

蛋白质中频繁发生的超二级结构模式

孙之荣

(清华大学生物科学与技术系, 北京 100084)

关键词 连接短肽、超二级结构、蛋白质

近年来, 高分辨 X 射线衍射的蛋白构象分析表明, 超二级结构的构象类型是有限的^[1,2]. 本文对 240 个分辨率在 0.25nm 以上的较高精度蛋白结构中的超二级结构进行了系统分析, 识别出频繁发生的超二级结构模式. 最普遍发生的共 11 种超二级结构, 依其在蛋白结构中的作用, 它们属于 4 类. 研究表明超二级结构模式与 α 螺旋和 β 折叠之间的连接短肽 (Loop) 的长度、与连接短肽残基的构象、与连接短肽所连接的二级结构元件的类型有关; 同时超二级结构的氢键模式、疏水性质等特征也都起着重要的作用. 本文的结果为超二级结构模型和预测的进一步研究建立了基础, 对在更高层次上揭示蛋白质的结构规律有一定意义, 对开拓蛋白质工程中的应用提供了参考.

1 材料和方法

1.1 蛋白数据库

在本文中, 分析了 326 个蛋白结构, 首先去除了分辨率在 0.25nm 以下的蛋白, 然后又运用 Procheck 程序^[3]去除了质量不高的 X 射线衍射结构的蛋白, 最后剩下 240 个高结构精度蛋白样本库. 蛋白的坐标数据从 Brookhaven 数据库得到^[4]. 本文又运用结构比较程序 Comparer^[5]对这 240 个蛋白三维结构进行比较, 确定共包含了 56 个同源蛋白家族. 蛋白结构显示软件在 IRIS 3 图形工作站上使用的是 Sybel 软件 (1992) 和 Setor 软件 (1993).

1.2 蛋白结构主链构象的分析

蛋白结构的千差万别与蛋白质主链上每一个氨基酸残基的构象有关. 本文确定从构象分析入手. 首先运用 Kabsch 和 Sander 的定义^[6], 辨别出螺旋 (H) 和折叠 (E) 结构. 然后遵照 Efimov^[7], Topham 等^[8]的方法, Topham 等从 84 个高分辨率晶体蛋白中分析了 12318 个残基的构象, 其中 5712 个残基构象落在由 Kabsch 和 Sander 的定义而判断的规则二级结构之外. 这 5712 个残基的扭角 (ϕ , ψ) 相对集中分布在 Ramachandran 图上的 7 个区域: a, b, e, g, l, p, t.

本文进一步简化 7 个构象状态为 5 个构象状态, 因为在 Ramachandran 图上, 区域 b 可以包括 p, 所以区域 b, p 合并为 b'; 同样, g, l 合并为 l', 这样 5 个构象区域为 a, b', e, l', t. 在本文研究中, 我们计算了总数为 38364 个残基的扭角, 确定了 240 个蛋白肽链上每一个残基的构象状态类型. 然后对每一蛋白都构建一个特征文件, 在特征文件中, 每一残基的构象状态分别用 H, E, a, b', e, l', t 来描述. 它们形成初始蛋白构象样本库.

1.3 超二级结构构象数据库的构建及其统计分析

本文建立了超二级结构数据库, 每一个超二级结构单元都是由连接多肽连接两个规则二

级结构元件组成。本文着重研究了 1—5 个残基长的短连接肽, 二级结构元件最少由 3 个连续的残基组成。例如 HHHlbbHHH 表示由三个构象为 lbb 的残基连接两个 α 螺旋形成的超二级结构, EEEllEEE 表示两个 β 折叠链被两个构象为 ll 的残基连接组成的超二级结构等。

表 1 各种类型的超二级结构发生频率的统计结果

H-loop-H 连接肽 频率 ^{a)}	H-loop-E 连接肽 频率	E-loop-H 连接肽 频率	E-loop-E 连接肽 频率
b 47 t 32 e 6 l 5	l 9 e 7		a 10
bb 30 tb 9 lb 6	ba 9	ab 18	ll 87 aa 28 ea 28 ba 14
lbb 15 bbb 8	lba 15	abb 11 eab 5	bl 16 abl 12 aaa 9 lla 9 abb 6 bae 6 aal 6
albb 22 blbb 6	alba 28 albb 10 bbaa 5	abbb 8 abbb 7 lbbb 7 allb 6	aal 28 baab 16 aalb 9 abaa 5 bblb 5 tllb 6
aalbb 16 taabb 10 llbbb 6	albab 5		baalb 35 bblbb 9 aabab 7

a) 频率单位为次/240 样本。

表 2 超二级结构单元 E-lla-E 序列模式

构象	序列模式	蛋白质	连接短肽
bEEE lla	EEEE LTNY LDS QYFG	4cms	11 13
EEEE lla	EEEE LTNY LNS QYYG	2ren	12 14
EEEE lla	EEEE LENY LDT EYFG	4pep	9 11
EEEE lla	EEEE LENY LDT EYFG	5pep	9 11
EEEE lla	EEEE EYSI NNA SLCS	2alp	11 13
bEEE lla	EEEE TMVI DGK MVKV	2gd1	301 303

