

# 染色体结构和基因组编

陈 汉 源

(中国科学院上海细胞生物学研究所)

一九八〇年八月下旬在英国牛津大学召开的第七届国际染色体会议上,重点地报告和讨论了在分子水平的染色体结构。这是当前染色体基础理论研究的首要任务,也是细胞生物学的中心问题<sup>[1,2]</sup>。由于染色体是遗传信息的载体,只有阐明染色体的精微结构,才能深入了解基因的表现及其调节。但是对此面临巨大困难,因为真核细胞染色体不仅组成复杂,结构巧妙,而且变化多端,功能各样。近年来由于应用了分子生物学的新技术和新成就,才能打开这个秘密王国的大门。现将目前取得主要进展分述如下:

## 一、染色体结构

1974年 Kornberg 发现构成真核细胞染色体的基本重复单位是核小体,电镜看来象是穿线的念珠。这是由 110 埃直径核心颗粒和连接其间的 DNA 组成。前者又由直径 60 埃的组蛋白球(各 2 分子 H2A、H2B、H3 和 H4),外绕 140np (核苷酸对) DNA 成 1.56 周。后者约为 60npDNA。每单位 DNA 共长 200 np<sup>[3]</sup>。至于从核小体构成染色质或染色体,目前有下列几种说法:

### (一) 超螺旋学说

从中子衍射和电镜观察认为,染色体是由染色纤维多级螺旋化所形成(图 1)。首先由核小体左旋绕紧,核心颗粒互相密集,其外各有 200npDNA 共绕成 2.2 周。H1 分子斜向组合在两粒之间,这就形成 100 埃核蛋白纤维, DNA 积压率为 1/7(图 2)。由此二级左旋绕成螺线管,每行约排列着 6 个核心颗粒, H1 分子位于螺线管内侧,从而形成 300 埃染色

纤维, DNA 压缩率 1/6。由此三级螺旋化就成 2000 埃的间期染色质, DNA 又压缩 1/20。复由此四级旋绕成为约 6000 埃中间染色体, DNA 再压缩 1/10<sup>[3,4]</sup>。在中期染色体, DNA 总共压缩 1/8000(1/5000—1/10000)。这学说未指明其他成分如 NHP (非组蛋白)和 RNA 的作用。

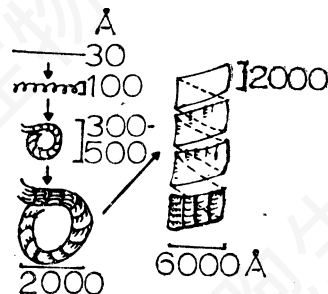


图 1 染色体结构的超螺旋模式

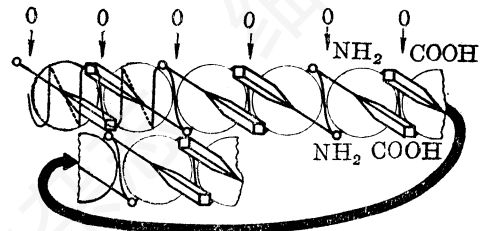


图 2 由核小体左旋成 100 埃核蛋白纤维  
组蛋白 H1 的两端 NH<sub>2</sub> 和 COOH 结合于连接 DNA 上,除左侧二个颗粒外,其余颗粒只画出连接 DNA

### (2) 骨架(Scaffold)学说

Laemmli(1977—1979)等报告<sup>[5,6,7,8]</sup>,染色体内部由 NHP 构成网状骨架,四周结合着许多长约 0.6 微米的侧环,后者为 300 埃染色纤维所形成。如果将分散的中期染色体以 2M NaCl 溶液或硫酸葡聚糖加肝素处理,除去组蛋白和大部分 NHP 之后,剩留部分体积增大

2—3倍,仍含有80%的DNA,8—10%的总蛋白,但仅有0.1%的组蛋白。这个结构可被胰凝乳蛋白酶、尿素或SDS的处理所解体,但对RNase处理无影响。该结构中部为NHP构成的网状纤维骨架,四周为DNA构成的晕。后者乃由许多DNA侧环密集而成(图3)。侧环长约25(15—30)微米,包含75,000(45,000—90,000)np,侧环的基部结合在骨架中。可以推想当侧环DNA与组蛋白结合时,形成300埃染色纤维,就缩短42倍,则只有0.6微米长度了,而中期染色体直径大约为1微米(图4)。如果中期染色体预先以核酸酶消化,去除DNA,再经过上述同样处理,就只剩下骨架。这个结构只含有0.2%DNA,10%总蛋白和0.1%组蛋白,但仍有30%的NHP。由于染色单体的网状骨架仍于着丝粒

