

综述与专论

蛋白质结构与功能中的结构域^{*}

廉德君 许根俊

(中国科学院上海生物化学研究所, 上海 200031)
(中国科学院生物物理研究所生物大分子国家重点实验室, 北京 100101)

摘要 结构域是蛋白质亚基结构中的紧密球状区域。结构域作为蛋白质结构中介于二级与三级结构之间的又一结构层次, 在蛋白质中起着独立的结构单位、功能单位与折叠单位的作用。在复杂蛋白质中, 结构域具有结构与功能组件与遗传单位的作用。结构域层次的研究将会促进蛋白质结构与功能关系、蛋白质折叠机制以及蛋白质设计的研究。

关键词 结构域, 蛋白质, 结构, 功能, 模块

天然的蛋白质具有特征的三维空间立体结构。Linderstrøm Long (1959 年) 及其同事最先认识到蛋白质具有不同的层次结构并引入一级、二级和三级结构描述这一现象。Wetlaufer (1973 年) 根据对蛋白质结构及折叠机制研究提出介于二级与三级结构之间存在另一结构层次即结构域。球状蛋白质折叠过程中, 结构域可能起着独立的折叠单位的作用。X 射线衍射技术已提供了大量蛋白质的三维空间立体结构资料, 结合蛋白质结构与功能关系的研究, 发现结构域不仅是结构单位, 而且是功能单位, 由此产生蛋白质结构与功能的组件模型 (modular model)。结构域这一概念在阐明蛋白质结构与功能之间关系及折叠机制时显得越来越重要, 本文旨在对这些领域中与结构域有关的一系列研究进展做一大致的概括分析。

1 结构域的定义及确定方法

结构域是构成蛋白质亚基的紧密球状区域, 为介于二级与三级结构之间的又一结构层次^[1]。现在使用的结构域概念具有三种不同的涵义, 即独立结构单位、独立的功能单位和独立的折叠单位。这三种涵义具有内在联系, 因为结构与功能是紧密相关的, 而独立结构必然具有内在稳定性, 因而在蛋白质折叠过程中

可能起着稳定折叠中间物的作用。

Wetlaufer 最先提出的考查蛋白质结构域的方法是视觉观察法, 他定义结构域为蛋白质亚基结构中明显分开的紧密球状结构区域, 并通过考查蛋白质的空间立体结构模型来划分这些区域。这种方法在判断结构域时必然掺杂大量的主观因素在内。较为客观的方法是利用现有 X 射线衍射技术得到的原子坐标数据的系统算法。近 20 年来发展了许多这种算法^[2], 如 Rossmann (1973 年) 的间距图方法 (distance map), Rippen (1978 年) 集束算法 (clustering algorithms), Wodark 和 Janin (1980 年) 发展的界面最小 (interface minimization) 方法, Zehfus 和 Rose (1986 年) 的计算区域紧密度 (coefficient of compactness) 方法等。所有这些算法都有一个共同的特点, 即通过利用原子坐标计算某种结构参数, 从而给出关于结构因素的一个级阶排列, 通过考查级阶结构, 不仅可以确定结构域, 而且可以确定所谓结构亚域, 这些结构亚域也在某种程度具备结构域所谓独立紧密折叠单位的特征, 它们也可能作为蛋白质折叠的成核中心。

^{*} 本课题部分由中国自然科学基金重点项目资助, 部分由攀登项目资助。

收稿日期: 1996-09-20, 修回日期: 1997-04-08

2 结构域与蛋白质结构与功能

2.1 结构域作为蛋白质结构与功能单位

结构域作为蛋白质结构中的紧密球状结构单元，广泛存在于球状蛋白中，其域间联结从IgG 的松散连接到木瓜蛋白酶、溶菌酶等小分子蛋白的紧密联结。典型的结构域间联结是一些双域蛋白如许多脱氢酶、激酶等的“双叶瓣”式结构，两结构域间形成所谓“刚体铰链”式结构，结合底物或辅酶后有域间的相互运动，有些在蛋白酶部分水解时常常产生独立的折叠单位。

结构域作为功能单位表现在不同的结构域具有特定的功能，如结合底物、催化反应、亚基间相互作用及活性调节等。各功能域间的排列构成组件排列 (modular model)。丙氨酸 tRNA 合成酶是这种结构的典型例子 (图 1)。另一种典型的结构是脱氢酶，作为双域蛋白，其一个结构域结合辅酶，另一结构域与催化有关，分别称为辅酶结合结构域与催化结构域。对目前已知的 200 多种脱氢酶的结构分析表明，它们的辅酶结合结构域的空间结构均具有相似性，且与多种以核苷酸磷酸为辅酶的激酶具有相似性，表明这种独立结构功能单位存在的普遍性。



