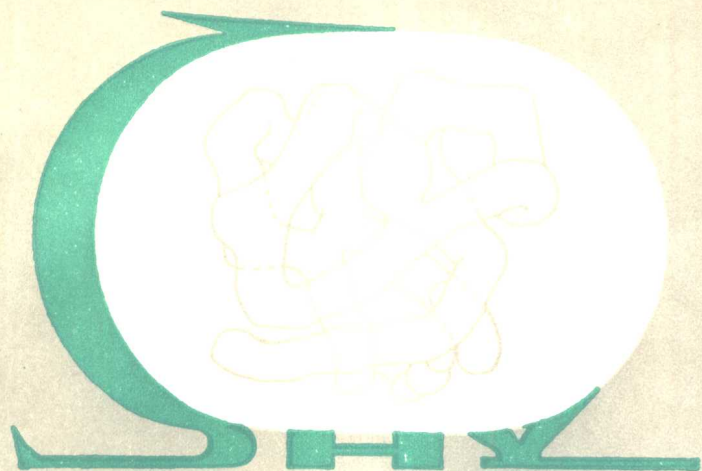




SHENGWUHUAXUEYICONG

生物化学译丛

第 3 辑

生物化学译丛

(第三辑)

《生物化学译丛》编辑部 编

•

上海科学技术文献出版社出版

(上海高安路六弄一号)

新华书店上海发行所发行

上海市印刷三厂印刷

•

开本787×1092 1/16 印张5.25 字数129,000

1981年4月第1版 1981年4月第1次印刷

印数: 1—3,700

书号: 13192·24 定价: 0.70元

《科技新书目》185-129

目 录

卵清蛋白基因区——受激素控制在鸡输卵管表达的三个结构基因的共同特征.....	(1)
利用带重组质粒的大肠杆菌 K12 合成卵清蛋白样蛋白质	(10)
DNA 修复——致癌的新线索	(15)
蛋白质合成的寡聚核苷酸抑制剂.....	(20)
干扰素、双链 RNA 及多型效应因子.....	(27)
细菌质粒的复制.....	(29)
兔肌磷酸果糖激酶的恒态动力学研究.....	(41)
单底物酶反应的综合速度方程.....	(46)
固定化酶活性的控制.....	(53)
利用水溶液两相系统提取酶.....	(60)
选择性抗体薄膜电极.....	(66)
作为生命现象基础的基因复制.....	(68)
术语解释.....	(79)
病毒致癌基因产物的蛋白激酶活性.....	(80)
RNA 排列顺序的化学测定法	(82)

卵清蛋白基因区—受激素控制在鸡输卵管表达的三个结构基因的共同特征

A. Royal 等

用分子克隆的方法,分离了选盖鸡卵清蛋白基因的两个大 DNA 片段。通过这些片段的分析,提出了 46 千碱基对鸡基因组 DNA 区段的图谱。这个区段包含一个完整的卵清蛋白基因(以及它的 mRNA 的前导编码序列)和至少两个其他未知功能的基因。三个基因按同一方向排列,它们在鸡输卵管内表达是受激素控制的。三个基因具有一些序列同源性,表明存在于卵清蛋白基因区的重复性可能是在进化过程中形成的。

在鸡的输卵管内,甾体激素能强烈地刺激包括卵清蛋白的几个基因的转录。认识这些基因结构,有助于了解高等生物中的激素调节和基因控制。在发现卵清蛋白基因是分散的这一现象后,我们用不同的方法,包括鸡 DNA 片段分子克隆(即无性繁殖)的方法,对它的结构作了分析,证明卵清蛋白基因有 7 个内含子(插入序列,见图 1c)。卵清蛋白 mRNA 5'端有一个 47 个核苷酸长的前导序列。前导编码序列在我们早期的基因克隆片段中未曾发现过。

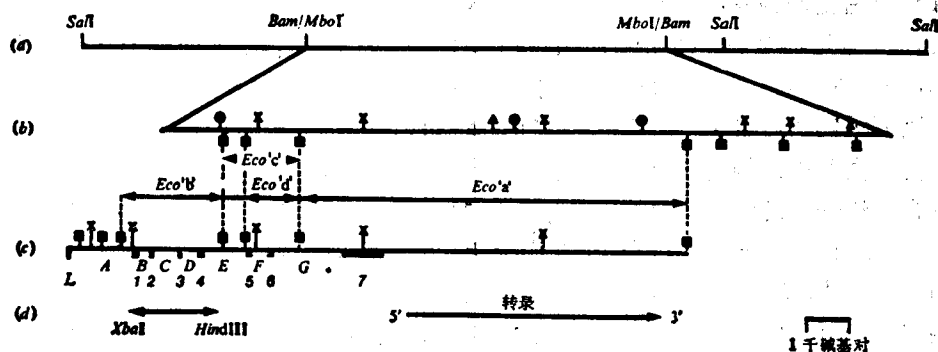


图1 pAR1 DNA 限制性内切酶图谱

pAR1 按以下程序组建:鸡红血球 DNA 用 *Mbo* I 进行部分水解,作蔗糖梯度超离心(10mM Tris-HCl pH 8.1, 1mM EDTA, 1M NaCl, 5~40%蔗糖, 25,000转/分, 4℃, 离心18小时)。DNA 各部分作琼脂糖电泳分析,将15~25千碱基对DNA片段部分合并,用乙醇沉淀浓缩。5 微克 DNA 与等量的事先经 *Bam*HI 水解的 pJC74 Km 连接。体外包装,产生250,000重组体,感染大肠杆菌1400,细菌在卡那霉素平板上分离,用卵清蛋白基因特异的³²P 标记的探针作原位杂交(*Xba*I-*Hind*III 探针如图中 d 行所示,由克隆的卵清蛋白基因 *Eco*RI "b" 片段衍生而来)。检查25,000菌落,得到一个克隆,质粒 pAR1 重新分离三次, pAR1 限制性酶解图谱是 pAR1 DNA 用限制性内切酶作水解片段分析而来的。(a) pAR1 线性图谱,表示17千碱基对的鸡 DNA *Mbo* I 片段通过 *Bam*HI 切点与 pJC74 Km 连接。(b)放大的17千碱基对的鸡 DNA 片段上不同限制性内切酶的切点: *Bam*HI (▲), *Eco*RI (■), *Kpn*I (●)和 *Xba*I (×)。(c) 卵清蛋白基因图,内含子(用 A~G 表示),外显子(用 1~7 表示),前导编码序列(L),分界线表示 *Eco*RI 片段 "a", "b", "c" 和 "d"。(d) 卵清蛋白特异的探针 *Xba*I-*Hind*III 相应的区域,转录方向, b 和 c 的标尺大小。

为了分离卵清蛋白 mRNA 的前导编码序列,我们确定了卵清蛋白基因邻近的结构,决定克隆迭盖卵清蛋白基因的大 DNA 片段。本文报道两个 DNA 片段的分离,这两个片段包含 46 千碱基对的鸡基因组。这个区段包括除卵清蛋白基因和它的前导编码序列以外的其他两个基因,它们在鸡输卵管里表达均受激素调节,并与卵清蛋白基因的一些序列有同源性。

迭盖卵清蛋白基因 DNA 片段的分离

我们先前克隆了四个 *Eco* R I 鸡 DNA 片段(“a”, “b”, “c”和“d”), 它们都带有卵清蛋白 mRNA 的大部分编码序列(见图 1c), 接着又试图用 λ WES 克隆从鸡 DNA *Eco* R I 部分水解片段分离的卵清蛋白基因的邻近区段。这一尝试失败以后,我们决定发展一种能够增殖大片 DNA 的克隆系统。在这些实验完成之前, Collins 和 Hohn 发表了名为“Cosmid”的同样克隆载体。Cosmid 是能够在体外包装到 λ 噬菌体头部蛋白内的质粒载体。在这一反应中,重组体分子优先于母体分子转变为感染颗粒。此种感染颗粒能将 DNA 注入细菌内,这就提高了包含重组片段的质粒克隆效率。有效的筛选方法是从复杂系统中分离基因的前提,我们发展了一种适宜的原位菌落杂交技术。

