧链的相互作用 [112]. Lazaridis 和 Paulaitis[113-114] 计算了简单小分子溶液水合作用的熵, 在此基础上, Ashbaugh 和 Paulaitis[115] 计算了甲烷和乙烷等分子水合作用的熵, 与实验结构吻合得很好. 在蛋白质配体结合过程中, 将溶解自由能分解为熵和焓两部分, 有利于进一步理解整个生物过程, 这种方法已经被用于配体结合自由能计算 [116-118]. 尽管如此, 统计热力学方法仍然面临着一些挑战, 例如, 结构自由能通常是熵和焓共同作用、相互补偿、相互抵消的结果, 因而往往变化量很小, 不易分析对整体结构稳定性的影响. 对蛋白质折叠过程中各种非共价作用的统计热力学分析, 可以帮助我们深入理解蛋白质各种结构状态的稳定性来源, 推动对折叠机理和结构预测研究的发展.

4 疏水作用与分子溶解自由能

4.1 疏水作用 —— 蛋白质折叠的主要驱动力

自然界存在 20 种氨基酸, 根据侧链的物理、化学属性不同, 可以分为疏水、极性、带电等多种类型. 大量研究表明 [36,101,107,119-123], 疏水作用是蛋白质结构在溶液环境中发生快速塌缩的重要驱动作用. 疏水残基具有类似于两性分子 (amphiphilic) 的结构, 残基侧链末端是非极性的碳氢 (烃) 基团, 另一端则是主链部分具有极性特征的碳氧 (羰基) 基团 C=O 和氮氢 (氨基) 基团 N—H. 在非极性端 “厌水” 和极性端 “亲水” 作用下, 疏水残基相互聚集, 在蛋白质结构内部形成紧密的疏水核心, 其中疏水末端构成蛋白质内部的疏水 “微小球团”(globule), 亲水一端多数形成彼此间氢键作用. 同时, 极性氨基酸残基也会在疏水球团外部形成亲水 “包裹层”. 因此, 蛋白质折叠态结构可以视为内部具有疏水球团核心、外部拥有亲水表面层的类球形结构. 图 3 给出了肌红蛋白 Myoglobin 的折叠态结构示意图. 外层亲水残基中, 小球状原子共价键连接模型 (红色) 表示带负电 (acidic) 残基 (包括 Asp Glu), 棒状模型 (蓝色) 表示带正电 (basic) 残基 (包括 Lys Arg), 球形模型 (绿色) 表示极性 (polar) 残基 (包括 Ser,Thr,Tyr,His,Cys,Asn,Gln,Trp); 内部球状 (白色) 残基 (具有范德华半径的实心球状模型) 表示疏水 (hydrophobic) 残基 (包括 Ala Val Phe Pro Met Ile Leu). 氨基酸缩写名称表示 Gly: 氨基乙酸, Ala: 丙胺酸, Val: 缬氨酸, Phe: 苯基丙氨酸, Pro: 脯氨酸, Met: 蛋氨酸, Ile: 异亮氨酸, Leu: 亮氨酸, Asp: 天 (门) 冬氨酸, Glu: 谷氨酸, Lys: 赖氨酸, Arg: 精氨酸 Ser: 丝氨酸 Thr: 苏氨酸 Tyr: 酪氨酸 His: 组氨酸 Cys: 半胱氨酸 Asn: 天冬酰胺酸 Gln: 谷氨酸盐 Trp: 色氨酸.

疏水作用是水溶液环境下球形蛋白质折叠的主要驱动力 [36,107,120], 疏水残基远离溶液环境并向蛋白质核心的埋藏过程导致了结构的快速折叠. 1959 年, Kauzmann[107] 提出通过改变周围溶液可以测量疏水残基自由能变化, 并指出疏水残基进入蛋白质内部后, 自由能降低约为 −8 kcal/mol 至 −25 kcal/mol. 这个结果表明疏水作用是蛋白质折叠过程中最为重要的能量因素. 一般认为, 疏水作用来源于水分子氢键网络被疏水残基破坏后引起的熵减效应. 相对于氨基酸之间的直接相互作用而言, 水分子相互作用是疏水效应的更重要原因 [107]. 但是, 导致疏水作用的热力学因素十分复杂, 至今为止, 直接给出疏水作用的定量计算公式仍然十分困难.

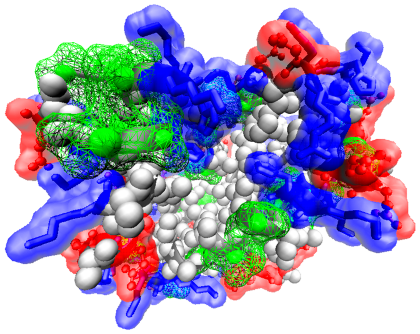


图 3 Myoglobin (2BLH) 的疏水核心和亲水外层

4.2 疏水作用的统计热力学机理

疏水基团溶解的热力学过程是理解疏水作用的关键, 许多学者研究了驱动疏水基团发生聚集的统计热力学机制 [101,119,124-128], 特别是溶液环境下水分子与疏水基团之间的交界面特征 [126]. 由于疏水基团在蛋白质内部的聚集, 引起结构中疏水相和亲水相的分离, 并形成分界面. 与均匀液体中分子之间相互作用的距离相比, 界面的尺度相当大, 足够形成类似于分子微团、双分子层等介观尺度的结构. 在溶液环境中, 水分子之间形成氢键网络, 分子间距离约为 3~5 Å. 在这个尺度上, 水分子的空间位置处于不断调整过程中, 彼此间的氢键作用也处于不断消失–生成状态.

通常认为, 在疏水基团侵入之后, 水分子会在

疏水基团周围形成具有规则结构的 “水笼”, 将疏水基团包围在中间. 水分子网络中规则结构的出现导致系统熵减少, 因而驱使疏水基团聚集成簇, 最大程度上降低水分子–疏水基团接触面积, 以期减少规则结构的出现. 但是, 这种刚性 “笼状” 结构的物理解释存在明显缺陷, 普通流体分子间的相关度较弱, 还不足以形成类似晶体结构的规则 “水笼”, 特别是在具有很大接触面积的疏水基团附近,“笼状” 结构在几何上更难以想象.

疏水基团在水溶液中阻断氢键网络, 排斥水分子并侵占近似球状的空间. 在这部分 “空腔” 内, 水分子无法存在, 更不能形成氢键作用, 因此通过研究 “空腔” 在水溶液中的溶解过程, 可以进一步了解疏水作用的物理基础 [101,125,129]. 溶液中 “空腔” 水合作用 (hydration) 的研究指出 [126,130], 对于较小的疏水基团 (如甲烷分子), 其体积尺度小于 5 Å, 相对于水分子之间氢键的作用尺度 (3~5 Å) 来说, 在无需打断水分子氢键的情况下, 这类小分子就能够存在于水分子氢键网络的空隙之中(见图 4).水分子之间通过调整彼此间相对位置, 使氢键分布在疏水小分子附近, 这个氢键位置重新分布过程近似于水溶液中氢键网络的自发动态调整过程. 对于具有较大体积的疏水基团来说, 基团表面积较大 (常超过 10 Å^2) 且曲率较低, 表面附近的水分子无法形成并维持完整的氢键网络. 在水溶液中每个水分子可以与周围 4 个水分子形成氢键, 而在大疏水基团附近, 每个水分子平均氢键个数只有 2~3 个. 因此, 水分子具有远离大疏水基团的趋势, 并在基团附近形成类似于液体–气体的交界面 [130].

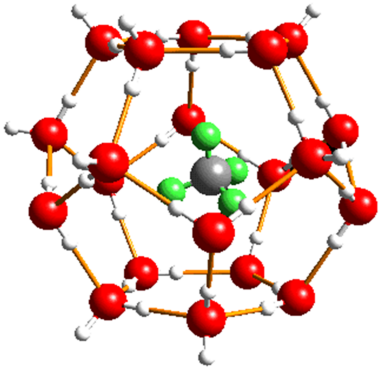


图 4　非极性小分子的疏水效应示意图

自由能是热力学系统状态发生改变的重要因素. 分子溶解自由能 ΔG 是指溶液结构重新调整并使溶质发生溶解的可逆功 [126,131], 包含焓的改变量 ΔH 和熵的改变量 ΔS

$$\Delta G = \Delta H - T\Delta S \tag{1}$$

式中 T 是绝对温度. 焓表示分子之间相互作用势能的平均强度, 熵表示了分子之间相关性或者分子结构的有序程度. 在热力学状态改变过程中, 当分子之间的相互作用发生明显改变时 (如氢键数量的减少和液体–气体界面的生成), 自由能 ΔG 主要受焓的影响, $\Delta G/T$ 随温度的升高而降低. 在热力学状态改变过程中, 当氢键在空间中产生特殊排布结构时 (如氢键网络中出现规则界面), 自由能主要受熵的影响. 室温下, 小疏水基团在溶液中的水合作用受到熵的控制 [101], 自由能 ΔG 随温度的升高而增加.

自由能给出了分子系统的变化趋势, 当分子系统变化过程导致 $\Delta G < 0$ 时, 该过程使系统趋向于平衡态, 是一个自发过程 (spontaneous process). 因此, 在焓改变量 ΔH 的最小化和熵改变量 ΔS 的最大化推动下, 分子系统到达平衡状态. 在蛋白质折叠过程中, 打开的蛋白质结构可取的空间构型数目远远大于折叠态结构, 即 $\Delta S > 0$; 疏水、氢键等非共价作用使折叠态蛋白质结构保持紧密, 破坏这些作用力并使蛋白质结构打开需要输入能量, 因此 $\Delta H > 0$. 根据自由能的定义, ΔG 依赖于焓 ΔH 和熵 ΔS 值的变化, 自由能改变量有可能是正值, 也有可能是负值.

通过 Boltzmann 因子, 分子在溶液环境中的溶解概率 P 与分子溶解自由能 ΔG 之间建立了明确的关系

$$P \propto \exp\left(-\Delta G/k_{\mathrm{B}}T\right) \tag{2}$$

式中 k_{B} 是 Boltzmann 常数.

在室温和标准大气压下, 液相水和气相水之间能够共存 (coexistence), 气、液两相的水分子之间存在一个交界面, 这是因为液相水分子进入气相水内部需要耗费较大热能. 理论分析 [130,132-133] 和模拟计算 [134-137] 发现, 液体水中的空腔周围存在交界面, 而且液体–空腔界面与液体–气体水分子交界面十分相似. 较大球形空腔的水合作用需要的自由能是 [126]

$$\Delta G \approx 4\pi R^2\gamma + \frac{4}{3}\pi R^3 p \tag{3}$$

式中, R 是球形空间半径, γ 是液体–气体表面的张力, p 是空腔内压力. 当空腔很大 (甚至达到宏观尺

度) 时, 压力项 (与体积成线性关系)$4\pi R^3 p/3$ 十分重要; 当空腔较小时, 张力项 $4\pi R^2\gamma$(与面积成线性关系) 十分重要, 而压力项可以忽略不计. 在标准温度和压力条件下, 由于体积排斥作用, 较大疏水残基侵占水分子空间位置, 形成无水空腔区域. 水分子之间的氢键被打断后, 在疏水溶质附近形成交界面. 考虑到上述液体–腔体的能量分析, 较大疏水溶质的溶解自由能主要包含与腔体面积 (R^2) 相关的能量项.

