

分子遗传学文集

童克忠 等 编

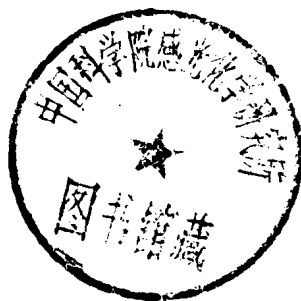
科学出版社

DFIP/05

DF1928

分子遗传学文集

童克忠 等 编



科学出版社

1982

内 容 简 介

本书着重讨论蛋白质和核酸的结构,遗传物质的复制、重组、突变,遗传密码的编码机理、转录和翻译过程,基因表达的调控,基因工程,真核生物的分子遗传学等内容。

可供生物学、分子生物学、遗传学方面的研究人员、高等院校师生和研究生参考。

分子遗传学文集

童克忠 等 编

责任编辑 刘 安

科学出版社出版

北京朝阳门内大街 137 号

中国科学院印刷厂印刷

新华书店北京发行所发行 各地新华书店经售

1982 年 12 月第 一 版 开本: 787×1029 1/16

1982 年 12 月第一次印刷 印张: 19 3/4

印数: 0001—5,600 字数: 462,000

统一书号: 13031·2082

本社书号: 2839·13—10

定 价: 3.10 元

5.11.1
2

目 录

第 一 章 绪论	1
第 二 章 核酸和蛋白质的结构	13
第 三 章 遗传密码	28
第 四 章 突变	51
第 五 章 DNA 复制	78
第 六 章 转录	99
第 七 章 翻译	134
第 八 章 基因表达的调节控制	173
第 九 章 遗传工程技术	191
第 十 章 染色质的结构与功能	215
第十一章 真核基因的纯系增殖	248
附 录 I 核苷酸顺序测定	265
附 录 II 氨基酸顺序分析	278

第一章 绪 论

童 克 忠

四百多年以前,西方有位哲学家认为他父亲遗传给他一种病——膀胱结石。但是他父亲是在他出生以后才得这病的,而且精液是液体,为什么能遗传膀胱结石呢?这位哲学家被这个问题难住了。

他所提出的问题,是一个典型的遗传学问题:亲代通过什么途径将遗传信息传给子代;在个体发育过程中,遗传信息如何表现出来。

1900年以来实验遗传学的发展,特别是五十年代以来分子遗传学的发展,部分地解答了这个问题。

遗 传 物 质

孟德尔通过豌豆杂交实验,在1866年提出遗传因子独立分离和自由组合的遗传法则,为遗传的粒子理论提供了科学依据。但孟德尔的遗传因子(基因),只是统计上的抽象单位。本世纪初,由于细胞学上有丝分裂、减数分裂、受精过程等一系列重要发现和实验遗传学的发展,并把这两方面的研究成果联系起来,建立了遗传的染色体学说。摩尔根用果蝇的许多突变体进行杂交实验,并且研究染色体的行为,认为染色体上呈直线排列的基因,才是遗传性状的物质载体,进一步把染色体学说发展为基因理论。然而无论染色体学说或早期的基因理论。都不能回答这种遗传物质载体的化学本性是什么的问题。

1928年Griffith发现肺炎球菌的遗传转化现象。他用60℃的高温杀死光滑型肺炎球菌,并将它和粗糙型肺炎球菌一起注射到小鼠体内,能使粗糙型定向地转化为光滑型。1943年Avery等分析了肺炎球菌转化因素的化学性质,证明转化因素就是脱氧核糖核酸(DNA),第一次确切地证实DNA是遗传信息的物质载体。

细菌病毒(噬菌体)的结构简单,由蛋白质和核酸两类物质组成。用 ^{32}P 标记T2噬菌体的DNA核心, ^{35}S 标记其蛋白质外壳,然后观察什么成分进入宿主细胞,什么成分遗传给后代,结果证明DNA是遗传物质。现在可以把纯净的病毒核酸引入细胞,使它在细胞内增殖,产生千百个子代病毒颗粒(转染)。

有些病毒没有DNA,只有核糖核酸(RNA),对于这些病毒来说,RNA就是它的遗传物质。例如烟草花叶病毒(TMV),除了一条含有6,400核苷酸的RNA核心以外,还有大约2,150个蛋白质分子组成的外壳。将甲、乙两品种的RNA和蛋白质分别提纯,相互交换后,重建成新的病毒颗粒,其后代的性状,取决于其RNA来源,证明RNA是其遗传物质。

在高等动植物方面,所有动植物都有染色体,凡染色体都有DNA;其DNA含量与

11059.23

• 1 •

染色体的倍数成比例；其他有自主遗传特性的细胞器，如线粒体、叶绿体，也都有 DNA；核酸在代谢上比较稳定，只要细胞健康地活着，就不会失去。虽然目前还不能象细菌那样，用高等动植物来进行转化、转染实验，但在无细胞体系中，同样证明 DNA 携带着遗传信息。

在上半世纪，许多遗传学家相信蛋白质可能是遗传物质。现在持这种看法的人很少了。从蛋白质的结构上看，它不可能成为遗传物质：在 20 种氨基酸中，有些氨基酸的侧基相似，容易混淆；在多肽链的某些部位插入(或缺失)一个氨基酸，往往仍能保持很高的活性，遗传学上的“嗅觉”不灵敏；蛋白质缺乏适于精确复制的遗传表面。

一类具有致病能力的类病毒(如马铃薯纺锤块茎病原)，本身只是很小的一个 RNA 分子，根本没有蛋白质，而类病毒也有遗传现象。这是从另一个角度证明核酸是遗传物质。

遗传物质的结构

DNA 和 RNA 这两种遗传物质的化学结构相似，都是由 4 种核苷酸连接起来的很长很长的大分子。每一种核苷酸都由碱基、核糖和磷酸三部分构成。无论 DNA 或 RNA 其磷酸部分都是一样的。DNA 的核糖是 D-2-脱氧核糖，与 RNA 的 D-核糖不同。DNA 和 RNA 都有 4 种碱基：其中 3 种相同，即腺嘌呤(A)、鸟嘌呤(G)和胞嘧啶(C)；另一种不同，在 DNA 是胸腺嘧啶(T)，在 RNA 是尿嘧啶(U)。核苷酸与核苷酸之间通过 3', 5'-磷酸二酯键连接成为长链，称为多核苷酸。