用了两个 Cosmid pJC74 能接受大到 32 千碱基对的外源 DNA 片段;pJC 74 Km 是抗卡那霉素的 pJC74 的衍生体(见图 1 说明),能接受大到 25 千碱基对的 DNA 片段。鸡红血球 DNA 经 *Mbo* I 部分降解,用蔗糖梯度离心分离一定大小的片段,与 pJC74 Km 的 *Bam* H I 线形分子连接(*Bam* H I 和 *Mbo* I 具有相同的粘性末端)。在另外一些实验里,鸡红血球 DNA 经 *Eco* R I 部分降解,蔗糖梯度离心分离一定大小的片段与 *Eco* R I 作用的 pJC74 线形分子重组。体外包装以后感染 *recA*⁻1400 号细菌,以 ³²P 标记的相应于卵清蛋白 *Eco* R I 片段“b”DNA 为探针,用原位杂交技术进行筛选(见图 1d)。分离了两个克隆片段,并从中分别纯化分离了重组质粒 pAR1(pJC74Km 的衍生体)和 pAR2(pJC74 的衍生体)。

包含卵清蛋白基因的 46 千碱基对克隆区段的序列

图 1 和图 2 为几种限制性内切酶构成的 pAR1 和 pAR2 图谱。pAR1 DNA 被 *Eco* R I 消化,形成 8 个片段,包括一个 9.2 千碱基对的片段和一个 1.3 千碱基对的片段,它们相当于卵清蛋白基因 *Eco* R I “a”和“d”克隆片段(见图 1b, c)。这两个片段的部分图谱进一步证实了先前关于卵清蛋白基因结构的报道,未见到 *Eco* R I 片段“b”大小的片段。但是内切酶构成的图谱表明它有大约一半的序列是存在的。pAR1 就是用相当于那个片段作探针被分离的。这也为 pAR1 DNA 和卵清蛋白 mRNA 杂交实验所证实。在电子显微镜下,可以观察到相当于内含子(intron) D, E, F 和 G 的 DNA 环,以及相当于外显子(exon) 1 和 2 不杂交的 RNA 尾巴。从这些实验中,可认为 pAR1 包含一个 17 千碱基对的鸡 DNA 片段,即从卵清蛋白基因内含子 C 内的 *Mbo* I 切点到卵清蛋白 mRNA 编码序列 3' 端以外的 12.5 千碱基对的片段(图 1)。

pAR2 DNA 被 *Eco* R I 作用产生一系列片段,用卵清蛋白双链环 DNA 为探针,作 Southern 转移实验,证实有卵清蛋白基因 *Eco* R I “b”片段(包含外显子 1~4)的存在,而“a”, “c”和“d”片段(包含外显子 5, 6 和 7)未被检测出。出乎意料,预计不是卵清蛋白基因的 3.5 千碱基对 *Eco* R I 片段和 2.4 千碱基对 *Bam* H I 片段也能与 *Hba*ov 探针反应。同样的能与 *Hha*ov 探针杂交而不属于卵清蛋白基因的 *Eco* R I 和 *Bam* H I 片段先前在细胞 DNA 中也曾检测出。这表明 pAR2 包含有与卵清蛋白基因相关的序列。图 2 (a 和 b)限制性内切酶图谱表示 pAR2 包含位于卵清蛋白基因 5' 端以外的大约 30 千碱基对的鸡 DNA 片段。如后面指出, *Eco* R I “b”片

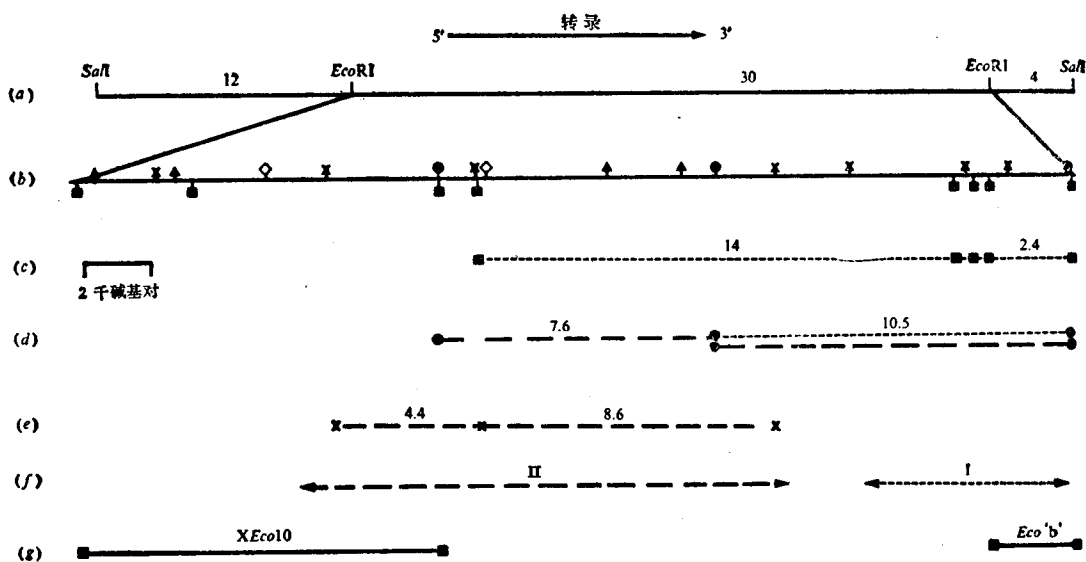


图 2 pAR2 DNA 限制性内切酶图谱

重组体的构建按图 1 程序进行。鸡 DNA 用 *EcoRI* 部分降解, 大于 25 千碱基对片段 (25 微克) 与 15 微克 pJC74*EcoRI* I 线形分子连接。反应总体积 140 微升, 将 2 微克重组体 DNA 体外包装, 感染 1400 细菌, 在氨基苄青霉素平板上分离了 20,000 个菌落, 用 *XbaI*-*HindIII* 探针 (见图 1) 杂交筛选, 获得一个克隆。限制性内切酶图谱是从一系列 pAR2 DNA 或分离片段的水解产物分析中确定的, 片段的鉴定用 *XEco10* 探针 (见后和正文) 或卵清蛋白探针 *XbaI*-*HindIII* (见图 1) 进行杂交。(a) pAR2 DNA 的 *SalI* 线形分子图谱, 表示 30 千碱基对片段插入载体的位置, 线上的数字表示载体或鸡 DNA 片段的千碱基数。(b) 克隆片段限制性内切酶图谱: *BamHI* (▲), *EcoRI* (■), *HpaI* (◇), *KpnI* (●) 和 *XbaI* (×)。(c)、(d) 和 (e), 细胞 DNA 的不同限制性内切酶片段: *EcoRI* (c), *KpnI* (d), 或 *XbaI* (e), 在 Southern 转移实验中, 用探针 I (.....) 或探针 II (——) 鉴定; 各片段上的数字表示千碱基数。(f) 比较 pAR2 和细胞 DNA 所用的探针: 探针 I 相当于卵清蛋白基因前导区的 *psf* I 4.5 千碱基对和 *EcoRI* 3.2 千碱基对的片段, 从 λC4-ov5 分离而来。探针 II 相当于 *HhaI* 14 千碱基对大片段, 与 pBR 322 平头连接, 再次克隆。(g) 卵清蛋白 *EcoRI* "b" 和 10.5 千碱基对 *EcoRI* 片段 *XEco10* 的位置, 后者用 λ WES 单独克隆, 用卵清蛋白 *Hbaov* 探针杂交筛选。(b)~(g) 标尺大小列在 (e)。