小疏水分子在水中所占体积很小, 无需打断水分子间的氢键, 也不会形成液体–空腔交界面. 在体积为 v 的溶液中, 发现 N 个水分子的概率表示为 $P_v(N)$, 则小疏水分子溶解自由能与 $P_v(N)$ 有关, 可以表示为 [138]

$$\Delta G \approx \frac{k_{\mathrm{B}}T\rho^2 v^2}{2\chi_v} + \frac{k_{\mathrm{B}}T\ln(2\pi\chi_v)}{2} \tag{4}$$

式中, ρ 是溶液密度, χ_v 表示分子数 N 的随机涨落的均方值 (mean-square fluctuation)[138]

$$\chi_v = \langle(\delta N)^2\rangle_v = \rho v + \rho^2\int_v \mathrm{d}\boldsymbol{r}\int_v \mathrm{d}\boldsymbol{r}'\left[g(|\boldsymbol{r}-\boldsymbol{r}'|)-1\right] \tag{5}$$

式中 $g(\boldsymbol{r})$ 是径向分布函数 (radial distribution function). 在近似情况下, χ_v 与体积 v 呈线性关系. 研究表明 [126], 小疏水分子的溶解自由能与所占体积 (R^3) 成比例, 而不是面积 (R^2). 小疏水分子溶解自由能随体积增大而近似线性增长; 大疏水基团溶解自由能随面积增大而近似线性增长. 因此, 疏水分子的溶解自由能存在尺度效应.

基于溶解自由能的尺度效应, 疏水分子在水溶液中将发生聚集. 设水中有相互分离的 n 个疏水小分子 (半径为 r), 这些分子的总溶解自由能表示为

$$\Delta G_1 = nkv = nk\left(\frac{4}{3}\pi r^3\right) \tag{6}$$

式中 k 表示溶解自由能与分子体积之间的线性系数. 当这些小分子紧密聚集在一起时, 形成较大疏水单元 (体积为 V, 表面积为 S, 半径为 R), 此时溶解自由能与单元面积呈线性关系, 分子溶解自由能表示为

$$\Delta G_2 = KV = K\left(4\pi R^2\right) \tag{7}$$

式中 K 表示溶解自由能与分子面积之间的线性系数. 当 n 足够大, 且小分子发生紧密聚集的条件下, 大疏水单元的体积近似等于小分子体积之和 [126], 即满足下式

$$V = \frac{4}{3}\pi R^3 = n\left(\frac{4}{3}\pi r^3\right) \tag{8}$$

因此得到

$$r = \left(\frac{1}{n}\right)^{1/3} R \tag{9}$$

则疏水分子发生聚集前后的溶解自由能之比可表示为

$$\frac{\Delta G_1}{\Delta G_2} = \frac{nk\left(\frac{4}{3}\pi r^3\right)}{K\left(4\pi R^2\right)} = \frac{kV}{KS} \tag{10}$$

从式中可以看出, 当大疏水单元的体积/面积比 (V/S) 很大时, 处于聚集状态的疏水分子具有相对更低的溶解自由能, 因而该状态更稳定. 对于小疏水分子, 溶解自由能主要受系统熵的影响 (由水分子氢键结构调整引起); 对于大疏水单元, 溶解自由能主要受到系统焓的影响 (由界面附近水分子氢键数目的减少引起). 在水溶液环境中, 疏水作用的尺度效应来源于熵与焓之间的相互竞争、相互平衡, 这也是使疏水分子发生聚集的驱动力.

除了体积排斥作用外, 疏水基团与水分子之间还存在范德华力作用, 同时, 疏水残基的羰基、氨基一端具有极性, 与水分子之间还会发生亲水作用. 尽管这些作用相对较弱, 仍然影响水分子 – 疏水基团界面的位置和形状.

5 氢键作用的强度与结构稳定性

5.1 氢键 —— 蛋白质结构稳定的重要作用力

氢键是氢原子与其他电负性 (electronegative) 原子 (如氧原子、氮原子等) 之间形成的吸引作用, 该氢原子通常已经与其它电负性原子形成共价键 (covalent bond). 氢键一般被描述为电偶极子 (dipole) 之间的静电作用, 同时, 氢键也具有部分共价键的性质. 例如, 氢键具有方向性, 只有当羰基 (C═O) 和氨基 (N—H) 之间满足一定的角度时, 才能形成氢键, 氢键的方向性使相邻水分子之间只能存在 4 个氢键; 氢键具有较大的强度, 形成氢键的两个原子 (H 和电负性原子) 之间的距离小于二者的范德华半径, 这表明存在电子云的部分重合. 在氢键中, 提供氢原子的一方常称为供氢者 (donor), 另外一方称为受氢者 (acceptor).

氢键在蛋白质折叠过程中发挥着关键作用, 早在 1951 年, Pauling 和 Corey 等 [12-13,139-140]

学者就指出氢键是形成 α 螺旋 (helix) 和 β 片层 (sheet) 的重要作用力. 通过对比氨基酸与水分子和氨基酸与氨基酸之间氢键能量的差异, 可以估计出氢键对蛋白质结构稳定的贡献 (见图 5). Pauling 等 [12]学者给出了氨基酸功能基团 N—H···O=C 之间氢键能量 (约 −8 kcal/mol); Schellmann 首次测量了蛋白质折叠反应中氨基酸之间氢键的强度, 结果表明氢键带来的焓变化量 ΔH 是形成螺旋结构的驱动力, 水溶液中 ΔH(enthalpy change) 约为 −1.5 kcal/mol[141-142]. 直到 1959 年, Kauzmann[107] 指出非极性基团的疏水作用是蛋白质折叠的主要驱动力 (这个结论得到不断证实并被普遍接受), 氢键作为蛋白质折叠驱动力的地位才逐渐被疏水作用取代.

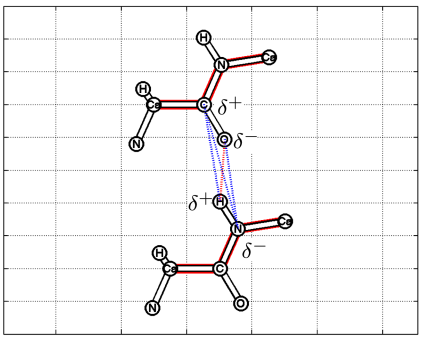


图 5　蛋白质主链上氨基酸残基之间氢键作用示意图

尽管疏水作用在蛋白质结构塌缩和结构自组装 (self-assembly) 中扮演着重要角色, 然而蛋白质折叠绝非简单的掩埋疏水残基、暴露亲水残基过程. 在疏水作用的驱动下, 普通油脂也会在水中发生聚集, 但是并不具有稳定的特殊空间结构. 因此, 仅仅依靠疏水作用还不足以解释为何蛋白质具有稳定的自然态结构. 包括氢键在内的多种作用力是蛋白质形成特殊结构的重要原因, 在这些作用力的共同驱使下, 残基侧链在蛋白质内部形成三维拼图游戏 (jigsaw puzzle) 一样的相互填补, 并使蛋白质的紧密程度近似于晶体结构 [143]. 非极性残基在蛋白质内部的聚集被认为是形成三级结构的推动力, 残基之间氢键作用则决定了蛋白质二级结构 [144]. 例如, 在 α 螺旋结构中, 第 i 个残基的羰基 (C=O) 与第 $i+4$ 个残基的氨基 (N—H) 之间形成氢键, 规则的氢键分布使蛋白质主链形成螺旋状结构; 与 α 螺旋相似, 不同的氢键排布方式还导致其他类型螺旋结构, 如 3_{10} 螺旋结构 (第 i 和第 $i+3$ 个残基之间形成氢键) 和 π 螺旋 (第 i 和第 $i+5$ 个残基之间形成氢键).

判断氨基酸残基之间是否存在氢键的法则包括距离–角度、氢键能量等方面. 较为简单的判别准则认为, 如果在 N—H 和 C=O 之间存在氢键, 原子 N—H—O 形成的角度 θ 和原子距离 r_{HO} 应该满足条件 [145-147]

$$\theta > 120^\circ \quad \text{且} \quad r_{\mathrm{HO}} < 2.5\ \text{Å} \tag{11}$$

根据这个法则, 大约 90%的蛋白质主链 N—H 和 C=O 基团之间存在氢键 [145]. 在更为细致的氢键判断准则中, 通过计算带电原子之间的的 Coulomb 能量 (图 5), 也可以判断残基之间是否存在氢键 [148]

$$E = f\delta^+\delta^- \left(\frac{1}{r_{\mathrm{NO}}} + \frac{1}{r_{\mathrm{HC}}} + \frac{1}{r_{\mathrm{HO}}} + \frac{1}{r_{\mathrm{NC}}} \right) \tag{12}$$

式中, $f = 332\ \text{kcal}/e^2$, δ^+ 和 δ^- 是原子的残余电荷, 单位是电子电量 e, r_{NO} 是 N—O 原子之间距离, r_{HC} 是 H—C 原子之间距离, r_{HO} 是 H—O 原子之间距离, r_{NC} 是 N—C 原子之间距离. 当氢键库仑能量小于阈值时 (如 $E < -0.5$ kcal/mol), 可以认为残基之间形成了氢键 [148]. 基于这个法则, 大约有 62%的蛋白质主链 N—H 和 C=O 基团之间存在氢键 [149].

5.2　蛋白质结构中氢键的能量计算

一般认为, 氢键包含电偶极子 (electric dipole) 的静电 (electrostatic) 作用和部分共价 (covalent) 作用. 研究表明, 库仑静电作用在氢键强度中占重要地位 (约 75%)[147,150], 共价作用所占比例少于 5%. 通过对晶体结构中氢键几何特征的研究, 可以导出氢键能量 E_{hb} 的经验公式 [151-153]

$$E_{hb} = E_{\mathrm{r}} E_p E_t \tag{13}$$

式中 E_{r} 表示原子 N—O 间距离 r 对能量的影响 [154]

$$E_{\mathrm{r}} = \left(\frac{4r_m^6}{r^6} - \frac{3r_m^8}{r^8} \right) E_m \tag{14}$$

式中, r_m 表示 N—O 原子间的平衡距离, E_m 是该距离对应最低能量. 对于蛋白质主链的氢键, 一般取 $r_m = 3$ Å.