绝大多数的 DNA 由两股多核苷酸构成，两股都以同一个轴为中心，向同一方向(右旋)盘绕成为规则的双螺旋，其中磷酸和核糖组成双螺旋的主链，碱基成对排列在两股之间。由于连接在嘌呤环和嘧啶环上的氮(N)经常是氨基(NH₂)的形式，只有在极少数情况下，才是亚氨基(NH)的形式；而和 G、T 的 6 位 C 连接的氧经常是酮式(C=O)，只有在极少数情况下才是烯醇式(COH)。所以 A 总是通过两个氢键和 T 配对，G 总是通过三个氢键和 C 配对。碱基成对配合，意味着 DNA 两条链上的碱基并不相同，而是互补。知道一条链上的核苷酸顺序，就可以准确指明另一条链的相应顺序。一条磷酸-核糖主链的走向和另一条主链的走向相反，把其中一条颠倒过来，则和另一条完全一样。1953 年 Watson 和 Crick 阐明了 DNA 的双螺旋结构，是生物科学上非常突出的成就，它开创了分子遗传学的新纪元，也开辟了分子生物学的新领域。

RNA 结构的基本特征和 DNA 相似，不过绝大部分都是单股的。单股 RNA 病毒的种类很多；但也有少数 RNA 病毒是双股 RNA，如呼吸道肠道病毒中，至少有两种是双股 RNA。

DNA 也有单股的，例如噬菌体 ϕ X174，M13、S13、f1 和 fd 的 DNA，都是环状单股 DNA。植物病毒中也有单股 DNA 病毒，例如玉米条纹病毒和木薯潜伏病毒。

在核酸分子上，除编码各种遗传信息的基因之外，还有调节控制基因活动的区域和信号。核酸的两端，以及基因与基因之间，甚至基因之内，都有长短不同的、无编码意义的核苷酸顺序。所有这些，都属于遗传信息的范畴。在真核生物中，DNA 与组蛋白等组成染色质，染色质和非组蛋白蛋白质组成中期染色体，在其 DNA 分子中，除单一顺序的结构

基因外,还有大量的重复顺序。重复顺序的生物学意义如何,现在正在探索之中。

核酸的分子很长,而细胞或病毒的外壳却很小,所以核酸分子必须折叠起来,才能包装到细胞或病毒颗粒中去。细菌的 DNA 折叠以后,还要在中心部分用 RNA 束捆起来。高等动、植物染色质、染色体的结构就更为复杂。当前正在集中研究核小体(nucleosome 或“nu” particles 或 Nu body)的精微结构。从不同的生物、不同的组织制备的核小体,与葡萄球菌或小球菌的核酸酶消化所得的核心颗粒,其结构基本一致:呈扁柱形,直径约 100 Å,高约 50 Å;含组蛋白八聚体和 140 bp(碱基对)的 DNA;组蛋白在内,DNA 在外;内部结构不明。

核酸的复制

组成双股 DNA 的两条多核苷酸链是互补的,碱基之间只以氢键相联。氢键是弱键,所以两股容易分开;碱基之间的氢键又是高度特异的。碱基之间形成氢键的能力,是理想的模板力量。互补的两股分开以后,以每一股为模板,根据碱基之间形成氢键的能力,并且受到两条主链之间空间的限制,在 4 种碱基之中,只能选择一种正确的碱基,由 DNA 聚合酶合成互补的多核苷酸链,和亲本股共同组成子代 DNA 分子。以 ^{15}N 标记和密度梯度离心的方法,证明了 DNA 的复制机理,是一种所谓半保留的复制。

DNA 分子很长,在子代 DNA 分子完全合成之前,两股并不完全分开,而只是部分解缠。每一股多核苷酸链上大约结合了 200 个解缠蛋白质,每个蛋白质约占 8—10 个核苷酸的长度,成为总长度约 2,000bp 的复制叉。随着复制过程逐渐向前推进,复制叉逐渐向前移动。

在原核生物中,复制都由特定的起点开始,向一个方向或同时向两个方向进行。T7 噬菌体直线 DNA 分子的复制起点,并不在其一端,而在距离左端大约 17% 处。噬菌体 ϕX174 、 λ 和质粒 ColE1 复制起点附近的核苷酸顺序已经测定。真核生物 DNA 的复制,往往有许多起点,在一条染色体上,有许多复制子。

已知大肠杆菌和枯草杆菌都有 3 种 DNA 聚合酶,即 DNA 聚合酶 I、II 和 III。在低温条件下合成 DNA,并且用放射性同位素标记新合成的约 20 个核苷酸组成的多核苷酸。大肠杆菌 3' 核酸外切酶能迅速消除快速标记的放射性;枯草杆菌的 5' 核酸外切酶则否。这说明 DNA 聚合酶只能按 5'→3' 方向催化合成。由于 DNA 的两条多核苷酸方向相反,所以两条子链的合成方向必然也是相反的。然而从总体上看,复制叉在两股之间进。粗看起来,一股是 5'→3',另一股却象是 3'→5'。后来知道后者是先反向合成长约 1,000 个核苷酸的片段,然后由 DNA 连接酶连成整体,所以并不矛盾。

大肠杆菌体系的 DNA 合成,主要由 DNA 聚合酶 III 负责。片段之间的空隙由聚合酶 I 负责填补。聚合酶 II 似乎并不负担专一性任务。

任何一种 DNA 聚合酶都只能在已有的多核苷酸链上添加核苷酸,而不能启动 DNA 合成。新合成的 DNA,其 5' 端往往有一小段 RNA。所以 DNA 复制的起点,不一定由 DNA 聚合酶去识别,而是由 RNA 聚合酶去识别的。DNA 合成往往不是由 DNA 牵头、而是由 RNA 牵头的。特异的 RNA 聚合酶以 DNA 为模板,转录一条 RNA,然后由 DNA 聚合酶在此 RNA 的 3' 端继续添加脱氧核糖核苷酸。起牵头作用的 RNA,以后

