

用人造微小染色体技术研究着丝粒和端粒的 DNA 结构与功能¹⁾

潘惟钧

(北京大学生物系)

为了在细胞世代中保持其稳定性,染色体起码应具备 3 个结构要素,那就是有一个 DNA 复制起点;一个着丝粒 (centromere) 使细胞分裂时两个姊妹染色单体能平均分配到子细胞里;最后,在染色体的两个末端必须有端粒 (telomere),使 DNA 能完成复制。近年来人们采用分子克隆技术把真核细胞染色体的复制起点、着丝粒和端粒的 DNA 片段分别克隆成功。并且把它们互相搭配或改造而构成所谓“人造微小染色体” (artificial minichromosomes),以研究这 3 种成分的结构与功能。

一、染色体复制起点

大肠杆菌质粒 pBR322 不能转化酵母细胞,因为 pBR322 上的 DNA 复制起点不能被酵母系统所识别, DNA 不能复制。1979 年 Stinchcomb^[13] 和 Carbon^[8] 实验室分别把带有遗传标记,例如 Trp⁺ 的酵母 DNA 的 EcoRI 片段插入 pBR322,用来转化 urp⁻ 酵母,获得了带有质粒并能传代的 Trp⁺ 细胞。它们所含的质粒比 pBR322 多出 1.4kb,若把这 1.4kb 片段用 EcoRI 重新切下,颠倒过来接回 pBR322,仍然能使 urp⁻ 酵母转化成 Trp⁺ 酵母。其他外源 DNA 或基因若与这 1.4 kb 片段相接并插入 pBR322,便能在酵母中复制、表达和传代。例如,接上含有 his3 和 DNA 片段后就能使 his⁻ 细胞转化成 His⁺ 等等。可见这 1.4kb 片段包含了 Trp⁺ 基因和一个酵母染色体的复制起点,后者被称为 ARS1,即能够起始染色体复制的 DNA 序列。这样的 ARS1-pBR322 环状质粒由于兼含酵母和大肠杆菌系统的两种复制起点,因此成为既能在酵母细胞也能在大肠杆菌中进行复制和传代的所谓“穿梭质粒”。

用 ARS1 片段还可以测定酵母染色体组中各个复制起点 DNA 序列的同源性。已知真核细胞染色体具有多个复制起点,根据酵母基因组中 DNA 的总长度, DNA 链合成的延伸速度以及 S 期所需时间,可以粗略地估计出大约有 70 个 DNA 复制起点分布在酵母的 17 条染色体上,但它们不是同步启动复制的。用上述 ARS1 片段作探针,与酵母总 DNA 的 EcoRI 消化物作 Southern 杂交,结果只显示 1 条 1.4kb 的杂交带。由此

推论酵母基因组中其他的复制起点 (ARS) 与 ARS1 序列相差甚远。这也许就是各复制起点不同步启动的分子基础之一。

二、着丝粒

上述插有 ARS1 的质粒虽然能在酵母细胞中复制和表达,但它和正常染色体的根本区别在于没有着丝粒。因此在有丝分裂时它不能被均等地分配到两个子细胞中去。酵母是以“出芽”方式分裂的,复制后的两个 ARS1 质粒总爱留在“母”细胞中,往往不进入“子” (“芽”)细胞。这样,经过若干个细胞世代,“母”细胞中 ARS 质粒拷贝数逐渐积累以至多达 50 个拷贝,同时整个群体中不含质粒的 urp⁻ 细胞比例迅速增加。

