

线粒体基因组结构研究的新进展

刘 蓉

(中国科学院生物物理研究所)

线粒体遗传系统基因组结构的研究是近年来比较活跃的一个领域。今年4月9日英国“Nature”杂志发表了人线粒体DNA分子全序列测定的结果^[1]。这是迄今为止测出全序列的最大的一个完整DNA分子,也是在分子水平上初步搞清基因结构的第一个真核细胞染色体以外的完整细胞器遗传单位。

人线粒体DNA(以下写作mt-DNA),是一个自主复制的双链环状分子,全长16,569个碱

基对。内含两种rRNA(以下写作mt-rRNA)、22种tRNA(以下写作mt-tRNA)和13种蛋白质编码序列。其特点是基因排列紧凑,遗传信息使用高度节约。不同基因之间很少或根本没有不参与编码的碱基。由线粒体DNA编码的两种mt-rRNA(16S和12S)比核仁DNA编码的rRNA(28S,18S)为小,而且没有相应的小rRNA。由于有的mt-tRNA可以通过U:N配对识别同族的四个密码子,因此mt-tRNA的

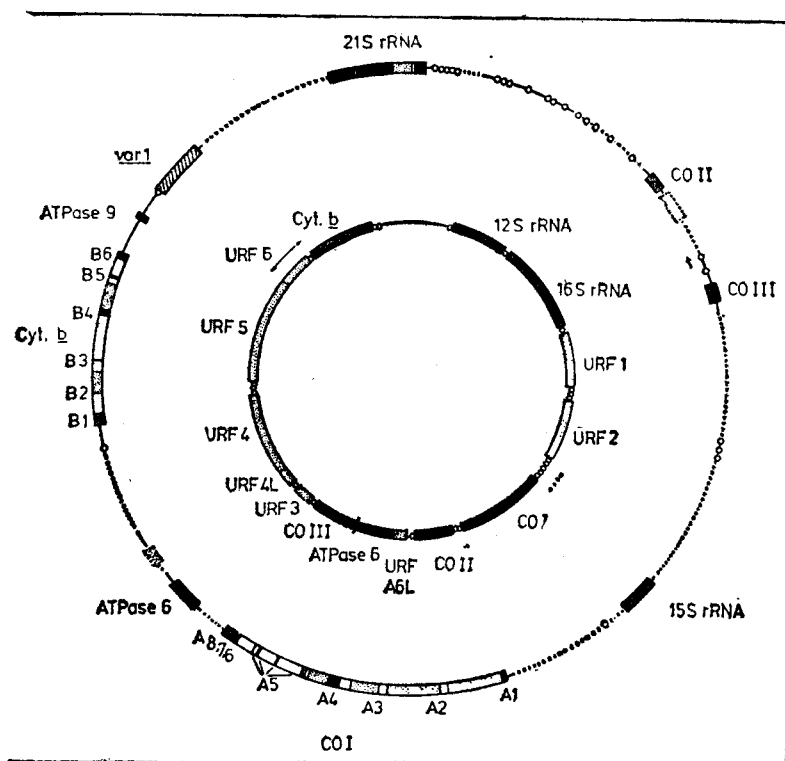


图1 酵母和人线粒体基因组结构的比较

内圈为人线粒体DNA(16,569bp)上基因排列,外圈为酵母线粒体DNA(~78,000bp)上基因排列。为便于比较,制图时把人线粒体DNA相对酵母线粒体DNA放大2倍。图中黑方块为已知蛋白质的编码序列;内有小点的方块是尚未鉴定的蛋白编码区(URF);小圆圈为tRNA基因。转录除用箭头特别标明的区域外都沿顺时针方向进行。鉴于酵母DNA序列尚未全部测出,图中外圈虚线为尚未测序部分,实线为已测出序列部分。空白方块为基因内部不含URF的插入序列(Intron)区。带斜线的方块为变种var1的遗传决定因子,此蛋白编码序列的准确位置尚未确定。(本图引自Nature 290, 443, 1981)