由核糖核酸酶 H 或 DNA 聚合酶 I 的 5'→3' 核酸外切酶活性切去,再由聚合酶 I 的 5'→3' 聚合酶活性催化填补空隙。已知噬菌体 M13、 ϕ X174、fd, 大肠杆菌素因子 (Col E1), 多瘤病毒 DNA, 其 DNA 复制都用 RNA 牵头。噬菌体 G4 也用 RNA 引发 DNA 合成,但其 RNA 引物是由大肠杆菌 dna G 蛋白质催化聚合的。人体淋巴细胞的 DNA 合成,也以 RNA 为引物。

由于 A 和 T 配对、G 和 C 配对这种极好的模板力量,在复制过程中,一般不容易发生错误。由于错误碱基不能和亲本股形成氢键,不能建立起互补结构,因此能够活动,这就成为配对错误的标志。3 种 DNA 聚合酶都有 3'→5' 核酸外切酶活性,这种外切酶活性就是优先对付配对错误的碱基的。一旦加上了错误的碱基,会有很高的机率在加上下一个碱基之前就把它切除,起着“校对员”的作用。从化学上看,碱基误配的机率不会低于 10^{-5} ,但同一个聚合酶参与两次反应,一次催化合成,一次辨别正误,所以总的错误机率只有 $10^{-5} \times 10^{-5} = 10^{-10}$ 。据以往观察,大肠杆菌 DNA 复制碱基误配的机率,的确只有 $10^{-9} - 10^{-10}$ 。

以上所述,只是有关 DNA 复制的梗概,实际情况要复杂得多。现在已经能用从原核生物提取的酶,在试管中重演 DNA 复制的许多反应。这些研究结果说明 DNA 复制是相当复杂的过程,其间有许多蛋白质参加。DNA 复制之所以如此复杂,其目的在于保证复制品高度精确。复制需要消耗能量,ATP 是提供能量的分子。

尽管 DNA 是遗传物质,但在 DNA 复制过程中,已知有十几种蛋白质参加,还要由 RNA 牵头;而这些蛋白质和 RNA,又都是由 DNA 编码的。这就显示出 DNA、RNA 和蛋白质三者在生命物质运动中固有的辩证关系。

单股 DNA 先以亲本链即(+)链为模板,合成一股(-)链,成为双股中间体,再以(-)链为模板,合成大量子代(+)链。 ϕ X174 就是如此。

RNA 复制的原理和 DNA 复制相同,也是根据碱基配对的原则,先以亲代(+)链为模板,合成(-)链;再以(-)链为模板,合成大量子代(+)链;但不形成双股中间体。RNA 噬菌体 MS2 和 R17 都是如此。

双股 RNA 复制时,先以其一股为模板,合成一股互补链;新链和旧链分开;再以新链为模板,合成另一股,从而成为双股子代 RNA。

某些单股 RNA 肿瘤病毒,先由反转录酶反转录为与其互补的单股 DNA,进一步复制得到双股 DNA,再由双股 DNA 转录成 RNA,作为子代的遗传物质,装配到子代病毒颗粒中。

中 心 法 则

恩格斯在《反杜林论》中写道:“生命是蛋白体的存在方式”。可见在生命活动中,蛋白质起着非常重要的作用。在简单的生物中,除核酸以外,另一个成分就是蛋白质。蛋白质构成生物体的许多部件,生物体的许多部件由酶催化合成,而所有的酶都是蛋白质。

蛋白质是由 20 种氨基酸通过肽键连接起来的大分子。由于氨基酸的排列次序不同,蛋白质的种类千变万化。一个由 150 个氨基酸组成的蛋白质,理论上有 $20^{150} \approx 10^{195}$ 种排列。在 20 种氨基酸中,约有一半是疏水的,它们在水相中趋于集中;另约一半是亲水的,

水在相中趋于分散。因此,氨基酸顺序一定的多肽链,在水相中会自动折叠起来,成为具有一定立体结构的蛋白质。在正常生理条件下,蛋白质的一级结构(氨基酸顺序)一旦确立,其三级结构也就相应地建立起来。蛋白质分子之间,或蛋白质分子和其它组分之间,通过极为专一的识别过程,发生非共价键的相互作用,构成更大的、更为有序分子集体,乃至构成细胞器和细胞。例如把几十种核糖体蛋白质和三种核糖体 RNA 一起放在试管中,它们会自动组装成为功能正常的核糖体。细菌细胞有几千种大分子,但构成噬菌体的各种分子,仍然能彼此识别,按严格的顺序,组装成为与其前代完全一致的噬菌体颗粒。这些都是在分子水平和亚显微水平渐成过程的例证。复杂的多分子的建造方案,可能包含在其各个组件的结构之中,但只是在它们组装起来以后,才能成为现实。细胞的分化和形态的建成,很可能也是蛋白质多重微观相互作用的结果。因此,生物的结构和功能,原则上都可以用蛋白质的氨基酸顺序加以分析说明。

生物的遗传信息贮存在 DNA 分子之中,而生物的种种性状,却是通过蛋白质表现出来的。然而 DNA 和蛋白质并不发生直接的联系,无核的细胞仍然能够短期合成蛋白质,说明 DNA 本身不直接参与蛋白质合成的过程。在蛋白质合成的无细胞体系中,加入 F2 噬菌体 RNA,可以合成 F2 外壳蛋白质,足见 RNA 是合成蛋白质的信使。在生命过程中,信息由 DNA 经 RNA 流向蛋白质的法则,称为中心法则。致癌病毒的 RNA,能通过反转录酶转录成 DNA,说明信息可以倒流,丰富了中心法则的内容。目前,由中心法则阐明的基本思想,已经渗入分子生物学的各个领域。1977 年有人从理论上讨论了反翻译的可能性,但尚未经实验证明。

遗 传 密 码

组成核酸的碱基虽然只有 4 种,但是核酸分子很长,这 4 种碱基可以按各种顺序排列,其所携带的遗传信息,是非常丰富的。例如马铃薯纺锤块茎病原,虽然只是~225 个核苷酸的一条 RNA,但按理论计算,225 个 4 种不同的碱基,应该有 $4^{225} \approx 10^{135}$ 种排列顺序。生物的遗传信息,用 4 种核苷酸作为 4 种符号,以密码的方式,记录在核酸分子上。遗传信息通过蛋白质表现出来,蛋白质相当于 20 个汉字编写的句子。在基因表达过程中, DNA 密码先转录成 mRNA (信使核糖核酸)密码,然后把 mRNA 密码解读为蛋白质的氨基酸顺序,最后表现为性状。研究遗传密码的编码机理,就是要了解 4 种符号的密码如何为 20 种氨基酸编码的问题。