图 1 丙氨酸 tRNA 合成酶的结构域结构

结构域作为结构功能单位，对于功能复杂的蛋白质的构成具有重要意义。在这些蛋白质中结构域往往以组件方式相互组合从而构成具有多种不同功能的分子。原核或真核生物的基因表达调控因子以及与信号传导有关的蛋白质因子往往以这种方式构成。表 1 列举了一些结构域模块^[3]。

结构域作为独立结构功能单位的极端情况是多功能酶。DNA 聚合酶 I 为这种情况的典

型例子 (图 2)。鼠肝 PFK2 作为一个典型多功能酶，分子中激酶结构域与酯酶结构域分别与果糖 6-磷酸-1-激酶及磷酸甘油酸变位酶相似，表明多功能酶的结构域作为独立的功能单位可能有其深刻的进化渊源。

表 1 脊椎动物蛋白中一些常见的结构模块

结构模块	特点
Gla	含多个 γ-羧基谷氨酸残基
G	类似表皮生长因子结构
K	“环饼”结构域
C	经常出现于补体蛋白中
F1、F2、F3	最早发现于纤维粘连蛋白
I	免疫球蛋白超家族结构域
L	出现于 LDL 受体及补体蛋白中
T	发现于血小板结合蛋白和补体蛋白备解素
N	发现于一些生长因子受体中
E	EF 手指型结构
LB	一些细胞表面蛋白的凝集素模块
SH2、SH3	2 类及 3 类 src 同源结构

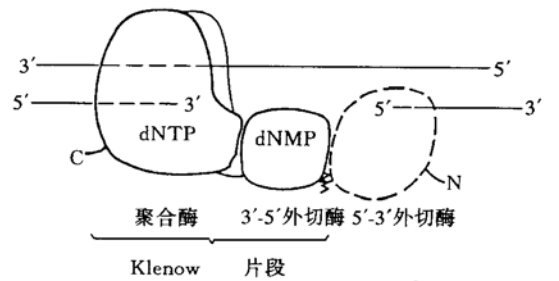


图 2 大肠杆菌 DNA 聚合酶的结构域结构示意图

2.2 结构域间运动、相互作用与蛋白质功能

结构域间的相互松散连接为域间运动提供了可能性，这种结构调整往往与蛋白质整体功能的行使相关。马肝乳酸脱氢酶的 apo 与 holo 酶的 X 射线衍射结构分析表明，该酶结合辅酶后经历了一个显著的构象变化 (图 3)。域间的相对运动导致乳酸脱氢酶由 apo 酶的封闭形式转换成 holo 酶的开放形式，有利于底物进入活性部位。这种作用方式有些类似于“活化”酶分子。类似的构象变化普遍存在于其他

脱氢酶、一些激酶、柠檬酸合成酶、阿拉伯糖结合蛋白等许多酶及蛋白质中。这种运动是酶的诱导契合模型的重要结构依据，其作用可能在于使酶蛋白的结构能够适合于其催化等功能的不同阶段。

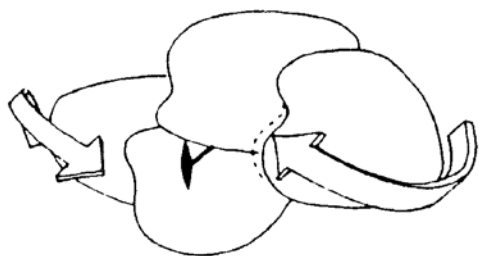


图3 马肝醇脱氢酶的结构域间相互运动

3 结构域与蛋白质折叠机制

3.1 蛋白质的折叠模式

蛋白质的二级结构堆砌具有一定的模式即折叠模式。对已知的蛋白质的结构研究表明它们可归属于数量有限的折叠模式，其数量预计在700至1500之间^[4]。尽管众多蛋白质的结构在原子水平上错综复杂，在结构域层次上的折叠模式却非常简单和有规律。Chothia (1995年)把蛋白质的折叠模式归纳为几大类：全 α 、全 β 、 α/β 及 $\alpha+\beta$ 。多个二级结构结合形成结构域，其性质一方面由多肽链的手性决定，另一方面也由于组装需要分子内部紧密堆积而决定的。对于小或中等规模的蛋白质，有限的折叠模式已能精确地描述其结构。对于大而复杂的蛋白质，其结构可进一步归结于亚结构——即结构域层次的折叠模式。对于结构域层次的折叠模式的深入研究有助于我们对越来越多新发现的复杂蛋白的结构与功能研究。

3.2 组件模型 (modular model)

蛋白质一级结构如何决定其高级结构，即所谓蛋白质折叠问题是现代分子生物学研究的热点。对于蛋白质如何形成其三维立体结构目前尚无定论。现在流行的融球态模型等其主要实验依据是蛋白质体外变性/复性实验结果，