与原子共价键之间角度有关的能量项为 [154]

$$E_p = \cos^2\theta \tag{15}$$

$$E_t = \begin{cases} [0.9 + 0.1\sin(2t_i)]\cos t_0, & 0^\circ < t_i \leqslant 90^\circ \\ K_1[K_2 - \cos^2(t_i)]\cos t_0, & 90^\circ < t_i \leqslant 110^\circ \\ 0, & 110^\circ < t_i \end{cases} \tag{16}$$

式中, 角度 θ, t_i, t_0 的含义见图 6.

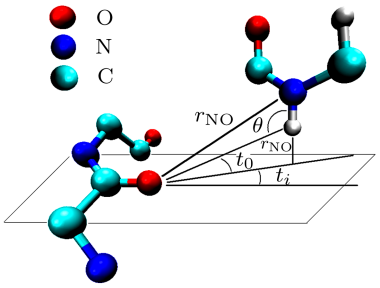


图 6 氢键中原子之间几何关系示意图

5.3 氢键对自由能和结构稳定性的影响

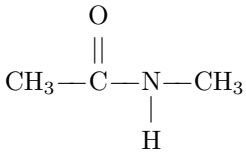
蛋白质折叠发生在水溶液中, 周围存在大量水分子并形成氢键网络. 蛋白质的极性残基和带电残基与水分子形成氢键、静电作用, 因而受到水分子的吸引; 疏水残基则阻碍水分子间形成氢键, 因此受到水分子的排斥. 通过将疏水残基掩埋到蛋白质内部, 水分子可以把缺失氢键引起的自由能改变量降低到最小. 但是, 当疏水残基进入蛋白质内部时, 残基主链部分的极性基团 N—N和C=O也被带入远离水分子的内部环境, 丧失与水分子形成氢键的机会. 主链极性基团与水分子之间氢键的断开导致自由能增加. 为了使结构更加稳定, 身处蛋白质内部的主链极性基团必须彼此配对, 重新形成氢键作用, 这样就催生了规则的二级结构.

氢键的有规律分布对自然态蛋白质结构起到固定作用. 多数氢键处于蛋白质内部, 其余则分布在蛋白质表面, 受到水分子从四面八方的不断冲击. 在一定溶液条件下, 当暴露在水溶液中的氢键被打断时, 水分子就可以逐步深入到蛋白质内部, 并试图断开内部结构的氢键, 直至最终打开蛋白质结构中所有氢键. 这时候, 蛋白质结构就会发生去折叠.

与氢键本身的能量相比, 氨基酸之间形成氢键的自由能变化量更具意义. 在蛋白质分子中, 氨基酸残基之间存在氢键作用 ($CO\cdots HN$); 同时, 氨基酸残基也能与水分子形成氢键 ($CO\cdots HOH$ 和$NH\cdots OH_2$, 见式 (17)). 因此, 水溶液对氨基酸之间的氢键有强烈的破坏作用, 蛋白质结构稳定性取决于氨基酸–氨基酸之间的氢键是否比氨基酸–水分子之间的氢键更稳定 [155-158].

$$CO\cdots HOH + NH\cdots OH_2 = CO\cdots HN + HOH\cdots OH_2 \qquad (17)$$

蛋白质分子中氢键热力学稳定性的研究仍然存在不一致的观点 [156,159-160]. 20 世纪五六十年代, Schellman 等 [141-142] 学者研究了包含氢键 (hydrogen-bonded) 的多肽链从折叠态转变为随机长链 (random chain) 过程中自由能的改变, 对聚氨基乙酸 (polyglycine) 形成的螺旋 (helix) 和片层 (sheet) 结构自由能估算表明 [141], 水溶液中极性基团之间氢键的稳定性较为微弱, 残基形成一个氢键带来的焓 (enthalpy) 降低约为 −1.5 kcal/mol, 自由能最多降低约为 −0.6 kcal/mol[141-142]. 不同类型的氨基酸侧链结构对残基氢键稳定性也有影响, 依据侧链性质和空间分布的不同, 会加强或削弱折叠态蛋白质结构的稳定性 [141]. 1962 年, Klotz 和 Franzen[161] 研究了在不同溶液环境下 N-甲基乙酰胺 (N-Methylacetamide) 的红外吸收光谱 (infrared spectroscopy). N-甲基乙酰胺是一种小分子化合物, 分子式为 C_3H_7NO, 分子结构是



, 能够很好表示残基之间氢键作用. 对 N-甲基乙酰胺 (N-Methylacetamide) 的二聚化 (dimerization) 实验研究发现, 在 25 °C 条件下, 极性溶液 (水溶液) 中 C=O 和 N—H 形成氢键的自由能变化为正值 (约 3.1 kcal/mol), 因此氨基化合物之间 (interamide) 的氢键十分不稳定 [161]; 在二氧杂环乙烷 (dioxane) 溶液中, 尽管介电常数 (dielectric constant) 很低 (大约为 2), 氨基化合物之间的氢键稳定性仍然较弱 (形成氢键的自由能约 0.39 kcal/mol)[161]; 只有在四氯化碳 (carbon tetrachloride) 有机溶液中, 氨基化合物之间的氢键才能稳定存在 (形成氢键的自由能约 −0.92 kcal/mol)[161]. 因此他们认为, 除了少数介电常数非常低的局部区域外, 在极性溶液中残基氢键对蛋白质分子的稳定作用贡献不大 [161].

另一种观点认为氢键是自然态蛋白质结构稳定的最主要原因 [162]. Fersht 等 [155] 学者研究了水溶液中酪氨酰–tRNA 合成酶 (tyrosyl-tRNA synthetase) 的突变体与基质 (substrate) 之间的氢键作用, 通过突变合成酶的某些氨基酸侧链, 可以使酶与基质之间的氢键断开, 并计算出氢键断开前后的自由能变化. 实验发现, 极性基团之间形成氢键带来的自由能降低大约为 −0.5 kcal/mol 至 −1.8 kcal/mol, 当氢键的供氢者或受氢者带电荷时, 自由能降低量进一步达到 −3.5 kcal/mol 至

−4.5 kcal/mol[155]. 单聚物 (monomer) 的二聚化 (dimerization) 过程研究表明 [157,163],分子间氢键的形成增加了双分子系统的约束, 使分子的平移和旋转自由度丧失, 导致了很大的熵减少, 正是这部分熵减 (而非氢键本身的能量) 使二聚物不稳定. 从熵损失角度说, 两个相互独立的自由分子之间形成氢键的二聚化过程, 并不等价于蛋白质分子内部氢键的形成过程. 这是因为蛋白质中绝大部分氢键是在分子内部形成的, 该过程引起的熵损失很小, 反应平衡常数 (equilibrium constant) 远远大于单分子之间的二聚化 [157]. 因此, 自由分子二聚化实验低估了氢键对蛋白质结构的稳定作用. 即使二聚物在极性溶液 (水) 中不稳定, 也不能得到 “氢键不是维持蛋白质结构稳定性的重要驱动作用” 的结论. 通过消除分子平移、旋转熵损失的影响, Doig 和 Williams[163] 研究发现, 无论是在极性溶液 (水) 还是在非极性溶液 (四氯化碳 (tetrachloromethane)) 中, 氢键对结构的稳定作用都比预想的要大许多, 因此氨基化合物之间的氢键作用为折叠态蛋白质结构稳定性提供了强大的驱动力. Shirley 等 [164] 研究了核糖核酸酶 T1(ribonuclease T1) 突变体的氢键变化和结构稳定性, 给出了氢键对结构稳定性的平均能量贡献 (约为 −1.3 kcal/mol), 并指出氢键对蛋白质结构的稳定作用与疏水作用同样重要. 如果在水环境下残基之间形成一个氢键带来的焓改变量为 $\Delta H = -1$ kcal/mol, 那么, 对于包含 100 个氢键的蛋白质分子来说, 能量降低约为 −100 kcal/mol. 普通小蛋白质分子折叠的能量势垒 (energy barrier) 约为 −20 kcal/mol 至 −60 kcal/mol[165], 而自由能的涨落 (fluctuation) 甚至更小 (约小于 5 kcal/mol)[166], 因此, 氢键能量对结构稳定性的影响将是巨大的.

蛋白质结构中的氢键强度与周围环境的相对介电常数 ε_{r} 也有关系. 在一个大气压、温度为25 °C条件下, 纯水的相对介电常数 (relative dielectric constant) 为 $\varepsilon_{\mathrm{r}} = 78.5$; 在蛋白质内部, 相对介电常数为 $\varepsilon_{\mathrm{r}} = 4$. 考虑到氢键的 Coulomb 作用力, 在蛋白质内部的氢键作用具有较大的强度, 使蛋白质结构更加稳定. 这些相对较强的氢键往往被疏水残基的侧链包裹, 周围较少出现水分子, 形成远离水分子的 “干燥” 区域.

对氢键热力学稳定性的研究仍然面临着许多挑战, 特别是缺少合适的化合物分子模型, 使氢键对蛋白质结构稳定性的能量分析更加复杂, 导致不同的结论. 水分子之间氢键的断开导致系统焓的增加, 同时也导致系统熵的增加, 这两部分能量变化具有相互抵消作用. 熵和焓对自由能变化的影响有一部分相互抵消后, 剩余部分使蛋白质结构处于稳定与不稳定的边缘, 任何微小的自由能变化都会带来很大影响 [101].

在极性溶液 (水) 环境中, 水分子与残基之间存在氢键作用, 约束了水分子的空间平移和旋转自由度, 因此氢键稳定性主要是水分子熵变化驱动的. 水分子不断从四周碰撞蛋白质分子, 倾向于使熵损失降低到尽可能小的程度. 在非极性溶液环境中, 氢键对结构的约束作用源于焓和熵两部分, 与氢键数量及电荷分布密切相关 [155,161]. 面对折叠态蛋白质分子内部残基–残基氢键和去折叠态蛋白质残基–水分子氢键, 如果比较两者的稳定性, 至少需要理解下述复杂的过程 [158]: 在水溶液中 (高介电常数) 和蛋白质内部 (低介电常数) 形成氢键时焓 (enthalpy) 的定量计算; 残基进入蛋白质内部并形成氢键引起的去溶剂化 (desolvation) 能量损失; 残基间氢键引起的侧链和水分子构型熵 (configuration entropy) 的变化; 残基间氢键强度对蛋白质分子几何构型的依赖关系. 目前, 关于蛋白质分子内部氢键的热力学稳定性问题仍然在持续不断进行 [167].

6　静电力对蛋白质结构稳定的影响

6.1　静电作用 —— 与多种因素有关的复杂作用力

静电力 (electrostatic interaction) 在蛋白质折叠和结构稳定中扮演着重要的角色, 例如, 在极端的酸碱度 (pH) 环境下, 蛋白质结构发生去折叠. 然而, 对静电力的精确描述仍然十分困难, 静电作用如何影响蛋白质折叠过程也是不易理解的. 这其中的部分原因来自静电力对溶液环境以及结构几何属性的依赖等.