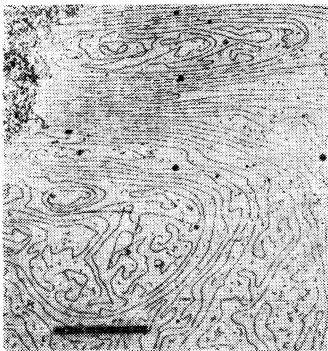


图3 去组蛋白后染色体骨架的DNA侧环  
线段代表1微米长度

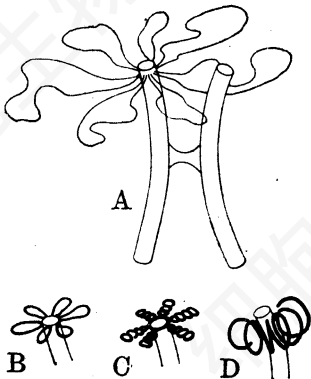


图4 染色体结构的骨架模式

a. 去组蛋白染色体骨架的DNA侧环  
b.c.d. DNA和组蛋白结合后凝缩成染色纤维

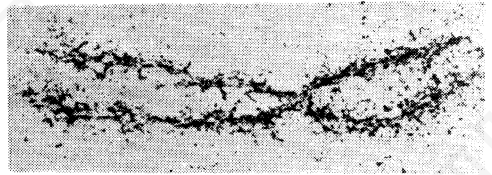


图5 染色体的NHP骨架

入染色体经过细球菌核酸酶和2MNaCl处理

区连接着,因而仍能辨认原来染色体的形状,整个骨架长度也约与原来染色体相近(图5)。对于骨架究竟是染色体原有的天然结构,还是人工矫作物,尚有争议。

### (3) 区段(Domain)学说

Igo-Kemenes(1977—1978)等<sup>[9,10]</sup>发现间期核内染色质不能任意地自由运动,由于染色纤维是按一定长度的区段互相交联,或与染色体原有的支架连接着,在交联点或连接点上,以RNA或蛋白固定住,因此保持紧张状态。这些固定点之间的染色纤维长度称为Domain(图6)。当将大鼠间期染色质经过核酸酶轻度切割后,凡其染色纤维长度在75kp(kp为千核苷酸对)以上的都成沉淀,只有长度在75—40kp的区段才能溶解。如果每一DNA区段只有一点被酶切断时,这种染色质仍全部沉淀。当每一区段有二点被酶切断,则有一半染色质可以溶解。被酶切断10处的区段,有90%溶解,被酶切断100处的,99%可以溶解。这种区段最大的为75kp,最小的为34kp。Benyciali等<sup>[11]</sup>认为果蝇间期染色纤维通常形成许多个超螺旋侧环,每侧环有85kp,约相当于400个核心颗粒的长度。果蝇间期染色

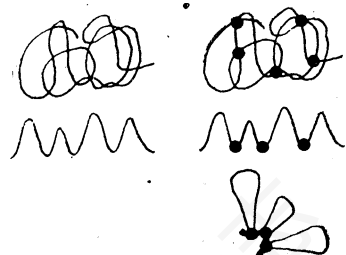


图6 染色质结构的区段模式

染色纤维非交联(左),互相交联(右上中)或与支持物连接(右下)

质经过 Nanidet P40 解体后,用蔗糖密度梯度离心所得到 5000S 沉淀物,就是如此。经过 DNase 作用使其中一股 DNA 形成缺刻,则超螺旋侧环失去张力,而舒张开来,使密度下降,并于 2600S 沉淀。继续以溴化亚乙基去除组蛋白,则于 1000S 就沉淀了。

## 二、基因的组编

研究真核细胞基因表现和调节遇到很大的困难,在于真核细胞基因组十分复杂。例如人类细胞的 C 值比  $\lambda$  噬菌体要大 100,000 倍,比原核细胞大肠杆菌也要大 800 倍。人双倍体细胞含有的双股 DNA 长达 1.74 米。直到近五年以来,由于应用了下列几种新方法才能逐步解开这个谜团<sup>[12,13]</sup>。

