

试 论 空 间 编 码

张 德 安 张 今

(吉林大学化学系)

60 年代阐明的遗传密码,即通用的、三联体的、简并性的密码,弄清楚了核酸的核苷酸排列顺序与其编码的蛋白质的氨基酸排列顺序之间的线性关系,故可称之为“线性编码”。然而,仅就基因的表达而言,尽管基因载有其编码的全部蛋白质信息,倘若没有核酸与蛋白质在空间结构上的直接关系,单凭线性编码,还是不可能实现的。其原因在于特异功能的实现还要靠两者能够相互识别、相互作用。因此,关于核酸-蛋白质相互识别的研究,日益为人们所重视,近几年已成为迅速发展的一個领域,初步揭示了核酸-蛋白质在空间关系上的一些规律,得出了“空间编码”的概念。倘若这种空间编码也象线性编码那样具有通用性,那么,对深入了解许多生命过程的机理,将有极重要的意义。本文试图对空间编码的概念、由来、性质及应用等知识,作一简要的介绍,并提出一些粗浅的设想。

一、何谓空间编码

空间编码,指的是核酸与蛋白质的特异相互作用中,核酸的控制位点与其相应的蛋白质立体特异位点之间的直接结构互补性。就目前的研究结果而言,空间编码比线性编码复杂得多,既要求有立体特异位点中适当位置的 AT-与 GC-碱基对编码的氨基酸存在,也需要有同反平行 β -结构一致的所有氨基酸残基的存在,均受到立体化学限制。

表 1 总结了强加于立体特异蛋白位点中的 R_1 和 R_2 的残基的编码规则及立体化学限制。从编码来看,由 AT 碱基对编码的氨基酸残基有六种: Ser, Thr, Asn, His, Cys, Gln; 由 GC 碱基对编码的氨基酸残基有九种: Gly, Ala, Val,

表 1 强加于立体特异蛋白位点中 R_1 和 R_2 残基的编码规则*与立体化学限制**

碱基对	多肽链段的类型	向外指向的侧链 (R_1)	向内指向的侧链 (R_2)
A : : : T 或 : : : A	t	* Ser Thr Asn His Gln Cys	
		g ** 除 Pro 外任一残基 Glu, Asp, Lys, Arg, His 未必可能存在	
C : : : : : : : G	t	* Gly, Ala, Val, Leu, Phe, Ile, Met, Tyr, Trp, Glu, Asp, Lys, Arg 未必可能存在	** 除 Pro, Asp, Glu 外任一残基
		g 除 Pro 外任一残基, Glu, Asp, Lys, Arg, His 未必可能存在	
G : : : : : : : C	t	这种情况下不利于准确识别,并且很少存在	
		g	

* 强加于 t-链段残基 R_1 上的编码限制

** 强加于 R_1 和 R_2 侧链及与 t-和 g-链段中带电荷残基和 Pro 存在有关的立体化学限制

Leu, Ile, Met, Phe, Trp, Tyr。这种在空间位置上由一个碱基对与其所编码的一个氨基酸残基相对应,称之为“对联体密码子”。这种编码所受到的限制是: 只有 t-链段的 R_1 残基参与编码,而所有的 R_2 残基,无论在 t-链段上还是在 g-链段上,均以其骨架 NH 与 C=O 参与同酸碱基的特异相互作用。至于其它一些立体化学限制,则与带电荷残基及 Pro 的存在有关,即带负电荷的 Asp 与 Glu 以及 Pro,可能不存在于立体特异蛋白位点中;而带正电荷的 Lys, Arg 与 His 侧链,以及 Tyr 极性侧链,可能存在于 R_2 位置上,能同磷酸基团形成氢键,当以 R_1

位置存在时,这种氢键便不能形成。总之,通过上述 t-链段 R_1 与其密码子相对应, R_2 与核酸碱基的特异相互作用, 以及一些严格的立体化学限制,即可保证蛋白质以其立体特异位点,精确地识别出核酸的特定位置,与之密切地相互作用。

