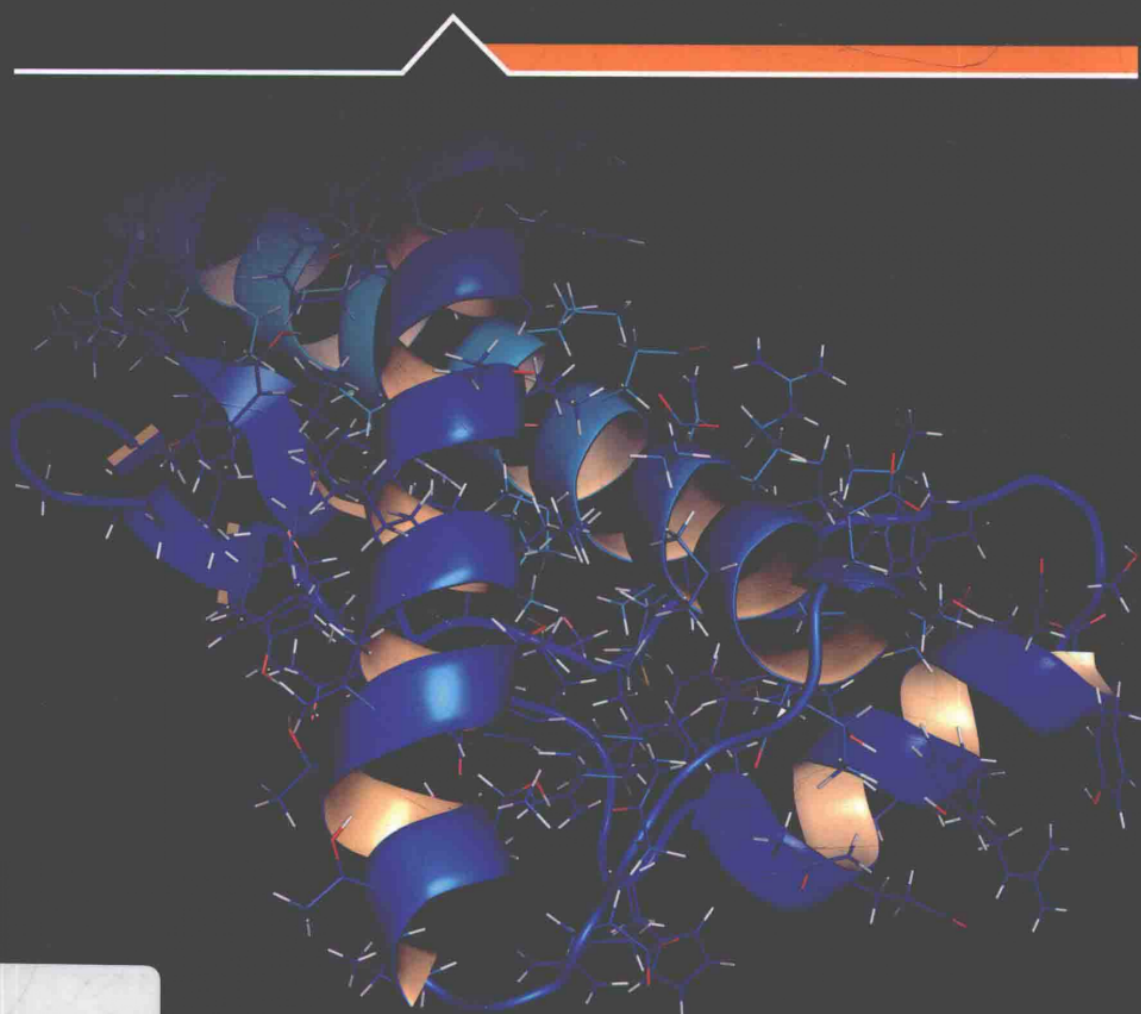


李瑞芳 编著

蛋白质折叠速率 与 mRNA



上海交通大学出版社
SHANGHAI JIAO TONG UNIVERSITY PRESS

内蒙古师范大学学术著作出版基金资助出版

蛋白质折叠速率 与 mRNA

李瑞芳 编著



上海交通大学出版社
SHANGHAI JIAO TONG UNIVERSITY PRESS

内容提要

本书是作者近年来研究蛋白质折叠速率所得成果的系统总结。主要内容包括：蛋白质折叠速率研究进展、蛋白质结构的预测方法、基因序列信息参数对蛋白质折叠速率的影响、回文结构及其统计分析、回文结构对蛋白质折叠速率的影响、同义密码子的使用偏性及其对肽链折叠速率的影响、mRNA 二级结构及其对蛋白质折叠速率的影响等。

