

“锌指”结构:蛋白质和 DNA 相互作用的一种模式

胡红雨 鲁子贤

(中国科学院上海生物化学研究所,上海 200031)

提 要

锌指结构是 DNA 结合蛋白的基本模型之一,它广泛存在于真核细胞与基因调控有关的蛋白质中。文章综述了锌指蛋白的发现、存在、结构模型、与 DNA 结合特点和生物功能,以及近期的研究重点。

关键词 锌指结构, DNA 结合蛋白, 转录因子 IIIA

蛋白质与核酸相互作用是基因表达和调控的基本环节。与基因调节有关蛋白质的结构和功能研究已成为真核细胞分子生物学的重要内容。现已发现三种 DNA 结合蛋白的基本结构模型:“螺旋-转折-螺旋”(helix-turn-helix),“亮氨酸拉链”(leucine zipping)和“锌指”(zinc finger)结构^[1]。其中,锌指蛋白是最近几年才发现并很快受到广泛重视的一类金属蛋白质。J. Burg^[2]认为锌指蛋白是真核细胞中最普遍的一类 DNA 结合蛋白,且在生物发育过程中,可能是调节基因表达的一种基本模式。

1 锌指的发现及存在

1983年, Hanas 等^[3]证实爪蟾卵母细胞转录因子 IIIA(transcription factor IIIA, TFIIA)与 DNA 的位点专一性结合需要 Zn^{2+} 离子参与。接着 Miller 等^[4,5]分离并且测定了 TFIIA 的 cDNA 序列,推测出蛋白质的一级结构,发现其氨基酸序列中含有九个依次排列的重复单位。每个单位约含 30 个氨基酸残基,通常为如下序列:

(Tyr, Phe)-X-Cys-X_{2,4}-Cys-X₃-Phe-X₅-Leu-X₂-His-X_{3,4}-His-X₂₋₆

(其中 X 为相对可变的氨基酸)

TFIIA-RNA 复合物经蛋白酶水解得到一

系列相差 3—4kD 的多肽片段^[6],证实这种重复单位的存在。基因分析结果显示前六个结构单位由分开的外显子编码^[7],很可能这些重复单位是通过基因重复形成的。由于每个 TFIIA 蛋白分子含有 7—10 个 Zn^{2+} 离子,并且已知 Cys, His 是锌蛋白中 Zn 的最常见配体,很自然推想到每段结构单位络合一个 Zn^{2+} 。 Zn^{2+} 和四个 Cys 或 His 配位形成四面体结构,称为“锌指”结构。EXAFS(延伸 X 射线吸收精细结构)分析也表明,TFIIA 中的 Zn 确实与 Cys, His 中的 S 和 N 原子配位^[8]。

随着研究的深入,借助计算机同源序列分析,在越来越多的核酸结合蛋白中发现类似 TFIIA 中 [Cys(His)-X_{2,3,4}-Cys(His)]_n 的结构同源单位,它们具有以下特征:

(a) 蛋白质中含有一个或多个类似 TFIIA 中的同源结构单位,每个单位通过 Cys 或 His 与 Zn^{2+} 配位,形成相对独立的结构域。这些结构域称为锌指结构域(zinc finger domain)。

(b) 两个 [Cys(His)-X_{2,3,4}-Cys(His)] 连接部分约 12 个残基形成完整的结构,或为扭曲的 β 折叠,也可能为 α 螺旋。此肽段形成的突起(锌指)与 DNA 相互作用。