段 3' 端的卵清蛋白基因与存在于 pAR2 的另外两个基因 X 和 Y 序列同源。

由于插入 pAR2 是一个比较大的 DNA 片段, 它是用 Cosmid 克隆的第一个真核 DNA 片段, 因此证明这个克隆片段在基因组内的地位是很重要的。用 Southern 转移实验, 对 pAR2 和鸡输卵管 DNA 限制性内切酶片段杂交图谱作了比较。用两种杂交探针 (图 2 f, g); 探针 I 包含卵清蛋白 *EcoRI* "b" 片段的 6 千碱基对的区段; 探针 II 包含 pAR2 克隆 DNA 片段中间的 14 千碱基对的区段。如图 3 所示, 用探针 I 与 pAR2 和输卵管 DNA 的 *EcoRI* 水解片段 (2 和 6 行), *KpnI* 水解片段 (4 和 8 行), *EcoRI* 和 *BmaHI* 双酶水解片段 (1 和 5 行) 或 *EcoRI* 和 *KpnI* 双酶水解片段 (3 和 7 行) 杂交都是同样的。pAR2 和输卵管 DNA 存在两个能与探针 I 杂交的大小相同的 *XbaI* 片段 (10 和 12 行), 探针 II 也能与两种 DNA 的三个 *KpnI* 水解片段反应, 其中两个片段 (b 和 c) 大小一样。在 pAR2 中, 第三个片段大约 26 千碱基对, 其中 16 千碱基对为载体序列 (11 行 a), 相应的细胞 DNA 的 a 片段大约为 18 千碱基对, 其中有 8 千碱基对在 pAR2 内是不存在的 (9 行 a)。如图 2 c, d 和 e 所示, 能与探针 I 和 II 杂交的 pAR2 和输卵管 DNA 共同的片段是一个 24 千碱基对的片段。另外又发现有一个在卵清蛋白基因序

列远端的与 24 千碱基对重迭的鸡 DNA 克隆片段 *XEco10*(图 2g)。从这些结果里,我们认为在 pAR2 上克隆的 30 千碱基对的片段相当于细胞 DNA 内的连结单一片段,而不是在克隆过程中重排造成的。

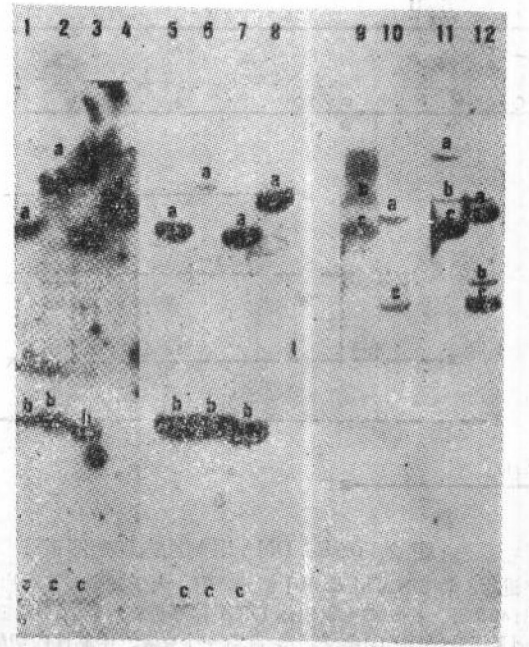


图 3 克隆的 pAR2 DNA 和染色体 DNA 相适应的区段限制性出切酶图

pAR2 DNA 和输卵管 DNA 用不同限制性内切酶水解,作 0.7% 琼脂糖凝胶电泳 (每条胶细胞 DNA 10 微克, pAR2 DNA 10 毫微克), 然后转移到硝酸纤维素滤纸上, 将同一种探针分析的水解产物在同一个胶上电泳, 但细胞 DNA 和 pAR2 DNA 分别转移到不同的滤纸上。与探针 I (图 2) 杂交的输卵管 DNA (1~4 行) 或 pAR2 DNA (5~8 行), 片段大小以千碱基为单位。1 行和 5 行: 用 *Bam*H I 和 *Eco*R I 双酶水解, $a=7.9$, $b=2.4$, $c=0.5$; 2 行和 6 行: 用 *Eco*R I 水解, $a=14$, $b=2.4$, $c=0.5$ (片段 b 上的显影点属污染点); 3 行和 7 行: 用 *Eco*R I 和 *Kpn*I 双酶水解, $a=7.2$, $b=2.2$, $c=0.5$; 4 行和 8 行: 用 *Kpn*I 水解 $a=10.5$ 。用探针 II (见图 2) 与输卵管 DNA (9 行和 10 行) 或 pAR2 DNA (11 行和 12 行) 杂交, 9 和 11 用 *Kpn*I 水解, a (9 行) = 18, p (11 行) = 26, $b=10.5$, $c=7.6$; 10 行和 12 行: 用 *Xba*I 水解, $a=8.6$, $b=5.1$, $c=4.4$; b (12 行) 相当于邻近 c 的 *Xba*I 片段 (在它的左侧见图 2b、c)。由于片段 b 仅仅与探针 II 的很小部分杂交, 在细胞 DNA 中是检测不出的。

pAR2 含有卵清蛋白前导编码序列

如果 pAR2 包含一个卵清蛋白基因 5' 端以外的 DNA 片段, 则预期它具有卵清蛋白 mRNA 的前导编码序列。为此用电镜分析了卵清蛋白 mRNA 和经 *Sal*I 切开热变性的 pAR2 DNA 杂交分子 (图 4d)。卵清蛋白 mRNA 杂交的 620 ± 67 个核苷酸被四个单链 DNA 环分割为 5 段, 其中三个环相当于卵清蛋白基因内含子 B、C 和 D, 从而可以确定分子的转录方向。第四个环位于卵清蛋白 mRNA 的 5' 端, 长约 1.606 ± 150 个核苷酸 (图 4d)。这个环相当于内含子 A, 它就是所预期的从编码 mRNA 前导的序列分割的外显子。实际上 pAR2 上的 A 环长度和几个限制性内切酶切点位置与另外两个单独克隆的相应片段是一样的, 该片段在环的 5' 端存在有前导编码序列已被 DNA 序列分析所证实。mRNA 游离的尾巴长约 1.200 核苷酸, 相当于外显子 5、6 和 7, 它不存在 pAR2 中。

pAR2 包含有在生蛋母鸡输卵管里表达的另外两个基因

在 pAR1 和 pAR2 上克隆的 46 千碱基对 DNA 区段里, 卵清蛋白基因只占约 7.7 千碱基

对,这可能在较大的间隔里存在有其他基因。为搞清楚鸡输卵管里是否有在同一个激素控制下表达的成串基因,将从生蛋母鸡输卵管里分离的含聚A RNA与pAR1和pAR2 DNA杂交,发现除卵清蛋白 mRNA外没有别的RNA可以和pAR1杂交。相比之下,杂交分子的电泳照片显示,除卵清蛋白 mRNA外,还有另外两种RNA能与pAR2的同一条链杂交(图4a)。这表明存在有转录方向相同的三个基因(图4a),转录方向由卵清蛋白 mRNA·DNA杂交区域所确定,而杂交区则通过相当于内含子B、C和D的DNA环的特有排列是很容易确定的(图4a, d,图4a, p 内含子A不存在)。这两个基因分别定为X和Y。基因的排列顺序为:5'-X-Y-卵清蛋白-3'。

测量位于克隆分子中间的DNA·RNA杂交区,Y RNA长度为 $2,220 \pm 175$ 核苷酸,DNA环表明在这个基因中存在有7个内含子(A~G)分成8个外显子(1~8)(图4c和图5a, b)。用Y RNA和用卵清蛋白 mRNA一样,在所有杂交分子中都没有见到近5'端的DNA环存在(实际上在图4a中这两个基因近5'端的DNA环是不存在的)。在卵清蛋白 mRNA情况下,由于前导序列短(47个核苷酸),在RNA·DNA杂交条件下,不能形成稳定的杂交。Y RNA可能是同样的情况,实际上外显子1在电泳照片上是无法分辨的,它的存在只能从A环的存在推断而来(图4c)。Y基因的总长(外显子加内含子)大约为6.5千碱基对,它的3'端到卵清蛋白 mRNA前导编码序列约11.5千碱基对。

Y RNA大约长2,400核苷酸,它与pAR2 DNA杂交长约 $1,945 \pm 200$ 核苷酸,被两个DNA环(内含子)分为三个外显子(图4b和图5a, b)。存在大约400核苷酸长的游离RNA尾巴,表明pAR2不存在这种RNA 5'端的编码区。X基因3'端和Y基因5'端之间长约5.5千碱基对。根据这些资料,X和Y基因如图5c和图6B(1)所示。

X, Y和卵清蛋白基因共有序列同源性

在pAR2的图谱(见前)中,已发现除了卵清蛋白基因序列以外,DNA其他一些部分也能与卵清蛋白特异的Hhaov探针杂交。基因X和Y的发现,促使我们去研究这些基因间能否交叉杂交。

pAR2 DNA用Hpa I和EcoR I或Hpa I和BamH I水解。Hpa I把Cosmid DNA切成很小的片段,有利于只被Hpa I切两次的鸡DNA克隆片段的分析。pAR2用Hpa I和EcoR I双酶水解后产生的DNA片段中有一个与Hhaov探针强杂交显影带(图6A1行c),另两个带杂交显影较弱(1行a和b)。从pAR2限制性酶解图谱中,我们能鉴定这些片段,并确定其在图6B1行上的位置。同样在Hpa I和BamH I双酶水解中,与Hhaov探针杂交的一个强显影带(图6A2行a),为卵清蛋白编码区,另两个显影带(2行b和c)分别表示基因X和Y的编码区(见图6B2行)。用基因X区段作探针(探针V,图6B5行),对这些同源性作进一步分析,在Hpa I和EcoR I水解片段里,获得两个杂交带(图6A3行),a带包含基因Y的一部分,b带含有制备探针的部分区段。用同样的探针与Hpa I和BamH I水解片段杂交,发现三个杂交带(图6A4行),其中两个带(c和d)来源于基因X,而b带相当于基因Y的一部分(见图6B4行)。这些实验表明基因X和Y间有同源性存在,这两个基因也含有一些与卵清蛋白外显子相同的序列。尽管如此,这三个基因和它们的RNA转录产物也有明显的不同,如电泳观察的结果,以及从卵清蛋白EcoR I“b”片段制备的Xba I-Hind I探针不能与Y基因杂交等。

激素控制下X和Y基因的表达

因为卵清蛋白基因的转录是受雌激素控制的。为了解X和Y的表达是否也同样受雌激素

核。这可能在等大的间隔里存在其基因。为清楚起见，我们是否有一个激酶基因

下表中的电报的基因，其基因在等大的间隔里存在其基因。为清楚起见，我们是否有一个激酶基因

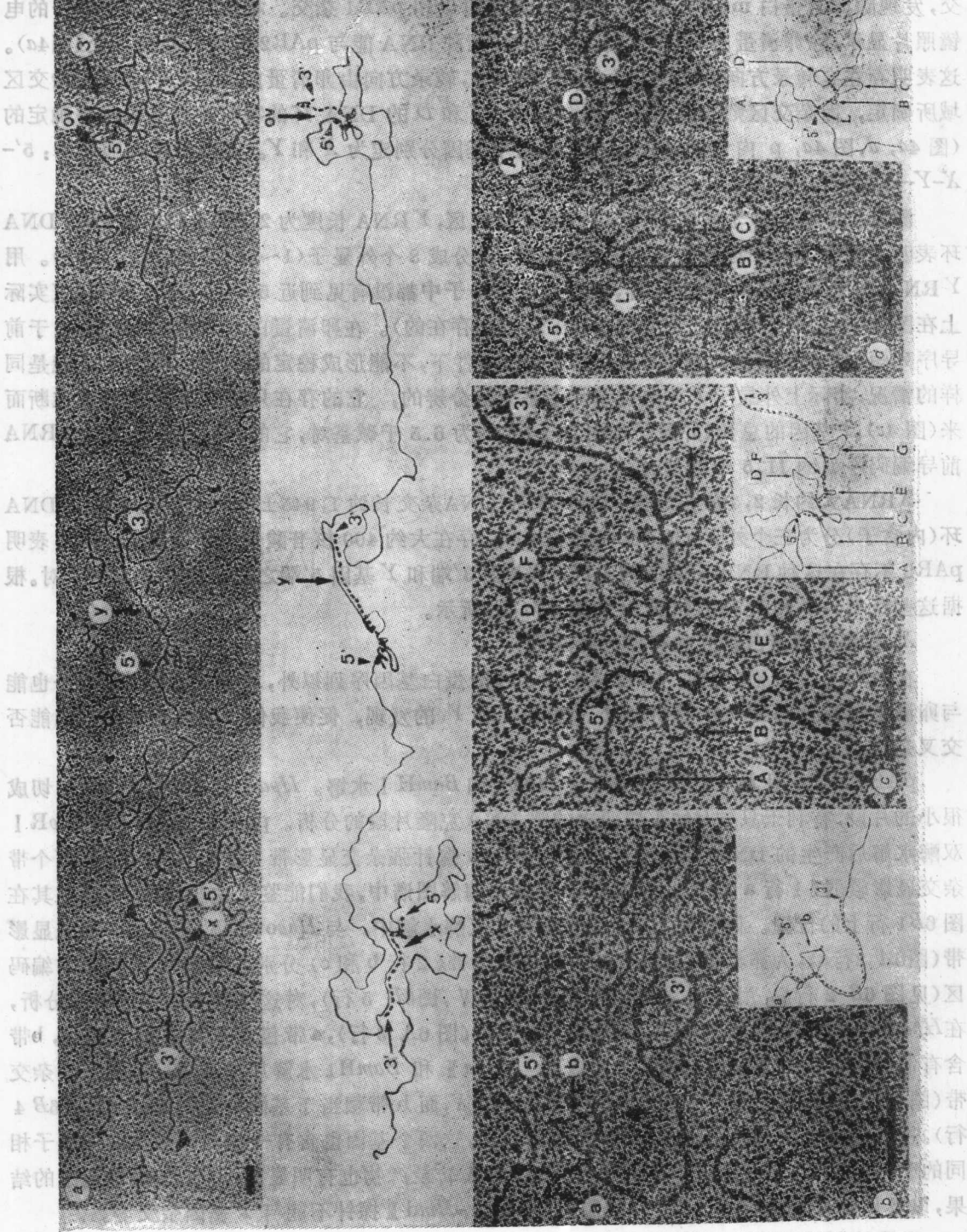


图 1. 激酶基因在等大的间隔里存在其基因。为清楚起见，我们是否有一个激酶基因

的控制,分别从生蛋母鸡或除去雌激素母鸡的输卵管里分离的 RNA,作氢氧化甲基汞琼脂糖凝胶电泳,转移并固定在重氮苯氧甲滤纸上,滤纸再分别与三种基因特异的探针杂交。

如图 7 所示,卵清蛋白基因特异的探针 IV (见图 5d)能与存在于生蛋母鸡输卵管内大约 1,900 核苷酸长的 RNA 杂交(3 行),而不能在去激素的输卵管里检测出(1 行),此 RNA 就是卵清蛋白 mRNA。探针 II (见图 5d)是 X 基因特异的,不能与 Y 基因和卵清蛋白基因杂交,它检测出存在于生蛋母鸡输卵管里的一种 RNA (图 7, 7 行),并不存在于去激素的输卵管里(6 行)。如前面电泳观察到那样, RNA 的大小为 2,300 核苷酸。当用包含 Y 基因的探针 I (图 5d)时,在去激素输卵管里未检测出杂交的 RNA 存在(图 7, 4 行),表明去激素输卵管里的 Y 基因是不表达的。在生蛋母鸡输卵管里,用探针 I 杂交,发现有大约 2,000 核苷酸长的 RNA(5 行),但是我们不能确定它就是 Y RNA,因探针 I 也能与卵清蛋白 mRNA 交叉杂交,其大小也一样。但是电泳观察结果表明, Y RNA 确实存在于生蛋母鸡输卵管里。“去激素”鸡注射雌二醇后,在输卵管里发现有 X 和 Y RNA 的存在,也表明 X 和 Y 基因的转录是依赖于雌激素的。然而图 7 显示的结果和电泳观察的 RNA·DNA 杂交频率都表明, X 和 Y RNA 浓度是大大低于卵清蛋白 mRNA 的。

卵清蛋白基因邻近的鸡基因组的结构

用 Cosmid 载体克隆真核 DNA 片段以前未见报道。本实验结果表明,这一系统配合适当的原位杂交方法是成功的。常用的噬菌体载体最多只能增殖 20 千碱基对的 DNA 片段。用 Cosmid pJC74 克隆了 30 千碱基对的鸡基因组 DNA 片段。小的 Cosmid 载体能允许 40~50 千碱基对 DNA 片段的克隆。由于 DNA 片段很大,故具有较大的包含重复序列的机率,且在克隆过程中不太稳定,为此采用一种 *recA⁻* 的细菌作寄主(有些噬菌体载体,包括 λ Charon 4A 在这种寄主细胞里是不能生长的)。Cosmid 载体的优点是用很少量的微生物便可以组成真核 DNA 的基因文库。克隆前富集特定的基因片段似乎是不必要的,甚至是不合理的,因为这会导致预料不到的片段如序列相关的基因的丢失。

在一个克隆中,发现有另两个基因(X 和 Y)紧连着卵清蛋白基因。当用电镜观察具有聚 A⁺ 的总输卵管 RNA 和克隆 DNA 的杂交分子以显示基因 X 和 Y 的存在时,还不能检测这两个

图 4 寡聚胸腺嘧啶核苷酸(dT)-纤维素纯化的生蛋母鸡输卵管 RNA 和 pAR2 DNA 杂同分子电泳图

RNA·DNA 杂交是用热变性的线形 pAR2 DNA(1.5 微克/毫升)和寡聚(dT)-纤维素纯化的 RNA(100 微克/毫升)在 54.3℃ 条件下形成的,并制作电泳样品。pAR2 DNA 用 *Sal* I 酶解变成线形分子。图中, RNA 以虚线表示, DNA 以实线表示。标尺长度为 0.1 微米。

(a)完整的单链 pAR2 DNA 分子与三个 RNA 分子杂交电泳图。箭头 X, Y 和 ov 分别代表相应的基因。箭头 5' 和 3' 代表 RNA 的 5' 端和 3' 端。pAR2 DNA 的总长为 45,940 ± 6,600 碱基对(20 个分子的平均数)。X RNA 杂交的 DNA 序列到 pAR2 DNA 的最近端约为 10,460 ± 1,150 碱基; X 基因 3' 端到 Y 基因 5' 端为 5,560 ± 600 碱基; Y 基因 3' 端到卵清蛋白基因 5' 端为 11,580 ± 1,350 碱基对;卵清蛋白 mRNA 杂交的 DNA 序列到分子的末端为 5,330 ± 540 碱基对。卵清蛋白基因(图 4d)和 Y 基因(图 4c)的 A 环在图上未见,估计是由于杂交不稳定的缘故(见正文)。

(b) X RNA·DNA 杂交分子电泳图。杂交区长 1,945 ± 243 碱基对(测量 20 个分子)。游离的 RNA 尾巴 450 核苷酸长。两个单链 DNA 环“a”和“b”分别为 253 ± 47 和 812 ± 120 核苷酸长,杂交区域(从 5' 到 3' 端)分别为 175 ± 39, 181 ± 35 和 1,613 ± 219 碱基对。其他符号同(a)。

(c) Y RNA·DNA 杂交分子电泳图(测量 20 个分子)。杂交区为 2,020 ± 175 碱基对,单链 DNA 环 A~G 的长度分别为 1,729 ± 180, 492 ± 130, 303 ± 88, 812 ± 135, 206 ± 86, 55 ± 20 和 890 ± 134 核苷酸。杂交区域(从 5' 到 3' 端)分别长 218 ± 71, 93 ± 30, 165 ± 40, 153 ± 33, 167 ± 33, 183 ± 41, 1,320 ± 148 碱基对。其他符号同(a)。

(d) 卵清蛋白 mRNA·DNA 杂交电泳图(测量 22 个分子)。杂交区长 620 ± 67 碱基对,游离 RNA 尾巴长 1,100 核苷酸。环 A~D 分别长 1,606 ± 150, 204 ± 47, 578 ± 95 和 386 ± 85 核苷酸长。杂交区域(从 5' 到 3' 端)分别长 213 ± 45, 76 ± 23, 171 ± 33 和 162 ± 46 碱基对。L 箭头指示前导编码片段,它由环 A 存在而显示出来。其他符号同(a)。

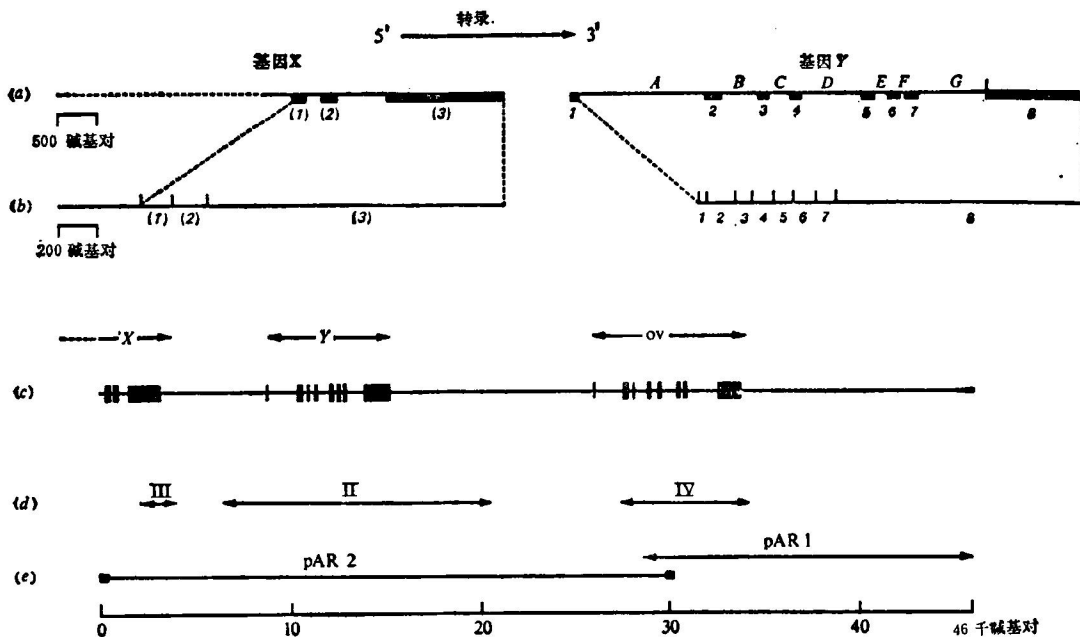


图 5 卵清蛋白基因区图谱

(a)和(b),基因 X 和 Y 图谱,表示内含子(用字母标明)和外显子(粗线和数字)各自的位置,以及用数字标明的相应的 RNA 图谱(b)(表示基因 X 外显子编号的括号内的数字需要了解该基因的全结构)。

(c),包含在 pAR1 和 pAR2 之内的 46 千碱基对 DNA 区段的图谱,表示基因 X, Y 和卵清蛋白的外显子(粗线)的位置。

(d),图 7 实验中所用的探针:探针 I 如图 2 说明;探针 II 是从 X Eco10 分离的 1.8 千碱基对的 Hind III 片段(见图 2 说明);探针 IV 相应于卵清蛋白基因 EcoR I “b”“c”片段和一部分“a”片段(见图 1c)。

(e),插入 pAR1 和 pAR2 片段的界限,c, d 和 e 行的标尺(千碱基对)。

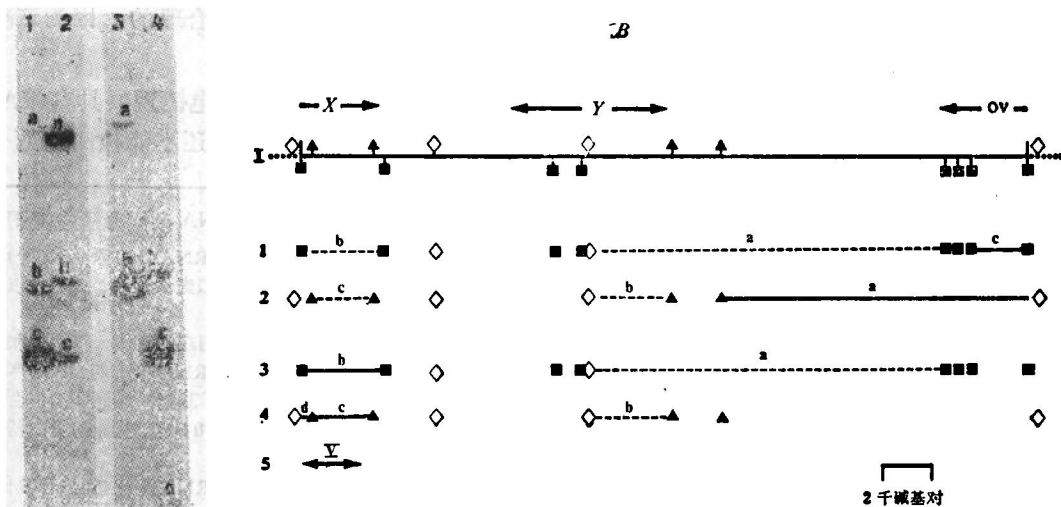


图 6 X, Y 和卵清蛋白(OV)基因的序列交源性

A, pAR2 DNA 的 EcoR I 和 Hpa I (1 行和 3 行)或 BamH I 和 Hpa I (2 行和 4 行)水解产物作 0.7% 琼脂糖凝胶电泳,转移到硝酸纤维素滤纸上,用相应于卵清蛋白基因的 Hhaov 探针(1 行和 2 行)或 X 基因的探针 V (3 行和 4 行)进行杂交。后一探针如图 B 所示,是一个 2.3 千碱基的 X Eco10 第二次克隆 Hind III 片段(见图 2 说明)各行的片段大小(千碱基对): 1 行和 3 行, a=14, b=3.4, c=2.4; 2 行和 4 行, a=12.5, b=3.6, c=2.4, d=0.9。B, 图 A1~4 杂交片段分别列在 B 中的 1~4, 实线表示包含探针一部分或全部序列的片段,虚线表示与探针交叉杂交的片段。I 表示 X, Y 和卵清蛋白基因的 BamH I (▲), EcoR I (■)或 Hpa I (◇)切点的位置,圆点线代表载体的序列。5, 探针 V 的位置, 图的标尺。

基因编码的是输卵管里未成熟的 RNA, 还是在其他组织里表达的 RNA。假定在克隆 DNA 上没有其他基因存在, 基因插入序列的长度(5~10 千碱基)和基因本身的长度,(6~8 千碱基)相当于每 10~15 千碱基的 DNA 有一个基因。

如果 *X*, *Y* 和卵清蛋白 mRNA 是已分化细胞产生的, 那么这一基因组区的结构是否也象免疫球蛋白基因一样是在分化过程中发生的呢? 经比较了在 pAR2 上克隆的红血球 DNA 和输卵管 DNA (图 3), 及红血球 DNA (结果未列出) 的限制性酶解图谱, 结果表明没有明显的基因重排。

在同一条链上转录, 同受雌激素控制的这些邻近基因在结构上有很多相似的地方 (图 4, 5c)。所有三个基因的 3' 端有一半以上的编码 RNA 的序列未被隔断。卵清蛋白和 *Y* 两个基因所有 7 个内含子和 8 个编码成熟 RNA 的编码区在大小上也很相似。另外在这两个基因上的第一个内含子 (*A*) 大约都是 1,600 核苷酸长, 它从邻近的外显子分隔了一个很短的 RNA 编码序列。确定基因 *Y* 的外显子 1 就是相当于卵清蛋白 mRNA 的前导编码序列是很有趣的。上述这些相同性以及基因 *X* 和 *Y* 和卵清蛋白基因具有部分序列同源性, 这些均有力地表明, 它们只少部分是来自共同的前体。在我们先前工作中, 检测了一定数量的鸡基因组的细胞 DNA 片段, 它们能与 *Hha* *ov* 卵清蛋白 cDNA 探针杂交, 而不属于卵清蛋白基因。这些片段的大部分, 如果不是全部的话, 都出现在基因 *X* 和 *Y* 中, 因此, 在鸡基因组内是不存在具有与卵清蛋白基因同样程度同源性的其他基因的。

卵清蛋白基因区的结构, 使人想到类 β -珠蛋白基因的发现。 β 和 δ 珠蛋白基因是很靠近

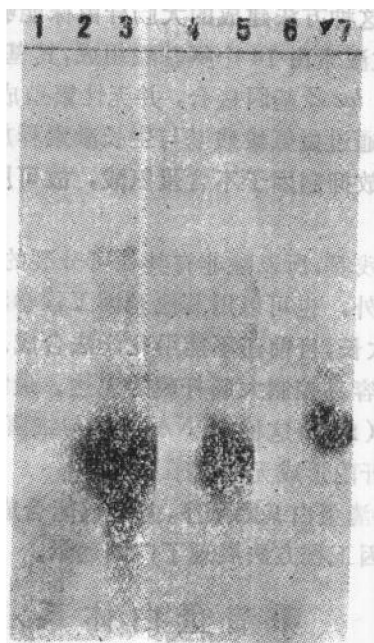


图 7 生蛋母鸡和阉割鸡输卵管内 *X* 和 *Y* RNA 的鉴定

5 天的小母鸡每天注射 1 毫克苯甲酸雌二醇, 注射 10 天 (先刺激), 激素停注后四周, 除去输卵管的大部分 (割除输卵管), 按照 F. Rougeon 改进的方法从生蛋母鸡或激素-割除输卵管制备总 RNA。RNA 样品 (各个胶 20 微克) 在 1% 琼脂糖含 10mM 氢氧化甲汞进行凝胶电泳分离, RNA 转移到重氮苯氧甲滤纸上, 用 ^{32}P 标记的探针杂交, 放射自显影用 Kodirex 底片, 曝光 4 天。1~3 行, 用 2.5×10^6 衰变/分 ^{32}P 标记的探针 IV (见图 5d) 杂交; 4 行和 5 行用 2.5×10^6 衰变/分 ^{32}P 标记的探针 I (见图 2f 和 5d) 杂交; 6 行和 7 行, 用 4×10^6 衰变/分 ^{32}P 标记的探针 II (见图 5d) 杂交。1、2、4 行和 6 行: 20 微克阉割鸡输卵管总 RNA, 2 行内加入 0.1 微克生蛋母鸡输卵管 RNA; 3、5 和 7 行: 20 微克生蛋母鸡输卵管总 RNA。

(下转第 26 页)

利用带重组质粒的大肠杆菌K12 合成卵清蛋白样蛋白质

Odile Mercereau-Puijalon 等

将卵清蛋白 mRNA 的无性繁殖 DNA 复制品从起始密码处切去少量核苷酸,再与大肠杆菌 β -半乳糖苷酶基因的开头相接合,带有这种杂交基因的无性繁殖系大肠杆菌可以生产大量卵白蛋白样蛋白质。

现在已经可以利用重组 DNA 技术将不同来源的基因引给大肠杆菌 K12,并形成无性繁殖系。可是直到现在还不能证明在这种细菌环境中有外来基因的表达。从低等真核生物(酵母、脉孢菌和果蝇)分离的一些基因可在大肠杆菌中自发的表达。但是还没有见到从高等真核生物获得的无性繁殖基因能够自发转录或翻译的报道。将基因顺序整合到细菌或噬菌体基因中,就能有效地实现转录作用,看来促进翻译成多肽的最好方法是将真核生物的顺序与细菌基因接合,建立杂交基因。利用这种方法建成的大肠杆菌株能够生产杂交蛋白类型的小分子激素:生长激素释放抑制因子。这激素由 14 个氨基酸组成,其基因已经用化学法合成,将此基因与大肠杆菌的 β -半乳糖苷酶的 *lacZ* 基因接合,并无性繁殖成一种多复制质粒。 β -半乳糖苷酶部分约含 1,000 个氨基酸,通过蛋氨酸残基与生长激素释放抑制因子相联。蛋氨酸在某些条件下不稳定,而生长激素释放抑制因子不含蛋氨酸,故可用溴化氰处理后从杂交多肽中提出。

由于大多数多肽含蛋氨酸残基,所以除非有其他可分裂的蛋白质“联结剂”(linker),否则不可能普遍采用上述方法。另外,也可以用其他基因工程学法合成比生长激素释放因子大的多肽,但是由于基因编码顺序太长,目前还不能用化学法合成。最近发现某些真核生物基因可以断裂,使其从基因组中分离,容易控制大肠杆菌的表达。最常用的方法包括:(1)从 mRNA 酶促合成所需要的基因顺序;(2)将这种 cDNA 顺序与细菌基因接合,使真核生物顺序能够在适当位置识别;(3)使大肠杆菌形成无性繁殖系。

我们将这种方法用于鸡卵清蛋白基因顺序,并建成能合成大量卵清蛋白样蛋白质的菌株,结果证明方法切实可行,使基因工程方法展现了广阔远景。

卵清蛋白体系

卵清蛋白是蛋清的主要蛋白质,为 386 个氨基酸的多肽链,有一个链内二硫键。翻译后只进行少许修饰,例如,将第一个蛋氨酸残基分裂掉、*N*-乙酰化作用、糖苷化作用以及某些残基的磷酸化作用。有多种熟知的遗传变种。

卵清蛋白 mRNA (ov mRNA) 主要在鸡输卵管中,很容易分离到相当纯。卵清蛋白 mRNA 含 1,859 个核苷酸,其全部顺序已由 McReynolds 测出,含有 64 个核苷酸长的 5' 非编码区,接着是 1,158 个核苷酸长的编码区,然后是 637 个核苷酸长的 3' 非编码区(图 1)。

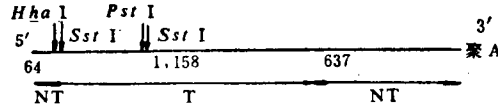


图 1 卵清蛋白 mRNA 图

根据 McReynolds 等卵清蛋白 mRNA 图解。T 为翻译区；NT 为非翻译区；AUG 是第 65~67 个核苷酸，是蛋白质合成开始信号；HhaI、PstI 及 SstI 部位在相应 DNA 顺序的位置。

卵清蛋白基因是从鸡的基因组中分离出来的，最近我们用分子无性繁殖法将各种编码胞浆卵清蛋白 mRNA 的基因组 DNA 碎片分离，分析这些无性繁殖碎片，发现至少有 6 个交叉顺序，都位于卵清蛋白 mRNA 的翻译区内。由于卵清蛋白基因的结构妨碍其在大肠杆菌中表达，因而我们把注意力转向卵清蛋白 mRNA 的 DNA 转录品。

Humphreis 等利用使珠蛋白 mRNA 转录品无性繁殖化的方法第一次分离出一种质粒——pCRIov2-1，它带有从卵清蛋白 mRNA 转录的 DNA 的 1,730 个碱基对。已经证明这是一个可靠的无性繁殖 DNA 复制品。因为卵清蛋白 mRNA 含 1,859 个核苷酸，所以 pCRIov2-1 不带有全部卵清蛋白 mRNA 顺序。不过，限制性图、DNA 顺序测定以及电泳分析等结果都证明丢失的部分是 mRNA 的非翻译区。

因为 pCRIov2-1 带有全部卵清蛋白顺序编码，我们首先想到这些基因是否在 C600^{r⁺} (pCRIov2-1) 中表达。已经证明能从正确的链转录卵清蛋白顺序，可是这个菌株不产生卵清蛋白样蛋白质，用放射免疫法证明为阴性，即使用能测出一个细胞有一个分子以下的方法也证明是阴性。我们推测 pCRIov2-1 中的卵清蛋白顺序偶然与一未知质粒基因接合，失去了适当的识码能力。

重组 *lac* ov 质粒的组建

我们决定将卵清蛋白基因顺序与大肠杆菌 *lacZ* 基因的开始端接合，大肠杆菌 *lacZ* 基因是编码 β -半乳糖苷酶的基因。Fuller 曾经制成一个由 205 个碱基对组成的小 DNA 碎片，两侧带有两个 *Eco*RI 区，它除了带有 L8UV5 *lac* 操纵基因促进子区外，还带有 β -半乳糖苷酶头 8 个氨基酸的信息 (L8UV5 突变使这体系对分解代谢物的阻抑作用不敏感)。Charney 等将这个碎片用遗传工程法引进噬菌体 λ I857 ϕ lac5 Δ ZL8UV5 中，这种噬菌体 DNA 在其 *lac* 顺序 β -半乳糖苷酶头 8 个氨基酸的后面有一个 *Eco*RI 区。我们用 *Eco*RI 及 *Hind*III 消化这种 DNA，制成一个 2.85 千个碱基的 DNA 碎片，后者含 *lac* 顺序的开端，将这种碎片在试管中与经上述两种酶消化的 pBR322 DNA 进行重组 (见图 2)，并使大肠杆菌无性繁殖化，生成的质粒叫 pOMP0，可做为再重组实验的引导 (Vector)。pOMP0 的体积大约为 7.1 千个碱基，它使受体宿主细胞抗氨苄青霉素，但不抗四环素，并有 *lac* 成分 (因为 *lac* 阻抑子被多复制质粒所带的操纵子所过分滴定)，而且在靠近 *lacZ* 基因的开端只有一个 *Eco*RI 区 (图 2)。

用 *Hha*I 消化质粒 DNA 得出一种含 2,430 个碱基对的碎片 *Hha*ov，可以将质粒 pCRIov2-1 带的大多数卵清蛋白顺序分离出来。分析卵清蛋白基因的无性繁殖基因组时，证明在近 5'-末端有一个对 *Hha*I 敏感的部位，与 pCRIov2-1 的 *Sst*I 部位很近 (大约 30 个碱基对)。当 McReynolds 等查清此卵清蛋白 mRNA 区的核苷酸顺序后，看来在顺序中没有 *Hha*I 区 (GCGC)，不过使我们想到，在靠近 5'-末端的 *Sst*I 部位 (第 116~120 个核苷酸) 之前 35 个核

A

EcoRI *lac* *SmaI* *HindIII*

B

EcoRI *HindIII* *pBR322*

SmaI *pOMPO*

C

HhaI *ova* *lac* *SstI* *HindIII* *PstI* *pOMP2*

Charnay 等组建的 λ CI857 *plac*5L8UV54Z 是先将 *lac* 顺序的 205 碱基对插入位于 *lacZ* 基因一端 *EcoRI* 区, *lacZ* 基因是 λ CI857 *plac*5L8UV5 的一种衍生物。然后在细胞内进行重组产生在 *lacZ* 端及 *lac* 顺序 205 碱基对之间只有一个 *EcoRI* 区的噬菌体 (见图 A)。取 500 毫微克 λ CI857 *plac*5L8UV54Z DNA (见图 A, 只画出带 *lac* 顺序的区域) 和 50 毫微克 pBR322 DNA (见图 B), 用 *EcoRI*, *Hind* III 消化, 再用 T4 DNA 联结酶联结, 最终体积 15 微升。取一部分使 CaCl_2 处理的 478 种 ($\text{C600rk}^+ \text{mk}^+ \text{recBC-lac}^-$) 转化, 转化剂选用含氨基青霉素 (20毫微克/毫升) 及 5-溴-4-氯嘧啶半乳糖苷 (X-gal, 40 微克/毫升) 的蛋白陈平板, 在这种平板上, 带有重组质粒的菌株产生兰色菌落。再用提纯 pOMP0 组建带卵清蛋白顺序的质粒。代表性质粒 pOMP2 图是用限制性分析法建成的。pOMP0 长约 7.1 千个碱基, 含 *EcoRI*, *Hind* III, *Sst* I, *Pst* I 区各一个, 所有实验都是在极纯的水中完成的。

[illegible]

• 12 •

在 *Hhaov* 中,卵清蛋白顺序与 *Z* 基因开端接合的方式见图 3。简单地说,经 *EcoRI* 处理过的 pOMP0 DNA 和 *Hhaov* 碎片用 S_1 核酸酶或者 DNA 聚合酶 I、或者 S_1 核酸酶和 DNA 聚合酶 I 处理,释出单股末端后,用过量的 T4 DNA 联结酶将粘合端联结,在 *Hhaov* 碎片按适当方向整合成的那些菌株中产生图 3 那种结构。预期这种杂交基因能产生一种 389 个氨基酸的卵清蛋白样蛋白质,在这种蛋白质中,卵清蛋白的头五个氨基酸换成 β -半乳糖苷酶的头 8 个氨基酸(包括第一个蛋氨酸残基)。

将重组分子通过转化引入大肠杆菌 1398,这是一种溶原性宿主(lysogenic host),适合进行在原位菌落杂交法,这个方法是最近发明的。然后用 ^{32}P 标记的 *Hhaov* 筛选抗氨苄青霉素的菌落, ^{32}P -*Hhaov* 只与带有卵清蛋白顺序的重组质粒反应。这样我们得出 8 种不同菌株,都表现 *lac* 成分(在 5-溴-4-吡啶半乳糖苷 x-gal 平板上显蓝色菌落),然后将分离的质粒(pOMP 1, 2, 4, 5, 7, 9, 10, 11)转给非溶原性宿主(600r_gm_g⁺)。

重组质粒的鉴定

被整合卵清蛋白顺序的方向可用下法测定:用 *Pst* I 将重组质粒在两个部位切开,一个在 pBR322 的氨苄青霉素基因中,另一个在卵清蛋白顺序中。如果后者方向正确,用 *Pst* I 消化质粒 DNA,则产生大约含 6.9 和 2.7 千个碱基的两个 DNA 碎片;如果方向不正确,则用 *Pst* I 消化产生大约含 8.5 和 1.1 千个碱基的碎片。

为了确定那些菌株生产卵清蛋白样蛋白质,我们用放射免疫法将碘化卵清蛋白与细菌提取液混合,然后使之与抗卵清蛋白抗体反应,结果完全与重组质粒的结构一致,即质粒 pOMP1, pOMP2 及 pOMP4 表现很强的竞争性,但是,其他质粒(包括 pOMP0)则否。任意选择质粒 pOMP2 作进一步的研究。

利用 pOMP2 生产卵清蛋白样蛋白质

进一步用放射免疫法定量测定大肠杆菌的卵清蛋白样产品的量。用纯的未标记卵清蛋白竞争作用与带 pOMP2 大肠杆菌提取物的竞争作用作比较,发现后者的竞争作用只有 80%,说明细菌制造的蛋白质失掉一个或数个抗原部位。此外竞争作用曲线可以认为是多相曲线,这表明大肠杆菌的某些抗原部位更多一些,或者有一些抗原部位比天然卵清蛋白抗原部位的亲和力低。由于卵清蛋白样产品的大小和预测的相似,我们倾向于第二种解释。正象我们在前面提到的那样,由于卵清蛋白在输卵管中进行一些翻译后的修饰作用,这种修饰作用在抗原决定簇结构中可能起重要作用。大肠杆菌不进行这种修饰作用,因而与卵清蛋白相比,卵清蛋白样产品的某些抗原部位的竞争性可能更差些,卵清蛋白样产品与天然卵清蛋白至少在 N-端有少量氨基酸差别。根据这一假设,多相曲线第一部分定量证明, 2.5×10^6 个细胞可以生产相当于 15 毫微克卵清蛋白的产品,也即每个细胞大约产生 90,000 个卵清蛋白样蛋白分子;如果按单相曲线考虑,则每个细胞产生 45,000 个分子(不同实验结果不同,其范围约 5,000~45,000,平均 30,000)。

我们还不能完全证明在 C600 (pOMP2) 中卵清蛋白样产品的合成是否受 *lac* 阻抑子调控(后者至少部分的被操纵子所过分滴定)。因此我们将 pOMP2 引入能产生过量 *lac* 阻抑剂的 1442 株中,证明用 *lac* 诱导剂 IPTG 可使合成增加 50 倍。

卵清蛋白样蛋白的大小可以用下法测定。将带有 pOMP2 的细胞提取液在活体内用 ^{35}S -

蛋氨酸标记,用抗卵清蛋白抗体沉淀,将解离的沉淀物在变性情况下进行聚丙烯酰胺凝胶电泳。荧光电泳谱证明有一条主要蛋白区带与卵清蛋白一齐泳动,或迁移率稍快于卵清蛋白。在 1442 株 (pOMP2) 中,这种蛋白质的合成与加入 IPTG 有关。在免疫沉淀反应中加入过量未标记的卵清蛋白可以消除大多数放射活性区带,这说明它和卵清蛋白样产品一样。在未进行免疫沉淀的总标记提取液中,有时可以直接观察这个区带。总之,卵清蛋白样产品的大小与卵清蛋白差不多。

用大肠杆菌合成异体多肽

我们已经组建成一些能合成卵清蛋白样产品的大肠杆菌株,已有的证据表明:(1)这个产品的合成是受 *lac* 阻抑剂调控的;(2)进行放射免疫测定时,产品能与碘化卵清蛋白竞争;(3)产品的大小与卵清蛋白极相似。

此外这些资料还有力的提示,我们的大肠杆菌菌株所合成的卵清蛋白样产品,其成分与所预期的极相似,也即 8 个氨基酸是 β -半乳糖苷酶的,381 个氨基酸是卵清蛋白的(来自天然卵清蛋白的 386 个氨基酸)。杂交蛋白的确切结构尚待研究。

带有重组质粒的细菌,每个细胞可以合成 30,000 或 60,000,甚至高达 45,000 或 90,000 个分子(与对放射免疫竞争曲线的解释有关),此值大约相当于大肠杆菌细胞总蛋白重量的 0.5~1%,这个数值比较接近理论产量(从带有功能性 *Z* 基因多复制质粒细胞合成的 β -半乳糖苷酶单体数目计算约为 $1\sim 2\times 10^6$ 分子/细胞)。

大肠杆菌株能生产大量卵清蛋白样蛋白。这将引起许多有兴趣的问题。首先,这些菌株十分稳定,而且不会失去重组质粒;其次,卵清蛋白是一种分泌蛋白,初步实验证明在 C600 (pOMP2) 培养基上清液中一点卵清蛋白样蛋白质也没有;第三,大肠杆菌制造的卵清蛋白样蛋白质在细菌环境中一定比较稳定,不过还没确切测定过;第四,卵清蛋白 mRNA 的密码不是随机分布的,而且其分布频率与大肠杆菌密码的频率显著不同。某些密码(AUC、GUG、CCA、GAC、GAA、GCA)在卵清蛋白中的频率比已知排列顺序的大肠杆菌噬菌体(如 MS2 或 ϕ X 174)更高,大肠杆菌可能缺少相应的微量 tRNA,但对卵清蛋白基因顺序表达的影响似乎不大。

据了解,这是用大肠杆菌合成高等真核生物的一条长肽链的第一个例子。我们所造的重组合基因还不完全正确,因为失去天然蛋白质的 5 个氨基酸,代之以来自细菌蛋白的少量其他氨基酸。没有什么理由认为 *lac* 和卵清蛋白顺序的联结部位不能挪到 ATG 开始密码处,以便合成正确的产品。这也是第一次用无性繁殖 mRNA 互补的 DNA 合成大量长链多肽一个例子。总之,我们的实验证明,利用遗传工程生成的细菌方便易行,而且有大量应用的潜力。卵清蛋白是氨基酸含量齐全平衡的蛋白质,常用做营养价值标准,因而可以相信由细菌产生的卵清蛋白样蛋白质可以用做饲料。

《Nature》275(5680):505~510,1978

(刘殿珍译)