本书可供生物物理专业研究生及相关研究者参考。

图书在版编目(CIP)数据

蛋白质折叠速率与 mRNA / 李瑞芳编著. —上海：
上海交通大学出版社，2017
ISBN 978 - 7 - 313 - 16292 - 2

I. ①蛋… II. ①李… III. ①蛋白质—分子折叠—研究②mRNA—研究 IV. ①Q51②Q522

中国版本图书馆 CIP 数据核字(2016)第 308843 号

蛋白质折叠速率与 mRNA

编 著：李瑞芳

出版发行：上海交通大学出版社

邮政编码：200030

出 版 人：郑益慧

印 制：凤凰数码印务有限公司

开 本：710 mm×1000 mm 1/16

字 数：196 千字

版 次：2017 年 6 月第 1 版

书 号：ISBN 978 - 7 - 313 - 16292 - 2/Q

定 价：58.00 元

地 址：上海市番禺路 951 号

电 话：021 - 64071208

经 销：全国新华书店

印 张：12.75

印 次：2017 年 6 月第 1 次印刷

版权所有 侵权必究

告读者：如发现本书有印装质量问题请与印刷厂质量科联系

联系电话：025 - 83657309

前 言

20 世纪 70 年代初, Anfinsen 提出蛋白质的氨基酸序列包含足够的信息决定其空间结构的假说, 并因此获得诺贝尔奖。但是, 一些研究者发现存在氨基酸序列相同但构象不同和构象相似而氨基酸序列相差显著的蛋白质。这说明氨基酸序列并不是决定蛋白质结构的充分条件, 同时意味着除氨基酸序列以外的其他信息对蛋白质空间结构的形成也是非常重要的。但是这种信息到底储存在哪里? 是通过什么途径来传递的? 这些问题一直困扰着相关研究者们。如果我们能够找到 mRNA 除了编码氨基酸之外影响蛋白质结构的其他定性或定量结果, 那就意味着我们找到了两者之间的一些联系, 这是一项极具挑战性的工作。

如果除氨基酸之外确实还有一些影响蛋白质结构和功能的信息, 这种信息很有可能存在于 mRNA 序列和结构中。但是, 现有的报道中没有找到相关的定量描述。考虑到折叠速率包含一些蛋白质结构和功能的信息, 那么蛋白质折叠速率与其相应 mRNA 序列之间的关系就会表明, 蛋白质对应的 mRNA 序列除了编码氨基酸之外确实还携带一些影响蛋白质结构和功能的信息。而蛋白质折叠速率又与蛋白质结构有着非常密切的相关性。如果能找到有关 mRNA 序列和结构对蛋白质折叠速率影响的一些定量结果, 也就意味着找到了 mRNA 序列和结构与蛋白质结构间相关性的定量结果。这是我们工作思路的出发点。

从最简单的单细胞生物到最高等的人类, 其最基本最重要的组成物质是 DNA、RNA 和蛋白质, 这些组成物质是生物体遗传信息的携带者。由于人类基因组计划的顺利实施以及各种后基因组计划的开展, 人们获得了大量的生

物分子数据。这些生物数据具有丰富的内涵,其背后隐藏着许多人类尚不清楚的生物学知识。利用数学、物理、化学等各种方法,通过大量数据分析来揭示生物分子数据中蕴含的生物信息是一项非常有意义的工作,也为我们的工作提供了研究方法。

基因组中 mRNA 序列被认为是一种遗传语言,它的词汇组成和语法结构中包含了绝大部分的遗传信息。作为遗传信息的使者,mRNA 序列包含的遗传信息也绝非编码规则这一项信息,如密码子第三位点存在的自由度和 mRNA 序列中复杂的高级结构均是编码规则之外的信息。挖掘这些遗传信息的生物学功能是后基因组时代研究的主要任务。人们发现 mRNA 序列中存在非常复杂的回文结构分布,而且回文结构可能形成发夹结构和十字形结构并具有重要的生物学功能。