类型是有限的。它们并非是不可能构象的随机组合。对于长度为 1—5 个残基的短连接肽, 这种可能的随机组合是 15620, 50 远远小于 15620。

这些频繁发生的超二级结构的类型模式与连接短肽所连接的二级结构元件的类型、与连

首先根据连接多肽所连接的二级结构元件的情况, 分成 4 组在 240 个蛋白构象样本库中搜索超二级结构, 这 4 组是: α -loop- α , α -loop- β , β -loop- α , β -loop- β 。然后再按照连接多肽的长度细分成更小的组来统计分析。

表 1 所示是超二级结构的统计分析结果。这些超二级结构的发生频率都在 5 次以上。

1.4 超二级结构的序列模式特征的搜索研究

本文研制了 PATTERNI 程序用以研究超二级结构的序列模式。表 1 中的全部超二级结构对应的序列模式被导出。从序列模式可以清楚地看到超二级结构的残基构象和序列的关系, 从而可以推导出每一构象位置各氨基酸的概率。本文对所有超二级结构类型中每一构象位置各氨基酸的概率进行了统计分析。表 2 所示是超二级结构单元 E-lla-E 序列模式的一个例子。

2 结果和讨论

2.1 超二级结构分类统计结果的分析

表 1 给出了在 240 个蛋白样本中发生次数多于 5 次的超二级结构的类型, 共有 50 种。这一结果证实了超二级结构的构象

接短肽的长度、与连接短肽残基的构象有关。它们都是蛋白质肽链依一定规则折叠的结果。

2.2 超二级结构的结构比较, 图形叠合研究

对于 E-loop-E 类型的超二级结构, 本文利用 IRIS 2 计算机图型工作站, 按照 Sibanda 等^[9]的方法, 根据连接短肽的长度和氢键模式, 将它们模型化。发现其中一些 E-loop-E 类型的超二级结构氢键模式和结构模型相同: EEE-aa-EEE (2:2), EEE-aal-EEE (3:3), EE-baa-b-EE (2:4), EE-Eaalb-EE (3:5), EE-baalb-EE (3:5), 它们分别属于同一个 β 发夹模型 (2:2) 和 (3:3)。

在图形工作站上将它们对应的肽链叠合比较, 发现叠合的很好。进一步我们计算了这些肽链叠合后对应的 C_{α} 原子的 RMSD, 全部 RMSD 数值都小于 0.2nm。说明它们可以分别合并为 EEE-aa-EEE, EEE-aal-EEE。

在超二级结构库中, 我们注意到, 一些超二级结构的连接短肽紧邻“H”的残基构象是“a”, 紧邻“E”是“b”。对于所有这些类型的超二级结构, 都在分子图形工作站上仔细分析, 进行叠合研究。一些类型的超二级结构可以被合并。

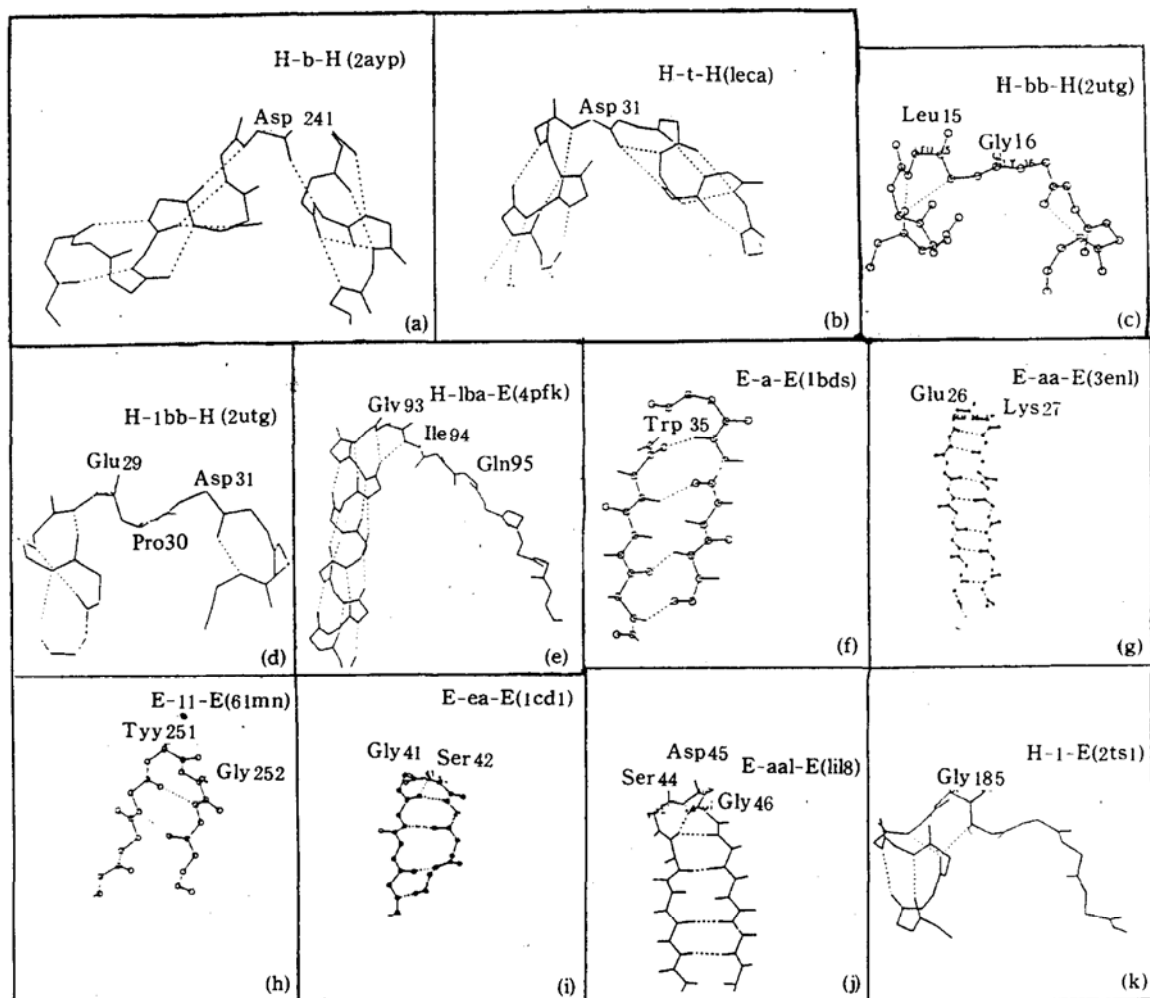


图 1 11 种频繁发生的超二级结构的构象及其氢键模式图

表 3. 频繁出现的超二级结构发生频率的统计结果

H-loop-H 连接肽 频率 ^{a)}	H-loop-E 连接肽 频率	E-loop-H 连接肽 频率	E-loop-E 连接肽 频率
b 68	l 74		a 10
t 48	e 7		
l 20			
e 6			
bb 41	ba 9	ab 18	ll 87
tb 9			aa 44
lb 6			ea 28
lbb 55	lba 48	abb 11	aal 47
bbb 8		eab 5	bll 16
			abl 12
			aaa 9
			lla 6
blbb 6	bbaa 5	abbb 8	aaal 28
		lbbb 7	abaa 5
		allb 6	
taabb 10			
llbbb 6			

a) 频率单位为次/240 样本。

角(corner), α 螺旋经此链的走向弯曲 90° 到另一层次. H-bb-H 和 H-lbb-H 是 α 发夹(haipin); 通过连接短肽 bb 和 lbb, 形成两个近似的反平行螺旋: E-l-E, E-aa-E, E-ea-E, E-aal-E 则是 β 发夹, 肽链在此形成两条反平行 β 折叠. H-l-E 和 H-lba-E 则是一拱桥结构(arch), 不同空间层次上的两种不同的二级结构被连接短肽所连接.

本文所得到的超二级结构模式的结果对于在更高结构层次上认识蛋白质的结构规律, 对于结构模型化及结构预测都有一定意义.

致谢 衷心感谢英国皇家学会院士、Birkbeck 学院 Tom Blundell 教授的帮助和指导. 感谢国家科学技术委员会生物工程中心的支持和帮助.

参 考 文 献

- [1] Efimov, A. V., *Current Opinion in Structural Biology*, 1993, 3: 379—384.
- [2] Thornton, J. M. et al., *Bioessays*, 1988, 8: 63—70.
- [3] MacArthur, M. W. et al., *J. Appl. Cryst.*, 1993, 26: 283—291.
- [4] Bernstein, F. C. et al., *J. Mol. Biol.*, 1977, 112: 535—542.
- [5] Sali, A. et al., *J. Mol. Biol.*, 1990, 212: 403—428.
- [6] Kabsch, W., Sander, C., *Biopolymers*, 1983, 22: 2577—2637.
- [7] Efimov, A. V., *J. Mol. Biol. (Moscow)*, 1986, 20: 250—260.
- [8] Topham, C. M. et al., *J. Mol. Biol.*, 1993, 229: 194—220.
- [9] Sibanda, B. L. et al., *J. Mol. Biol.*, 1989, 206: 759—777.

例如:

HHH-lbb-HHH, HHH-t-HHH,
HH-albb-HHH, HH-at-HHH,
H-aalbb-HHH, HHH-ta-HH.

在一些类型的超二级结构被合并后, 分类被简化, 表1变成表3. 其中最频繁出现(频率高于 25 次)的超二级结构有 11 种类型. 图1是这 11 种类型的超二级结构的构象及氢键模式图.

2.3 频繁发生的超二级结构的结构作用

根据超二级结构中的连接多肽在蛋白质三级结构中的作用, 11 种频繁发生的超二级结构组可以分成 4 种模式来讨论. H-b-H 和 H-t-H 是 α - α 弯