构成蛋白质的 20 种氨基酸中包含带正电 (如 Lys, Arg) 和带负电 (如 Asp, Glu) 的氨基酸, 还包含具有极性但是不发生电离的氨基酸 (如 Asn, Gln,Ser,Thr,Tyr); 同时, 由于肽键 (peptide bond) 具有共振结构 (resonance structure)[168-169], 蛋白质主链上肽键两侧的 C═O 和 N—H 基团也具有残余电荷 (O 原子 −0.28 电子电荷,N 原子 0.28 电子电荷), 使肽键称为一个永久电偶极子. 当具有相反电荷的残基或带电基团相互接触时 (如距离小于

5 Å), 彼此之间会产生静电作用, 这种非共价键作用称为盐桥 (salt bridge)[170-172]. 盐桥静电作用对结构自由能的贡献可以从三方面考虑 [173]: 带电残基之间相互作用, 电荷与永久电偶极子之间相互作用, 电荷去溶剂化引起的能量改变.

盐桥对蛋白质结构的稳定作用十分复杂, 因而使静电自由能受多种因素影响 (如溶液 pH 值、蛋白质几何形状等). 对于水溶液中的球形蛋白质 (globular protein) 来说, 周围是充满极性分子的溶液环境, 蛋白质内部是远离溶液的疏水环境. 一方面, 带电残基倾向于蛋白质外部的极性环境, 但折叠过程中部分带电残基进入蛋白质内部, 电荷在内部的聚集产生去溶剂化和规则分布效果. 去溶剂化引起焓增加, 规则分布引起熵减少, 导致自由能增加和蛋白质结构不稳定, 这部分能量变化称为去溶剂化消耗 (desolvation cost). 另一方面, 带电残基与结构内部的其他带电残基 (或永久电偶极子) 形成盐桥的静电吸引作用, 能够使蛋白质结构更加稳定. 因此, 静电作用使蛋白质结构更加稳定还是更加不稳定, 取决于占主导地位的是去溶剂化消耗还是电荷静电吸引作用. 蛋白质内部盐桥吸引作用带来自由能降低, 如果这部分自由能降低无法抵消去溶剂化引起的自由能增加, 那么蛋白质结构变得更加不稳定. 带电残基的去溶剂化和电荷吸引是结构稳定的两个对立面, 两者都依赖于多种因素, 如蛋白质结构改变引起的带电残基重新分布, 溶液 pH 值对残基电荷的影响, 蛋白质内部介电常数的估计等. 这些因素的任何微小变化都会改变双方的地位, 使结构稳定性发生逆转. 因此, 静电作用的自由能处于临界状态, 介乎稳定和不稳定之间.

蛋白质结构稳定性依赖于折叠态结构自由能 G_{F} 与去折叠态结构自由能 G_{U} 之间的差异 $\Delta G_{\mathrm{U}} = G_{\mathrm{U}} - G_{\mathrm{F}}$, ΔG_{U} 是打开折叠态蛋白质结构所需的自由能. 如果 $\Delta G_{\mathrm{U}} > 0$, 那么折叠态结构更稳定. 与此类似, 盐桥静电作用对蛋白质结构稳定性的影响可以表示为 [173]

$$\Delta\Delta G_{\mathrm{U}} = \Delta G_{\mathrm{U}}^{\mathrm{sb}} - \Delta G_{\mathrm{U}}^{\mathrm{non\text{-}sb}} \tag{18}$$

式中 $\Delta G_{\mathrm{U}}^{\mathrm{sb}}$ 表示打开具有盐桥静电作用的折叠态蛋白质结构所需自由能, $\Delta G_{\mathrm{U}}^{\mathrm{non\text{-}sb}}$ 表示打开缺少盐桥静电作用的折叠态蛋白质结构所需自由能, $\Delta\Delta G_{\mathrm{U}}$ 表示盐桥静电作用对打开折叠态蛋白质结构所需自由能的贡献. 如果 $\Delta\Delta G_{\mathrm{U}} > 0$, 那么盐桥静电作用使蛋白质折叠态结构更加稳定; 反之, 则静电作用使结构更加不稳定.

6.2 静电作用的主要影响因素

6.2.1 溶液环境 pH 值

静电作用的自由能改变与溶液 pH 值密切相关, 根本原因在于 pH 值对带电残基电荷的影响. 可电离 (ionable) 残基的带电状态由酸度系数 (acid dissociation constant)K_{a} 或其对数 pK_{a} 表示. 在氨基酸的离解 (dissociation) 过程 [174]

$$\mathrm{HA} \rightleftarrows \mathrm{A^-} + \mathrm{H^+} \tag{19}$$

氨基酸 HA 发生去质子化 (deprotonated), 成为酸 (acid)$\mathrm{A^-}$. 酸度系数 K_{a} 是离解化学反应的平衡常数 [175]

$$K_{\mathrm{a}} = \frac{[\mathrm{A^-}]\,[\mathrm{H^+}]}{[\mathrm{HA}]} \tag{20}$$

K_{a} 定量描述了溶液中酸性物质的强度. 由于 K_{a} 的数值范围很大, 通常使用对数表示酸度系数

$$pK_{\mathrm{a}} = -\lg K_{\mathrm{a}} \tag{21}$$

pK_{a} 的值越大, 氨基酸离解的程度越弱. 弱酸在水中的 pK_{a} 值通常处于 $[-2, 12]$, $pK_{\mathrm{a}} < -2$ 表示强酸. 当氨基酸去质子化 (丢失一个 $\mathrm{H^+}$) 之后, 带有负电荷; 同时, 被质子化的氨基酸得到一个 $\mathrm{H^+}$, 带有正电荷. 氨基酸的带电状态通过 Henderson Hasselbach 方程描述 [176]

$$pH = pK_{\mathrm{a}} + \lg\frac{[\mathrm{A^-}]}{[\mathrm{HA}]} \tag{22}$$

当 pH 值比 pK_{a} 小 2 时, Henderson Hasselbach 方程可以写为 $[\mathrm{A^-}]/[\mathrm{HA}] = 0.01$, 表示氨基酸 99% 被质子化 (具有正电荷); 当 pH 值比 pK_{a} 大 2 时, 方程可以写为 $[\mathrm{A^-}]/[\mathrm{HA}] = 100$, 表示氨基酸 99%被去质子化 (具有负电荷). 带电氨基酸残基的 pK_{a} 值代表该氨基酸的电离程度, 即所带电荷的强度. 理论和实验研究表明 [177-178], 与孤立状态的氨基酸相比, 处于蛋白质结构中的氨基酸残基所带电荷会发生改变, 称为 pK_{a} 漂移 (pK_{a} shift). 这种变化源于残基与结构中其他带电基团的相互作用, 也包括与极性溶液环境的相互作用. 利用 Poisson–Boltzmann 差分方法计算静电自由能发现, 蛋白质结构稳定性与环境 pH 值之间存在依赖关系 [179]. 因此, 溶液环境的 pH 值决定了氨基酸带电状态, 也对蛋白质折叠过程中静电作用的自由能改变 (即结构稳定性) 具有重要影响.

6.2.2 蛋白质内部的介电常数

如果将蛋白质内部看作具有较低介电常数的均匀连续体, 将外部水环境看作具有较高介电常数的均匀连续体, 同时设游离带电离子满足 Boltzmann 分布, 可以建立 Poisson–Boltzmann 模型计算静电作用的能量 [180-181]

$$-\nabla\cdot\varepsilon\nabla\phi=\rho+\sum_i q_i n_i\exp\left(\frac{q_i\phi}{k_{\mathrm{B}}T}\right)\tag{23}$$

式中, ϕ 是静电作用势能, ε 是介电常数, q_i 是第 i 种带电离子的电荷, n_i 是第 i 种带电离子的浓度, k_{B} 是 Boltzmann 常数, T 是绝对温度. 在得到 ϕ 之后, 可以计算出带电残基能量 G_{ele}

$$G_{\mathrm{ele}}=\frac{1}{2}\sum q_i\phi_i\tag{24}$$

式中, q_i 是第 i 个残基的残余电荷, ϕ_i 是第 i 个残基所处位置的静电势能.

在 Poisson–Boltzmann 模型中, 残基静电能与介电常数 ε 成反比例, 因此, 将带电残基从极性 (高 ε 值) 水环境中移至非极性 (低 ε 值) 需要消耗能量, 即去溶剂化消耗. ε 的变化程度直接影响能量消耗的多少. 为了计算折叠过程中带电残基位置变化引起的静电自由能改变, 需要给出蛋白质内部的介电常数 ε.

电荷周围环境的介电常数可以表示为 $\varepsilon=\varepsilon_{\mathrm{r}}\varepsilon_0$, 这里 $\varepsilon_0=8.854\,187\,817\times10^{-12}$ F/m 是真空中介电常数, ε_{r} 是相对介电常数. 通常认为, 水溶液中 $\varepsilon_{\mathrm{r}}\approx80$, 蛋白质内部 $\varepsilon_{\mathrm{r}}\approx4$, 但是研究表明, 将蛋白质内部的相对介电常数设为 $\varepsilon_{\mathrm{r}}=20$ 能够得到更好的结果 [182], 蛋白质的 pK_{a} 计算值与实验结果更加吻合. 一个可能的解释是, 由于 Poisson–Boltzmann 模型忽略了结构改变引起的 pK_{a} 漂移现象, 给计算结果带来误差, 而较高的介电常数可以抵消这部分误差 [182].

电荷产生的静电场会使电场中其他电荷重新排布 (图 7), 如果考虑蛋白质结构和溶液中每个原子的极化现象, 可以精确计算静电作用的自由能 [183-185].另一种较为简单方法是通过介电常数表示电荷产生的屏蔽作用 (shielding). 在静电力计算中, 通常把蛋白质看作一个均匀介质, 利用单一的介电常数 ε 表示其极化程度. 然而, 蛋白质空间结构是非均匀的, 不同区域具有不同的极化程度. 在蛋白质与水环境的接触表面附近, 电荷引起的极化程度较高, 对应较大介电常数; 在远离水分子的蛋白质内部, 极化程度较低, 对应较小的介电常数. 同时, 考虑到蛋白质结构在折叠过程中不断发生变化, 带电基团在空间的重新分布也会改变蛋白质内部介电常数. 均匀介电常数模型无法考虑蛋白质与水环境之间静电屏蔽作用, 引起静电作用自由能计算误差, 甚至导致蛋白质结构在稳定与不稳定状态之间发生逆转. 因此, 可以建立随空间位置变化的蛋白质介电常数模型. 空间非均匀介电常数模型的研究表明, 电荷分布对介电常数 ε 的影响非常复杂, 在有些情况下, 介电常数 ε 与电荷距离之间存在近似线性关系 [186-187], 但是多数情况下, 目前仍无法用单一函数表示蛋白质结构中介电常数的变化规律 [188].