种类比胞质 tRNA (31 种) 为少, 只有 22 种就可满足线粒体内蛋白质合成的要求。人线粒体内为蛋白质编码的基因比真核细胞核内的蛋白质基因简单, 内部没有插入序列 (Intron), 转录和加工过程也与核内的蛋白质基因不同, 表现在转录分别从 mt-DNA 的两条链上、各自的启动区 (promotor) 开始, 全长对称转录。原初转录产物是多顺反子 RNA 链, 由 mt-tRNA 序列充当加工时的“标点”, 断裂成不同 RNA 组分。成熟的 mt-mRNA 中没有 5' 帽子结构和引导序列, 也没有 3' 拖尾区的非编码序列, 但是有 3' poly(A) 尾巴, 因此加工过程中似乎无需加帽 (capping) 和剪接 (splicing) 步骤。此外, 转译时, mt-mRNA 上遗传密码的使用也与胞质 mRNA 有一些不同, 首先是 UGA 不作终止密码而被译作色氨酸 (Trp), 其次是可以利用 AGA 和 AGG 作终止密码, 在有些场合还可利用 AUA 和 AUU 作 mt-mRNA 上的起译密码。

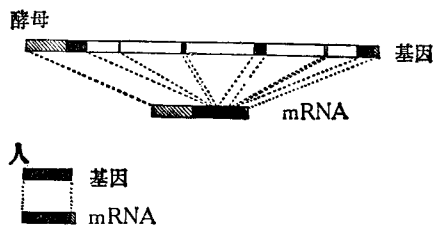


图 2 酵母线粒体和人线粒体内脱辅基细胞色素 b 基因和 mRNA 的比较

图中黑方块为编码序列, 空白方块为插入序列 (Intron)。带斜线的区域在酵母 mt-mRNA 中代表非编码序列, 在人 mt-mRNA 中为 poly(A) 尾巴。(本图引自 *Nature* 290, 440, 1981)

在人线粒体 DNA 上基因组结构研究取得重大突破的同时, 酵母线粒体 DNA 的研究也取得很大进展, 虽然还没有测出全部序列 (全长约 78,000 个碱基对), 就现有的部分序列测定的结果也可看出, 酵母线粒体基因组结构与人线粒体基因组有很大差别, 主要表现是酵母线粒体 DNA 上, 基因排列不那么紧凑, 基因之间间隔较大, 基因内部也有插入序列 (Intron)。转译分别从 5 个启动区开始, 成熟 mt-mRNA 中有 5' 引导序列和 3' 拖尾区非编码序列, 但是似乎没有 3' poly(A) 尾巴。遗传密码的使用情况既与胞质 mRNA 有所不同, 又与人 mt-mRNA

有一定差别, 其中突出的是 UGA 编码色氨酸, CUN 编码苏氨酸 (Thr)。

为了便于说明不同生物线粒体遗传单位基因组结构的区别, 现在把人和酵母线粒体内基因组结构研究的已知结果比较如下:

1. 基因组 DNA 分子全长

人线粒体	酵母线粒体
16,569 个碱基对	~78,000 个碱基对

2. 基因结构和排列: (参见图 1, 图 2)^[2]

人线粒体	酵母线粒体
(1) 含两种 mt-rRNA、22 种 mt-tRNA、和 13 种蛋白的编码序列。其中 5 种膜蛋白为已知, 它们是: 细胞色素 c 氧化酶亚基 I、II、III (Co I-III); ATPase 复合物亚基 6 (ATPase 6); 脱辅基细胞色素 b (Cyt. b)。另外 8 种蛋白尚未鉴定, 以 URF 表示。(牛线粒体 DNA 中也保留这 8 种 URF)。	(1) 含两种 mt-rRNA、25 种 mt-tRNA、5 种已知膜蛋白和若干种尚未鉴定的蛋白的编码序列 (URF)。主要膜蛋白编码序列与人线粒体膜蛋白编码序列有 50% 同源性, 但排列和结构不同; URF 则与人线粒体内 URF 无重要同源性。
(2) 基因排列紧凑, 基因之间很少或根本没有无编码作用的碱基间隔, 基因内部没有不表达的插入序列 (Intron)。	(2) 基因之间有大段的非编码序列间隔, 基因内部有表达时被删节的插入序列 (Intron)。以脱辅基细胞色素 b 为例, 就含有 5 个不表达的插入序列 ^[3] 。
(3) 多数 mt-tRNA 基因位于 mt-rRNA 基因和 mt-mRNA 基因之间, 并直接与这些基因邻接。	
(4) mt-mRNA 基因的 5' 端与 mt-tRNA 基因邻接, 3' 端与另一个 mt-tRNA 或 mt-mRNA 基因邻接 ^[4,5] 。	

3. 线粒体内 RNA 产物特点:

人线粒体	酵母线粒体
(1) 两种 mt-rRNA { 16S mt-rRNA (1559 Np) 12S mt-rRNA (954 Np)	(1) 两种 mt-rRNA { 21S mt-rRNA (3200 Np) 15S mt-rRNA (1660 Np)
(2) mt-mRNA 特点:	(2) mt-mRNA 特点:

(i) 没有 5' 引导序列, 也没有 3' 端拖尾区非编码序列。	(i) 有 5' 端引导序列和 3' 端拖尾区非编码序列。以脱辅基细胞色素 b 为例, 5' 端就有长达1000个核苷酸的引导序列, 3' 端有 50 个核苷酸的非编码拖尾区。
(ii) 半数以上的 mt-mRNA 3' 端无完整的终止密码子, 而是以 U 或UA结束, 在加接 poly(A) 以后, 形成 UAA 终止密码子。	(ii) 编码序列3'端有终止密码子。
(iii) 成熟 mt-mRNA 有 3' poly(A) 尾巴。	(iii) 成熟 mt-mRNA 无 3' poly(A) 尾巴。
(3) mt-tRNA 共 22 种, 结构不正规, T4C 环长度不等, 有些 mt-tRNA 上反密码子的第一个碱基是 U, 许可 U: N 配对而能识别同族的四个密码子。	(3) mt-tRNA 共 25 种。有些可通过三个碱基中的两个碱基配对规律而识别同族所有四个密码子。同功受体较人 mt-tRNA 为多。

4. 线粒体遗传系统基因表达的特点

人线粒体	酵母线粒体
(1) 转录特点	
(i) 两条 DNA 链上各有一个启动区, 从各自的启动区开始全长对称转录。	(i) 分别从 5 个启动区开始转录。
(ii) 原初转录产物是一条多顺反子 RNA 链。链上 mt-rRNA 与 mt-mRNA 之间由 mt-tRNA 隔开, 接点紧密邻接。	
(2) 转录调控与转录后加工特点	
(i) H 链在转录时或许边加工断裂成不同 mt-RNA 组分。	(i) 由成熟酶参与转录调控和转录后加工 ^[3] 。
(ii) 加工时断裂发生在 mt-tRNA 的前后, mt-tRNA 序列的二级结构可能是加工酶的识别信号。由此 mt-tRNA 在原初转录产物的加工过程中就起着“标点”的作用 ^[5]	(ii) 未见 mt-tRNA 的“标点”作用。
(iii) mt-mRNA 在 3' 端的切点上立即加接 poly(A) 尾巴而形成成熟 mt-mRNA。	

5. 线粒体 mRNA 中遗传密码使用的特殊情况

人 mt-mRNA	酵母 mt-mRNA
(1) 除 AUG 以外尚可用 AUA 和 AUU 作起译密码子	(1) 用 AUG 作起译密码子。AUA 和 AUU 与胞质 mRNA 的用法一样, 编码异亮氨酸 (Ile)
(2) 可用胞质 mRNA 中编码精氨酸 (Arg) 的密码 AGA 和 AGG 作终止密码子	(2) AGA 和 AGG 在酵母 mt-mRNA 中仍用于编码精氨酸 (Arg)
(3) 用胞质 mRNA 中的终止密码子 UGA 编码色氨酸 (Trp)	(3) 用 UGA 编码色氨酸 (Trp)
(4) AUA 也可编码甲硫氨酸 (Met)	(4) AUA 与胞质 mRNA 用法一样, 编码异亮氨酸 (Ile)
(5) CUN 仍用于编码亮氨酸 Leu (用法与胞质 mRNA 相同)	(5) 用 CUN 编码苏氨酸 (Thr)

从上述资料的比较来看, 线粒体遗传系统既和原核生物的遗传系统有区别, 又和真核生物的遗传系统不同, 而且不同生物来源的线粒体遗传系统也有明显的不同。以人和酵母的线粒体遗传系统为例, 不但遗传密码的使用情况不同, 而且基因排列和结构以及信息使用的节约程度都有很大区别。这些区别或许可以用线粒体内 rRNA 和 tRNA 进化迅速加以解释。例如, 不同生物 mt-rRNA 和 mt-tRNA 序列近似程度的比较表明, 即使在哺乳动物中, 大鼠和小鼠的 16S mt-rRNA 就只有 68% 同源序列; 而人和牛的 mt-tRNA^{Thr} 也不过含 74% 同源序列, 而且其中 TΨC 环长度不同, 人 mt-tRNA^{Thr} 的 TΨC 环由 3 个碱基组成, 牛 mt-tRNA^{Thr} 的 TΨC 环却有 8 个碱基。目前人线粒体的研究中还有许多问题有待解决, 首先是还需要搞清楚 8 种 URF 所编码的蛋白质的性质和用途, 其次是 mt-mRNA 在转译过程中如何与线粒体内的核蛋白体 (mt-ribosome) 相互识别并结合的问题也需要进一步探讨。由于人 mt-mRNA 在起译密码子前没有 5' 引导序列, 类似于原核生物或真核生物中 mRNA 通过起译密码子前的核苷酸序列与 16S rRNA (原核) 或 18S rRNA

(真核) 3' 端核苷酸序列中部分碱基相互作用的识别方式显然不足以解释人线粒体内 mt-mRNA 与核蛋白体的相互作用。这一问题有待今后的研究加以阐明。

- [2] Borst, P. & et al.: *Nature*, **290**, 443, 1981.
- [3] Borst, P. & et al.: *Nature*, **289**, 439, 1981.
- [4] Montoya, J., et al.: *Nature*, **290**, 465, 1981.
- [5] Ojala, D., et al.: *Nature*, **290**, 470, 1981.

[本文于 1981 年 7 月 13 日收到]

参 考 文 献

- [1] Anderson, S. et al.: *Nature*, **290**, 457, 1981.

核酸序列分析与蛋白质结构功能研究的关系

王 培 之

(中国科学技术大学 生物系)

自从生物学的研究发展到分子水平以后,科学家们竟相从新的高度在生物大分子结构、遗传学和生物膜等各个研究领域中开展研究,并且很快都取得了十分重要的成果。但在开始时,他们相互间的联系和渗透却不多,正如 J. C. Kendrew^[1]所指出的那样:当时,实际上存在着两类分子生物学家——结构学家(structurists)和信息学家(informationists)。有一个时期,这两者几乎是完全分立的两大学派。以 M. Delbrück 和 S. E. Luria 为代表的“噬菌体分子生物学家”,他们的主要兴趣是用一维的、亚显微的信息实体来分析、解释病毒和细菌的遗传学,这个信息实体最终归之于 DNA 分子;另一个学派是以 W. T. Astbury、J. D. Bernal 和他们的学生为代表的结构学派,他们的兴趣在于发展测定生物大分子,尤其是蛋白质的三维结构的技术,但对遗传学等却不关心,或者很不感兴趣。

现在这两个学派之间已有了较大的渗透,因为无论如何,遗传现象及其信息实体——核酸,与蛋白质的结构功能总是不可分割地相互联系的。正如 Crick 在 1958 年就提出的那样:生物信息的传递方向是 DNA → RNA → 蛋白质,以后又进一步发展成为分子生物学的所谓“中心法则”。到 1966 年提出了遗传密码

之后,则从根本上解决了信息联系的具体方式和确切内容,亦即基因和蛋白质的同一线性关系,或者说解决了核酸一级结构和蛋白质一级结构的专一对应关系。

1954 年 Sanger 发表了胰岛素的全部氨基酸排列顺序之后,至今已有近千个蛋白质的一级结构被测出,其中最大的蛋白质包含了一千多个氨基酸残基,成百种蛋白质的空间结构已得到确定或正在研究,并据此解释这些蛋白质是怎样利用自身的结构来行使其功能的。

但是,核酸的顺序测定工作长期以来落后于蛋白质的顺序测定,直到 1975 年,还是由 Sanger 等建立了测定 DNA 顺序的“加”、“减”法,接着 Maxam 和 Gilbert 又建立了“化学法”之后,核酸的顺序测定才变得非常快速而简便,只要几天功夫就可以测定一个含有一、二百个核苷酸的 DNA 片段,测定四、五千个核苷酸长度的 DNA 分子也只要花半年到一年的时间,以至于许多研究人员认为,这些技术的建立预告了分子生物学的新纪元的到来^[2]。而从近几年所取得的成果来看,这样的说法确实是一点也不过份。这方面已有不少专文综述,本文想就核酸序列分析与蛋白质结构功能研究的关系方面作一些介绍。