且只适用于单域或简单蛋白。对于多域蛋白，现在倾向于所谓组件模型^[5]，即新生肽通过局部成核及组装逐渐形成空间结构，而成核中心或组装单位往往是结构域。

结构域作为折叠单位这已为通过酶解或化学裂解获得的蛋白质片段的体外变性/复性实验证实。基因工程方法的发展已使定向分离多肽片段成为可能。应用这种技术研究分离的结构域的独立折叠已给出一些成功的例子。如大肠杆菌丙氨酸 tRNA 合成酶、DNA 聚合酶 I、磷酸甘油酸激酶、三磷酸甘油醛脱氢酶等。这些结果显示蛋白质结构及折叠途径信息在多肽链中也是成段排列的。这为进一步从蛋白质一级结构预测其三级结构提供了新的思路。

4 结构域与蛋白质分子进化

4.1 蛋白质的分子进化

生物进化过程的本质是分子进化，蛋白质的进化在这一过程中起了中心作用。理论计算表明由个别氨基酸的随意选择决定蛋白质的演变过程是不可能的。类似的结构域出现在不同的蛋白质分子中使人们有理由相信它们来源于共同的祖先基因。通过编码结构单位的基因片段的融合的蛋白质分子进化，是阐明蛋白质多样性产生的一种可能而有吸引力的假设。

结构相似性比较研究表明，结构域可能作为一种遗传单位在蛋白质分子进化过程中起着中心作用。相似的结构域可能出现在同一分子中，亦可能出现于不同的分子中，表明这种进化过程可能是通过基因复制，融合或缺失完成。需要阐明的是这种进化过程的机制。真核生物基因的内含子/外显子结构为实现这一过程提供一种合理的模式，而深入研究外显子与结构域的内在联系，将有助于进一步阐明蛋白质的进化机制。

4.2 结构域与真核基因内含子/外显子的结构

自从1977年发现基因的内含子/外显子结构以来，其来源与作用一直处于争论之中。那种认为外显子编码完整结构域的观点已成为广为传播的误解。目前发现多数编码结构域的基

因本身亦为内含子所分隔。一种进化观点认为外显子在进化早期的原始生物中可能是编码蛋白元件的微基因, 在以后的进化过程中这些微基因作为结构元件组合成编码复杂功能蛋白的现代基因——即所谓原生内含子 (intron early) 假说。支持这一理论的依据主要是磷酸三糖异构酶的基因结构研究: 磷酸三糖异构酶的外显子与蛋白质结构单元的关系极不同于按内含子随机分布理论所预测的结果。与此相反, 后生内含子 (intron late) 理论认为原始的基因是连续的, 内含子可能是由一些基因片段插入而形成的。转座子的发现为这种观点提供了间接证据。目前更倾向于后生内含子假说^[6]。

尽管如此, 内含子在蛋白质分子进化中的作用仍难以否定。内含子的存在可能极大程度上促进了结构单元特别是以结构域为单位的元件重组 (module shuffling), 从而加速了生物进化过程。这也可能是后生动物兴旺发展的内在原因。

5 结 语

结构域作为介于二级与三级结构之间的蛋白质结构层次, 在阐明蛋白质结构与功能关系中起着十分重要的作用。对于结构域的组合规律和形成机制的研究有助于我们对以下问题的深入研究:

a. 蛋白质结构域作为结构与功能单位, 广泛存在于参与基因转录、翻译调控以及与信号传导有关的蛋白质因子等结构功能复杂的蛋白质中。对于复杂蛋白, 目前一种比较成功的研究战略是解剖 (dissect) 完整蛋白, 将其按结构单元——主要是结构域划分为相对独立的模块 (module), 研究这些模块的结构与功能及它们之间的相互作用。随着新的蛋白质的不断发现, 这种研究的前景可观。而对结构域层次的结构、功能、折叠模式及相互作用的深入研究必将为这一研究战略提供理论依据。

b. 如何从蛋白质的一级结构预测其三级结构一直是蛋白质研究的目标。结构预测的一种尝试是从已知的蛋白质结构出发, 希望发现

蛋白质结构组合的内在规律。利用这种经验方法, 预测蛋白质的结构特别是二级结构的努力已给出相当满意的结果。目前预测二级结构的方法至少有 15 到 20 种, 成功率可达 60% ~ 75%, 同源性比较方法也为预测新发现的蛋白质的空间结构提供有益的信息。结构域作为蛋白质的结构单位, 在了解蛋白质结构形成和组合规律方面具有重要意义。拓扑学研究表明, 结构域在蛋白质中的分布具有某种规律性。对其规律的深入研究必将为预测蛋白质的空间结构提供有益的帮助。