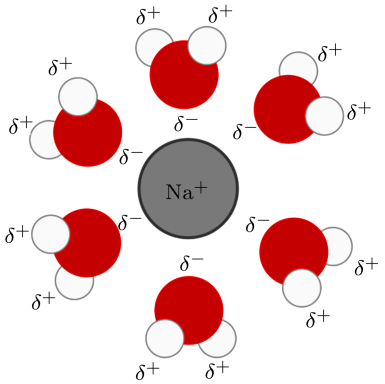


图 7　带电离子在水溶液中的极化和屏蔽效应

6.2.3 盐桥几何特征

在蛋白质结构中, 带电残基侧链之间形成盐桥静电作用, 盐桥对结构稳定性的影响与侧链之间的相对位置 (即盐桥的几何特征) 有关. 通常盐桥几何特征可以从两方面刻画: 带电残基距离 r 和残基相对角度 θ. 残基距离 r 表示两个残基侧链上带电基团之间的距离, 相对角度 θ 表示由残基 C_α 原子指向带电基团的矢量形成的角度.

在折叠过程中, 蛋白质结构处于不断运动状态, 氨基酸残基位置的变化导致盐桥之间距离 r 和相对角度 θ 发生改变. 同时, 蛋白质结构上电荷分布位置发生改变, 引起结构内部介电常数、电场强度变化, 导致盐桥带电残基的静电作用强度变化. 对核磁共振 (NMR) 蛋白质结构数据的研究表明, 离子间静电作用强度与蛋白质结构变化有显著关系 [172]. 在不同几何特征下, 盐桥的离子静电作用在稳定/不稳定状态之间发生逆转 [189], 这表明静电作用对蛋白质稳定性的影响依赖于空间结构变化和盐桥几何特征. 进一步研究指出, 当盐桥几何特征同时满足 (1) 残基带电基团 (电性相反) 中心距离 $r<4.0$ Å; (2) Asp/Glu 侧链氧原子与 Arg/Lys

侧链氮原子距离小于 4.0 Å时, 盐桥静电吸引作用最强 [190].

7 范德华力对自由能的影响

范德华 (van der Waals) 力是分子之间的相互吸引/排斥作用力. 与共价键和离子键不同, 范德华力是相互接近的分子 (原子) 间发生极化耦合 (polarization correlations) 引起的, 可以通过 Lennard-Jones 势表示 [191-192]

$$E=\varepsilon\left[\left(\frac{r_{\min}}{r}\right)^{12}-2\left(\frac{r_{\min}}{r}\right)^{6}\right] \tag{25}$$

虽然简单模型的分子间范德华力已经研究得很清楚, 但是如何准确计算蛋白质分子中范德华力对自由能的影响仍然很困难. 一般认为, 范德华力对稠密流体 (dense fluid) 的结构影响很小 [193], 但是, 在疏水基团与水溶液的相互作用中, 范德华力的影响与基团尺寸有关. 在较小尺寸 (半径约 4 Å) 疏水基团附近, 水分子通过调整局部位置以维持氢键网络, 使贴近疏水基团表面的水分子密度增大 (约两倍), 这个过程被称为 “浸润 (wet)”[126]. 在较大尺寸 (半径大于 10 Å) 疏水基团附近, 由于基团和水溶液的接触面曲率低、延伸广, 贴近疏水基团表面的水分子之间难以维持氢键作用, 氢键网络受到破坏, 因此水分子倾向于远离疏水表面, 使疏水表面附近水分子密度降低, 这个过程被称为 “不浸润 (dewetted)”[126]. 在较大疏水基团 “不浸润” 表面附近, 由于液体水分子稠密程度相对较低, 水分子分布比较稀疏, 水分子密度和密度涨落受范德华力的影响力较大. 尽管在范德华力吸引作用下, 大尺寸疏水基团表面附近的水分子密度有所增大, 但是这不同于小尺寸疏水基团表面的 “浸润” 情况, 两者的重要区别在于水分子密度的涨落. 在疏水基团–水溶液交界面附近, 涨落的大小决定了范德华力对平均水分子密度的影响程度 [193], 同时也决定了疏水基团发生聚集的动力学过程 [194]. 研究表明: 范德华力对溶解自由能的影响可以表示为 [195]

$$\Delta G=\Delta G_V+\rho\int_V g_V(\boldsymbol{r})\,u(\boldsymbol{r})\,\mathrm{d}\boldsymbol{r} \tag{26}$$

式中 ρ 是溶质 (solute) 密度, $g_V(\boldsymbol{r})$ 是溶剂平均密度的相对值, 表示溶剂中存在和不存在溶质两种情况下平均密度的比值. $u(\boldsymbol{r})$ 表示在 $\boldsymbol{r}$ 处范德华力对溶剂的势能作用, ΔG_V 表示溶质体积为 V 的溶解自由能. 这里溶质是疏水基团, 溶剂是水溶液. 在疏水基团–水溶液界面附近, 范德华力表现为吸引力, $u(\boldsymbol{r})$ 为负值, 因此水分子与疏水基团之间的范德华力有利于溶解过程. 在较小疏水基团附近, 上式中 ΔG 近似与溶质体积 V 成线性关系 [126,195-196]; 在较大疏水残基附近, ΔG 近似与溶质表面积成线性关系. 因此, 范德华力对溶解自由能的影响, 类似于溶质体积对溶解自由能的影响 [126].

在蛋白质结构内部, 范德华力对结构稳定性的影响十分复杂. 折叠态蛋白质内部结构往往十分紧密, 残基侧链之间形成类似晶体结构的排布 [197].当原子之间近距离接触时, 范德华力占有十分重要的地位, 对蛋白质结构的稳定性产生显著影响 [198]. 对葡萄球菌核酸酶 (staphylococcal nuclease) 的突变数据分析表明, 蛋白质结构的稳定性与分子内部范德华作用力之间存在很好的相关性 [199]. 但是, 范德华力和氢键作用往往相互耦合. 一方面范德华力的吸引作用使残基相互接近, 拉近了氢键的 NH 和 CO 基团, 导致氢键作用力增强 [200]; 另一方面残基间氢键的吸引作用也会缩短原子的距离, 增强范德华作用力, 使蛋白质内部结构更加稳定. 因此, 在定量分析范德华力对蛋白质结构稳定性的贡献时, 如何区分范德华力和氢键对自由能的影响仍然是一个挑战.

8 结 语

蛋白质在水溶液中发生折叠, 折叠过程依赖于微观状态的结构特征和自由能分布, 需要以微观结构系综的熵/焓变化规律为研究对象, 建立统计热力学模型. 疏水作用、氢键、静电力和范德华力是折叠过程中的重要非共价作用力, 对结构稳定具有举足轻重的影响. 非共价作用的自由能变化机理研究表明, 驱动蛋白质发生折叠的原因是水溶液–蛋白质、蛋白质内部残基之间相互作用引起的熵/焓竞争. 对主要非共价作用自由能变化的分析, 为进一步研究蛋白质折叠的统计热力学原理奠定了基础,

致 谢

论文撰写得到了林家翘先生的有益指点, 与林先生的讨论使作者获得很大帮助, 在此表示感谢.

参 考 文 献

1 Kendrew J C, Bodo G, Dintzis H M, et al. A three-dimensional model of the myoglobin molecule obtained by X-ray analysis. *Nature*, 1958, 181(4610): 662-666

2 Wuthrich K. NMR of Proteins and Nucleic Acids. New York: Wiley, 1984

3 Wuthrich K, Wider G, Wagner G, et al. Sequential resonance assignments as a basis for determination of spatial protein structures by high resolution proton nuclear magnetic resonance. *J Mol Biol*, 1982, 155(3): 311-319

4 Wuthrich K, Billeter M, Braun W. Polypeptide secondary structure determination by nuclear magnetic resonance observation of short proton-proton distances. *J Mol Biol*, 1984, 180(3): 715-740

5 Havel T F, Wuthrich K. An evaluation of the combined use of nuclear magnetic resonance and distance geometry for the determination of protein conformations in solution. *J Mol Biol*, 1985, 182(2): 281-294

6 Qian Y Q, Billeter M, Otting G, et al. The structure of the Antennapedia homeodomain determined by NMR spectroscopy in solution: comparison with prokaryotic repressors. *Cell*, 1989, 59(3): 573-580

7 Bottcher B, Wynne S A, Crowther R A. Determination of the fold of the core protein of hepatitis B virus by electron cryomicroscopy. *Nature*, 1997, 386(6620): 88-91

8 Conway J F, Cheng N, Zlotnick A, et al. Visualization of a 4-helix bundle in the hepatitis B virus capsid by cryo-electron microscopy. *Nature*, 1997, 386(6620): 91-94

9 Mancini E J, Clarke M, Gowen B E, et al. Cryo-electron microscopy reveals the functional organization of an enveloped virus, Semliki Forest virus. *Mol Cell*, 2000, 5(2): 255-266

10 Zhou Z H, Dougherty M, Jakana J, et al. Seeing the herpesvirus capsid at 8.5 Å. *Science*, 2000, 288(5467): 877-880

11 Medalia O, Weber I, Frangakis A S, et al. Macromolecular architecture in eukaryotic cells visualized by cryoelectron tomography. *Science*, 2002, 298(5596): 1209-1213

12 Pauling L, Corey R B, Branson H R. The structure of proteins; two hydrogen-bonded helical configurations of the polypeptide chain. *Proc Natl Acad Sci U S A*, 1951, 37(4): 205-211

13 Pauling L, Corey R B. The pleated sheet, a new layer configuration of polypeptide chains. *Proc Natl Acad Sci U S A*, 1951, 37(5): 251-256

14 Linderstrom-Lang K, Schellman J A. Protein Structure and Enzymatic Activity. In: Lardy H, Myrback K. eds. The Enzymes 2nd ed. New York: Academic Press, 1959. 443-510

15 Chiang Y S, Gelfand T I, Kister A E, et al. New classification of supersecondary structures of sandwich-like proteins uncovers strict patterns of strand assemblage. *Proteins-Structure Function and Bioinformatics*, 2007 68(4): 915-921

16 Tanford C. Protien denaturation. *Advances in Protein Chemistry*, 1968, 23: 121-282

17 Miller W G, Brant D A, Flory P J. Random coil configurations of polypeptide copolymers. *Journal of Molecular Biology*, 1967, 23(1): 67-80

18 Dobson C M, Evans P A, Williamson K L. Proton nmr-studies of denatured lysozyme. *Febs Lett*, 1984, 168(2): 331-334

19 Griko Y V, Privalov P L, Venyaminov S Y, et al. Thermodynamic study of the apomyoglobin structure. *Journal of Molecular Biology*, 1988, 202(1): 127-138

20 Privalov P L, Tiktopulo E I, Venyaminov S Y, et al. Heat-capacity and conformation of proteins in the denatured state. *Journal of Molecular Biology*, 1989, 205(4): 737-750