蛋白质是生命活动的主要承担者,是一类最重要的生物大分子,在生物体内占有特殊的地位,蛋白质的一级结构即肽链是没有生物活性的,只有折叠成一定的空间结构,才能具有生物活性。蛋白质的折叠就是指一个蛋白质从它的变性状态转变到它的特定生物学天然构象的过程。异常的蛋白质空间结构很可能导致其生物活性的降低、丧失,甚至会导致疾病。疯牛病、阿尔茨海默病等都是由蛋白质折叠异常引起的疾病。蛋白质折叠问题是分子生物学的核心问题之一,其中有很多谜团尚未解开。蛋白质折叠问题是生命科学领域的前沿课题之一,并且被列为“21 世纪生物物理学”的重要课题之一,它是分子生物学中心法则尚未解决的一个重大生物学问题。基于蛋白质一级序列或 mRNA 预测蛋白质折叠并进一步预测其功能,是极富挑战性的工作。研究蛋白质折叠,尤其是折叠早期过程,即新生肽段的折叠过程是全面的最终阐明中心法则的一个根本问题,是蛋白质分子设计及蛋白质工程的需要,也是越来越多基因工程产物复性复活的需要,更是理解与错折叠相关的疾病起源的需要。人们正从不同角度入手来研究蛋白质的折叠问题。

蛋白质折叠问题非常复杂,具体可分解为两个问题:① 蛋白质折叠的热力学(thermodynamics)问题,即什么是蛋白质折叠的驱动力?② 折叠的动力学(kinetics)或折叠速率问题,即为什么新生或变性肽链可以迅速折叠成天然构象状态?这两个问题相辅相成,只有同时得到解决,才能彻底揭开蛋白质折叠之谜。我们一直重点关注蛋白质折叠速率问题。

目前,越来越多的研究者加入到蛋白质折叠速率的研究中,提出了诸多蛋白质折叠速率的影响因素,随之提出了一些蛋白质折叠速率的预测方法。但是,还有诸多影响因素,特别是基于蛋白质一级序列的一些因素尚未提出,蛋白质折叠速率的预测精度尚待提高。另外,人们普遍认为决定蛋白质折叠速率的因素主要来自蛋白质的氨基酸序列和各级结构以及环境和温度。虽然有关于氨基酸序列包含了决定蛋白质三维结构全部信息的假说,但决定正常功能蛋白质信息的组分和来源是不完善的,也是非常复杂的。我们坚信这些信息的一部分来自核酸序列,并将提取或定义刻画 mRNA 特征的信息参数,探讨它们对蛋白质折叠速率的影响。我们认为,决定蛋白质折叠速率的信息还来自 mRNA 序列,它们在调节蛋白质折叠过程中起着重要的作用。

本书大部分内容是笔者及其团队近年来的研究工作成果。本研究分析了蛋白质氨基酸序列的特征,以期从中挖掘蛋白质折叠速率的影响因素,指出 mRNA 序列中复杂的二级结构、同义密码子使用、回文结构等所包含的信息是影响 mRNA 运输、基因表达调控乃至蛋白质结构的重要根源之一。通过分析 mRNA 序列这种遗传语言的词汇组成、语法结构信息、同义密码子使用偏好以及回文结构、二级结构的组成和分布,研究它们对相应蛋白质折叠速率的影响,从而挖掘它们蕴含的新的生物学意义。笔者期望通过这些研究工作推进相关学者对 mRNA 序列以及蛋白质折叠等生物过程的认识,并为青年学者提供研究思路和方法。

感谢为本书研究成果共同奋斗过的几位学生,感谢上海交通大学出版社杨迎春博士对本书的修改和指正,感谢内蒙古大学李宏教授的指导,感谢内蒙古自然科学基金(2016MS0362)对本书的资助。感谢内蒙古师范大学的领导和老师们对我学习的支持和鼓励。感谢我的家人对我的支持和关心!非常感谢我的父母和姐姐,他们始终鼓励我、支持我,帮我做生活琐事,让我有更多的时间投入到科研工作中。感谢我的弟弟和弟妹在英文写作上的有益帮助和支持。感谢我的丈夫和女儿,是他们的爱给予我无穷的动力和战胜困难的勇气。

书中存在的不足与缺点,恳请读者批评和指正。

目 录

第 1 章 绪论	001
1.1 研究背景	001
1.2 本书结构	003
参考文献	006
第 2 章 蛋白质的折叠速率	007
2.1 蛋白质的折叠速率研究进展	007
2.1.1 基于蛋白质三级结构的预测方法	008
2.1.2 基于蛋白质二级结构的预测方法	009
2.1.3 基于蛋白质一级结构的预测方法	009
2.2 氨基酸片段的极性对相应肽链折叠速率的影响	011
2.2.1 材料准备	011
2.2.2 统计分析	011
2.2.3 研究结果	013
2.2.4 讨论	014
2.3 mRNA 对蛋白质结构的影响	015
2.3.1 mRNA 对蛋白质结构影响的必然性分析	015
2.3.2 数据获得方法	017
2.3.3 数据分析方法	018
参考文献	021

第 3 章 基因序列信息参数对蛋白质折叠速率的影响	025
3.1 全 β 类蛋白质 mRNA 序列的 GC 含量对蛋白质折叠速率的影响	025
3.1.1 材料	025
3.1.2 统计分析	026
3.1.3 结果	028
3.1.4 讨论	030
3.2 核苷酸统计关联	031
3.3 mRNA 序列信息参数对蛋白质折叠速率的影响	033
3.3.1 材料	033
3.3.2 统计分析	036
3.3.3 结果与讨论	039
参考文献	041
第 4 章 回文结构	043
4.1 回文结构及其生物学意义	043
4.1.1 回文结构的意义	043
4.1.2 回文结构的获得	045
4.2 SARS 病毒全序列回文结构的统计分析	046
4.2.1 材料的选取	046
4.2.2 研究方法	047
4.2.3 结果	048
4.2.4 讨论	051
4.3 艾滋病病毒全序列回文结构的统计分析	052
4.3.1 回文结构的统计	053
4.3.2 结果与讨论	053
4.3.3 结论	056
参考文献	057

第 5 章 回文结构对蛋白质折叠速率的影响	061
5.1 病毒 mRNA 序列中回文结构对蛋白质折叠速率的影响	061
5.1.1 材料	062
5.1.2 统计分析	062
5.1.3 结果	063
5.1.4 讨论	067
5.2 全 β 类蛋白质相应 mRNA 序列中的回文结构对蛋白质折叠速率的影响	069
5.2.1 材料	070
5.2.2 统计分析	071
5.2.3 结果	072
5.2.4 讨论	074
5.3 mRNA 序列中回文结构对蛋白质折叠速率的影响	075
5.3.1 数据	076
5.3.2 统计分析	080
5.3.3 结果	081
5.3.4 讨论	086
参考文献	087
第 6 章 同义密码子的使用偏性	089
6.1 mRNA 同义密码子的使用偏性	089
6.1.1 同义密码子使用偏性	089
6.1.2 同义密码子使用偏性分析方法	090
6.2 大肠杆菌核蛋白的同义密码子使用偏性对肽链折叠速率的影响	091
6.2.1 材料	092
6.2.2 统计分析	097
6.2.3 结果	097
6.2.4 讨论	100

6.3 枯草芽孢杆菌核蛋白同义密码子使用偏性对肽链折叠速率的影响	101
6.3.1 材料	102
6.3.2 统计分析	107
6.3.3 结果	107
6.3.4 讨论	112
6.4 人类核蛋白同义密码子使用偏性对肽链折叠速率的影响	113
6.4.1 材料	113
6.4.2 结果	121
6.4.3 讨论	122
6.5 同义密码子的使用偏性对蛋白质折叠速率的影响	122
6.5.1 材料	124
6.5.2 统计分析	127
6.5.3 结果	127
6.5.4 讨论	133
参考文献	134
第7章 mRNA 二级结构及其对蛋白质折叠速率的影响	139
7.1 RNA 二级结构	139
7.1.1 RNA 二级结构预测算法	139
7.1.2 RNA 二级结构的元件	140
7.1.3 二级结构的图形表示方法	141
7.1.4 RNA 二级结构特征量提取	143
7.2 A 型流感病毒同义密码子的使用偏性对 mRNA 二级结构的 影响	145
7.2.1 材料	146
7.2.2 分析方法	150
7.2.3 结果	151
7.2.4 讨论	157

7.3 冠状病毒同义密码子的使用偏性对 mRNA 二级结构的影响	158
7.3.1 材料	158
7.3.2 统计分析	160
7.3.3 结果	160
7.3.4 讨论	168
7.4 mRNA 的二级结构对蛋白质折叠速率的影响	169
7.4.1 数据集	169
7.4.2 分析方法	173
7.4.3 结果	174
7.4.4 讨论	176
参考文献	178
第 8 章 总结与展望	182
索引	186

1.1 研究背景

蛋白质(protein)取自希腊文“proteios”,是“首要”“最原始”的意思,可见蛋白质在生命活动中的重要性。蛋白质是决定细胞性质的最主要物质,它是一类重要的生物大分子,在生物体内占有特殊的地位,是生命活动的主要承担者,是生命现象的主要物质基础。蛋白质不仅是构成一切细胞和组织的重要组成成分,而且也是生物体细胞中含量最丰富的高分子有机化合物。蛋白质与生命的基本现象密切相关,许多重要的生命现象和生理活动都是通过蛋白质来实现的,如催化反应、代谢调节、免疫保护、物质运输、机体运动、遗传控制、生长发育等^[1]。

然而,蛋白质只有折叠成正确的三维空间结构才能具有正常的生物学功能^[2]。错误的折叠、过慢的折叠不仅会导致其生物学功能的丧失,甚至会引起疾病,如可传播性海绵状脑病(CJD)、疯牛病(BSE)、肌萎缩性脊髓侧索硬化症(ALS)、帕金森综合征(PD)、阿尔茨海默病(AD)^[3]等。可见蛋白质折叠过程对所有的生物体系来说都是最重要和最基本的过程,这个过程是一个既复杂又深奥的基础理论问题,也是生命科学中最基本的问题之一,但对人类而言仍然是个未解之谜。

蛋白质折叠是一个生命自组织过程,当一个蛋白质自发地经历一个“混乱↔有序(disorder↔order)”转变时,称为蛋白质“折叠↔解折叠(folding↔unfolding)”,其中有很多谜团尚未解开。这是一个非常复杂的过程。在这一过程中,除二硫键外,主要是氢键、盐键等共价键的断裂和形成。这一折叠过

程遵循自由能减小的原理,即蛋白质在一定的时间范围内沿着某些或某一特定的路径到达其自由能最小或极小的生物学天然构象。

目前有关蛋白质折叠的研究主要包括以下几个方面:①对蛋白质折叠编码问题的研究,即天然结构是如何选择它的氨基酸序列的,该问题属于蛋白质折叠的热力学问题;②对蛋白质折叠过程的研究,即蛋白质是如何在一定的生物学时间范围内从一条伸展的多肽链到达其特定的天然构象的,该问题涉及蛋白质折叠的动力学和能量变化规律问题,以此为基础可以更好地对蛋白质结构进行预测。蛋白质结构预测是指从蛋白质的氨基酸序列预测其三维空间结构。对于理解蛋白质的结构与功能关系,并在此基础上进行蛋白质复性、突变体设计以及基于结构的药物设计具有重要意义,因而进行蛋白质的结构预测成为一种迫切需要。近年来,随着高精度的蛋白质折叠数据的不断积累,使得从蛋白质折叠速率角度研究蛋白质折叠机制的理论工作者,迎来了前所未有的机遇和挑战。

蛋白质的折叠速率体现了蛋白质折叠的动力学特征,在蛋白质的折叠过程中,有些蛋白质,特别是一些小蛋白质,没有明显的中间态,遵循“全”或“无”的过程,而有些蛋白质则需要中间态的不断积累才能最终完成其折叠过程。在这些蛋白质中,它们的折叠速率又有着很大的差异(通常情况下,折叠速率用 k_f 表示,单位为 s^{-1}),有些蛋白质在几微秒内就能完成其折叠过程,而有些则需要几个小时^[4]。另外,蛋白质序列的极小差别或溶剂环境的细微变化,都会引起蛋白质折叠动力学的巨大改变。为了更好地解释这一差异,真实地模拟蛋白质的折叠过程,并计算出它们的折叠速率,近年来发展了许多实验方法来研究蛋白质的折叠过程,如各种光谱技术(圆二色、荧光、红外)、质谱技术、氢交换以及核磁共振等^[5-9]。迄今为止,蛋白质动力学研究方面已经积累了相当可观的数据,并出现了相应的蛋白质折叠数据库^[10,11],为我们数据库的建设奠定了深厚的基础。

目前,关于蛋白质折叠速率问题,大都是从蛋白质的高级结构、二级结构,乃至氨基酸序列出发进行研究的。然而,我们研究表明,除此之外,RNA也包含了影响蛋白质折叠速率的因素。这是生物学的一个重要问题,如果解决了这一问题,等同于回答了以下多年来一直困惑生物学家们的问题:分子生物学“中心法则”揭示了遗传信息从DNA到RNA到蛋白质的传递过程,在此过

程中,通过翻译,由成熟的 mRNA 序列得到相应的蛋白质氨基酸序列。除了这种翻译信息之外, mRNA 是否携带决定蛋白质构象的更为重要的因素呢? 1973 年, Anfinsen^[12] 等提出一个原理认为蛋白质的氨基酸序列包含了决定蛋白质结构的全部信息。但是, 随后的研究者发现存在着氨基酸序列相同但构象不同的蛋白质, 以及构象相似而氨基酸序列相差显著的蛋白质^[13]。这说明氨基酸序列并不是决定蛋白质空间结构的充分条件, 即除氨基酸序列以外的其他信息对蛋白质空间结构的形成也是重要的。这种信息很有可能存在于 mRNA 的序列和结构中。基于这个思路, 我们就相关问题开展了系列研究。

1.2 本书结构

本书内容安排如下: 第 1 章介绍我们工作的研究思路及本书内容的整体安排。第 2 章介绍蛋白质折叠速率的研究进展及我们关于氨基酸片段的平均极性与蛋白质折叠速率关系的研究。具体做法是按照蛋白质的二级结构类型, 从每条氨基酸序列中截取所有的 α 螺旋和 β 折叠片段作为研究对象, 计算各片段的折叠速率和平均极性, 分别在各物种的 α 螺旋和 β 折叠两类二级结构片段中分析折叠速率和平均极性的相关性。研究表明, 不论是病毒蛋白质还是大肠杆菌中的蛋白酶, 其中两类氨基酸片段的平均极性与折叠速率都是极显著相关的: 对于所有的 α 片段, 二者呈线性正相关; 而对于所有的 β 片段, 二者呈线性负相关。结果证实了在蛋白质折叠中, 氨基酸的极性起着非常重要的作用。

第 3 章中以罗辽复先生提出的基因序列信息参数^[14]为基础, 研究了 RNA 序列对相应蛋白质折叠速率的影响。选取了描述遗传语言词汇组成的信息参数一阶信息冗余(D_1)、描述遗传语言语法结构的信息参数二阶信息冗余(D_2)及其衍生的信息参数作为刻画编码序列的特征参数, 并基于一个较大的蛋白质数据集分析它们与蛋白质折叠速率之间的线性关系。结果表明, 对于二态蛋白质, D_2 和 $X\left(X = \frac{D_2}{D_1 + D_2}\right)$ 这一参量组与全 α 蛋白质和全 β 蛋白质的折叠速率均有极显著的相关性, 尤其对于全 α 蛋白质, 相关系数达到 0.84。而对于多态蛋白质, 相应的 mRNA 序列的 GC 含量对蛋白质折叠速率的影响较

为突出,进一步分析表明对于这种 GC 含量的影响,其中一部分来自密码子的第三位点,又一次证实了同义密码子的使用对蛋白质折叠速率的影响。为了从多个角度印证 mRNA 序列的 GC 含量对蛋白质折叠速率的影响,先选择编码序列的序列参数——GC 含量来做初步分析。基于 Gromiha 给出的 13 个全 β 类蛋白质,得到相应编码序列的 GC 含量并将它加入到 Gromiha 基于蛋白质序列 4 个参数的预测折叠速率回归方程中来检验 GC 含量在蛋白质折叠过程中的作用。与 Gromiha 的结果比较,预测值和实验值之间的相关系数得到了提高,结果表明 GC 含量对预测蛋白质折叠速率是有效的,意味着蛋白质编码序列的 GC 含量确实对蛋白质折叠速率有影响。进一步分析也表明这种影响主要来自密码子的第三位点及其与第二位点的关联,而不是来自从密码子到氨基酸的翻译信息。

第 4 章详细介绍了回文结构,并介绍了我们关于一些病毒的回文结构研究结果。统计分析和比较了艾滋病病毒(HIV-1)、丙型肝炎病毒、SARS 病毒及其他几种冠状病毒的回文结构 GC 含量的特征和分布,发现了其中一些回文结构的特殊分布。综合考虑几种高致病病毒序列的一些特殊回文结构,如在 GC 含量,回文长度以及位置等方面特殊分布的回文结构,发现这些特殊回文结构往往分布在病毒序列中较为关键的位置。所以,它们一定不只是普通的序列,而应该是具有某种生物功能的重要元件。我们猜想这些特殊回文一定携带一些影响病毒蛋白质功能的重要信息。既然回文结构与蛋白质的功能有非常紧密的联系,而且它们在 mRNA 序列中有非常广泛的分布,那么,回文结构完全可以作为研究 mRNA 与蛋白质之间关系的一个很好的参量。

第 5 章从不同的角度研究了回文结构对蛋白质折叠速率的影响。首先定义了 mRNA 序列片段的回文 GC 含量及回文密度两个参量。基于 α 螺旋片段及 β 折叠片段的折叠速率与其平均极性之间的相关性,分析了回文 GC 含量及回文密度对两者之间关系的影响,研究发现这两个参量确实对氨基酸片段折叠速率和平均极性之间的关系有影响。统计分析表明这两个来自 mRNA 序列的参量与氨基酸片段的平均极性无显著关系,这间接说明相应 mRNA 序列片段的回文 GC 含量及回文密度影响了蛋白质的折叠速率,且蛋白质折叠速率与回文两参量之间均呈正相关。进一步分析表明影响的根源来自回文结构的复杂性或可变性和同义密码子的使用偏好,而不是来自从密码子到氨基酸

的翻译信息。这说明 mRNA 序列确实携带有影响肽链折叠速率乃至蛋白质结构的信息。虽然在第 3 章中加入 mRNA 序列的 GC 含量后预测蛋白质折叠速率的结果有所改善,但回归方程的 Jackknife 检验结果表明 mRNA 全序列的 GC 含量并未完全涵盖影响蛋白质折叠速率的主要信息。我们分析,出现这样的结果是由于 mRNA 序列的 GC 含量不能反映编码序列的结构信息。于是,第 5 章定义了一个既能涵盖序列又能包含序列结构的参数,称为回文 GC 含量。用 mRNA 序列中的回文 GC 含量代替全序列 GC 含量做同样的分析。与 Gromiha 的结果比较,蛋白质折叠速率的预测值与实验值之间的相关系数得到了进一步的提高,并且卡方检验、Jackknife 检验也表明 mRNA 序列中回文 GC 含量对蛋白质折叠速率有更大的影响。这意味着 mRNA 序列的结构和序列组成是影响蛋白质折叠速率的主要因素。进一步分析也表明这种影响一部分来自回文结构的复杂性和可变性,基于这一思路,我们整理了一个较大的包含蛋白质折叠速率实验值、相应的 mRNA 序列中回文结构参数等信息的蛋白质数据集,分析了回文结构各参量与蛋白质折叠速率之间的关系,又一次证实了回文结构的复杂性和可变性是影响蛋白质折叠速率的一个重要因素。