五十年代对编码机理有过各种各样的推测。六十年代初,以原黄素诱发大肠杆菌 T4 噬菌体 rII 区突变系统的“+”(插入一对碱基)、“-”(缺失一对碱基)突变;用各种“+”、“-”突变体进行重组实验,观察重组体的性状表现,发现邻近部位的“+”、“-”突变,可以相互抑制突变表型;邻近部位 3 个“+”突变或 3 个“-”突变,可以部分地恢复野生表型。这表明密码子由 3 个碱基组成,每一组碱基三联体决定一个氨基酸,此即所谓三联密码。

三联体是不重叠的。假如是重叠的,那么一个碱基的变化,将会引起蛋白质中相邻 3 个氨基酸的变化。实验证明一个碱基的变化,只会引起一个氨基酸的变化,证明密码是不重叠的。

六十年代初,建立了离体合成蛋白质的实验技术。这样就可以用已知其碱基顺序的

人工信使,来研究编码问题。最初发现的是 PolyU(多聚 U) 特异地促进苯丙氨酸掺入多肽链。证明 UUU 是苯丙氨酸的密码子。其后数年间,将全部 64 种密码子一一破译:其中有 3 种不编码任何氨基酸,而作为蛋白质合成的终止信号(终止密码子);其余 61 种三联体,分别为 20 种氨基酸编码。为一种氨基酸编码的密码子,少的只有一种,多的可达 6 种。几种密码子为同一种氨基酸编码,称为简并。

氨基酸由特异的氨酰基合成酶催化两步化学反应,以共价键连接在特异的 tRNA(转移 RNA) 分子上。tRNA 的反密码子根据碱基配对原理,识别 mRNA 的密码子,将它解读为正确的氨基酸。反密码子 5' 端碱基的空间位置并不十分固定,能够非特异性地和密码子 3' 端的两种、甚至 3 种碱基配对。这是密码简并的分子基础,称为摆动假说。

有两个密码子兼负起始密码子的作用,它们是 AUG 和 GUG。AUG 编码甲酰甲硫氨酸和甲硫氨酸。原核生物的蛋白质合成从甲酰甲硫氨酸开始,甲酰甲硫氨酸 tRNA ($\text{tRNA}^{\text{f-Met}}$) 识别起始密码子 AUG。甲酰基是在甲硫氨酸接上 $\text{tRNA}^{\text{f-Met}}$ 以后才加上去的。甲硫氨酸 tRNA (tRNA^{Met}) 识别起始密码子 AUG 以外的、在 mRNA 内部的其他 AUG。真核生物也用 AUG 作为起始密码子,但甲硫氨酸的氨基不用甲酰基封阻。真核生物也有两种甲硫氨酸 tRNA。在原核生物中,究竟是 $\text{f-Met-tRNA}^{\text{f-Met}}$ 还是 mRNA 先和核糖体小亚基结合,还不清楚。在真核生物中, $\text{Met-tRNA}^{\text{Met}}$ 先和 40-S 亚基结合成复合体,再和 mRNA 结合。真核生物的 $\text{Met-tRNA}^{\text{Met}}$ 能被细菌甲酰化而成为 $\text{f-Met-tRNA}^{\text{Met}}$ 。GUG 编码缬氨酸,也可作为起始密码子,但起始的效果较差。

许多诱变剂能诱发遗传密码发生歧义突变,因而所编码的氨基酸也随着发生变化。从这两种变化的对应性,可以验证我们理解编码机理的正确性。发生无义突变,则使合成蛋白质中途停止,未完成的多肽往往没有活性。正常的终止密码子突变为氨基酸的密码子,则使正常的多肽链进一步加长。例如正常人血红蛋白 α 链共 141 个氨基酸,第 142 位三联体 UAA 是终止密码子;倘若此三联体的 U 因突变而成为 C,则第 142 位译为谷氨酰胺;顺次往下译,多读了 93 个核苷酸,多译出 31 个氨基酸,得到异常的血红蛋白 α 链,这就是“常春”血红蛋白。

tRNA 的反密码子也可以发生突变。如果某 tRNA 的反密码子突变以后,能和终止密码子配对,则能读出无义密码,从而抑制无义突变的突变表型。如果由于突变而得到四联反密码子,则可抑制相隔较近的插入突变。迄今尚未获得抑制缺失突变的突变反密码子。这说明 tRNA 可以一次解读 3 个或 4 个碱基的顺序,但不能一次解读两个碱基的顺序;也说明碱基按三联一组在核糖体上移动,是因为 tRNA 的反密码子由 3 个碱基构成的缘故。

各种生物,从病毒一直到人,凡是其编码机理经过详细研究的、证明都使用同一种密码。遗传密码很可能是生物界通用的。

遗传信息不单指密码子。密码子只有联系上下文才有意义。核酸分子上除为各种氨基酸编码的区域以外,还有各种调控区域和接收各种控制信号的区域。这些也属于遗传信息的范围。广义的遗传密码学说,不仅包括遗传物质的化学结构及其所负载的遗传信息等等内容,而且也包括形态学、生理学上表现遗传信息的分子机理,因此是生命过程的普遍性理论。

前已提及,密码子是不重叠的,但遗传信息却是可以重叠的。同一截多核苷酸,根据

阅读框架的不同,可以解读为不同的意义。就目前所知,遗传信息重叠有两种情况:一是甲基因内包含着另一个基因乙;二是前一基因终止密码子 3' 端作为后一基因起始密码子 5' 端。

基 因 表 达

基因表达的第一步是转录,由 DNA 指导 RNA 聚合酶,把 DNA 遗传密码转抄成 RNA 密码。转录的基本原理和 DNA 复制相似, DNA 双螺旋部分解缠,以其中一股为模板,根据碱基配对的原理选择碱基,由 RNA 聚合酶把核糖核苷酸以 3',5' 磷酸二酯键连接在一起,合成多核苷酸的长链。