21 Creighton T E. Protein folding. *Biochem J*, 1990, 270(1): 1-16

22 Weikl T R, Palassini M, Dill K A. Cooperativity in two-state protein folding kinetics. *Protein Sci*, 2004, 13(3): 822-829

23 Dill K A, Fiebig K M, Chan H S. Cooperativity in protein-folding kinetics. *P Natl Acad Sci USA*, 1993, 90(5): 1942-1946

24 Chan H S, Bromberg S, Dill K A. Models of cooperativity in protein-folding. *Philos T Roy Soc B*, 1995, 348(1323): 61-70

25 Anfinsen C B. Principles that govern the folding of protein chains. *Science*, 1973, 181(96): 223-230

26 Anfinsen C B, Scheraga H A. Experimental and theoretical aspects of protein folding. *Adv Protein Chem*, 1975, 29: 205-300

27 Wu H. Studies on denaturation of proteins. XIII. A theory of denaturation. *Chinese Journal of Physiology*, 1931, 5: 321-344

28 Anson M L, Mirsky A E. The equilibria between native and denatured hemoglobin in salicylate solutions and the theoretical consequences of the equilibrium between native and denatured protein. *J Gen Physiol*, 1934, 17(3): 399-408

29 Northrop J H. Crystalline trypsin IV. reversibility of the inactivation and denaturation of trypsin by heat. *J Gen Physiol*, 1932, 16(2): 323-337

30 Friedman H L. A Course in Statistical Mechanics. New York: Prentice Hall College Div, 1985

31 Lesk A M, Chothia C. How different amino-acid-sequences determine similar protein structures - structure and evolutionary dynamics of the globins. *Journal of Molecular Biology*, 1980, 136(3): 225-270

32 Dill K A, Ozkan S B, Weikl T R, et al. The protein folding problem: when will it be solved? *Curr Opin Struc Biol*, 2007, 17(3): 342-346

33 Mirny L A, Abkevich V I, Shakhnovich E I. How evolution makes proteins fold quickly. *P Natl Acad Sci USA*, 1998, 95(9): 4976-4981

34 Kim D E, Gu H, Baker D. The sequences of small proteins are not extensively optimized for rapid folding by natural selection. *Proc Natl Acad Sci U S A*, 1998, 95(9): 4982-4986

35 Bryngelson J D, Onuchic J N, Socci N D, et al. Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins*, 1995, 21(3): 167-195

36 Privalov P L. Stability of protein-structure and hydrophobic interactions. *Biol Chem H-S*, 1988, 369(4): 199-199

37 Ooi T, Oobatake M. Effects of hydrated water on protein unfolding. *J Biochem-Tokyo*, 1988, 103(1): 114-120

38 Fandrich M, Fletcher M A, Dobson C M. Amyloid fibrils from muscle myoglobin–even an ordinary globular protein can assume a rogue guise if conditions are right. *Nature*, 2001, 410(6825): 165-166

39 Dobson C M, Sali A, Karplus M. Protein folding: a perspective from theory and experiment. *Angew Chem Int Edit*, 1998, 37(7): 868-893

40 Bryngelson J D, Wolynes P G. Spin-glasses and the statistical-mechanics of protein folding. *P Natl Acad Sci USA*, 1987, 84(21): 7524-7528

41 Dill K A. Polymer principles and protein folding. *Protein Sci*, 1999, 8(6): 1166-1180

42 Dill K A, Alonso D O V, Hutchinson K. Thermal stabilities of globular-proteins. *Biochemistry-Us*, 1989, 28(13): 5439-5449

43 Dill K A, Chan H S. From levinthal to pathways to funnels. *Nat Struct Biol*, 1997, 4(1): 10-19

44 Leopold P E, Montal M, Onuchic J N. Protein folding funnels —— a kinetic approach to the sequence structure relationship. *P Natl Acad Sci USA*, 1992, 89(18): 8721-8725

45 Bryngelson J D, Onuchic J N, Socci N D, et al. Funnels, pathways, and the energy landscape of protein-folding - a synthesis. *Proteins-Structure Function and Genetics*, 1995, 21(3): 167-195

46 Moult J. Rigorous performance evaluation in protein structure modelling and implications for computational biology. *Philosophical Transactions of the Royal Society B-Biological Sciences*, 2006, 361(1467): 453-458

47 Chou P Y, Fasman G D. Prediction of protein conformation. *Biochemistry-Us*, 1974, 13(2): 222-245

48 Chou P Y, Fasman G D. Empirical predictions of protein conformation. *Annu Rev Biochem*, 1978, 47: 251-276

49 Rost B, Eyrich V A. EVA: large-scale analysis of secondary structure prediction. *Proteins-Structure Function and Genetics*, 2001, (Suppl 5): 192-199

50 Dor O, Zhou Y. Achieving 80% ten-fold cross-validated accuracy for secondary structure prediction by large-scale training. *Proteins*, 2007, 66(4): 838-845

51 Bowie J U, Luthy R, Eisenberg D. A method to identify protein sequences that fold into a known three-dimensional structure. *Science*, 1991, 253(5016): 164-170

52 Jones D T, Taylor W R, Thornton J M. A new approach to protein fold recognition. *Nature*, 1992, 358(6381): 86-89

53 Chowdhury S, Lee M C, Xiong G M, et al. Ab initio folding simulation of the Trp-cage mini-protein approaches NMR resolution. *Journal of Molecular Biology*, 2003, 327(3): 711-717

54 Klepeis J L, Floudas C A. Astro-fold: a combinatorial and global optimization framework for ab initio prediction of three-dimensional structures of proteins from the amino acid sequence. *Biophys J*, 2003, 85(4): 2119-2146

55 Murzin A G, Brenner S E, Hubbard T, et al. Scop —— a structural classification of proteins database for the investigation of sequences and structures. *Journal of Molecular Biology*, 1995, 247(4): 536-540

56 Taylor W R, Flores T P, Orengo C A. Multiple protein structure alignment. *Protein Sci*, 1994, 3(10): 1858-1870

57 Notredame C, Higgins D G, Heringa J. T-coffee: a novel method for fast and accurate multiple sequence alignment. *J Mol Biol*, 2000, 302(1): 205-217

58 Sali A, Blundell T L. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol*, 1993, 234(3): 779-815

59 Bradley P, Misura K M S, Baker D. Toward high-resolution de novo structure prediction for small proteins. *Science*, 2005, 309(5742): 1868-1871

60 Zhang Y. I-Tasser: fully automated protein structure prediction in CASP8. *Proteins-Structure Function and Bioinformatics*, 2009, 77: 100-113

61 Zhang Y, Arakaki A K, Skolnick J R. Tasser: an automated method for the prediction of protein tertiary structures in CASP6. *Proteins-Structure Function and Bioinformatics*, 2005, 61: 91-98

62 Baker D. Prediction and design of macromolecular structures and interactions. *Philosophical Transactions of the Royal Society B-Biological Sciences*, 2006, 361(1467): 459-463

63 McCammon J A, Gelin B R, Karplus M. Dynamics of folded proteins. *Nature*, 1977, 267(5612): 585-590

64 Tsai J, Levitt M, Baker D. Hierarchy of structure loss in MD simulations of src SH3 domain unfolding. *J Mol Biol*, 1999, 291(1): 215-225

65 Tirado-Rives J, Jorgensen W L. Molecular dynamics simulations of the unfolding of apomyoglobin in water. *Biochemistry-Us*, 1993, 32(16): 4175-4184

66 Roccatano D, Amadei A, Di Nola A, et al. A molecular dynamics study of the 41-56 beta-hairpin from B1 domain of protein G. *Protein Sci*, 1999, 8(10): 2130-2143

67 Pande V S, Rokhsar D S. Molecular dynamics simulations of unfolding and refolding of a beta-hairpin fragment of protein G. *Proc Natl Acad Sci U S A*, 1999, 96(16): 9062-9067

68 Mark A E, van Gunsteren W F. Simulation of the thermal denaturation of hen egg white lysozyme: trapping the molten globule state. *Biochemistry-Us*, 1992, 31(34): 7745-7748

69 Voelz V A, Bowman G R, Beauchamp K, et al. Molecular simulation of ab initio protein folding for a millisecond folder NTL9(1-39). *J Am Chem Soc*, 2010, 132(5): 1526-1528

70 Day R, Daggett V. All-atom simulations of protein folding and unfolding. *Adv Protein Chem*, 2003, 66: 373-403

71 Daggett V, Levitt M. Protein unfolding pathways explored through molecular dynamics simulations. *J Mol Biol*, 1993, 232(2): 600-619

72 Daggett V. Protein folding-simulation. *Chem Rev*, 2006, 106(5): 1898-1916

73 Brooks C L 3rd. Characterization of "native" apomyoglobin by molecular dynamics simulation. *J Mol Biol*, 1992, 227(2): 375-380

74 Adcock S A, McCammon J A. Molecular dynamics: survey of methods for simulating the activity of proteins. *Chem Rev*, 2006, 106(5): 1589-1615

75 Langevin P. On the theory of Brownian motion. *C. R. Acad. Sci*, 1908, 146: 530-533

76 McCammon J A, Wolynes P G, Karplus M. Picosecond dynamics of tyrosine side chains in proteins. *Biochemistry-Us*, 1979, 18(6): 927-942

77 McQuarrie D. Statistical Mechanics. New York: Academic, 1976

78 Scheraga H A, Khalili M, Liwo A. Protein-folding dynamics: overview of molecular simulation techniques. *Annu Rev Phys Chem*, 2007, 58: 57-83

79 Moritsugu K, Smith J C. Temperature-dependent protein dynamics: a simulation-based probabilistic diffusion-vibration Langevin description. *J Phys Chem B*, 2006, 110(11): 5807-5816

80 Moritsugu K, Smith J C. Langevin model of the temperature and hydration dependence of protein vibrational dynamics. *J Phys Chem B*, 2005, 109(24): 12182-12194

81 Horger I, Velasco E, Mingorance J, et al. Langevin computer simulations of bacterial protein filaments and the force-generating mechanism during cell division. *Phys Rev E Stat Nonlin Soft Matter Phys*, 2008, 77: 011902

82 Franklin J, Doniach S. Dynamic bond constraints in protein Langevin dynamics. *J Chem Phys*, 2006, 124(15): 154901

83 Metropolis N, Ulam S. The Monte Carlo method. *J Am Stat Assoc*, 1949, 44(247): 335-341

84 Fishman G. Monte Carlo: Concepts, Algorithms, and Applications. New York: Springer Verlag, 1995,

85 Robert C, Casella G. Monte Carlo Statistical Methods. second edition. New York: Springer-Verlag, 2004

86 Mosegaard K, Tarantola A. Monte-Carlo sampling of solutions to inverse problems. *J Geophys Res-Sol Ea*, 1995, 100(B7): 12431-12447