于是上述两个实验室又进一步设法把酵母染色体上的着丝粒 DNA 序列插入这个 ARS1 质粒中^[9,7,14,11]。由于人们已经积累了十分详尽的酵母遗传学资料,在许多染色体的着丝粒附近都标明了可供筛选的标记基因。以这些基因为靶子,就可把有关的着丝粒 DNA 片段克隆出来。例如染色体 III 上紧挨着丝粒有个“CDC10”基因,它是温度敏感突变基因“cdc10”的等位基因。用上述 ARS1 质粒克隆 CDC10 基因片段,由此重组的质粒转化 cdc10 突变株后,在 37°C 培养便可筛选出含 CDC10 的转化子。凡插入了 CDC10 基因的质粒在酵母细胞中的行为便类似于正常染色体了。在有丝分裂后“母”、“子”两个细胞各得一个质粒拷贝。可见染色体 III 的着丝粒 DNA 片段 (命名为 CEN3) 也随着 CDC10 基因一起插入了这质粒,使之在有丝分裂时能够“挂”到纺锤微管上去,从而被均等分配。

用单个细胞后代的谱系追踪分析法可测得每经一

Pan Weijun: Study on DNA Structure and Function of Centromere and Telomere by an Artificial Micro-chromosome Technique

- 1) 中国遗传学会教育工作委员会推荐稿。文中所用缩写符号为: ARS=autonomously replicating sequence=染色体中起始复制的 DNA 序列; CEN=着丝粒 (centromere) 的 DNA 序列; TEL=端粒 (telomere) 的 DNA 序列。

本文于 1985 年 9 月 10 日收到,1986 年 4 月 9 日修回。

次细胞分裂两个新细胞中有一个丢失了质粒的机率, 在无 CEN3 的 ARS1 质粒为 30%—60%, 一旦插入了 CEN3, 便降至 1%。同时, 在那些含有质粒的细胞中, 质粒拷贝数也由平均每个细胞 20—50 个 ARS1 质粒下降到 1—2 个 ARS1-CEN3 质粒。用四分体分析法还证明带 CEN 的质粒在减数分裂时也能按孟德尔法则分离。

在自然情况下, 酵母细胞含有称为“2 μ ”的内源性小质粒。每个酵母细胞平均带 50—100 个 2 μ 拷贝。一旦把 CEN 插入 2 μ 质粒, 它们就成为单拷贝的了^[16]。看来 CEN 除了保证染色体在有丝分裂时均等分配外, 似乎还在控制拷贝数目方面起作用。

根据同样的原理, 他们还克隆了酵母其他染色体 (ch IV, VI 和 XI) 的 CEN, 分别为 CEN4, 6 和 11。它们都具有 CEN3 的功能。用重叠杂交技术 (overlap hybridization or walking mapping) 可以把各个 CEN 的功能片段局限到很小, 因而测出了它们的 DNA 序列^[7]。它们的共同特点是有一个 AT 比例极高 (90% 以上) 的核心区 II 被夹在两个比较短的、相当保守的核心区 I 和 III 之间 (表 1)。

通过 CEN 缺失损伤试验, 即用限制酶切去一段 CEN, 或在 CEN 中插入几个 bp 或者用外切酶 BAL31 逐渐缩短 CEN 等, 然后检查其着丝粒功能。发现一旦伤及表 1 中核心区 II 和 III, CEN 即失去生物学功能。

用克隆的 CEN 序列作探针, 与酵母染色体的小球菌核酸酶的水解产物作 Southern 杂交, 显示了清楚的“核小体梯带”。证明染色体的 CEN 也有核小体结构。而且 CEN 核心区 I、II 和 III 对 DNA 酶 I 是高度抵抗的, 这可能是由于有某种与 CEN 紧密结合的蛋白质的保护作用有关。这种蛋白质的功能也许就是使着丝粒 CEN 与纺锤微管蛋白相结合^[3]。于是人们试图用克隆的 CEN DNA 作亲和探针, 通过亲和层析法, 特异性地吸附和纯化在有丝分裂时专门与 CEN 结合的蛋白质^[4]。

人们还克隆成功了衣藻和线虫的 DNA 片段, 这些片段能在酵母中起 CEN 的作用。可是线虫 DNA 片段的功能很弱。它们是否就是衣藻和线虫的 CEN, 还有待进一步研究^[3]。