c. 蛋白质工程是本世纪 80 年代初诞生的一个新兴领域, 其主要内容是通过有控制的基因修饰和基因改造, 产生符合人们意愿的蛋白质分子。定点突变技术为这一研究提供了强有力的工具, 目前已发展得相当成熟。另一更为新颖的战略是利用蛋白质组合规律及结构与功能关系的研究结果, 设计新的蛋白质分子。鉴于结构域是蛋白质的独立结构功能单位, 本实验室于 1986 年提出通过结构域作为结构与功能元件组装的方式构建功能全新的蛋白质分子, 这种方式有些类似于建筑工程中的预制件组装方式。通过不同的结构与功能元件的不同组合, 可以形成数目众多, 功能各异的蛋白质分子, 从而为蛋白质分子设计提供更多的可能性。

参 考 文 献

- 1 Richardson J S. The anatomy and taxonomy of protein structure. *Adv Protein Chem*, 1981, **34**: 168~ 340
- 2 Rossmann M G, Argos P. Protein folding. *Annu Rev Biochem*, 1981, **50**: 497~ 532
- 3 Camphell I, Bowning A K. Building protein structure and function from modular units. *Trend Biotech*, 1994, **12**: 168 ~ 172
- 4 Chothia C, Finkelstein A V. The classification and origins of protein folding patterns. *Annu Rev Biochem*, 1990, **59**: 1007~ 1039
- 5 Jaenicke R. Protein folding: local structures, domains, subunits and assemblies. *Biochemistry*, 1991, **30**: 3147~ 3160
- 6 Doolittle R F. The multiplicity of domains in proteins. *Annu Rev Biochem*, 1995, **64**: 287~ 314

Structure Domains in Protein Structure and Function. LIAN Dejun, XU Genjun (Shanghai

Institute of Biochemistry, The Chinese Academy of Sciences, Shanghai 200031, China; National Laboratory of Biomacromolecules, Institute of Biophysics, The Chinese Academy of Sciences, Beijing 100101, China).

Abstract A protein with long polypeptide chain always folds into more than one distinct, compact globular regions that have been designated as domains. Being one of the structural hierarchy, domain is a structural level situating in between secondary and tertiary structure. It is a basic unit to compose three-dimensional structure

of protein, to execute the biological function as well as to fold a functional protein. As structural modules, domains may have advantage to compose complex proteins during the process of protein evolution by insertion, deletion and excising. Studies on the structure and interaction of domains can throw light on the study of the relationship between protein structure and function and the design of novel proteins.

Key word domains, protein, structure, function, module

用于转基因的阳离子脂质体

王 颀 林其谁

(中国科学院上海生物化学研究所, 分子生物学国家重点实验室, 上海 200031)

摘要 通过直接导入外源基因来治疗人类疾病的方法需要一种有效、安全并且可重复进行的载体, 阳离子脂质体是基本满足这些条件的有限几种载体中的一员。目前已经有十几种阳离子脂质体。这些脂质体通过外周的电荷与 DNA 相结合, 静电吸引形成复合物在与细胞膜相互作用后, 通过细胞的内吞或融合作用使复合物进入细胞内, 从而将外源基因导入细胞, 这种 DNA 传递技术的有效性和安全性已经确立。二例利用阳离子脂质体的人体基因治疗临床试验也已开始实施, 将来会有更多的临床试验得到开展, 阳离子脂质体在基因治疗领域有较好的前景。

关键词 阳离子脂质体, DNA, 传递, 表达, 基因治疗, 转染

基因治疗是通过对一个需要长期表达基因的设计, 将其植回宿主细胞并成功地得到表达。发展安全合适、高表达的 DNA 介导传递技术是基因治疗研究极其重要的方面, 而阳离子脂质体转染技术的运用是一次飞跃。

1 阳离子脂质体/脂质形式

阳离子脂质分子含有四种结构域形式: 带正电荷的极性头; 能改变长度的隔离区; 连接键和疏水的锚着区。除一种脂质含脒基外, 其余阳离子脂质的极性头都包含胺类基团, 从简单的氨基到被甲基或羟乙基团取代的季铵盐。

阳离子脂质的极性头起着脂质体与 DNA、脂质体-DNA 复合物与细胞膜或细胞内其他组分的相互结合的作用。在阳离子胆固醇衍生物中, 带有叔胺基团的阳离子胆固醇化合物比季铵盐化合物有更佳的转染活力, 并且毒性小得多^[1]。带有多价极性头基团的阳离子脂质形成脂质体时比那些单价阳离子脂质排列得更紧密, 这可能与转染能力较高有关。间隔垫充区的链长和键长也影响转染活力, 带有长间隔垫充区的阳离子脂质体能显著增强与粘膜表面的相互作用。

收稿日期: 1996-09-20, 修回日期: 1997-01-20