### (1) Southern 吸附法

这能把基因 DNA 从整个 DNA 中分离出来。限制性核酸内切酶能专一地识别包括 4-6bp 的 DNA 顺序区域,并切断其中一点。这些顺序互相倒置成回文对称。在全部 DNA 中,平均每 256—4096bp 顺序才有一个识别位置,因此 DNA 被切割之点是有限的,通常产生几百到几千 bp 的片段,当用琼脂糖凝胶电泳按大小分离,经过变性,使双股 DNA 拆成单股,再吸附到硝酸纤维素滤纸上,与放射性标记的 mRNA 杂交,经曝光显影后,有银粒之处即存在着基因 DNA。就可从相对应的电泳带中获得基因 DNA。目前进一步应用亲和层析法,将 DNA 片段与柱内的 mRNA 结合,由此将基因 DNA 分离出来。

### (2) 基因纯系扩增(Gene cloning)

先将质粒的环状 DNA 以限制酶于一点上切开,同样把外源 DNA (包括已分离的基因 DNA 片段或从 mRNA 反转录的 cDNA) 经酶切割,两者的粘附末端互相配对,再以连接酶把缺口修好,就成为重组体质粒,然后,感染细菌。后者预先以氯化钙处理,使重组体 DNA 易于渗入进去。由于质粒常有抗某种抗生素的基因,当在培养物中加入抗生素之后,

其他细菌均被杀死,只有重组体质粒所感染的细菌能单独生长,形成群落。经过大量繁殖后,再从重组体质粒中分离出大量基因 DNA。或在细菌的平皿接种复制板上,以放射性标记的 cDNA 为探针,进行原位分子杂交,就能辨认含有重组体质粒的细菌群落。经纯系扩增的基因 DNA 可供分析核苷酸顺序,或研究基因在蛙卵质中的表现。

### (3) 电镜观察

在电镜下能直接观察到蛙卵核内核仁的核糖体基因转录图象,新合成的 RNA 的末端仍附着于 DNA 模版上,整个图形宛似一株桤树。当将 rDNA 经过部分变性后,其中 18S 和 28S 基因区段富于 G+C,因而这两股 DNA 仍旧配对。但间隔 DNA 含 A+T 较多,该两股易于分开。因此在电镜下,可以看到基因 DNA 和间隔 DNA 互相穿插前后成簇排列。如果将质粒 DNA 与重组体 DNA 变性后杂交,则后者的外源 DNA 区段不能配对,因而形成侧环。或将 mRNA 前体和重组体 DNA 杂交,由于 RNA/DNA 杂交体比 DNA/DNA 杂交体更为稳定,因而 mRNA 前体与重组体的外源 DNA 中一股互补顺序杂交,而不互补的一股顺序则脱离开去,形成眼状的 R 侧环,在电镜下就能看到基因 DNA 了(图 7a)。如将成熟的 mRNA 与基因 DNA 片段杂交,则后者的非编码顺序不能与 mRNA 配对,因而这段顺序单独凸出成为侧环(图 7b)。应用这些方法分析基因组编目前取得的结果如下:



图 7 小鼠  $\beta$  珠蛋白基因 DNA 片段与 mRNA 杂交的电镜图形

- 与 15S 珠蛋白 mRNA 前体杂交形成 R 侧环
- 与成熟的 9S 珠蛋白 mRNA 杂交,箭头指基因 DNA 的非编码顺序

(一) 拼接基因(Splicing gene)(割裂基因 Split gene 或嵌入基因 Insert gene) 真核细胞中, 目前除发现免疫球蛋白基因的 V(可变)和 C(恒定)区段经过重排外, 其他基因都有固定的 DNA 分子结构。但是真核基因常分成许多转录单位, 其中编码顺序被非编码顺序所分隔, 分别称为外子(结构基因顺序)和内子(嵌入顺序)<sup>[14]</sup>。1977 年 Glover 和 Hogness 首先发现果蝇 28SrRNA 基因嵌入 5.4kp 的静止顺序, 把编码顺序分隔成二段。其他如四翼虫的 rRNA 基因, 腺病毒 Adeno 2 protein Mr 基因, 家蚕丝蛋白基因, 人、小鼠和兔珠蛋白基因, 小鼠免疫球蛋白  $\gamma_1$  重链基因, 鸡卵白蛋白和粘蛋白基因均已证明如此(图 8)。但组蛋白基因和某些 tRNA 和 rRNA 基因则否。整个拼接基因一同转录, 大分子转录物经过加工, 切除非编码顺序后, 再拼接成为 mRNA, 如小鼠  $\beta$  珠蛋白 mRNA 前体为 15S, 而成熟的 mRNA 只有 9S, 前者比后者长 2—3 倍。前者的半衰期少于 2 分钟, 其中的嵌入顺序先形成发夹侧环, 随即切除, 其余编码顺序连接成 mRNA。酵母 tRNA 前体原无活性, 只在切除 15np 的嵌入顺序后, 成熟的 tRNA 才能胺基乙酰化。Mendel<sup>[16,18]</sup> 和 O'Mally<sup>[17]</sup> 等探明鸡卵白蛋白基因 DNA 全长约 7000np, 包含 8 个外子和 7 个内子(图 8, 9), 拼接而成的 mRNA 长 1,859np, 只及基因 DNA 的 1/3。这基因存在于鸡的所有组织中, 不仅限于表现活性的输卵管上皮。今

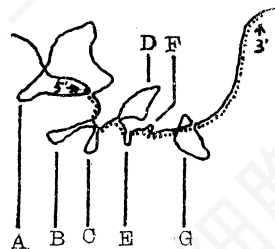


图 8 鸡卵白蛋白基因 DNA 与成熟的 mRNA 杂交的电镜图形  
实线 DNA, 虚线 mRNA, 内字 A、B、C、D、E、F、G 箭头指末端

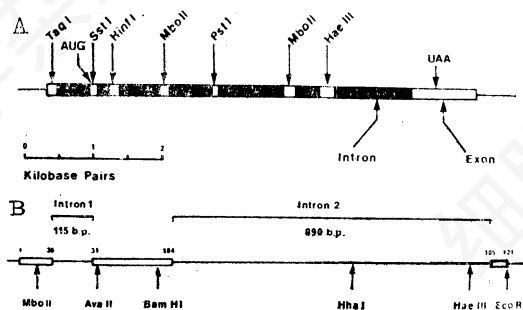


图 9 鸡卵白蛋白基因(A)和人胚胎球蛋白基因(B)的组编图解  
Intron 内子, Exon 外子, Kilobasepair 千碱基对, bp 碱基对

已分离到长 46kp 的基因 DNA, 包括三部分 X-Y-Ov(卵白蛋白结构基因), 前两者乃由后者重复而成, 三个基因的内子和外子类型基本相同。外子是保守的, 内子可有变化, 但 X 和 Y 基因的转录速度比 Ov 基因较慢。

(二) 移动基因 (Mobile gene) [移动 DNA, (mobile DNA) 或滑动 DNA (Shuffling DNA)] 真核基因大多是单一的, 而核糖体和组蛋白等重复基因前后成簇排列, 集中于染色体一定位置上。如爪蟾核糖体基因位于核仁组织区, 有 450 个复本, 其间被不转录的顺序分隔, 在 18S, 5.8S 和 28S 基因之间则有可转录的间隔顺序连接。此外, Finnegan (1978)<sup>[18]</sup> 和 Geogiev (1979)<sup>[19]</sup> 从果蝇 DNA 分离到经过纯系扩增的顺序, 这能与 Poly(A) RNA 杂交, 故视为基因。这些重复基因单个地散布在染色体许多不同位置上, 由于基因可以移动位置。如三类重复顺序 Copia, 412 和 297, 前两者各长 5 和 7.3kp (图 10)。在 Oregon R 品系果蝇中, 三者各有 60、40 和 30 个复本, 这与它们在多线染色体的原位杂交位点数目相近 (图 11)。至于在两株培养细胞中, 三种基因的复本数目各为 90 和 70、120 和 40、110 和 170, 表明各自独立变化<sup>[20]</sup>。而且培养细胞所新增加的基因复本都座落于新的位点上。同样, 来自酵母的纯系扩增 DNA 顺序 Ty1 长 5.6kp, 其功能未知, 有 35 复本。这些重复顺序能够移动, 散布各处, 可与基因组