如前所述, 带有酵母 ARS-CEN 的环形质粒可以近似于正常染色体那样在酵母细胞中复制、表达和传代。

但是, 当用限制性内切酶把这种质粒切成线形 DNA 分子后, 尽管没有伤及 CEN 和 ARS, 它也不再能在酵母中复制了, 因为它的两端没有端粒结构。

三、端粒

早在 40 多年前的经典遗传学时期, McClintock 和 Muller 就分别以玉米和果蝇为材料发现经过各种途径的损伤而断裂的染色体断口是“粘性”的, 各断口之间可以彼此重新接起造成易位、倒位、缺失等畸变, 但是染色体的天然末端似乎从来不与其他断口连接, 也不彼此相接。看来染色体的两个末端有某种称之为“端粒” (telomere) 的特殊装置。此后端粒这个符号就被画在许多教科书的染色体模式图上。然而几十年来对端粒的实际结构仍然一无所知,

70 年代, 随着分子生物学对 DNA 复制机制的研究, 从分子水平上重新提出了端粒问题。所有已知的各种 DNA 聚合酶都不能无引物地起始 DNA 链的聚合, 而只能使既有的 DNA 或 RNA 链的 3' 末端向前延伸。这样, 在 DNA 复制时, 新合成的 DNA 链的 5' 末端是如何起始的? 或者当新链 5' 端的 RNA 引子被消除后, DNA 链 5' 端的这一段空缺如何补上? 这一问题不解决, DNA 势必每复制一次就要缩短一截 (图 1)。

已经探明某些线形的病毒 DNA 分子复制时能采取多种形式来解决这一问题, 例如在复制过程中暂时地环化 (如噬菌体 λ) 或形成多连体 (如 T4, T7); 或者在核酸链末端共价键地与一特殊蛋白结合, 后者能起始新链的合成 (腺病毒, ϕ 29 和脊髓灰质炎病毒 RNA); 或者其 DNA 分子末端形成发夹结构 (痘病毒) 等等^[3, 10]。那么真核细胞染色体的末端结构又是如何呢?

可惜真核细胞染色体对研究端粒是极其不利的。假如每条染色体平均大小为 10⁷kb, 这意味着每 10⁷kb DNA 中才有 2 个端粒供研究, 于是人们很自然地注意到一种单细胞纤毛虫 (四膜虫) 的 rDNA 分子, 四膜虫编码核糖体 RNA (rRNA) 的基因在大核中扩增成为脱离染色体的、大小仅 20kb 的独立复制单位, 称作 rDNA^[6, 9]。每个细胞中这种 rDNA 分子多达 10⁴ 拷贝, 而且很容易提纯。这样, 四膜虫 rDNA 便成为研究真核生物端粒的理想材料, 因为每 20kb 中就有 2 个

表 1 酵母着丝粒 CEN3、4、6 和 11 三个核心区的序列

I	II	III
CEN3	ATAAGTCACATGAT.....88bp, 93%AT.....TGATTTCGGAA	
CEN4	AAAGGTCACATGCT.....82bp, 93%AT.....TGATTACGGAA	
CEN6	ATCACGTGCTAT.....87bp, 94%AT.....TGTTTTTCGGAA	
CEN11	ATAAGTCACATGAT.....89bp, 94%AT.....TGATTTCGGAA	

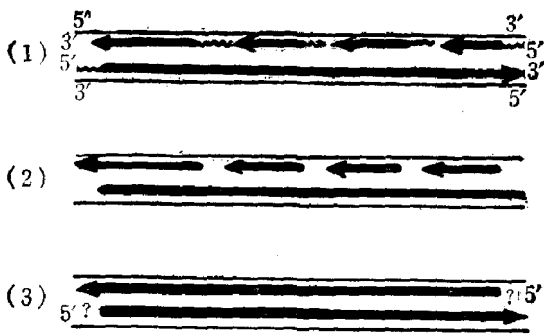


图1 设想线形 DNA 分子末端复制所遇到的问题
(1) 新的 DNA 链从各段 RNA 引子的 3' 端向前延伸。
(2) RNA 引子被消除。(3) 各段 DNA 的 3' 末端继续向前延伸并与前面一个片段的 5' 末端连接。但是新链的 5' 端由于 RNA 引子的消除而留下的空缺如何填补?
——DNA 旧链; ——DNA 新链; ~~~RNA 引子。

端粒。

1978 年 Blackburn 和 Gall 首次阐明了四膜虫 rDNA 分子的末端结构^[2],从图 2 可见,此结构的特点是:第一,仅由 2 种脱氧核苷酸构成的简单序列的大量重复,即重复 20—70 次的 (CCCCAA) 序列。第二,在这些重复序列的链上分布了许多自然的缺刻 (nicks)。第三,分子的最末端是由迴文形成的发夹结构。此后,各种不同的实验室先后分析了十多种真核生物 rDNA 或染色体 DNA 的末端结构,发现它们都具有四膜虫 rDNA 末端的那 3 个基本特征^[13](表 2)。端粒的 DNA

序列 (TEL) 包括表 2 所示的那些简单序列的大量重复区段和与其相接的向内延伸若干 kb 的一段结构较复杂的 DNA 序列^[13]。

如前所述,把插有酵母 ARS 的环形质粒用限制酶如 *Bgl* II 切成线形分子后就不能在细胞中复制了。

表 2 低等生物 TEL 中 DNA 重复序列^[13]

生物名称	TEL 的重复序列 5' → 3'
<i>Holotrichous ciliates</i> (全毛类纤毛虫)	
<i>Tetrahymena</i> (四膜虫)	CCCCAA
<i>Glaucoma</i>	CCCCAA
<i>Paramecium</i> (草履虫)	CCCCAA
<i>Hypotrichous ciliates</i> (下毛类、纤毛虫)	
<i>Stylonychia</i>	CCCCAAAA
<i>Oxytricha</i>	CCCCAAAA
Flagellates (鞭毛虫)	
<i>Trypanosoma</i> (锥虫)	CCCTAA
<i>Leptomonas</i>	CCCTAA
<i>Leishmania</i> (利什曼虫)	CCCTAA
<i>Crithidia</i>	CCCTAA
Slime molds (粘菌)	
<i>Physarum</i>	CCCTA _n
<i>Dictyostelium</i>	C ₁₋₈ T
Fungus (真菌)	
<i>Saccharomyces</i> (酵母)	C ₁₋₃ A

是, Szostak 和 Blackburn 在切口处接上两段从四膜虫

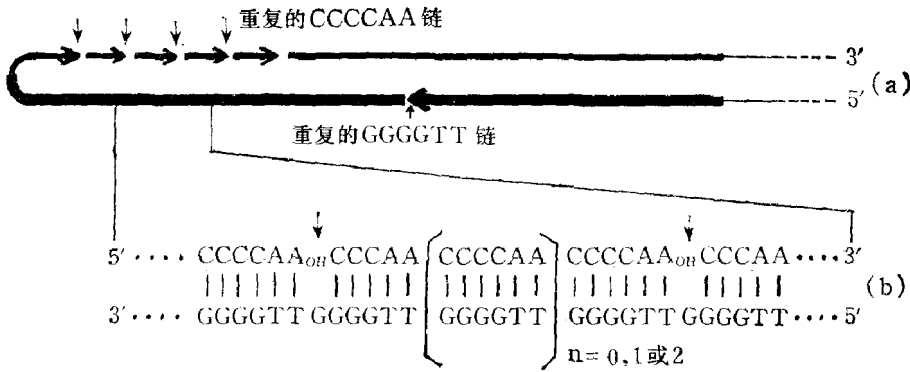


图 2 四膜虫 rDNA 末端结构^[2]
(a) 两种粗线分别代表 TEL 中 (C₄A₂) 和 (G₄T₂) 的重复序列区。↓ 表示 DNA 链上的自然缺刻。(b) 为 (a) 的局部放大。

rDNA 切下来的包含 (CCCCAA)_n 重复序列的 TEL 片段,这种插有 ARS 其两端又接上了四膜虫 TEL 的线形质粒便能在酵母中复制并表达其抗药性了^[13](图 3)。看来,四膜虫和酵母这两种分类关系极远的生物竟能共用一套 TEL。

假如用限制酶 *Pvu* I 切去这种线形质粒一端的 TEL,它便不能复制了。但切去以后再与酵母染色体

DNA 的 *Pvu* I 消化片段随机拼接并转化酵母,凡是由因此而能在酵母中复制的重组质粒显然是偶然地接上了酵母某条染色体的 TEL 片段,从而形成了一端为四膜虫 TEL,另一端为酵母 TEL 的线形质粒(图 3)。用这个技术“钓”出了不同的酵母 TEL,对它们进行序列分析,证明也是 C 和 A 的简单序列的大量重复(表 2)。

用由此获得的酵母 TEL 作探针,与酵母总 DNA

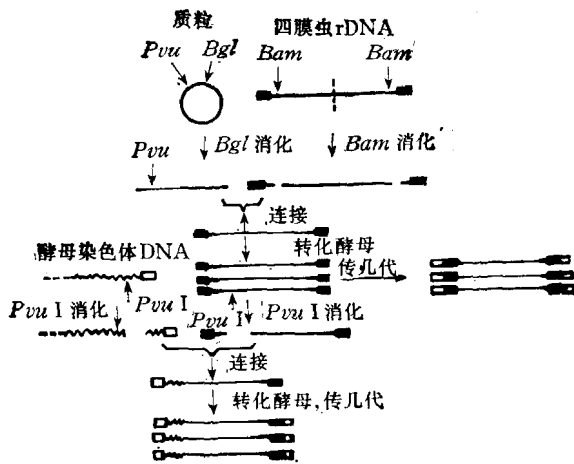


图3 四膜虫与酵母 TEL 序列的拼接

—示插入了 ARS 的 pBR322; —四膜虫 rDNA;
 ~~~酵母染色体 DNA; ■四膜虫 TEL 中 (CCCCAA)<sub>n</sub> 序列;  
 □酵母 TEL 中 (C<sub>1-3</sub>A)<sub>n</sub> 序列; |四膜虫迴文结构的 rDNA 分子的对称轴。

的各种限制酶消化物作 Southern 杂交, 证明酵母染色体组中有许多与此同源的序列。用 BAL31 外切酶消化, 证明它们中大部分也都分布在 DNA 分子的两端<sup>[12]</sup>。

Szostak 等还发现人造染色体的大小与其在有丝分裂中的稳定性关系很大。含有 ARS、CEN、TEL 和标记基因的线形染色体大小若降至 7—15kb, 它在有丝分裂中的稳定性明显下降, 即分裂后丢失质粒的细胞频率以及那些含质粒的细胞内质粒拷贝数都大为增加。当用噬菌体 λDNA 来构建人造质粒, 使其长度增加到 55kb, 则其在有丝分裂中的稳定性显著提高, 更接近于天然染色体的行为<sup>[11]</sup>。

#### 四、TEL 的复制

上述两端都接有四膜虫 TEL 的线形质粒在酵母中多次传代后被重新提取出来并分析其末端序列, 证明它依然保持着四膜虫 TEL 所特有的序列, 即重复几十次的 (CCCCAA)<sub>2</sub> 序列的外端 (末端) 竟又添加了酵母自己特有的 TEL 序列 (C<sub>1-3</sub>A)<sub>n</sub>, 而且长达 288bp 以上<sup>[12]</sup>。看来酵母能够在无模板的情况下把自己特有的 TEL 加到四膜虫 TEL 上去 (图 3)。

用下毛类纤毛虫 *Oxytrichia* 的 TEL (CCCCAAAA)<sub>n</sub> 代替四膜虫的 TEL 在酵母中复制, 其末端也被加上了酵母的 TEL 序列<sup>[13]</sup>。此外, 锥虫表面抗原基因紧挨着染色体末端, 以此基因片段作杂交探针, 可以测定锥虫该染色体末端限制酶切片的大小。人们发现, 克隆化的同步分裂的锥虫群体中, 随分裂代数的增加, 其染色体末端的限制酶切片也逐渐加长, 平均每个细胞世代增长 7—10bp, 然后又突然缩短到原来的长度<sup>[13]</sup>。被显微注射到非洲爪蟾卵中的四膜虫 rDNA 仍能在卵

中复制, 但由此回收的 rDNA 分子在其两末端被加上了非四膜虫 TEL 所特有的序列。这些外加的序列是否就是爪蟾的 TEL, 尚有待进一步工作<sup>[13]</sup>。纤毛虫分子生物学工作者们都早已熟知的一个现象是, 由纤毛虫非同步分裂群体所提取的 rDNA 末端的限制酶消化片段的大小是十分参差不齐的, 在凝胶电泳中总是呈现为一条宽而边缘不清的带, 说明各 rDNA 分子末端的 (CCCCAA) 的重复次数可以有很大出入。看来, 在染色体复制过程中 TEL 序列长短发生变化, 在生物界可能是个比较普遍的现象。