从整个 DNA 分子来说,两股都可作为转录 RNA 的模板。但就各个基因来说,则有些基因用这一股为模板,另一些基因却用另一股。在环状 DNA 分子中,有些基因按顺时针方向转录,另一些按逆时针方向转录。正确选择 DNA 链和正确选择转录方向的机理,还不清楚。

细菌 RNA 聚合酶由 5 个亚基组成;其中 4 个(即 $\alpha_2\beta\beta'$) 结合得比较紧密,称为核心酶;另一个 σ 因子容易分开;可能还有其他因子。 σ 因子负责识别 DNA 模板上起始转录的信号。转录开始以后, σ 因子脱下,再去为别的核心酶寻找起始部位。核心酶催化形成磷酸二酯键,把一个个核糖核苷酸加到 RNA 的 3' 末端上。所有 RNA 都按 5'→3' 方向延伸。细菌受噬菌体侵染后,其 RNA 聚合酶的亚基结构,往往发生规律性的变化。

DNA 分子上启动子 5' 端的十几对核苷酸是 RNA 聚合酶的识别部位。至于是由 σ 因子独立负责识别,抑或 σ 与核心酶结合,改变核心酶的构型,间接起到识别的作用,还不清楚。RNA 聚合酶在此处和 DNA 结合以后,向前进入一段 A-T 碱基对特别多的结合部位,将 DNA 局部解缠,结合得更为紧密。已知 T7、 λ 噬菌体 SV40、病毒和大肠杆菌操纵子等 9 个启动子的紧密结合部位,都有 5' TAT Pu ATG 3' 的共同顺序。由此再往前 5—6 对核苷酸,就是转录的起点。

转录终止有两种办法,一种是由 ρ 因子认出终止信号。 ρ 必须处于多聚状态,才能起作用,可见它所识别的是对称顺序。另一种办法是由 RNA 聚合酶自己识别终止信号;这类终止信号,往往是一连串的 A。

已知真核生物有 3 种 RNA 聚合酶。其中 RNA 聚合酶 I 负责转录核糖体 RNA (rRNA), RNA 聚合酶 II 负责转录 mRNA, RNA 聚合酶 III 负责转录 tRNA。据目前已有的资料看来,最初转录的 mRNA 很长,称为前 mRNA。转录完了以后,立即将其 5' 端绝大部分降解为游离核苷酸,只将其 3' 端约 10% 保留。经过切割后的 5' 端,还以 5'—5' 缩合的方式加上鸟嘌呤,形成一个一般结构为 $5'G^p p p 5'N^p p \cdots$ 的帽子,并把这个 5' 末端鸟嘌呤和下一个核苷酸的 2'-OH 都甲基化。5' 端以 $m^7G p p p G^m$ 封阻的真核生物 mRNA,其稳定性好,翻译效率高;而用 m^7GMP 或 m^7GDP 封阻都能抑制翻译。此外,在其 3' 端要由 PolyA 聚合酶加上 100—200 个腺嘌呤核苷酸组成的 PolyA。PolyA 可能与 mRNA 的寿命有关,没有 PolyA 的 mRNA 易受破坏,作业时间短。不过已知组蛋白的 mRNA 是没有 PolyA 的。前 mRNA 经过这么切割加工之后,成为 mRNA,穿过核膜进入细胞质,行使功能。原核生物的 mRNA,毋须这种切割加工过程。

在真核生物中,有些结构基因并不连续;一个基因内部往往有若干无编码意义的“间隔”区把结构基因分割为几截。而成熟的 mRNA,却不包括那些“间隔”区。

真核生物 mRNA 和原核生物 mRNA 的另一个重要区别,在于有两种蛋白质和前者结合,一种和 5' 端结合,一种可能和 3' 端结合。原核生物则无。

基因表达的第二步是把 mRNA 的核苷酸顺序,解读为蛋白质的氨基酸顺序(翻译)。

翻译在核糖体上进行。氨基酸不直接和 mRNA 接触,而先以共价键和特异的 tRNA 相接,形成氨酰基 tRNA,由 mRNA 的密码子,根据碱基配对的原理,选择正确的反密码子,从而选择正确的氨酰基 tRNA,选择正确的氨基酸。氨基酸通过 tRNA 结合在核糖体上,由相邻两个密码子编码的氨基酸,在核糖上彼此以肽键连接起来。mRNA 每次移动 3 个核苷酸,就在肽链的羧基末端加上相应的一个氨基酸,直到 mRNA 上出现终止密码子,完成一条多肽链为止。解读遗传密码的工作,如果由 mRNA 和氨基酸通过直接“对话”来完成,那末由于 mRNA 和氨基酸的结构迥然不同,在技术上将是十分复杂的;通过“译员”tRNA 来完成,那末只要为某一种氨基酸配备一种特异的氨酰合成酶和一种 tRNA,把识别氨基酸的任务交给 tRNA 的反密码子,问题就迎刃而解了。

mRNA 长短不一,随基因大小而定。翻译由 mRNA 的 5' 端开始,向 3' 端进行。在 5' 端起始密码子之前,常有一段“先导”顺序,此处并不译出。“先导”顺序之后,紧挨着起始密码子 AUG (或 GUG) 之前,有核糖体结合部位,核糖体在此处和 mRNA 结合。此处约 8—13 个核苷酸中,有一部分(如 AGGA 以及其他若干核苷酸)可以和 16-S rRNA 3' 端部分顺序配对,或许可以解释这种结合过程。

核糖体大、小亚基的界面上有两个物理功能部位,即氨酰基 A 位和肽基 P 位。据称可能还存在第三个部位,称为识别 R 位。氨酰基 tRNA 先进入识别部位,经过一个转辙,才能到达 A 位。但在转辙过程中,核糖体和 mRNA 并不发生相对运动。头一个进入译码位置的密码子是起始密码子 AUG (或 GUG)。随着 AUG (或 GUG) 移动到 P 位处, tRNA^{f-Met} 直接进入 P。此后,第二个以及其余各个氨酰基 tRNA 鱼贯进入 A 位。处于 P 位的氨酰基 tRNA (或肽基 tRNA) 和处于 A 位的氨酰基 tRNA,其氨基酸组分由肽基转移酶(很可能是 50-S 亚基的一个蛋白质)催化形成肽键,于是肽链延伸一个单位。下一步是 P 位的 tRNA 释出, P 位腾空;延伸因子 G(EF-G, 常称移位酶)把肽基 tRNA 从 A 位带到 P 位, A 位空出;再下一个密码子和再下一个氨酰基 tRNA 就座。按照这样的程序,根据 mRNA 所负载的密码,把氨基酸依次连接起来,完成译码工作。

在翻译过程中,除上述各种成分外,还需要起始因子、延伸因子、释放因子等多种因子参加,并都起着重要的作用。翻译过程需要消耗能量,鸟苷三磷酸是提供能量的分子。

原核生物的一条 mRNA 分子上,有许多核糖体顺次同时工作,形成多核糖体。核糖体完成一条多肽之后,立即降解为大、小两个亚基,再去从头开始新一轮工作。新合成的多肽,往往在未脱离核糖体以前,就已折叠成形,可以表现特异的功能。mRNA 完成任务之后,迅即被核糖核酸酶降解。

在蛋白质合成过程中,同一氨酰合成酶先催化合成氨酰腺苷酸,再催化合成氨酰 tRNA,有两次选择的机会,起了“校对员”的作用。延伸因子 EF-T 促进氨酰 tRNA 进入 A 位也分两步:第一步和氨酰 tRNA、鸟苷三磷酸结成复合物;第二步才把氨酰 tRNA 植入 A 位。进行这两步反应时,反密码子都和密码子结合着。如果第一步反应有误,进

行第二步反应时还有纠正错误的机会,也起着校对员的作用。由于有这两次“校对”,所以很少错误。

许多蛋白质在合成以后,往往还要经过修饰,才能表现活性。例如胰岛素原有 81 个氨基酸,后切去中间第 31—60 位的 30 个氨基酸,才成为胰岛素。其他多肽在译完以后切去一截的例子很多。还有一些蛋白质,是由几条各自独立转录、翻译得到的多肽共同组成的。

基因表达的调节

任何一种生物,其遗传信息非但要通过转录、翻译,才能表达出来;而且要根据时间、位置、条件的不同,按严格的程序依次表现出来,才能成其为统一的整体。简单的受精卵发育为高度复杂的哺乳动物;一粒种子长成叶茂根深的大树;细菌因所处环境条件不同而改变其酶的合成;噬菌体在其生活周期的不同时期,参与活动的基因各不相同等等,都是基因表现受到精巧调节的结果。对这类带有控制论意义的调节机理,特别是真核基因表达的调节机理,我们已经知道的,还只是极少的一部分。

大肠杆菌乳糖操纵子的调节基因,编码一种调节蛋白质。这种调节蛋白质和操纵区结合,封阻 RNA 聚合酶转录结构基因的去路,因此不能合成 β -半乳糖苷酶,不能利用乳糖。当培养基中葡萄糖减少以致耗尽时,环腺苷酸 (cAMP) 的量增加。cAMP 和降解物基因活化蛋白 (CAP) (分子量~44,000d 的二聚分子) 结合形成复合物,再和启动子结合,促进结构基因的转录。此时如在培养基中加入乳糖(诱导物),则乳糖作为一个信号分子,和调节蛋白结合,使调节蛋白发生变构而失去阻遏的活性,结构基因的转录就能顺利进行。

乳糖操纵子的启动子和操纵区,其核苷酸顺序已得到鉴定。在启动子上, CAP 结合部位在前, RNA 聚合酶结合部位在后。RNA 聚合酶结合部位和操纵区有小部分重合。调节蛋白已经分离纯化,其 343 个氨基酸的顺序,已经鉴定。操纵区中一段由 35 对核苷酸组成的、单股褶叠、自我配对后成为十字形结构的对称顺序,是和调节蛋白结合的部位。切去调节蛋白 N 端 59 个氨基酸,则其残留部分只能和乳糖结合,而不再和操纵区 DNA 结合。

调节蛋白和 CAP 都是控制基因表达的,乳糖操纵子的调节蛋白能阻遏转录,称为负控制。CAP 和启动子结合,能促进转录,称为正控制。CAP 如果因为突变而失去功能,那么即使操纵区没有调节蛋白阻遏转录,仍然不能利用乳糖。其他如半乳糖、阿拉伯糖、麦芽糖也是如此。可见 CAP-cAMP 和促进区结合,不但传达了葡萄糖短缺的信号,而且的确起了促进转录的作用。已经分离到无须 CAP 就能转录 β -半乳糖苷酶基因 (z 基因) 的突变体,由此可见 CAP 或许能够补偿 RNA 聚合酶和启动子相互作用不够完善的缺陷,但补偿的实质不明。

阿拉伯糖操纵子的调节与此有所不同。环境中葡萄糖短缺的信号,通过 cAMP 按乳糖操纵子的方式传达到启动子;阿拉伯糖作为诱导物,和调节蛋白结合成复合物,再和启动子结合,不但传达了培养基中有阿拉伯糖的信号,而且起到促进转录的作用。这个调节蛋白质和乳糖操纵子的调节蛋白(阻遏物)不同,它是正控制的因子;如果因突变而失去机

能,阿拉伯糖操纵子就不能诱导。

在上面所举的两例操纵子中,调节蛋白只有调节功能,并无酶学活性。已知另外一些调控系统,酶本身兼负着调节蛋白的功能。例如鼠伤寒沙门氏杆菌的组氨酸利用操纵子,在有组氨酸时,组氨酸作为诱导物,和调节蛋白(阻遏物)结合,解除它的负控制作用。此时如果培养基中正常碳、氮来源枯竭,则缺碳的信号通过 CAP-cAMP 的复合物传达到启动子;缺氮,则谷氨酰胺合成酶成为非腺苷化的活性形式,和启动子结合,传达第三个信号;组氨酸利用操纵子于是开动起来。谷氨酰胺合成酶本身是把氨(NH_3)同化为有机分子的一个重要的酶。

在这个例子中,必须先由组氨酸解除调节蛋白的阻遏作用,碳、氮短缺的信号才能传达。假如环境中没有组氨酸,开动这个操纵子,生产分解组氨酸的酶,就是毫无意义的了。

色氨酸操纵子的 5 个结构基因连续排列,负责色氨酸的生物合成。操纵区和结构基因之间,有一截 16 bp 的“先导”,“先导”内第 123—150 bp 是削减区,削减区的 5' 端连续有 10 对 GC,其后连续有 7 对 AT。当色氨酸足量时,色氨酸作为辅阻遏物和阻遏物结合成复合物,在操纵区阻断转录,在这种情况下,合成色氨酸的量,比无色氨酸时的合成量降低 70 倍。当色氨酸不够充分时,阻遏解除,转录开始;但能通过削减区、转录下完整 mRNA 的 RNA 聚合酶,只约占 10%;其余约 90% 不能通过削减区,只转录了约 140 个核苷酸的 RNA 片段。当色氨酸严重缺乏时,削减区开放,转录才能顺利进行。所以色氨酸操纵子有两套控制装置,可以根据色氨酸短缺的严重程度,灵活作出反应。削减区的作用机理还不清楚。

把大肠杆菌培养在葡萄糖低限培养基中,它必须开动异亮氨酸生物合成的一系列基因,使苏氨酸经过一系列的酶促反应,合成为异亮氨酸,才能生活下去。异亮氨酸积累多了,过量的异亮氨酸和催化第一步反应的苏氨酸脱氨酶结合,使它发生变构而失去催化活性,其他的酶也就因缺乏底物而依次停止活动。当异亮氨酸减少后,异亮氨酸和第一个酶脱离,又开始合成异亮氨酸。象这样用终产物来抑制第一个酶的催化活性的调节方式,称为反馈抑制。

在噬菌体的生长和溶菌过程中,一般都采用正控制的方式,由前一组基因活动的产物启动下一组基因。溶源状态却采用负节制的方式,调节蛋白质和操纵区结合,封阻大部分基因的转录。对 λ 噬菌体基因表达的调控机理,已经有了相当深入的了解。RNA 聚合酶亚基的更替,在噬菌体基因表达的调节控制上,也起着重要的作用。

RNA 噬菌体如 MS2、R17 等,其 RNA 本身就是信使,基因表达的调节,完全在翻译这一层次进行。MS2 和 R17 都只有 3 个基因,即 A 基因(编码吸附蛋白)、CP 基因(编码壳蛋白)和 REP 基因(编码复制酶)。RNA 进入细胞时,A 基因和 REP 基因的核糖体结合部位,都由于 RNA 的次级结构而被包裹在内,不能启用。第一个译出的基因是 CP。由于翻译 CP,逐渐把 REP 的核糖体结合部位暴露出来,REP 才开始翻译。壳蛋白积累到一定数量以后,和 REP 的核糖体结合部位结合,又把 REP 的翻译封阻。因为所需要的复制酶不多,此时已能满足需要了。A 蛋白质是从子代 RNA 翻译的。当子代 RNA 正在(一)链模板上合成、尚未褶皱,A 基因的核糖体结合部位依然外露时,立即翻译。在子代噬菌体颗粒形成过程中,每个颗粒装配进 1 分子的 A 蛋白质,供下次吸附之用。

还有许多蛋白质,其合成速率不受负、正控制系统的控制,称为组成型合成。但不同

的蛋白质,其合成速率仍然相差很大。这种差别,可能是由于其启动子的核苷酸顺序不同,因而和 RNA 聚合酶的亲和力也不相同造成的。mRNA 半衰期的长短,也是调节蛋白质产量的办法。

λ 噬菌体后期基因转录速率相似,但所合成的蛋白质,在数量上相差达 870 倍之巨。乳糖操纵子 3 个结构基因转录在同一条 mRNA 上,但 z 、 y 、 a 3 个基因的产物,数量上并不相等,呈 10:5:2 之比。

高等动植物细胞的分化,是胚胎发育过程中基因选择性表达的结果,其调控机理如何,目前正在进行研究。

基 因 工 程

所有的生物,不论其结构繁简,体制高低,其遗传密码都是通用的,都根据同一本密码来给各式各样的遗传特性编码。所以原则上可以利用分子遗传学的技术,把不同生物的遗传物质重组在一起,创造新的生物类型,解决工业、农业、医疗卫生和国防上的重大问题。20多年来分子遗传学理论研究上的成就,终于开始应用到实际中来了。

有些限制性核酸内切酶(简称限制酶)能识别特异的对称顺序。对称部位的两条多核苷酸链,如果将其一条旋转 180° ,则这两条链的碱基顺序完全相同。所以限制酶在每一条上都有一个切点,所产生的 DNA 片段具有单股的、碱基互补的粘性末端。不同生物的 DNA 分子,假如具有共同的对称顺序,那末用同一种限制酶切割后,所得到的 DNA 片段,将具有相同的粘性末端。把这些 DNA 片段混合在一起,由于互补的粘性末端会根据碱基配对的原理相互粘合,不同生物的 DNA 片段有可能组合在一起。此时加入 DNA 连接酶,连接起片段之间的共价键,就会得到完整的人工杂种分子。如果用这种办法把甲种生物的基因,粘接在质粒或病毒的核酸分子上,由质粒或病毒带到乙细胞中去,乙细胞就获得了甲种生物的遗传信息。这是当前基因工程经常采用的一套技术。

在不同的原核生物之间,基因工程已经使用得相当普遍,在理论研究和实际应用上都取得了丰硕的成果。在真核生物和原核生物之间,已经把胰岛素基因移入大肠杆菌细胞中。高等植物的固氮基因工程,也在进行之中。生长抑制素(简称 SRIH)是下丘脑中产生的一种 14 肽激素,将人工合成的 SRIH 基因和质粒 DNA 组合在一起,由质粒带进大肠杆菌细胞中去,终于使得大肠杆菌产生 SRIH。这是基因工程在实际应用上的第一次成功。

应用基因工程技术,还为研究真核生物基因的结构、功能,提供了有效的实验手段。已经把几种真核生物的 rDNA(为核糖体 RNA 编码的 DNA)、hDNA(为组蛋白编码的 DNA)、珠蛋白 DNA、免疫球蛋白 DNA、线粒体 DNA、叶绿体 DNA 重组在大肠杆菌质粒中,在大肠杆菌细胞中增殖。通过质粒把果蝇 DNA 片段带到大肠杆菌小细胞中去,有的片段能转录、翻译。

基因工程是七十年代出现的一个新领域。总的来说,还不完善,在方法、手段方面有许多问题有待解决,还不符合严格的工程的意义。当前,首先要发展基因工程技术;其次,基因工程的研究要和分子遗传学、分子生物学、细胞生物学的研究相结合,积累更多有关生命过程的基本知识。这样才有可能按人类的需要在分子水平加工或转移遗传物质,创

造生物的新物种。

结 束 语

生物的宏观结构千变万化,具有无限的多样性,而其微观结构却是惊人地一致的。原核生物的结构虽然比较简单,但是具备生命过程的基因特点。因此,以原核生物为材料进行研究所得到的结果,如果采取正确的观点,进行适当的试验,可以相当准确地判断其中那些同样也适用于真核生物。四十年代以来,利用细菌和噬菌体研究基本的遗传规律,取得了丰硕的成果。基因工程问世以来,原核生物分子遗传实验体系又显示出新的生命力。真核生物的遗传问题,由于其遗传结构高度复杂,缺乏有效的研究手段,进展比较迟缓。发展基因工程技术,可以把真核生物的遗传物质切成小段,引入细菌体内,进行纯系增殖。这样就可以化繁为简,便于研究其结构和功能的关系,促进真核生物遗传学的发展。

分子遗传学虽然只有 20 几年的历史,但它所取得的成就,在生物科学史上是非常突出的。倘若继续保持这样的势头将会有重大的突破,真核生物的细胞分化、胚胎发育和肿瘤病因的问题,都有可能得到解决。

参 考 文 献

- Abelson, J. et al., (1977): *Science*, **196**: 159.
Alberts, B. et al., (1970): *Nature*, **227**: 1313.
Alberts, B. et al., (1977): *Nature*, **269**: 655.
Brullay, D. et al., (1971): *Proc. Natl. Acad. Sci.*, **68**: 2826.
Cook, N. D., (1977): *J. Theor. Biol.*, **64**: 113.
Drake, J. W., (1969): *Nature*, **221**: 1132.
Felsenfeld, G., (1978): *Nature*, **271**: 115.
Hamlyn, P., (1977): *Nature*, **270**: 662.
Kornberg, A., (1974): "DNA Synthesis" Freeman, San Francisco.
Lee, F. et al., (1977): *Proc. Natl. Acad. Sci.*, **74**: 4365.
Losick, R. et al., (ed.) (1976): "RNA Polymerase", Cold Spring Harbor Lab.
Malcolm, A. D. B., (1977): *Nature*, **268**: 196.
Max, J. L., (1977): *Science*, **196**: 638.
Monod, J., (1972): "Chance and Necessity" Collins, London.
Muthukrishnan, S., et al., (1975): *Nature*, **255**: 35.
Ptashne, M. et al., (1976): *Science*, **194**: 156.
Roman, R. et al., (1976): *Nature*, **260**: 359.
Sanger, F. et al., (1977): *Nature*, **265**: 687.
Termin, H. M., (1972): *Sci. Amer.*, **226**: 25.
Topal, M. D. et al., (1976): *Nature*, **263**: 285.
Tseng, B. Y. et al., (1977): *Fedn. Proc.*, **38**: 654.
Watson, J. D., (1976): "Molecular Biology of the Gene" (3rd ed.) Benjamin, W. A. Inc.

第二章 核酸和蛋白质的结构

孙永华

在经典遗传学研究时期,人们就力图探索遗传的物质基础,并且把决定遗传的因子称为“基因”。由于核酸和蛋白质的普遍存在及它们在细胞中的分布,人们自然而然地把注意力集中在这两类生物大分子上。细菌的转化、噬菌体的转导、病毒的重建等一系列实验,令人信服地证明了核酸是遗传的物质基础,在遗传中起着决定作用。首先,核酸携带着遗传信息,并且通过复制把整套遗传信息传给子代,一代一代传下去。其次,通过转录和翻译,把核酸所携带的遗传信息传递并翻译成蛋白质,使遗传信息得到表达。除了各种RNA外,许多蛋白质也参与了转录和翻译及其调节控制,在遗传信息的表达过程中,起着重要的作用。

核 酸

1868年, Miescher 由脓细胞的核分离出一种含磷极多的酸性物质。当时,他把这种化合物称为“核素”,即今天人们所说的核蛋白,从而发现了核酸。以后,发现核酸普遍存在于动物、植物中。Kossel 鉴定了核酸的成份为4种含氮的碱基和五碳糖和磷酸。1920年 Jones 证明了存在两种核酸,就是现在所说的脱氧核糖核酸(DNA)和核糖核酸(RNA)。碱基、糖和磷酸连结成核苷酸;核苷酸是构成核酸的基本单位。1930年左右,证明了核苷酸通过糖环间的磷酸二酯键(图2-6)连接起来,构成核酸。

碱基

DNA 和 RNA 两类核酸,都含有两种嘌呤碱和两种嘧啶碱。嘌呤碱和嘧啶碱是由嘌呤和嘧啶衍生而来的。图2-1给出了嘌呤和嘧啶的结构。DNA 和 RNA 所含的4种碱基中,有三种是共同的其中二种是嘌呤碱,一种是嘧啶碱,就是腺嘌呤、鸟嘌呤和胞嘧啶。

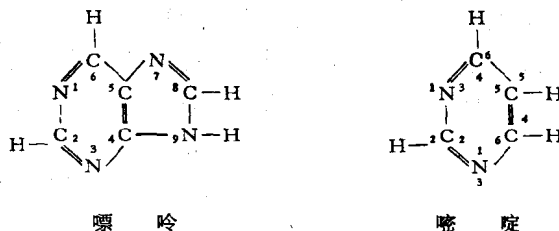


图2-1 嘌呤和嘧啶的结构。

嘧啶环上各原子的标法有两种系统:环内的标法是现在较为普遍采用的标法,嘧啶环外的数字是以前采用的标法