

dsDNA-蛋白质相互识别的结构模型

张 今 张红婴

(吉林大学酶工程实验室, 长春)

提 要

核酸同蛋白质之间的相互识别是生命的核心过程之一。核酸同蛋白质相互识别取决于核酸的状态。本文在不考虑核酸构象变化的情况下, 讨论了 dsDNA-蛋白质相互识别。根据 dsDNA-蛋白质相互识别的基本结构部件, 基本因素, 识别信号, 探讨了“一般识别”和“特殊识别”的结构模型和“识别词典”问题。

关于核酸同蛋白质相互识别, 1959 年 Zubay 和 Doty^[1] 提出, α 螺旋可能装进 B-DNA 的大槽。1972 年, Adler 等^[2] 报道 α 螺旋上的侧链可以同大槽中的碱基实现特异的相互识别。他们提出 Lac 阻遏蛋白的 α 螺旋同操纵基因 DNA 相互识别的模型。1974 年, Carter 和 Krah^[3] 提出蛋白质-RNA 相互作用模型。他们提出两个反平行的 β 折迭股的缠绕, 维度和对称性适于两股反平行 β 折迭同 RNA 相互识别。1977 年, Church 等^[4] 提出一对反平行股可以装进 B-DNA 的小槽。1981 年, Steitz 等^[5] 研究了 *E. coli* 中三个蛋白质, 即 CAP、Lac 阻遏蛋白和分解酶 (Resolvase, 催化特异位点结合的一种酶, 它是 *E. coli* 中易位单元所产生的两个酶之一, 分子量为 20,000) 同 dsDNA 特异相互识别。他们提出 CAP 同 dsDNA 相互识别的结构模型。总之, 核酸与蛋白质相互识别基本上有三种方式: (1) 蛋白质同暴露于 DNA 大槽中碱基对相互识别; (2) 蛋白质同暴露于小槽中碱基对互补; (3) 蛋白质识别糖磷酸骨架某些特异顺序。

我们这里讨论的是 dsDNA-蛋白质相互识别的一般结构模型。它可分为“一般识别”和“特异识别”。

一、“一般识别”的结构模型

蛋白质识别 DNA, 就总体而言, 在不考虑核酸构象变化的情况下, 它是识别核酸结构的基本特征。这就是说, DNA 和蛋白质都有一定的相互识别的结构部件, 即 $7\text{\AA}$ 重复距离和双重对称轴。它们分别示如图 1 和图 2。图

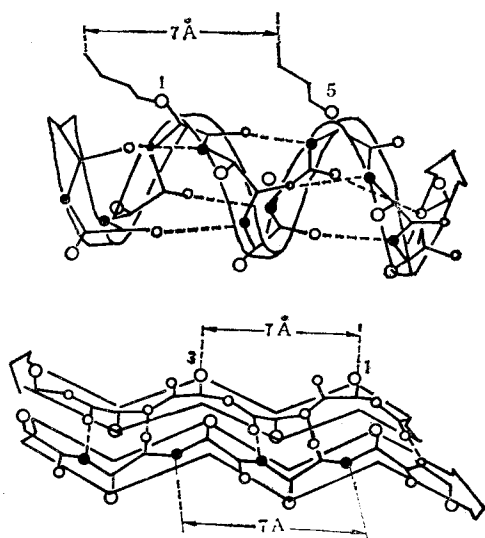


图 1 上部是 α 螺旋, 骨架用带箭头的右手螺旋带表示, 由 N 端到 C 端。●——N, ○——氧基氧, 大圈表示侧链 C_β 原子。氢键用虚线表示。残基 1 和 5 指排除 C_α 外第 5 个原子间距离。这个距离相应于 1 到 5 位残基上的精氨酸或赖氨酸正电荷之间的距离。底部是反平行 β 片层, 符号同上。

1 表示蛋白质的 α 螺旋和反平行 β 片层的重复距离;图 2 表示 dsDNA 和 β 片层的旋转对称。如图所示,反平行 β 片层中交替残基的酰胺基,位于同向,之间相距约 $7 \text{ \AA}$ 。同样,沿 dsDNA 骨架的相邻磷酸基之间的距离(按磷酸基外围测之)也约为 $7 \text{ \AA}$ 。这对于完全伸展的肽或反平行的 β 片层中交替残基的 C_β (与 C_α 相连的侧链碳原子)也是适用的。

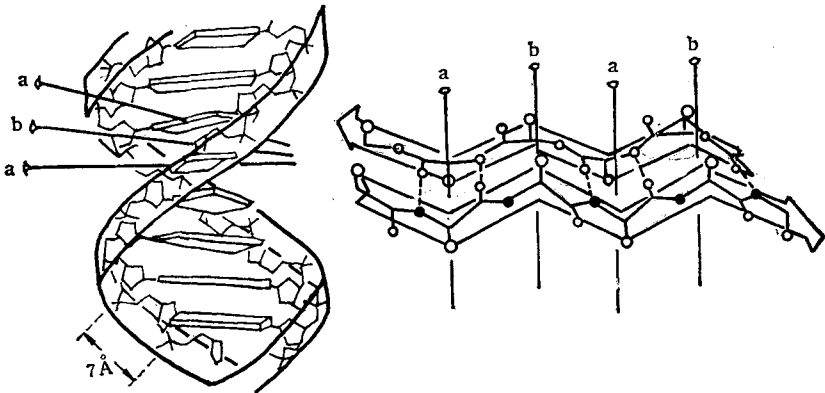


图 2 (左)是 B-DNA 两种类型的伪双重轴,一个是在每个碱基对的平面上(a),另一个在两个相邻碱基之间(b)。(右)是反平行 β 片层两种类型(a 和 b)伪双重轴。

在蛋白质的 α 螺旋结构中,肽的骨架上没有可比较的维度。然而,精氨酸或赖氨酸的两个正电荷基团之间距离约 $7 \text{ \AA}$ (当两个残基由三个其它的残基相隔时,即图 1 中带电荷基团 1 和 5)。这样,在 dsDNA 和三种类型的蛋白质二级结构(α 螺旋,反平行 β 片层和完全伸展的肽)之间有维度相容性。这些总汇于表 1。

表 1 dsDNA 和蛋白质二级结构的结构相容性

	旋转对称	外部重复距离
dsDNA(B)	2×2 折迭	$\sim 7 \text{ \AA}$ (PO_4-PO_4)
β 折迭($\uparrow\downarrow$)	2×2 折迭	$\sim 7 \text{ \AA}$ (残基 1 到 3)
伸 展	无	$\sim 7 \text{ \AA}$ (残基 1 到 3)
α 螺旋	无	$\sim 7 \text{ \AA}$ (残基 1 到 5)

当考虑旋转对称结构时,在 dsDNA 和反平行 β 片层之间也有相容性,如图 2 所示。A-DNA, B-DNA 的每个碱基对都有两种类型的伪双重轴,一个是在每个碱基对的平面上,一个位于两个相邻碱基对之间。

总之, $7 \text{ \AA}$ 重复距离和双重对称轴构成了 dsDNA-蛋白质相互识别的基本结构部件。

根据维度和对称性的相容性,可提出三种类型的“一般识别”的结构模型,示如图 3。

图 3 所示类型 I-a: B-DNA 大槽中含精氨酸或赖氨酸的 α 螺旋。两个相邻残基 (1 和 2) 可以横跨大槽。由三个其它残基所隔离的两个碱性残基 (1 和 5),可以与 DNA 的同一

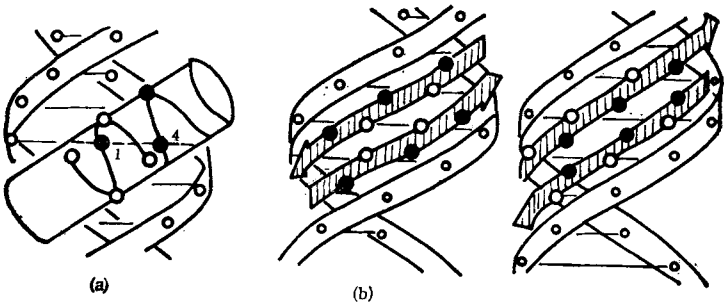


图 3 DNA-蛋白质“一般识别”的三类模型
类型 I: DNA 和 α 螺旋之间的相互识别

股的两个相邻磷酸基结合。赖氨酸和精氨酸残基的伸展构象有适当的长度,以达到能与磷酸基形成氢键以及中和电荷。肽骨架和 DNA 之间无相互识别。类型 I-b: 与 I-a 相似,但相互识别是在小槽,碱性残基的侧链被弯曲容进小槽。类型 II-a: 反平行 β -片层在 B-DNA 的小槽中。肽的极性反平行于和它最近的多核苷酸。肽的交替的 $-NH$ 基与 DNA 的磷酸基形成氢键。 β 片层的两种类型伪双重轴与 DNA 的伪双重轴恰好重合,以小点示出。如果交替的残基也是碱性的,它们可以中和磷酸电荷。类型 II-b: 与 II-a 相似,但在小槽中。类型 III-a: 富有碱性残基的伸展的肽同 B-DNA 大槽的两股相互识别。类型 III-b: 与 III-a 相似,但相互识别是在小槽中。

类型 II: DNA 和反平行 β 片层之间的相互识别。

图 3(II-a,b)反平行 β 片层正覆盖 dsDNA 的小槽。这里, β 片层的维度和对称性与 dsDNA 恰好重合。以这样的方式,肽骨架与核酸骨架以氢键相连。如果形成反平行 β 片层的侧链在交替顺序上有碱性残基,它们能与 DNA 骨架的磷酸基中和以及氢键相连。这类相互识别模型纯系根据模型构筑研究提出的^[6,7]。 β 片层可以假定取两种可能的方位。一是肽的极性与 DNA 骨架的极性平行;另一是彼此反平行。借助计算机构筑的模型表明,两者有惊人的互补性^[4]。

类型 III: DNA 与伸展多肽的相互识别^[7]

图 3(III-a) 是伸展的单一多肽链和 DNA 槽之间的相互识别。这里,多肽的交替残基同核酸的一股相互识别,另一交替残基同另一股相互识别。这类相互识别的特殊要求是肽的侧链必须是精氨酸或赖氨酸,因为它们足以伸展到 DNA 骨架上。这些残基的正电荷基团可以中和磷酸的电荷,并且可能同磷酸基或糖基以氢键相连。III-b, 精氨酸或赖氨酸残基的侧链弯曲容于小槽。类型 III-a 和 III-b,肽和 DNA 骨架之间无直接的相互识别。

二、“特异识别”的结构模型

在一般识别模型中,类型 I 和 II 可以认为是“特异识别”的固定阶段。类型 I 取决于碱基对和侧链之间“特异识别”的程度,精氨酸和赖氨酸数的要求可以放宽, DNA 螺旋轴和 α 螺旋之间距离可以变动。

蛋白质特异识别 dsDNA 的大槽可以考虑两个因素: 氢键和立体接触。

氢键识别,如图 4 所示。

在上面讨论的“一般识别模型”中,只有类型 I 可以用来表示识别 dsDNA 的大槽。DNA 的大槽同 α 螺旋的接触区,有一种有趣的结构组合,见图 5。 α 螺旋的残基 1 和 4 的 C_α 原子基本与碱基对平面共平面。因此,由其它两个残基所隔离的两个氨基酸侧链可以识别碱基对(除了 C_α 原子外,在 3 或 4 原子上,即 C_β 或 C_γ 位,大多数氨基酸侧链都可以形成氢键)。当然,在这种类型的识别中,一般假定 DNA 的构象变化很小,或不变化^[8]。在这种假定的情况下,一是可以研究在 α 螺旋上的 1 和 4 位的各种氨基酸,并检验它们对识别特殊碱基对的相容性。二是可以构筑“识别词典”。

除了在大槽中形成氢键外,立体要求似乎也起重要的作用。对于这类型的相互识别,最重要的基团是胸腺嘧啶的甲基。因为胸腺嘧啶的甲基最易接近大槽或小槽上任何基团的表面。这样,作为识别信号,立体接触大概取决于大槽上甲基的分布^[9]。

如图 4 所示,在 dsDNA 小槽上氢键形成的潜力较之大槽无区别。例如,小槽中氢键模型, A—T 碱基对实际上与 T—A 相同。这对于 G—C 和 C—G 也是如此,只不过程度不同而已。即氨基的氢键受体位在两个碱基对之间稍有差别。这样,出现于小槽中仅有两类模型: 一是 A—T 或 T—A, 另一是 G—C 或 C—G。换句话说,小槽的识别信号可看作是二元编码(Binary Code)组成的,大槽识别信号是四元编码(Tetrad Code)组成的。

对于小槽的识别,类型 I-b, 类型 II-a 或

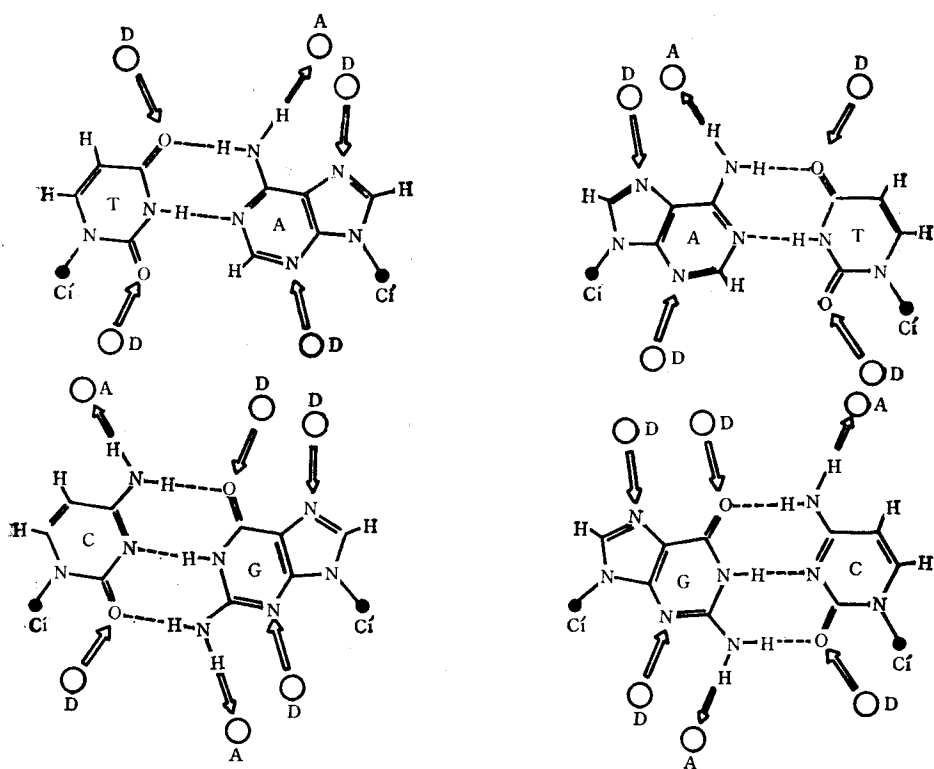


图4 同四个不同碱基对形成氢键的可能原子位置：A，氢键受体的蛋白质的原子；D，氢键供体的蛋白质的原子。

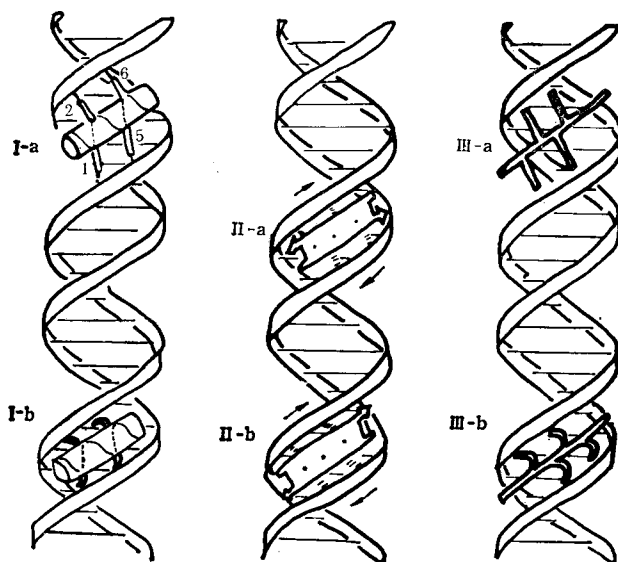


图5 “特异识别”的几个模型

(a) 类型 I: α 螺旋的残基 1 和 4 的 C_β 或 C_γ 原子落入碱基对的平面上。●——同 DNA 大槽接触的 C_β 或 C_γ , ○——离开槽的 C_β 或 C_γ 。在相应于 C_β 或 C_γ 位上有氢键供体或受体原子。(b) 类型 II: 多肽的 C_β 位置有如沿 DNA 螺旋轴的碱基对同样的隔离。●——面向小槽的 C_β 原子, ○——离开小槽的 C_β 原子。

类型 II-b 在结构上是适宜的。对于类型 II 模型,交替残基的 C_β 将面对双螺旋的小槽,并且这些残基的相隔与碱基对相隔是一致的,见图 5。

三、构筑“识别词典”初探

由 dsDNA-蛋白质的“一般识别”和“特异识别”的结构模型,似乎可以得出以下几点意见:

1. 核酸与蛋白质相互识别的基本结构部件是维度和对称性的相容性。

2. “特异识别”的基本因素是氢键和立体接触,后者取决于胸腺嘧啶的甲基分布。

3. 蛋白质的 α 螺旋可识别 DNA 的大槽, β 片层可以识别 DNA 的小槽。

4. 小槽的识别信号是二元编码组成的;大槽的识别信号是四元编码组成的。

上述意见意味着,核酸与蛋白质相互识别的核心问题是蛋白质的二十个字母和核酸的四个字母之间的关系,即氨基酸“词”和核苷酸“语言”之间的精确识别。这个问题的实质是构筑“识别词典”,即建立“空间编码”。

由 dsDNA 同蛋白质相互识别的结构模型指出,由 A—T 编码的氨基酸残基有六种: Ser, Thr, Asn, His, Gln, Cys。有时也为两种碱性氨基酸 (Lys, Arg) 编码;由 G—C 编码的氨基酸残基有九种: Gly, Ala, Val, Leu, Ile, Met, Phe, Trp, Tyr。有时也为两种酸性氨基酸 (Asp, Glu) 编码。由于 Pro 残基处多肽链构象发生纽结,不与核苷酸链作用,故未为其编码的碱基对。这样,我们似乎可能构筑出

“识别词典”的雏型^[10,11]。

表 2 “识别词典”

dsDNA		氨基酸残基		
槽的大小	碱基对	参与特异性识别	参与取决于位置的识别	不处于识别位点
小槽	A ⋮ T	Ser, Thr, Asn His, Gln, Cys	Lys, Arg	Pro
	C ⋮ G	Gly, Ala, Val Leu, Phe, Ile Met, Tyr, Trp	Asp, Glu	
大槽	结合位点顺序中双重轴上的四个碱基对			

参 考 文 献

[1] Zubay, G. et al.: *J. Mol. Biol.*, **1**, 1, 1959.
[2] Adler, K. et al.: *Nature (London)*, **237**, 322, 1972.
[3] Carter, D. et al.: *Proc. Natl. Acad. Sci. USA*, **71**, 283, 1974.
[4] Church, G. et al.: *ibid.*, **74**, 1458, 1977.
[5] Steitz, T. A. et al.: *Nature (London)*, **290**, 744, 1981.
[6] Warrant, R. W. et al.: *ibid.*, **271**, 10, 1978.
[7] Kim, S. H. et al.: In *Structure and Conformation of Nucleic Acids and Protein—Nucleic Acid interactions* (Sundaraligam and Rao, Eds), University Park Press, Baltimore, Maryland, P. 571, 1975.
[8] 张 今等:《吉林大学自然科学学报》, **3**, 91, 1983.
[9] Alden, C. et al.: *J. Mol. Biol.*, **132**, 411, 1979.
[10] Engelhardt, W. A.: *Macromolecules in the Functioning Cell* (Edited by Francescosalvatore et al.), Plenum Press, New York and London, p. 301—310, 1978.
[11] 张德安, 张 今:《生物化学与生物物理进展》, (6), **8**, 1981.

[本文于 1986 年 11 月 20 日收到]