87 Huang K. Lectures on Statistical Physics and Protein Folding. Singapore: World Scientific Publishing, 2005

88 Huang K. Conditioned self-avoiding walk (csaw): stochastic approach to protein folding. *Biophys. Rev. Lett.*, 2007, 2: 139-154

89 Sun W. Protein folding simulation by all-atom CSAW method. *IEEE International Conference on Bioinformatics and Biomedicine*, 2007, 2(4): 45-52

90 Duan Y, Kollman P A. Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science*, 1998, 282(5389): 740-744

91 Pitera J W, Swope W. Understanding folding and design: Replica-exchange simulations of "Trp-cage" fly miniproteins. *P Natl Acad Sci USA*, 2003, 100(13): 7587-7592

92 Zagrovic B, Snow C D, Shirts M R, et al. Simulation of folding of a small alpha-helical protein in atomistic detail using worldwide-distributed computing. *Journal of Molecular Biology*, 2002, 323(5): 927-937

93 Levintha C. Are there pathways for protein folding. *J Chim Phys Pcb*, 1968, 65(1): 44-45

94 Creighton T E. Effects of urea and guanidine.hcl on folding and unfolding of pancreatic trypsin-inhibitor. *Journal of Molecular Biology*, 1977, 113(2): 313-328

95 Creighton T E. Energetics of folding and unfolding of pancreatic trypsin-inhibitor. *Journal of Molecular Biology*, 1977, 113(2): 295-312

96 Creighton T E. Conformational restrictions on pathway of folding and unfolding of pancreatic Trypsin-Inhibitor. *Journal of Molecular Biology*, 1977, 113(2): 275-293

97 Karplus M, Weaver D L. Protein-folding dynamics. *Nature*, 1976, 260(5550): 404-406

98 Karplus M, Weaver D L. Diffusion-collision model for protein folding. *Biopolymers*, 1979, 18(6): 1421-1437

99 Branden C I, Tooze J. Introduction to Protein Structure. 2nd edition, New York: Garl. and Science, 1998

100 Englander S W. Protein folding intermediates and pathways studied by hydrogen exchange. *Annu Rev Bioph Biom*, 2000, 29: 213-238

101 Lee B. Solvent reorganization contribution to the transfer thermodynamics of small nonpolar molecules. *Biopolymers*, 1991, 31(8): 993-1008

102 Schröinger E. Statistical Thermodynamics. Cambridge: Cambridge University Press, 1948

103 Hill T L. An Introduction to Statistical Thermodynamics. New York: Dover, 1987

104 Fowler R H, Guggenheim E A. Statistical Thermodynamics. London: Cambridge University Press, 1939

105 Callender R H, Dyer R B, Gilmanshin R, et al. Fast events in protein folding: the time evolution of primary processes. *Annual Review of Physical Chemistry*, 1998, 49: 173-202

106 Fersht A R, Shi J P, Knill-Jones J, et al. Hydrogen bonding and biological specificity analysed by protein engineering. *Nature*, 1985, 314(6008): 235-238

107 Kauzmann W. Some factors in the interpretation of protein denaturation. *Advances in Protein Chemistry*, 1959, 14: 1-63

108 Cordes M H J, Davidson A R, Sauer R T. Sequence space, folding and protein design. *Curr Opin Struc Biol*, 1996, 6(1): 3-10

109 Hecht M H, Das A, Go A, et al. De novo proteins from designed combinatorial libraries. *Protein Sci*, 2004, 13(7): 1711-1723

110 Kamtekar S, Schiffer J M, Xiong H Y, et al. Protein design by binary patterning of polar and nonpolar amino-acids. *Science*, 1993, 262(5140): 1680-1685

111 Chen J M, Stites W E. Packing is a key selection factor in the evolution of protein hydrophobic cores. *Biochemistry-Us*, 2001, 40(50): 15280-15289

112 Matouschek A, Kellis J T, Serrano L, et al. Mapping the transition-state and pathway of protein folding by protein engineering. *Nature*, 1989, 340(6229): 122-126

113 Lazaridis T, Paulaitis M E. Entropy of hydrophobic hydration —— a new statistical mechanical formulation. *J Phys Chem-Us*, 1992, 96(9): 3847-3855

114 Lazaridis T, Paulaitis M E. Simulation studies of the hydration entropy of simple, hydrophobic solutes. *J Phys Chem-Us*, 1994, 98(2): 635-642

115 Ashbaugh H S, Paulaitis M E. Entropy of hydrophobic hydration: extension to hydrophobic chains. *J Phys Chem-Us*, 1996, 100(5): 1900-1913

116 Lamb M L, Jorgensen W L. Computational approaches to molecular recognition. *Curr Opin Chem Biol*, 1997, 1(4): 449-457

117 Gilson M K, Given J A, Bush B L, et al. The statistical-thermodynamic basis for computation of binding affinities: A critical review. *Biophys J*, 1997, 72(3): 1047-1069

118 Hermans J, Wang L. Inclusion of loss of translational and rotational freedom in theoretical estimates of free energies of binding. Application to a complex of benzene and mutant T4 lysozyme. *J Am Chem Soc*, 1997, 119(11): 2707-2714

119 Pratt L R, Chandler D. Theory of hydrophobic effect. *Journal of Chemical Physics*, 1977, 67(8): 3683-3704

120 Tanford C. The Hydrophobic Effect: Formation of Micelles and Biological Membranes. 2nd edition. New York: John Wiley & Sons Inc, 1980

121 Hummer G, Garde S, Garcia A E, et al. The pressure dependence of hydrophobic interactions is consistent with the observed pressure denaturation of proteins. *P Natl Acad Sci USA*, 1998, 95(4): 1552-1555

122 Charton M, Charton B I. The structural dependence of amino-acid hydrophobicity parameters. *J Theor Biol*, 1982, 99(4): 629-644

123 Pace C N, Shirley B A, McNutt M, et al. Forces contributing to the conformational stability of proteins. *Faseb J*, 1996, 10(1): 75-83

124 Bennaim A, Marcus Y. Solvation thermodynamics of nonionic solutes. *Journal of Chemical Physics*, 1984, 81(4): 2016-2027

125 Widom B. Potential-distribution theory and the statistical-mechanics of fluids. *J Phys Chem-Us*, 1982, 86(6): 869-872

126 Chandler D. Interfaces and the driving force of hydrophobic assembly. *Nature*, 2005, 437(7059): 640-647

127 Lazaridis T. Solvent size vs cohesive energy as the origin of hydrophobicity. *Accounts Chem Res*, 2001, 34(12): 931-937

128 Lee B. The physical origin of the low solubility of nonpolar solutes in water. *Biopolymers*, 1985, 24(5): 813-823

129 Hummer G, Garde S, Garcia A E, et al. An information theory model of hydrophobic interactions. *P Natl Acad Sci USA*, 1996, 93(17): 8951-8955

130 Stillinger F H. Structure in aqueous solutions of nonpolar solutes from the standpoint of scaled-particle theory. *J Solution Chem*, 1973, 2(2-3): 141-158

131 Eisenberg D, Mclachlan A D. Solvation energy in protein folding and binding. *Nature*, 1986, 319(6050): 199-203

132 Lum K, Chandler D, Weeks J D. Hydrophobicity at small and large length scales. *Journal of Physical Chemistry B*, 1999, 103(22): 4570-4577

133 Huang D M, Chandler D. The hydrophobic effect and the influence of solute-solvent attractions. *Journal of Physical Chemistry B*, 2002, 106(8): 2047-2053

134 Gerstein M, Lyndenbell R M. Simulation of water around a model protein helix .1. 2-dimensional projections of solvent structure. *J Phys Chem-Us*, 1993, 97(12): 2982-2990

135 Gerstein M, Lyndenbell R M. Simulation of water around a model protein helix .2. the relative contributions of packing, hydrophobicity, and hydrogen-bonding. *J Phys Chem-Us*, 1993, 97(12): 2991-2999

136 Wallqvist A, Berne B J. Computer-simulation of hydrophobic hydration forces on stacked plates at short-range. *J Phys Chem-Us*, 1995, 99(9): 2893-2899

137 Huang D M, Geissler P L, Chandler D. Scaling of hydrophobic solvation free energies. *Journal of Physical Chemistry B*, 2001, 105(28): 6704-6709

138 Garde S, Hummer G, Garcia A E, et al. Origin of entropy convergence in hydrophobic hydration and protein folding. *Phys Rev Lett*, 1996, 77(24): 4966-4968

139 Pauling L, Corey R B. Configurations of polypeptide chains with favored orientations around single bonds: two new pleated sheets. *Proc Natl Acad Sci USA*, 1951, 37(11): 729-740

140 Pauling L, Corey R B. Atomic coordinates and structure factors for two helical configurations of polypeptide chains. *Proc Natl Acad Sci USA*, 1951, 37(5): 235-240

141 Schellman J A. The stability of hydrogen-bonded peptide structures in aqueous solution. *C R Trav Lab Carlsberg Chim*, 1955, 29(14-15): 230-259

142 Schellman J A. The thermodynamics of urea solutions and the heat of formation of the peptide hydrogen bond. *C R Trav Lab Carlsberg Chim*, 1955, 29(14-15): 223-229

143 Richards F M. Interpretation of protein structures - total volume, group volume distributions and packing density. *Journal of Molecular Biology*, 1974, 82(1): 1-14

144 Hvidt A, Westh P. Different views on the stability of protein conformations and hydrophobic effects. *J Solution Chem*, 1998, 27(5): 395-402

145 Baker E N, Hubbard R E. Hydrogen-bonding in globular-proteins. *Prog Biophys Mol Bio*, 1984, 44(2): 97-179

146 Bordo D, Argos P. The role of side-chain hydrogen-bonds in the formation and stabilization of secondary structure in soluble-proteins. *Journal of Molecular Biology*, 1994, 243(3): 504-519

147 Jeffrey G A, Saenger W. Hydrogen Bonding in Biological Structures Telos. New York: Springer-Verlag, 1994

148 Kabsch W, Sander C. Dictionary of protein secondary structure - pattern–recognition of hydrogen-bonded and geometrical features. *Biopolymers*, 1983, 22(12): 2577-2637

149 Andersen C. Protein Structure and the Diversity of Hydrogen Bonds: [PhD Thesis]. Copenhagen, Denmark: The Technical Unversity of Denmark, 2001

150 Avbelj F, Grdadolnik S G, Grdadolnik J, et al. Intrinsic backbone preferences are fully present in blocked amino acids. *P Natl Acad Sci USA*, 2006, 103(5): 1272-1277

151 Boobbyer D N A, Goodford P J, Mcwhinnie P M, et al. New hydrogen-bond potentials for use in determining energetically favorable binding-sites on molecules of known structure. *J Med Chem*, 1989, 32(5): 1083-1094

152 Wade R C, Clark K J,, Goodford P J. Further development of hydrogen-bond functions for use in determining energetically favorable binding-sites on molecules of known structure 1. ligand probe groups with the ability to form 2 hydrogen-bonds. *J Med Chem*, 1993, 36(1): 140-147

153 Wade R C, Goodford P J. Further development of hydrogen-bond functions for use in determining energetically favorable binding-sites on molecules of known structure 2. ligand probe groups with the ability to form more than 2 hydrogen-bonds. *J Med Chem*, 1993, 36(1): 148-156

154 Frishman D, Argos P. Knowledge-based protein secondary structure assignment. *Proteins-Structure Function and Genetics*, 1995, 23(4): 566-579

155 Fersht A R, Shi J P, Knilljones J, et al. Hydrogen-bonding and biological specificity analyzed by protein engineering. *Nature*, 1985, 314(6008): 235-238

156 Rose G D, Wolfenden R. Hydrogen bonding, hydrophobicity, packing, and protein folding. *Annu Rev Biophys Biomol Struct*, 1993, 22: 381-415

157 Stahl N, Jencks W P. Hydrogen-bonding between solutes in Aqueous-solution. *J Am Chem Soc*, 1986, 108(14): 4196-4205

158 Bolen D W, Rose G D. Structure and energetics of the hydrogen-bonded backbone in protein folding. *Annu Rev Biochem*, 2008, 77: 339-362

159 Baldwin R L. In search of the energetic role of peptide hydrogen bonds. *J Biol Chem*, 2003, 278(20): 17581-17588

160 Fleming P J, Rose G D. Do all backbone polar groups in proteins form hydrogen bonds? *Protein Sci*, 2005, 14(7): 1911-1917

161 Klotz I M, Franzen J S. Hydrogen bonds between model peptide groups in solution. *J Am Chem Soc*, 1962, 84(18): 3461-3466

162 Privalov P L, Gill S J. Stability of protein-structure and hydrophobic interaction. *Advances in Protein Chemistry*, 1988, 39: 191-234

163 Doig A J, Williams D H. Binding-energy of an amide amide hydrogen-bond in aqueous and nonpolar-solvents. *J Am Chem Soc*, 1992, 114(1): 338-343

164 Shirley B A, Stanssens P, Hahn U, et al. Contribution of hydrogen bonding to the conformational stability of ribonuclease T1. *Biochemistry-Us*, 1992, 31(3): 725-732

165 Pace C N. Conformational stability of globular-proteins. *Trends Biochem Sci*, 1990, 15(1): 14-17

166 Gruebele M. Downhill protein folding: evolution meets physics. *Cr Biol*, 2005, 328(8): 701-712

167 Baldwin R L. Energetics of protein folding. *J Mol Biol*, 2007, 371(2): 283-301

168 Pauling L. The application of the quantum mechanics to the structure of the hydrogen molecule and hydrogen molecule-ion and to related problems. *Chemical Reviews*, 1928, 5(2): 173-213

169 Berg J, Tymoczko J, Stryer L. Biochemistry. New York: W. H. Freeman, 2002

170 Anderson D E, Becktel W J, Dahlquist F W. Ph-induced denaturation of proteins —— a single salt bridge con-

tributes 3-5 kcal mol to the free-energy of folding of T4-lysozyme. *Biochemistry-Us*, 1990, 29(9): 2403-2408
171 Horovitz A, Serrano L, Avron B, et al. Strength and cooperativity of contributions of surface salt bridges to protein stability. *Journal of Molecular Biology*, 1990, 216(4): 1031-1044
172 Kumar S, Nussinov R. Close-range electrostatic interactions in proteins. *Chembiochem*, 2002, 3(7): 604-617
173 Bosshard H R, Marti D N, Jelesarov I. Protein stabilization by salt bridges: concepts, experimental approaches and clarification of some misunderstandings. *J Mol Recognit*, 2004, 17(1): 1-16
174 Miessler G L, Tarr D A. Inorganic Chemistry. London: Prentice Hall, 1991
175 Popov K, Ronkkomaki H, Lajunen L H J. Guidelines for NMR measurements for determination of high and low pK(a) values. *Pure Appl Chem*, 2006, 78(3): 663-675
176 Henderson L J. Concerning the relationship between the strength of acids and their capacity to preserve neutrality. *Am J Physiol*, 1908, 21(2): 173-179
177 Gilson M K, Honig B H. Calculation of electrostatic potentials in an enzyme active-site. *Nature*, 1987, 330(6143): 84-86
178 Sternberg M J E, Hayes F R F, Russell A J, et al. Prediction of electrostatic effects of engineering of protein charges. *Nature*, 1987, 330(6143): 86-88
179 Schaefer M, Sommer M,, Karplus M. pH-dependence of protein stability: absolute electrostatic free energy differences between conformations. *Journal of Physical Chemistry B*, 1997, 101(9): 1663-1683
180 Davis M E, Mccammon J A. Calculating electrostatic forces from grid-calculated potentials. *J Comput Chem*, 1990, 11(3): 401-409
181 Honig B, Nicholls A. Classical electrostatics in biology and chemistry. *Science*, 1995, 268(5214): 1144-1149
182 Antosiewicz J, Mccammon J A, Gilson M K. Prediction of Ph-dependent properties of proteins. *Journal of Molecular Biology*, 1994, 238(3): 415-436
183 Warshel A, Russell S T. Calculations of electrostatic interactions in biological-systems and in solutions. *Q Rev Biophys*, 1984, 17(3): 283-422
184 Warshel A, Russell S T, Churg A K. Macroscopic models for studies of electrostatic interactions in proteins - limitations and applicability. *P Natl Acad Sci-Biol*, 1984, 81(15): 4785-4789
185 Warshel A, Russell S T, Churg A K. Microscopic and macroscopic models for calculations of electrostatic energies in proteins. *Biophys J*, 1985, 47(2): A20-A20
186 Warwicker J, Watson H C. Calculation of the electric-potential in the active-site cleft due to alpha-helix dipoles. *Journal of Molecular Biology*, 1982, 157(4): 671-679
187 Pickersgill R W. A rapid method of calculating charge-charge interaction energies in proteins. *Protein Eng*, 1988, 2(3): 247-248
188 Fersht A R, Sternberg M J E. Can a simple function for the dielectric response model electrostatic effects in globular-proteins. *Protein Eng*, 1989, 2(7): 527-530
189 Kumar S, Nussinov R. Fluctuations between stabilizing and destabilizing electrostatic contributions of ion pairs in conformers of the c-Myc-Max leucine zipper. *Proteins-Structure Function and Genetics*, 2000, 41(4): 485-497
190 Kumar S, Nussinov R. Salt bridge stability in monomeric proteins. *Journal of Molecular Biology*, 1999, 293(5): 1241-1255
191 Jones J E. On the determination of molecular fields III - from crystal measurements and kinetic theory data. *P R Soc Lond a-Conta*, 1924, 106(740): 709-718
192 Jones J E. On the determination of molecular fields - II from the equation of state of a gas. *P R Soc Lond a-Conta*, 1924, 106(738): 463-477
193 Chandler D, Weeks J D, Andersen H C. Vanderwaals picture of liquids, solids, and phase-transformations. *Science*, 1983, 220(4599): 787-794
194 ten Wolde P R, Chandler D. Drying-induced hydrophobic polymer collapse. *P Natl Acad Sci USA*, 2002, 99(10): 6539-6543
195 Chandler D. Introduction to Modern Statistical Mechanics. New York: Oxford University Press, 1987
196 Gallicchio E, Kubo M M, Levy R M. Enthalpy-entropy and cavity decomposition of alkane hydration free energies: numerical results and implications for theories of hydrophobic solvation. *Journal of Physical Chemistry B*, 2000, 104(26): 6271-6285
197 Richards F M. Areas, volumes, packing, and protein-structure. *Annu Rev Biophys Bio*, 1977, 6: 151-176
198 Dahiyat B I, Mayo S L. Probing the role of packing specificity in protein design. *P Natl Acad Sci USA*, 1997, 94(19): 10172-10177
199 Chen J M, Lu Z Q, Sakon J, et al. Increasing the thermostability of staphylococcal nuclease: implications for the origin of protein thermostability. *Journal of Molecular Biology*, 2000, 303(2): 125-130
200 Nelson R, Sawaya M R, Balbirnie M, et al. Structure of the cross-beta spine of amyloid-like fibrils. *Nature*, 2005, 435(7043): 773-778
201 Tsemekhman K, Goldschmidt L, Eisenberg D, et al. Cooperative hydrogen bonding in amyloid formation. *Protein Sci*, 2007, 16(4): 761-764

CNKI 优先出版编码: lxjz2010-107

STATISTICAL THERMODYNAMICS FOUNDATIONS OF PROTEIN FOLDING
——MECHANISM OF NON-COVALENT INTERACTIONS*

SUN Weitao†

Zhou Pei-Yuan Center for Applied Mathematics, Tsinghua University, Beijing 100084, China

Abstract Protein folding is a complex process involving physics, chemistry and biology. Multidisciplinary approaches have to be used, including thermodynamics, statistical mechanics, polymer dynamics. The attempts to develop protein folding models solely at macroscopic or microscopic scale still face huge difficulties. It is believed that statistical thermodynamics is a prior way for addressing protein folding problem. In this work, we reviewed the mechanism of important non-covalent interactions in water solution, such as hydrophobic effect, hydrogen bonding and electrostatic interactions. The effects of non-covalent interactions on protein structure stability is discussed. The understanding of relationship between non-covalent interactions and entropy/enthalpy/free energy is essential for developing statistical thermodynamics theories, as well as for proposing microscopic statistical interpretations for macroscopic thermodynamics.

Keywords protein folding, free energy, statistical thermodynamics

孙卫涛, 清华大学周培源应用数学研究中心副研究员. 1998 年在大连理工大学工程力学系获得工学学士学位, 2003 年在清华大学工程力学系获得工学硕士、工学博士学位. 2003~2005 年在清华大学作博士后研究工作, 2005 年至今在清华大学周培源应用数学研究中心工作, 主要研究方向包括蛋白质结构动力学、弹性波动力学、全局优化理论等. 2005 年以来, 在林家翘先生的亲自指导下, 主要从事蛋白质结构的动力学过程和统计力学理论研究. 作为课题负责人承担过国家自然科学基金 “复杂介质波动方程反演中的多尺度混合优化方法研究”(10402015), 清华大学骨干人才支持计划研究 (2007)“基于条件自回避行走的蛋白质折叠机理研究”, 清华大学自主科研计划 (2010~2012)“蛋白质折叠的统计热力学原理”, 以及清华大学校内基础研究基金和中国博士后科学基金等多个课题的研究工作. 在国内外学术期刊、学术会议发表学术论文 30 余篇, 著有《弹性波动方程的有限差分数值方法》一书.

* The project supported by Tsinghua University Initiative Scientific Research Program (20101081751) and C.C. Lin special fund

† E-mail: sunwt@tsinghua.edu.cn