

四、遗传密码的破译

遗传密码这一名词在 1944 年薛丁谔的《生命是什么？》小册中就已提出，但在 1953 年沃森和克里克建立 DNA 双螺旋结构以前，其真实内涵乏人知晓。60 年代遗传密码的破译，使遗传物质核酸和决定细胞功能的蛋白质两种重要的生物化学物质之间架起了一座桥梁。这对于从分子水平了解遗传和细胞的生物化学有着重大的影响，它大大推进了生物科学的进展，成为奠定分子生物学的重要基石之一。

生物界的遗传规律是生命活动中基本规律之一。早在 1909 年，W·L·约翰逊(Johnson)曾提出基因的概念，并应用了基因型和表现型两个术语。1944 年，O·T·埃弗里(Avery)用肺炎球菌作实验，证实遗传物质就是 DNA。1941 年，G·W·比德尔(Beadle)和 E·L·塔特姆(Tatum)用脉孢菌突变株作的实验表明基因可调节某些酶促反应过程。1949 年 L·鲍林发现镰刀形红细胞贫血症是由血红蛋白分子的链上发生了一个突变。一个基因决定一种蛋白质的概念逐渐受到人们的重视。

自从沃森和克里克于 1953 年夏在美国冷泉港学术讨论会上作 DNA 的结构和其遗传含意的报告后，围绕着 DNA 碱基序列和蛋白质的氨基酸序列之间的相互关系展开了热烈讨论，其中心议题是：四种不同的碱基怎样排列组合进行编码才能表现出 20 种不同的氨基酸。

第一个提出具体设想的是俄国出生的美国物理学家伽莫夫(G.Gamov, 1904—1968)。当他读到沃森和克里克在英国《自然》杂志上的文章之后，立即提出一种设想，发表在 1954 年 2 月该杂志上。伽莫夫把双螺旋结构中由于氢键生成而形成的空穴用氨基酸填上，就像钥匙和锁一样。每一个空穴的四角是 4 个碱基，也就是说四种碱基的排列组合形成了遗传密码。1955—1956 年间，伽莫夫又发表文章，从排列组合计算，认为二个碱基组成的密码太少，只有 $4^2 = 16$ 种，四个碱基组成的密码又太多，有 $4^4 = 256$ 种，而三联密码有 $4^3 = 64$ 种，足够所有蛋白质合成之需。他还以令人信服的证据证明这种密码并不重复。

对 DNA 的研究使伽莫夫的科学想象的激情得到满足。他建立了一个 RNA 俱乐部，并将该俱乐部的成员数限定为 20，每一个成员代表一种氨基酸——伽莫夫是“phe”（苯基丙氨酸的英语名称缩写）而克里克是“Tyr”（酪氨酸）。他们经常聚会，交流信息。这场讨论对揭开遗传之谜无疑产生了一定的影响。

1960 年，J·莫诺(Monod)和 F·雅各布(Jacob)发现了信使 RNA(mRNA)。这种信使 RNA 由 DNA 接受遗传信息，将它带到细胞内合成蛋白质的部位，再按遗传指令合成蛋白质。mRNA 实质上是遗传基因的翻版，它是特定 DNA 区段严格“转录”出来的。而 mRNA 和蛋白质之间的关系则靠遗传密码来揭示，这就是 mRNA 中核苷酸顺序和蛋白质氨基酸顺序的关系。所以，特定的遗传信息由特定的信使 RNA 传递，合成出特定的蛋白质。

60 年代初，对 DNA 或 RNA 中核苷酸顺序尚无法直接分析测定，对蛋白质中氨基酸顺序所知也有限，所以没有办法直接比较二者在顺序上的联系。

第一个用实验给遗传密码以确切解答的是德国出生的美国生物化学家尼伦贝格(M.W.Nirenberg, 1927—)。1961 年他和另一位德国科学家马太(Heinrich Matthaei)在美国国家卫生研究院的实验室内发现了苯丙氨酸的

密码是 RNA 上的尿嘧啶 (UUU)。他们在用大肠杆菌的无细胞提取液研究蛋白质的生物合成问题时发现：当向这个提取液中加进核酸，则合成了蛋白质；当用由单一的尿嘧啶组成的核酸长链加进这个提取液中，则产生了由单一苯丙氨酸组成的多肽长链。这个结果立即震动了科学界。但是测定其他氨基酸的密码需要各种各样的碱基组合，而当时这种组合并不是很容易得到的，它需要一种多核苷酸磷酸化酶。美国另一位西班牙血统的生物化学家奥乔亚 (Ochoa, Severo, 1905—) 于 1955 年发现了多核苷酸磷酸化酶 (PNP 酶) 帮助了尼伦贝格合成了同聚核苷酸——多聚 U (PolyU) (奥乔亚因发现此酶而获得 1959 年诺贝尔生理学或医学奖)。当他将多聚 U 作为模板加入到无细胞体系中时，那就是只有加有标记苯丙氨酸所产生的那一试管蛋白质沉淀具有放射性。而加其他标记氨基酸的各管则均无放射性进入沉淀。于是，第一个密码便被破译出来，即 UUU 是苯丙氨酸的密码子。用同样的方式以其他多聚核苷酸作为模板，又测出 CCC 是脯氨酸的密码子，AAA 是赖氨酸的密码子。多聚 G 的氨基酸密码子当时用此法测定时遇到困难，未能测出。

在取得第一阶段突破性成果之后，尼伦贝格用混合的核苷酸制备人工合成的 mRNA 模板，分别测试其作用。

用 2 种或 3 种不同的核苷酸制备 mRNA 模板时，PNP 酶合成的产物都是杂聚物，其中核苷酸的顺序是随机的，无法预测。但各种三联体出现的相对几率则是可以推算出来的。在测定了各种标记氨基酸参入蛋白质的量之后，将其相对参入量和三联体出现的几率加以比较，即可知道每种三联体相对应的是哪一种氨基酸。

此法的应用有局限性，密码子中不同核苷酸的比例固然可以推测出来，但是它们的排序却不能确定；尽管如此编码的范围还是大大缩小了。后来又发现有些密码子具有重复性，即一种氨基酸可以有多种密码子，但是每一种密码子只编码一种氨基酸。

为了搞清密码子中核苷酸顺序，尼伦贝格巧妙地设计了第三阶段的实验。他采用的是核糖体结合法新技术，并加入的模板一律改为具有一定顺序的单个三联体。实验仍在无细胞体系中进行。他们的小组合成了全部 64 种单个的、顺序固定的三联体密码。实验结果能使 50 种密码子所对应的氨基酸能确定下来。实验中发现，有三个密码子并不编码任何氨基酸，后来知道它们是终止信号。还知道甲硫氨酸的密码子可兼作起始信号。完整的密码子表，到 1963 年由与尼伦贝格共获 1968 年诺贝尔生理学或医学奖的霍拉纳 (Khorana, Har Gobind, 1922—) 利用其他技术加以确定的。至此遗传密码辞典问世了。

表 16-1 遗传密码表

	U	C	A	G	
U	苯丙氨酸	丝氨酸	酪氨酸	半胱氨酸	U
	"	"	"	"	C
	亮氨酸	"	终止号	终止号	A
	"	"	"	色氨酸	G
C	亮氨酸	脯氨酸	组氨酸	精氨酸	U
	"	"	"	"	C
	"	"	谷氨酰胺	"	A
	"	"	"	"	G
A	异亮氨酸	苏氨酸	天冬酰胺	丝氨酸	U
	"	"	"	"	C
	"	"	赖氨酸	精氨酸	A
	甲硫氨酸"	"	"	G	
G	缬氨酸	丙氨酸	天冬氨酸	甘氨酸	U
	"	"	"	"	C
	"	"	谷氨酸	"	A
	"	"	"	"	G

经克里克提议列出上面表格。表中 U, C, A, G 分别代表四种碱基的符号，左边的字母代表第一个核苷酸的碱基，顶端的四个字母代表第二个核苷酸的碱基，右边的字母代表第三个核苷酸碱基。很多科学家预计，这个表在生物学上的意义，可以同元素周期表在化学上的意义相比。

这些成就，使传统的基因概念得到了新的发展。过去认为，突变、重组、功能都同时是“基因”这一遗传基本单位的属性；根据 DNA 分子结构和座落在其上的密码的分子解释，基因的功能单位是指 DNA 大分子上一段多核苷酸序列，而突变和重组则是在核苷酸碱基对上的变化。这方面的成就，使过去必须在活细胞内才能进行的蛋白质生物合成得以在试管中重现，用简单明确的分子间具体的信息的传递和结构的变化，表达出过去认为十分复杂而神秘的遗传现象，这又一次沉重地打击了活力论的残留影响。以后的大量工作证明，生物界从最高等的人类到自己不能繁殖、必须寄住在活细胞内才能繁殖的病毒，在蛋白质生物合成的遗传密码上竟几乎完全一致！这意味着，遗传密码代表着生命现象所必须具备的基本条件，在追溯生命起源的阶梯上，科学又向前推进了一大步。