在前期的研究过程中发现,影响蛋白质折叠速率的另一部分因素来自同义密码子的使用或者说是密码子第三位碱基的使用。所以,我们认为,同义密码子的使用偏性是体现 mRNA 特性的另一重要参量。第 6 章以不同物种为样本,研究了同义密码子使用偏性对蛋白质折叠速率的影响。研究发现了一些关于蛋白质折叠速率与同义密码子使用度之间的相关性,一些密码子的使用与蛋白质的折叠速率显著相关。比较 3 类蛋白质折叠速率与同义密码子使用度之间的相关性后发现,同一个密码子对不同类蛋白质折叠速率的调节方向一般是不同的。比较二态和多态蛋白质,发现了一些同义密码子的使用对二态和多态蛋白质的区别,它们对蛋白质折叠速率的调节方式是不一样的,说明同义密码子对不同类蛋白质的影响是不同的。

第 7 章定义了 RNA 二级结构的信息参量,主要有茎结构含量、环结构含量、平均能量密度,并以此为基础,定义了 RNA 柔性;建立了一个较大的包括 mRNA 二级结构的信息参量的蛋白质折叠速率库;研究了 RNA 二级结构对蛋白质折叠速率的影响。统计结果表明,mRNA 二级结构中茎结构的含量与蛋白质折叠速率呈显著的负相关性,而环结构含量与蛋白质折叠速率呈显著

的正相关性。另外,结果显示,无论是否考虑蛋白质的结构分类或折叠类型,mRNA 柔性与其相应蛋白质折叠速率之间均呈现出极显著或显著的正相关性。因此我们认为,mRNA 柔性是影响蛋白质折叠速率的普适参量。在前面工作总结的基础上,第 8 章提出了我们以后工作的方向和思路。

参考文献

- [1] 朱玉贤,李毅,郑晓峰. 现代分子生物学(第三版)[M]. 北京: 高等教育出版社,2007.
- [2] 王镜岩,朱圣庚,徐长法. 生物化学(第三版)[M]. 北京: 高等教育出版社,2002.
- [3] 井明艳,孙建义. 分子伴侣与蛋白质折叠[J]. 科学通报,2004,20(5): 407 - 411.
- [4] Kubelka J, Hofrichter J, Eaton W A. The protein folding “speed limit” [J]. *Curr. Opin. Struct. Biol*, 2004, 14(1): 76 - 88.
- [5] Fabian H, Naumann D. Methods to study protein folding by stopped-flow FT - IR [J]. *Methods*, 2004, 34(1): 28 - 40.
- [6] Zeeb M, Balbach J. Protein folding studied by real-time NMR spectroscopy [J]. *Methods*, 2004, 34(1): 65 - 74.
- [7] Maity H, Maity M, Krishna M M G, et al. Protein folding: the stepwise assembly of foldon units [J]. *Proc. Natl. Acad. Sci. USA*, 2005, 102(13): 4741 - 4746.
- [8] Xiao H, Hoerner J K, Eyles S J, et al. Mapping protein energy landscapes with amide hydrogen exchange and mass spectrometry; I. A generalized model for a two-state protein and comparison with experiment [J]. *Protein Sci*, 2005, 14(2): 543 - 557.
- [9] Maxwell K L, Wildes D, Zarrine-Afsar A, et al. Protein folding: defining a “standard” set of experimental conditions and a preliminary kinetic data set of two-state proteins [J]. *Protein Sci*, 2005, 14(3): 602 - 616.
- [10] Fulton K F, Bate M A, Faux N G, et al. Protein folding database (PFD 2.0): an online environment for the International Foldon Consortium [J]. *Nucleic Acids Res*, 2007, 35(Database issue): D304 - D307.
- [11] Bogatyreva N S, Osypov A A, Ivankov D N. KineticDB: a database of protein folding kinetics [J]. *Nucleic Acids Res*, 2009, 37(Database issue): D342 - D346.
- [12] Anfinsen C B. Principles that govern the folding of protein chains [J]. *Science*, 1973, 181(4096): 223 - 230.
- [13] Telling G C, Parchi P, Dearmond S J, et al. Evidence for the conformation of the pathologic isoform of the prion protein enciphering and propagating prion diversity [J]. *Science*, 1996, 274(5295): 2079 - 2082.
- [14] 罗辽复. 生命进化的物理观[M]. 上海: 科学技术出版社,2000.