二、空间编码的由来

空间编码的知识,主要是来自 DM_3 (偏端霉

素 A)-DNA 复合物的立体化学模型的建立,然后以类推法对多肽-DNA 复合物的分子模型的研究,以及对蛋白质-核酸复合物的分子模型的分析。

1. DM_3 -DNA 的立体化学模型

DM_3 及其类似物,是一类能与 dsDNA 的富 AT 区相结合,同 E. Coli RNA 聚合酶竞争启动子顺序,从而抑制 RNA 合成的寡肽抗菌类。其化学结构如下:

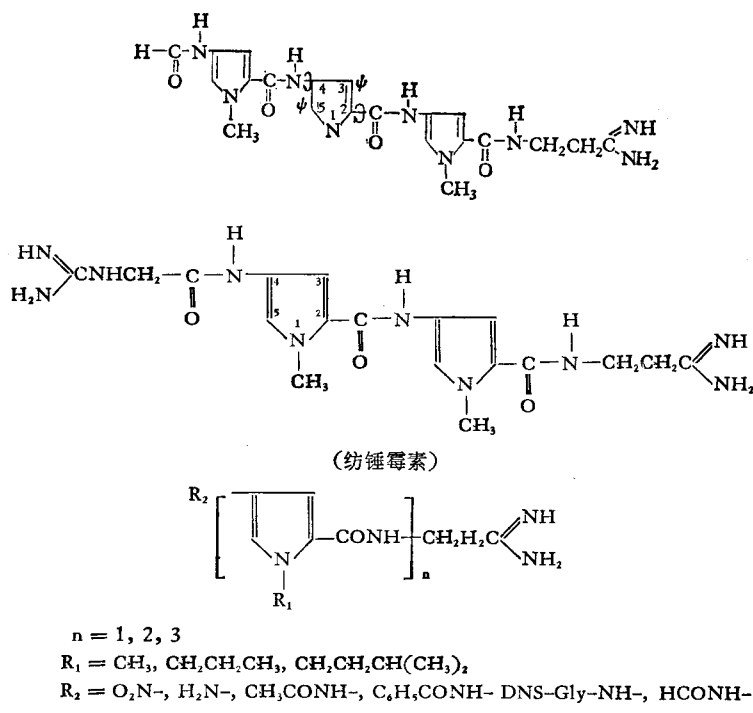


图 1 DM_3 及其类似物的化学结构

$\varphi = 0$, $C'-N$ 键对于 C_4-C_5 键是顺式; $\psi = 0$, $C'-N$ 键对于 C_2-C_3 键是顺式。
键的远端相对于近端顺时针旋转,构象角是正值。

图 2 示出了 DM_3 与 DNA 的富 AT 区结合的模型,由该模型可知: (1) DM_3 延伸于有五个碱基对长的 DNA 小槽中; (2) DM_3 以其酰胺基作为氢键的供体,分别与 AT 群的 TATT 形成氢键,因此,同一条多核苷酸链上的 TATT 为 DM_3 的结合位点,胸嘧啶的 $C_2=O$ 基与腺嘌呤的 N_3 原子则作为氢键的受体; (3) DM_3 上的丙脒基可与磷酸基的两个氧原子形成氢键,以稳定 DM_3 -DNA 复合物。基于上述 DM_3 -DNA 复合物分子模型的基本特征,可把 DM_3 视为同 DNA 小槽中恒定识别位点相互作用的

配位体。

2. 多肽-DNA 的分子模型

根据 DM_3 -DNA 立体化学模型的知识,可以把每个调节蛋白视为同 DNA 特异相互作用的具有若干反应中心的晶格配位体。反应中心的氨基酸排列顺序应能识别 DNA 上的特定碱基对顺序,也就是说,调节蛋白与 DNA 的特征晶格重复距离应该精确地一一对应。由此,建立多肽-DNA 复合物分子模型的基本程序是^[2]: (1) 建立多肽结构,使其螺旋参数与核酸一致,以便能识别 DNA 上具有特定几何形状的碱基

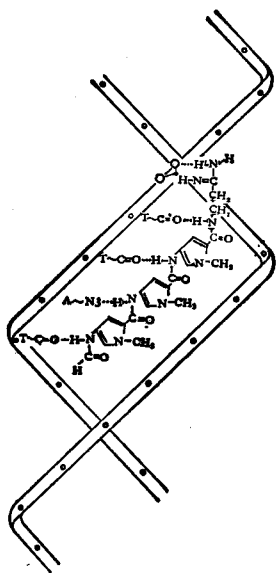


图2 DM₃-DNA 复合物的模型

对顺序；(2) 在多肽链的酰胺基与 DNA 的碱基对之间建立一组有规律的氢键，可根据 poly(dG) · poly(dC) 及 poly(dA) · poly(dT) 等均聚物的结合图形，对其中识别的个别碱基对稍加修饰得到；(3) 合理装填多肽与多核苷酸链，使 Van der Waals 接触不得短于其标准值。

由 DM₃-DNA 的立体化学模型特征类推，多肽应识别 DNA 小槽中的碱基对顺序，并与之形成氢键。其中，嘧啶的 O₂ 及腺嘌呤 N₃ 原子可作为氢键的受体，与肽链骨架的 NH 基形成氢键；而鸟嘌呤的 2-NH₂ 基可作为氢键的供体，与肽链骨架的 C=O 基形成氢键^[3]。这对于 poly(dG) · poly(dC) 来说，由于鸟嘌呤的 2-NH₂ 基与胞嘧啶的 O₂ 原子的空间位置不同，所以多肽链有两种有规律的构象，即与 poly(dG) 互补的多肽链为 g-构象，与 poly(dC) 互补的多肽链为 t-构象^[4]，多肽链以其 N—C^α—C' 顺序同多核苷酸链 C₃—C_{5'} 顺序相一致的方向排列成行。又对于双链多肽而言，发现仅 gt-和 gg-构象是可能的，即 gt-构象为变形的反平行 β-片层，其中一半氢键断裂，同碱基对形成一套新氢键 (图 3a)；而 gg-构象的所有氢键都正常地存在，通过 β-片层中心区的有规律弯曲，同横嵌在两条多核苷酸链中的鸟嘌呤形成氢键 (图 3b)。

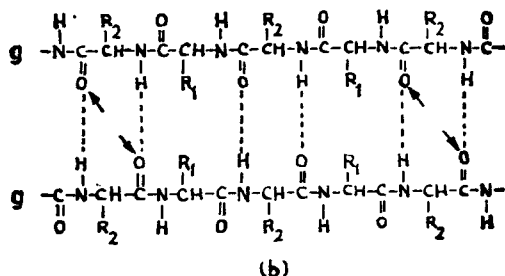
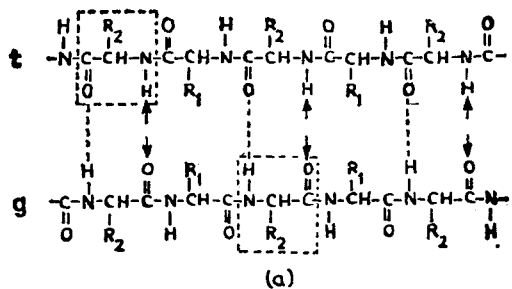


图3 同 DNA 复合的立体特异位点中二级结构模式图

两个反平行多肽链段以氢键结合，形成 gt(a) 和 gg(b) 双股多肽螺旋。有框者是同 poly(dG) · poly(dC) 中确定的碱基对形成氢键的 gt-结构中的氨基酸残基。R₁ 和 R₂ 是分别表示指向外和指向内的侧链。箭头指示骨架上分别与 poly(dG) · poly(dC) (a) 及 poly[d(G-C)]-poly[d(G-C)] (b) G-C 碱基对形成氢键的 NH 和 C=O 基团。

图4为部分缠绕 poly(dG) · poly(dC) 小

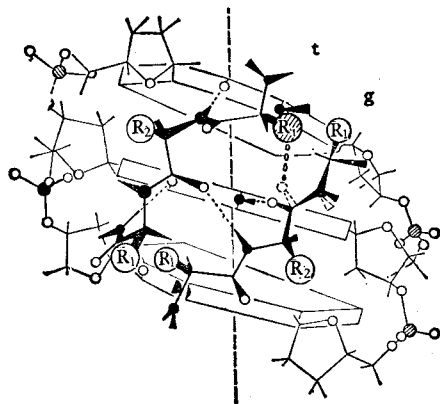


图4 gt-多肽螺旋——poly(dG) · poly(dC) 复合物几何形状模式图

该结构通过螺旋轴投影到垂直面上，并垂直于复合物的伪二重轴。符号：○(氧)；●(氮)，⊙(磷)。点线表示氢键。G 的 2-氨基和 g-多肽链的酰胺基间氢键受体位于 t-链段中 R₁ 位的氨基酸残基控制 (画影的环)。在该位有 AT 编码的残基 (例如 Ser) 时，同 G 形成的氢键减弱或断裂。在 Ser 之 OH 和一 CONH 之间形成新的氢键。虚线表示同 G 无氢键时，酰胺基的空间位置。

槽的 gt -多肽螺旋。可见,同 DNA 复合的立体特异蛋白位点的主要特点是:由两条反平行肽链组成,靠氢键维系在一起,形成一个右手扭转的反平行 β -片层。当与 DNA 结合时,其 t -链段作为氢键供体,同一条 DNA 链的所有碱基(嘧啶的 O_2 及腺嘌呤的 N_3 原子)形成氢键,互补于控制顺序;而 g -链段作为氢键受体,同另一条 DNA 链的鸟嘌呤形成氢键。也就是说,在调节蛋白与 DNA 结合时,伴随有立体特异蛋白位点和 DNA 控制位点的显著结构变化^[5],以致两者能够特异地相互作用,精确地相互识别。同时不难看出, DNA 控制位点应表现为鸟嘌呤在两条多核苷酸链间的不对称分布,以利于蛋白质-核酸的相互识别。

在上述研究中,由于发现 t -链段的某些氨基酸残基(如 Ser, Thr, Asn, His, Cys 及 Gln)的侧链,可同 g -链段的骨架 $C=O$ 基形成氢键,因此断裂或强烈地减弱鸟嘌呤与这些 $C=O$ 之间的氢键,并导致 g -链段的构象变化,所以,便把这些氨基酸残基视为由 AT 碱基对编码的;而不能按此机制相互作用的氨基酸残基,则视为由 GC 碱基对编码的。由此,便得出了空间编码的概念。

三、空间编码的性质

由表 1 可见,空间编码具有高度简并性,也就是说,一种 AT 对联体密码子可为六种氨基酸残基编码,一种 GC 对联体密码子可为九种氨基酸残基编码。这种高度简并性,似乎对特异地相互识别不利,但是,由于在核酸-蛋白质的相互作用中,不但有多个对联体密码子分别与其编码的氨基酸残基相对应,而且还有同反平行 β -结构相一致的氨基酸残基作保证,再加之严格的立体化学限制,还是可以精确地相互识别的。仅就前述编码限制而言, R_2 是侧链指向螺旋轴,位于分子表面上的氨基酸残基, R_1 是侧链向外指向,位于分子中间的氨基酸残基,并且只有 t -链段的 R_1 残基参与编码。如果 R_1 为 GC 编码的氨基酸残基,鸟嘌呤应位于同 g -链段互补的多核苷酸链中,与 g -链段的骨架

$C=O$ 基定向地形成氢键,倘若鸟嘌呤位于同 t -链段互补的多核苷酸链中,将会同 g -链段的 NH 基竞争与 t -链段的 $C=O$ 基形成氢键,使氢键大大减弱,即 $GC \rightarrow CG$ 的置换是不利的。这就是说,不但 AT 与 GC 碱基对能被 DNA 小槽中的立体特异蛋白位点识别,而且 GC 与 CG 碱基对也能被极好地识别。由此可见,鸟嘌呤在两条多核苷酸链间的不对称分布,应该是立体特异蛋白位点所识别的核酸控制位点的重要特征。

四、空间编码的应用

目前,空间编码的一些规律已被用于阐释更多的蛋白质-核酸的特异相互作用,以及由核酸的控制顺序预测与其相互作用的立体特异蛋白位点的结构。例如:

1. Lac 阻遏物-操纵基因结合的阐释

近来, Gursky 等根据已确定的 Lac 阻遏物与操纵基因相互作用的顺序^[6],提出了 Lac 阻遏物-操纵基因结合的详细模型,并用空间编码的规律作了较好的阐释。如图 5 所示,(1) 同操纵基因特异相互作用的阻遏物顺序为 Gly 14—Ala32 与 Ala53—Leu71;(2) 同操纵基因碱基对顺序相对应的肽段是 Thr19—Val30,其中 Thr19, Ser21, Asn25 及 His29 由 AT 编码, Val23 和 Ala27 由 GC 编码,这六个氨基酸残基分处于六个分离位置,可避免随机事件发生;(3) Ser31—Val52 含有三个带负电荷的 Glu 及一个 Pro,不能嵌入 DNA 小槽,为连接 t -与 g -链段的突环;(4) Val20—Val30 与 Ser61—Leu71 为反平行 β -片层, Gly14—Gln18 与 Ala53—Lys59 为两个粘性末端,可促进精确识别;(5) 阻遏物与 DNA 结合,伴随着 20—30 和 61—71 的 β -片层区的构象变化,即 β -片层的一半氢键断裂,转变成 gt -结构,被同操纵基因所形成的新氢键取代,为协同过程;(6) GC-编码残基取代 AT-编码残基的突变,可减弱阻遏物对操纵基因的亲合性^[7];(7) 操纵基因的二重旋转对称性,可指挥阻遏物与其对称连接^[7],即左侧结合的亚基 g -链段同 0 与 +1 位

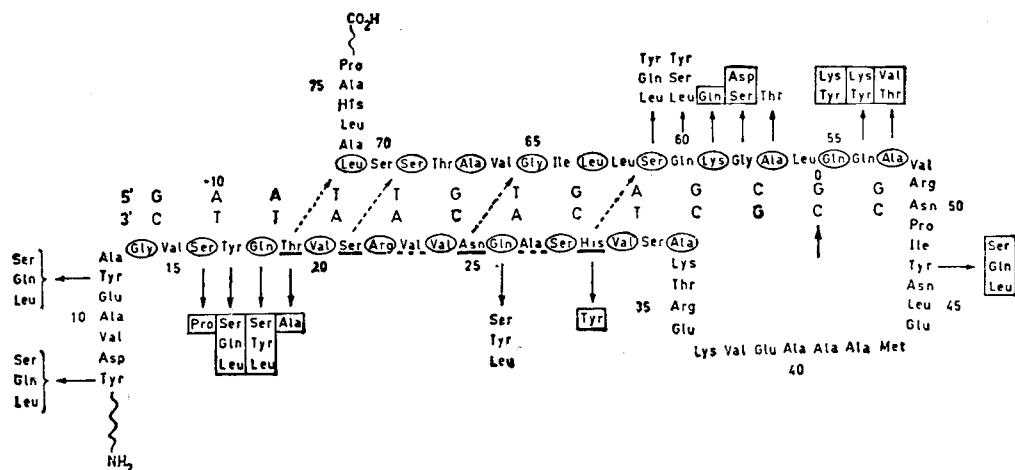


图5 立体特异阻遏物位点的氨基酸顺序和 Lac 操纵基因的碱基对顺序之间的对应关系

阻遏物多肽链顺序在 14—30 和 53—71 分别处于 t-和 g-构象。由 AT 和 GC 编码的氨基酸残基分别以底下划线连线和虚线示之。圈者为 R_1 型残基。斜箭头表示氢键状态受 AT 编码残基影响的 g-链段中的那些残基。多肽链顺序在 20—30 和 61—71 处形成 gt-双股多肽螺旋。粗的垂直箭头表示操纵基因中二重对称轴的位置。连续箭头表示阻遏物顺序中的取代位点。框者是破坏操纵基因-阻遏物亲和力的交换氨基酸。未标记的氨基酸残基可掺合于阻遏物之中,不失去操纵基因的结合活性。

的鸟嘌呤(5'链)相互作用,右侧结合的亚基 g-链段同 -1 位的鸟嘌呤(3'链)相互作用,两个 g-链段的重叠区可以氢键结合,形成 gg-多肽双螺旋,大概包括 Ala53—Lys59 区,其中两个 Gln 成二重对称关系,由侧链形成两个氢键,闭锁阻遏物亚基^[8]。大概只有 Asn 与 Gln 是参与识别二重对称部位的氨基酸残基。

2.2 阻遏物的预测

Gursky 等^[4,6]认为,根据上述 Lac 阻遏物原理,按照空间编码规律,可以预测所有调节蛋

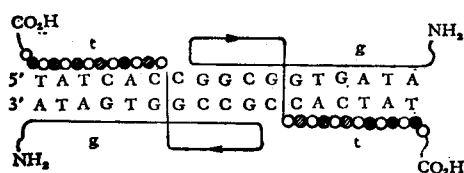


图6 预测 λ 阻遏物亚基的立体特异位点中 AT-和 GC-编码的氨基酸残基的排列

氨基酸残基类型		符号	氨基酸残基
R_1	AT 编码的残基	●	Ser, Thr, Asn, His, gln, Cys
	GC 编码的残基	⊙	gly, Ala, Val, Leu, Ile, Met, Phe, Tyr, Trp
R_2	非编码之残基	○	除 Pro, Asp, glu 外的任一残基

两个阻遏物亚基呈二重对称。DNA 的碱基顺序引自文献[11]。

白的立体特异位点,大体都能表示为 t-与 g-链段形成的短 β -片层区,可延续约五个碱基对,约旋转 180° 。Maniatis 等^[11]应用表 1 总结的编码规则,预测了能识别 λ 操纵基因特定碱基对顺序的阻遏物氨基酸顺序。如图 6 所示,两个阻遏物亚基的立体特异位点呈二重对称关系,其 g-链段延续 10 个碱基对。这种预测是否正确,有待测定 λ 阻遏物的氨基酸顺序来证明。

3. S8 与 16 S RNA 特异结合的分析

根据已确定的 S8 核糖体蛋白的氨基酸顺序,以及与其相互作用的 16S RNA 的核苷酸顺序分析^[12],预测了同 16S RNA 特定碱基顺序相对应的 S8 的多肽链区。如图 7 所示,在 Thr11-Asn20 区, AU-与 GC-编码的残基处于相应的位置,带正电荷的 Arg 与 Lys 残基处于 R_2 位置,非常类似于 Lac 阻遏物-操纵基因的结合。其主要特征是:未配对的 A 与 U,同未配对的 A 分处于相对的两条多核苷酸链段上,很可能以次级氢键相连,即特征的三级结构在相互识别中起作用^[13]。

五、结 语

尽管空间编码的知识主要来自复合物分子模型的建立与研究,但是,可较好地阐释一些核

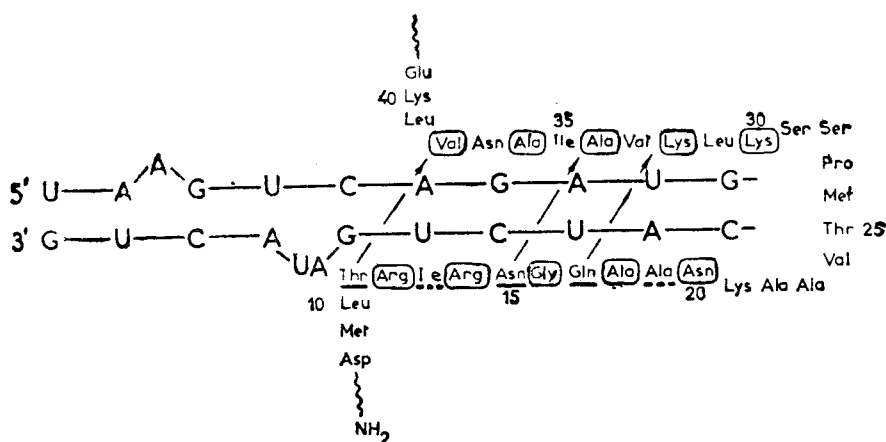


图 7 识别位点中 S8 核糖体蛋白的氨基酸顺序和 16S RNA 的核苷酸顺序之间的对应

酸-蛋白质的特异相互作用,并且由已知的核酸控制位点顺序,可预测与其特异相互作用的蛋白质结构。仅就现有的空间编码的概念而言,我们认为:

1. 空间编码的高度简并性,既是核酸-蛋白质的特征动态空间结构联系所致,也是两者特异功能所需。假若没有这种高度简并性,就有可能因为一个空间对联体密码子的滑动,而造成所有互补性的丧失。正因为这种空间编码的高度简并性, RNA 聚合酶才能既精确地识别 DNA 的初始识别序列,又足以向前滑动、既紧密结合 DNA 的牢固结合位点,确保转录在特定位点开始,又足以在结构基因上迅速移动,顺利地完成其转录作用,当然,这种不同程度的结合原因有多种可能,尚待进一步探讨。

2. 空间编码的通用性,是极可能存在的,根据现有的空间编码规则,无论对哪种特异相互作用,对核酸特定位点的共同要求是:鸟嘌呤应该在两条多核苷酸链间不对称分布,这应该是核酸特定位点的最重要特征,也可以说是空间编码通用性存在与否的重要标志。从目前所积累的 CAP 结合位点, RNA 聚合酶初始识别位点,牢固结合位点^[14]、核糖体结合位点^[15]等蛋白质特异结合的核酸序列来看,确实较广泛地存在着鸟嘌呤不对称分布之例证。因此,空间编码的通用性是极为可能存在的。同时可以预期,象阻遏物-操纵基因以及 CAP-CAP 结合位点之类的结合或解离的两个极端的相互作用,

不但蛋白质亚基只是部分对称。而且所作用的核酸特定位点也存在有极大的 180° 旋转对称性;但是,对于象 RNA 聚合酶-DNA 之类的具有不同结合程度的相互作用来说,不但蛋白质亚基只部分对称,而且所作用的 DNA 特定位点也不是完全对称的,这可能是沿着一定方向不对称转录的依据。

参 考 文 献

- [1] Zasedatele, A. S. et al.: *Dokl. Acad. Sci. USSR*, 231, 1006, 1976.
- [2] Pauling, L. et al.: *Proc. Natl. Acad. Sci. U.S.A.*, 37, 205, 1951. Rich, A. et al.: *Nature (London)*, 176, 915, 1955.
- [3] Arnott, S.: *Biochem. Biophys. Res. Commun.*, 47, 1504, 1972.
- [4] Gursky, G. V. et al.: *Mol. Biol. Rep.*, 2, 413, 1976.
- [5] Arnott, S. et al.: *J. Mol. Biol.*, 81, 107, 1973; 88, 509, 523, 551, 1974.
- [6] Gursky, G. V. et al.: *Mol. Biol. Rep.*, 2, 4217, 1976.
- [7] Gilbert, W. et al.: *Proc. Natl. Acad. Sci. U.S.A.*, 56, 1891, 1966.
- [8] Steitz, T. A. et al.: *ibid*, 71, 593, 1974.
- [9] Gilbert, W. et al.: in *Protein-Ligand interaction* (H. Sund and G. Blauer. eds.). 193—210, de Gruyter, Berlin, 1975.
- [10] Maniatis, T. et al.: *Cell*, 5, 109, 1975.
- [11] Ungewickell, E. et al.: *Eur. J. Biochem.*, 51, 165, 1975.
- [12] Kim, S. H. et al.: *Proc. Natl. Acad. Sci. U.S.A.*, 71, 4970, 1974.
- [13] 吉川宽:《真核生物の分子生物学》, 9, 1978.
- [14] Grunberg-Manago, M.: *FEBS. Letters*, 43, 5, 1978.

[本文于 1980 年 11 月 18 日收到]