(c) 每个锌指结构域与 DNA 螺旋作用区域为三个碱基对左右。

(d) 其它基因调节蛋白(见表 1)很可能存在与 TFIIIA 同源的序列和相似的结构和功能。

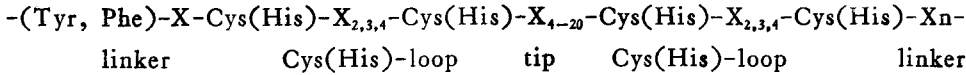
表 1 具有锌指结构域的调控蛋白

蛋白质及来源	数目	结构域氨基酸序列	推测的某些功能
转录因子类			
TFIIIA (爪蟾卵母细胞)	9	(Y,F)XCX _{2,4} CX ₁₂ HX _{3,4} HX ₂₋₆	专一识别 DNA, 5SRNA 转录必需因子
ADR1 (酵母)	2		顺式调节因子, 结合 ADH2 基因上游点
GAL4 (酵母)	1	CX ₂ CX ₆ CX ₆ CX ₂ CX ₆ C	转录激活因子, 可能为 Zn ₂ Cys ₆ 双核结构
Sp1 (哺乳动物细胞)	3	CX ₂ CX ₁₂ HX ₂ HX ₂ H	结合 GC 盒启动子, 激活 mRNA 合成
“gag”类基因产物			
Kruppel (果蝇)	4	CX ₂ CX ₁₂ HX ₃ H	与生物发育中细胞分裂有关
Serendipity (果蝇)	6 或 7	CX ₂ CX ₈ HX ₃ H	结合 RNA
Hunchback (果蝇)	4+2		
k1 和 k2 (小白鼠)	7 和 9		
Xfin (果蝇)	37		
核酸结合蛋白			
基因 32 蛋白(噬菌体 T4)	1	CX ₂ HX ₂ CX ₂ C	结合单链 DNA
R-M LVp10 和 R-ASVp12	3	CX ₂ CX ₂ HX ₂ C	“gag”类基因产物,
腺病毒 E1A 大蛋白		(CCHC box)	与病毒逆转录有关
甾体激素受体			
糖皮质激素受体(人, 鼠)	1	CX ₂ CX ₂ HX ₂ CX ₂ C---	结合 DNA, 调节基因表达,
盐皮质激素受体(人)	1		可能含锌指结构
孕酮受体(鼠, 鸡)	1		
维生素 D3 受体(鸡)	1	CX ₂ CX ₂ HX ₂ CX ₂ C---	
雌激素受体(鸡)	1		
修复蛋白			
uvrA (大肠杆菌)	2	CX ₂ CX _{11,22} CX ₂ C	细胞增殖, 分化的调节
多聚 (ADP-核糖)聚合酶 (哺乳动物)	2	CX ₂ CX _{25,30} HX ₂ HX ₂ C	DNA 修补
			修饰核蛋白
癌基因产物			
V-erb-A (人胎盘)	2	CX ₂ CX ₁₃ CX ₂ C	低分子量核酸结合蛋白
C-erb-A (人胎盘)			原癌基因产物
			编码甲状腺激素受体
性决定因子			
ZFY (人)	13	CX ₂ CX ₁₂ HX _{3,4} H	编码与性别决定有关的蛋白质
ZFX (人)	13		

上表列举了一些可能含有锌指结构域的蛋白质。它们都含有四个以上 Cys 或 His 相间组成的同源序列, 且多数已被证实与 Zn 结合。很有意义的是, 从酵母到人的每个真核细胞基因库中都能找到与 Kruppel-基因探针杂交的序列, 可是 *E. coli* 基因库却不能与之杂交, 表明锌指序列可能仅普遍存在于真核细胞基因库中^[9]。这对于研究锌指蛋白的分子进化很有意义。

2 锌指的结构模型

从表 1 中可看出, 含有锌指结构域的蛋白质中一般都含有 [Cys(His)-X_{2,3,4}-Cys(His)-] 的肽段, 长度为 14—48 个残基不等, 通常为 30 个氨基酸残基左右; 其序列可示意如下:



在 TFIIIA 中,九个直线排列的 30 个氨基酸残基左右的肽段通过中间约 10 个相当保守的残基连接起来,称为 H-C linker,一般为 His-Thr-Gly-Glu-Lys-Pro-(Tyr, Phe)-X-Cys. 这里 His 和 Cys 分别是前后两个锌指结构域的金属配体. 一种模型认为 His-Thr-Gly-Glu 四肽形成 II 型 β 转折,其中 Gly, Pro 是两个极易形成转折的残基,而 Thr 的羟基与前一结构域羧基形成氢键; Glu 的羧基和 His 的氢键以及 Lys 的盐键,共同稳定 β 转折的构象.

Cys-Cys-loop 与已知晶体结构的天冬氨酸转羧甲基酶 (aspartate transcarbamylase) 和还原氧化还原蛋白 (rubredoxin) 中部分序列同源,推测 Cys-Cys-loop 可能含有两段反平行的 β 折叠链; 而 His-His-loop 则与嗜热蛋白酶 (thermolysin) 和蚯蚓血红蛋白 (hemerythrin) 同源,推测含有螺旋结构. 将这推测的两部分亚结构合并,便得到 TFIIIA 中的一个锌指结构模型 (如图 1)^[10]. 同这两类蛋白质中一样,

并且得到合成的模型肽和 2D-NMR 结构解析的证实^[11].

金属离子 (主要为 Zn^{2+}) 是保持锌指结构域和核酸结合功能必不可少的. 通过基因工程方法合成并研究 30 个残基肽的同源结构,发现在没有 Zn^{2+} 或 Co^{2+} 存在时,这段肽为无规卷曲,而 Zn^{2+} 、 Co^{2+} 可促使其折叠成稳定的锌指结构,且降低了它对胰蛋白酶水解的敏感性. Zn^{2+} 结合于肽链比 Co^{2+} 更牢固,可能是锌指蛋白与 DNA 特异结合而行使 Zn^{2+} 的生物功能的重要因素.

一般来说,少于 50 个残基的多肽不易折叠成稳定的结构域. 金属离子与肽的氨基酸侧链基团配位络合,很可能起到促进肽链折叠并稳定结构域的作用. 其它例子还可见于金属硫蛋白 (metallothionein) 的 α, β 结构域^[13].

3 锌指蛋白与核酸相互作用

通过足印实验 (footprinting), Miller 等^[10] 最初认为 TFIIIA 中每个锌指结构结合 DNA 双螺旋的半圈,即 5.0—5.5 碱基对. 然而,很多锌指蛋白,如 Sp 1、Sw 15 的每个锌指结构却仅识别三个碱基对. Sp 1 含有三个锌指结构域,能识别并结合于 9—10 个碱基对的 DNA 片段^[14]. 另外,一些推测的锌指结构,由于指间连接处 (linker) 长度有限,如 TF III A, 很难构建其它结构模型,使每个结构域识别并结合于大于 3 个碱基对. Fairall 等^[15] 设计了 TF III A

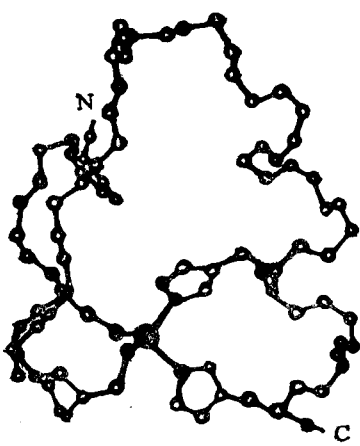


图 1 锌指结构模型(肽链走向)
 ● Zn; ● 配位原子 S 或 N; ○ C; N: N 端; C: C 端;
 Cys-Cys-loop (左): 反平行 β 折叠;
 His-His-loop (右): α 螺旋;
 Tip (中): β 转折

Zn^{2+} 与 2 个 Cys、2 个 His 的四个配位原子以四面体方式配位^[10]. NMR 技术研究表明 Xfin 中第 31 位结构域与上述预测的模型相类似^[11],

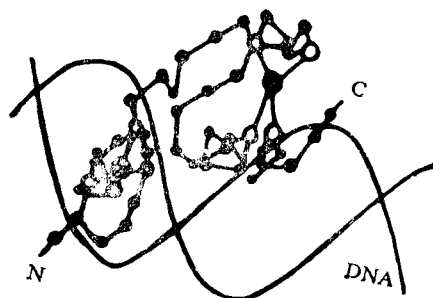


图 2 锌指结合于双链 DNA 模型
 ● Zn ○ 配位氨基酸 Cys 或 His ● 其他氨基酸
 N: N 端, C: C 端

与 5S RNA 基因复合物的两种结构模型。一种模型认为蛋白质沿着双螺旋大沟缠绕 DNA；另一种模型则主张蛋白质的一些氨基酸残基位于 DNA 螺旋的表面，而一些构成锌指的残基却镶嵌在 DNA 双螺旋的大沟中。

锌指结构域很可能嵌合于双链 DNA 的大沟而与 DNA 相互作用。如果将 5S RNA 基因的鸟嘌呤上 N₇ 甲基化，则阻止了 TF III A 结合于这段基因^[16]。基于单个锌指和 H-C linker 及双链 DNA 的结构特征，J. Berg 构建了如图 2 所示的模型^[17]。锌指的部分螺旋肽嵌于 DNA 的大沟，而 β 折叠则偏离 DNA 螺旋轴，与 DNA 螺旋表面接触。根据此模型可推测，锌指指端的氨基酸残基特异识别 DNA 大沟中的碱基对。

用羟自由基足印法研究 TF III A-DNA 复合物，发现不是每个锌指结构域与 DNA 作用的模式都完全相同^[18]。TF III A 含有九段锌指同源序列，其中五、六和六、七段间的连接处序列较短，分别只有四个和五个残基（一般为七或八个残基）^[4]。这样使得第六段锌指结构很难嵌于 DNA 的大沟。因此，1—5 结构域和 7—9 结构域很可能成为两组，而第 6 段正好位于其转弯处，导致 DNA 轴的弯曲（TF III A 诱导的 DNA 弯曲）^[19]。上述 TF III A-DNA 复合物模型得到了 DNase I 足印法实验的有力支持。若在第九个锌指内发生基因突变，则由七、八、九组成的多结构域单位就不能有效的结合于这些 DNA 区域，使其容易被 DNase I 水解，而 1—5 结构域仍然与 DNA 结合，并阻断 DNase I 的进一步水解^[18]。

锌指蛋白是一类与核酸相互作用的金属蛋白，是目前研究基因调控的重要内容之一。短短五、六年中已确定上百种蛋白质及其基因的同源序列，而且在基因结构，锌指蛋白质结构，中心离子 Zn^{2+} 的配位性质，DNA 结合及基因调控作用，化学合成和基因工程方法合成

等领域进行了广泛的研究。下面是一些有待今后进一步解决的基本问题：

- a. 继续寻找序列同源、功能相似的基因或蛋白质；
- b. 确证锌指结构模型和锌指蛋白与核酸相互作用的模型；
- c. Zn^{2+} 的生物配位化学（结构化学、热力学和动力学）；
- d. 原核生物是否存在与锌指蛋白结构和功能相类似的蛋白质；
- e. Zn^{2+} 除了结构上作用外，是否还有别的调节功能；
- f. 从分子进化观点，为什么选择了 Zn^{2+} ，而且又是多个结构域线性排列？

参 考 文 献

- 1 Struhl K. *Trends Biol Sci*, 1989; 14(4): 137
- 2 Burg J M. *Annu Rev Biophys Biophys Chem*, 1990; 19: 405
- 3 Hanas J S, Hazuda D J *et al.* *J Biol Chem*, 1983; 258: 14120
- 4 Miller J, McLachlan A D, Klug A. *EMBO J*, 1985; 4: 1609
- 5 Brovn R S, Sander C, *et al.* *FEBS Lett*, 1985; 186: 271
- 6 Frankel A D, Burg J M *et al.* *Proc Natl Acad Sci USA*, 1987; 84: 4841
- 7 Tso J Y, Van der Burg D J *et al.* *Nucleic Acids Res*, 1986; 14: 2187
- 8 Diakun G, Frairall L, Klug A. *Nature*, 1986; 324: 698
- 9 A Ruiz i Altaba, H Perry-O'keefe *et al.* *EMBO J*, 1987; 6: 3065
- 10 Berg J M. *Proc Natl Acad Sci USA*, 1988; 85: 99
- 11 Lee M S, Gippert G P *et al.* *Science*, 1989; 245: 635
- 12 South T L, Kim B *et al.* *J Am Chem Soc*, 1989; 111: 395
- 13 Furey W F. *Science*, 1986; 231(4739): 704
- 14 Kadonaga J T, Carner K R *et al.* *Cell*, 1987; 51: 1079
- 15 Fairall L, Rhodes D, Klug A. *J Mol Biol*, 1986; 192: 577
- 16 Gidoni D, Dynan W S *et al.* *Nature*, 1984; 312: 409
- 17 Smith D R, Jackson I J *et al.* *Cell*, 1984; 37: 645
- 18 Vrana K E, Churchill M E A *et al.* *Mol Cell Biol*, 1988; 8: 1684
- 19 Schroth G P, Cook G R *et al.* *Nature*, 1989; 340: 467