根据上述结果, Blackburn 和 Szostak 等对 TEL 的复制提出了一种假说<sup>[11,9,12]</sup>。认为把特定的重复序列加到染色体末端是 TEL 复制所固有的特性。既然酵母细胞能把自己的 TEL 序列加到四膜虫的 TEL 上, 看来这种加序是不需要模板的, 只要有某种适当的 TEL 结构预先存在即可。这样的 TEL 加序有可能是靠了类似于末端转移酶的活性在无模板的情况下从头合成的 (de novo synthesis)。也许正是靠了这种无模板合成, 抵消了前面所叙述的 (图 1) 染色体复制时其 DNA 末端 TEL 序列的丢失。

#### 参 考 文 献

- [1] Blackburn, E. H.: 1984. *Cell*, 37: 7—8.
- [2] Blackburn, E. H. & J. G. Gall: 1978. *J. Mol. Biol.*, 120: 33—53.
- [3] Blackburn, E. H. & J. W. Szostak: 1984. *Ann. Rev. Biochem.*, 53: 163—194.
- [4] Bloom, K. S., M. Fitzgerald-Hayes & J. Carbon: 1983. *Cold Spring Harbor Symp. Quant. Biol.*, 47: 1175—1185.
- [5] Clarke, L. & J. Carbon: 1980. *Nature*, 287: 504—509.
- [6] Engberg, J., P. Anderson, V. Leick & J. Collins: 1976. *J. Mol. Biol.*, 104: 455—470.
- [7] Fitzgerald-Hayes, M., L. Clark & J. Carbon: 1982. *Cell*, 29: 235—244.
- [8] Hsiao, C. L. & J. Carbon: 1979. *Proc. Nat. Acad. Sci., USA*, 76: 3829—3833.
- [9] Karrer, K. M. & J. G. Gall: 1976. *J. Mol. Biol.*, 104: 421—453.
- [10] Lewin, B.: 1985. *Gene*, 2nd ed, John Wiley & Sons, Inc. p. 553—555.
- [11] Murray, A. W. & J. W. Szostak: 1983. *Nature*, 305: 189—193.
- [12] Shampay, J., J. W. Szostak & E. H. Blackburn: 1984. *Nature*, 310: 154—157.
- [13] Stinchcomb, D. T., C. Mann & R. W. Davis: 1979. *Nature*, 282: 39—43.
- [14] Stinchcomb, D. T., C. Mann & R. W. Davis: 1982. *J. Mol. Biol.*, 158: 157—179.
- [15] Szostak, J. W. & E. H. Blackburn: 1982. *Cell*, 29: 245—255.
- [16] Tschumper, G. & J. Carbon: 1983. *Gene*, 23: 221—232.