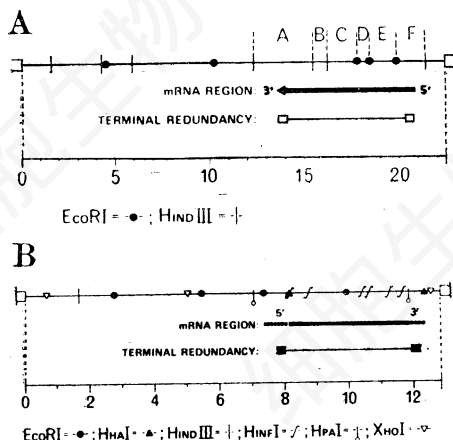


图 10 412(A)和 Copia(B)顺序  
被核酸酶的切割点

箭头指转录方向, 标线单位一千核苷酸对;  
mRNA region 信使 RNA 区域; Terminal  
redundancy 末端重复顺序

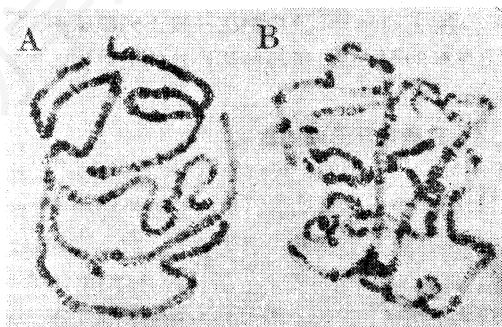


图 11 412(A)和 Copia(B)重复顺  
序在多线染色体上的定位  
银粒代表原位杂交位点

消化产物的 25 个带杂交。Tyl 末端的  $\delta$  顺序  
与细菌转位子(Transposon)的结构类似。

### 三、结构基因的倍增

真核细胞除重复基因之外, 在一定发育时期, 有些基因倍增。如尖眼蕈蚊科多线染色体的 DNA 胀泡和两栖类的核仁 rDNA。此外, Schimke<sup>[21,22,23]</sup> (1978—1979) 首先报道结构基因的倍增。小鼠或仓鼠体外培养细胞以抗癌药氨甲嘌呤处理后, 产生抗药性, 这对药物的耐受力比敏感型大 3000 倍。同时, 抗性细胞的二氢叶酸还原酶比活性增加几百倍。通过标记氨基酸渗入, 免疫沉淀和荧光标记测定表明, 抗

性细胞的蛋白质合成速度和酶蛋白含量均提高几百倍。甚至酶蛋白在细胞全部可溶蛋白中竟占到 10—11%。而且采用分子杂交和体外转录法证明, 细胞的 mRNA 含量和酶基因复本也增加 45—400 倍。进一步观察到在稳定的抗性仓鼠细胞中, 第 2 染色体出现额外均匀着色区 (HSR) (图 12), 这约占全部基因组的 2%。小鼠染色体的 HSR 包含  $2.3 \times 10^8$  np, 如果已倍增 300 倍, 则每一重复单位原有 800kp 而敏感型无 HSR。原位分子杂交表明 HSR 为基因所在位置。至于在不稳定的抗性小鼠 S180 细胞中, 出现若干微小双体染色体 (DMC), 竟多达 300 多个, 倍增的酶基因即位于此。当除去药物后, 抗药性降低, 则细胞内的 DMC 减少。至于细胞如何发生不稳定和稳定抗性的互变呢? DMC 可能整合到染色体中去, 或由染色体的原有顺序倍增形成 HSR, 然后再从染色体上游离出来, 形成 DMC。

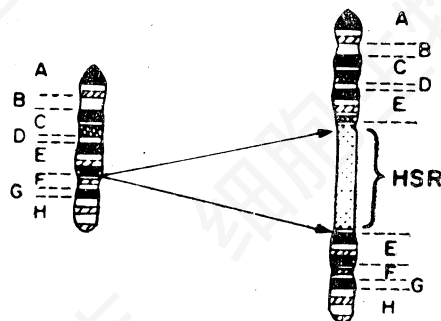


图 12 均匀着色区 (HSR) 在染色  
体上的位置图解  
染色体经 G 带染色

### 参 考 文 献

- [1] Crick, F. H. C., 1978 in Engberg, J., "Specific eukaryotic genes 17—24.
- [2] Lillery, D. M. T. et al., 1979 *Ann. Rev. Genet.*, 13:197—234.
- [3] Worcel, A., 1978 *Cold Spring Harb. Symp.*, 42:313—324.
- [4] Sedat, J. et al., 1978. *Cold spring Harb. Symp.* 42:331—350.
- [5] Paulson, J. R. et al., 1977 *Cell*, 12: