

## 动物线粒体DNA 的分子生物学研究进展

张 方 米志勇

(中国科学院发育生物学研究所 北京 100080)

绝大多数的真核生物中都含有线粒体(mitochondria, mt)这种细胞器,它自身携带DNA,可自我复制、表达,并有核基因编码的蛋白质和酶从细胞质输入线粒体,共同完成生物氧化的生理功能。通过多年来对各种生物线粒体基因组结构和功能的研究,人们推测线粒体起源于紫色光合细菌,这种细菌入侵真核生物,与真核生物形成内共生体,逐渐退化为细胞器。

1981年Anderson等人(1)测定了人线粒体基因组的全序列,到目前为止已有多种动物的线粒体基因组全序列被测定出来,如哺乳动物的牛、大鼠、小鼠、鲸、海豹、马等,鸟类的鸡,两栖类的爪蟾,以及鱼类的鲤鱼、泥鳅、虹鳟、七鳃鳗、肺鱼、矛尾鱼、Bichir等,还有一些动物的部分线粒体基因序列也已被测定。由于线粒体基因的碱基替代率(substitution rate)比单拷贝的核基因组快5—10倍(2)所以线粒体DNA一直是系统进化上一种很好的分子标记。近年来关于线粒体基因组复制与基因表达调控的研究也有了很大进展,多种参与复制、表达的酶及反式作用因子已被发现和分离,对复制与表达调控的机制也有了初步认识。线粒体基因组是分子遗传学和发育生物学研究中值得注意的一种模式体系(3)。

### 一、动物线粒体基因组的结构

已发表的全序列测定工作表明,脊椎动物线粒体基因组的长度大多在16kb左右,环状双链,根据碱性氯化铯密度梯度离心中双链密度不同分为重链(H-链)和轻链(L-链),由13个蛋白质基因、2个tRNA基因、22个tRNA基因、控制区(D-环区)和轻链复制起始区组

成。除一个蛋白质基因(ND6)和8个tRNA基因由L-链编码外,其余的大部分基因都由H-链编码。各基因间排列紧密,非编码序列比例小,基因排列的顺序基本一致(只有鸟类(如鸡)稍有改变),基因内不含内含子,碱基的使用节约、高效。线粒体基因组的这种长度短小、结构紧密的特点被认为是适应于细胞器快速复制的选择的结果。

#### 1. 蛋白质基因

脊椎动物线粒体基因组含有13个蛋白质基因,包括细胞色素b基因(Cyt b)、细胞色素C氧化酶三个亚基的基因(CO I, CO II, CO III)、NADH氧化还原酶七个亚基的基因(ND1, ND2, ND3, ND4, ND4L, ND5, ND6)和ATP酶两个亚基的基因(ATPase6, ATPase8)。这13个蛋白质或亚基都是线粒体内膜呼吸链的组分。不同脊椎动物相应基因序列的比较显示, Cyt b、CO I、CO II、CO III基因最保守,同源性最高, ATPase6、ATPase8基因、ND基因变异比较大。大多数蛋白质基因间都有tRNA基因间隔,只有ND4L/ND4、ATPase8/ATPase6、ATPase6/CO III、ND5/ND6(分别由H/L链编码)基因紧密相连,甚至彼此重叠。重叠基因的阅读框彼此相对位移一个碱基。对基因组转录产物的分析发现,人的重叠基因ND4L/ND4和鸡的重叠基因ATPase8/ATPase6分别只有一个转录本,推测其他动物也可能存在同样的情况(5)。

在已测序的线粒体蛋白质基因中,翻译起始密码子的使用比较一致。ATG是最常用的起始密码,如鱼类和鸡,除CO I基因使用GTG外,其余12个基因都使用ATG(4, 5, 6, 7, 9,

10)。GTG 是真核生物线粒体基因组中特有的起始密码子, 它的使用并不鲜见, 鼠的ND1、ATPase8 基因, 果蝇的ND5 基因以及原核生物的某些基因也以 GTG 为起始密码子。

线粒体蛋白质基因的终止密码子分两类:

1. 完全终止子, 为三联体密码子; 2. 不完全终止子, 只有 T 或 TA, 另外所需的 A 由初级转录本的 poly A 尾提供。终止密码子的使用在不同种类中存在一些差异, 完全终止子 TAA、TAG 较常见, AGA、AGG 比较少见, 以 T 或 TA 不完全终止子结尾的基因各物种各有不同。

脊椎动物线粒体基因组遗传密码的使用存在偏倚性(bias)。首先, 密码子的第三位碱基为 A 和 C 的比例明显高于 G 和 T, 其中 A 的比例最高, 在 33.4—42.0% 之间, G 最低, 在 3.8—8.9% 之间。其次, 密码子的第二位碱基为嘧啶的比例( $\% (G + C) = 68.0 \pm 0.3$ )明显高于嘌呤, 由于 NYN 类型的密码子多编码疏水氨基酸, 所以嘧啶的偏倚性直接反映了线粒体基因组编码的蛋白质氨基酸组成中疏水氨基酸的偏倚性, 这些蛋白质都是与线粒体内膜结合的呼吸链组分, 因此疏水氨基酸的偏倚性与其功能是相适应的。

## 2. tRNA 基因

脊椎动物线粒体基因组含有 22 个 tRNA 基因, 可以满足线粒体蛋白质翻译中所有密码子的需要。其中 tRNA-Glu, Ala, Asn, Cys, Tyr, Ser(UCN), Gln, Pro 由 L-链编码, 其余由 H-链编码。H-链编码的 tRNA 基因散布于蛋白质基因和 rRNA 基因之间, 相邻基因相隔 1—30 个碱基或紧密相连, 甚至也发生重叠。一般说来, 脊椎动物的 mt tRNA-Met, His, Leu(CUN)等序列保守性最高(一般达 80% 左右), 而 tRNA-Ser(AGY)的变异最大, 同源性只有 54—64% (4)。mt tRNA 也能形成典型的三叶草形二级结构, 含有 7bp 的氨基酸接受臂, 5bp 的 T $\Psi$ C 臂、反密码子臂及 4bp 的双氢尿嘧啶(DHU)臂, 各臂碱基对存在高比例的错配。mt tRNA 3'-端与核 tRNA 一样也不编码-CCA,

而在转录后由酶的作用加上。氨基酸接受臂和反密码子环较保守, 反密码子环由 7 个碱基组成, 其中包括 3 个碱基的反密码子。T $\Psi$ C 环和 DHU 环变异大, T $\Psi$ C 环缺少 T54- $\Psi$ -C-Pu-A 序列, DHU 环碱基数目明显减少, 如七鳃鳗 tRNA-Lys 的 DHU 环只有一个碱基(8), 这可能是正确终止仅含有两个 tRNA 基因的转录本转录所必需的(见下文)。但例外的是 tRNA-Leu(UUR)的 DHU 环最保守, 甚至虹鳟的与人的序列完全相同。tRNA-Ser(AGY)的 DHU 臂在鱼中保守, 而在其他大多数动物中已丢失(6)。鲤鱼、爪蟾、鲸、鸡、小鼠、牛和人等 6 种脊椎动物的 tRNA-Phe 的 DHU 臂存在一 13bp 的保守区, 而这一区域在其他 21 种 mt tRNA 中最不保守, 暗示了 tRNA-Phe 在功能上可能存在特殊性(11)。

## 3. rRNA 基因

线粒体的 12s 和 16s rRNA 基因定位于 H-链 tRNA-Phe 和 tRNA-Leu(UUR) 基因之间, 中间以 tRNA-Val 基因为间隔, 12s rRNA 基因比 16s rRNA 基因更保守。

rRNA 的二级结构很保守, 形成多个大小不一的茎环结构。环的核苷酸替代率高于茎, 并且 C-T 转换是一种常见的形式。茎区的碱基替代大多为补偿突变(compensatory mutations), 以保证碱基对的存在, 维持二级结构。茎和环的碱基组成表现出偏倚性, 即环富含 A, 而茎富含 G。这可解释为环是 rRNA 与蛋白质结合的部分, 和蛋白质发生疏水相互作用, 而 A 是四种碱基中极性最小的, 多位于与蛋白质结合的位点上; G-C 碱基对间的氢键强于 A-U 和 G-U 对, 在茎区更有利于维持 rRNA 的二级结构(12)。

对线粒体基因组转录产物的研究发现, rRNA 的量是其他基因转录产物的 15—100 倍, 这种差别可能是某种蛋白质因子结合于特定序列上, 造成转录在 rRNA 基因后提前终止的结果。有研究显示, 这种特定序列由 13 个碱基组成, 特别是其 5'-端的 7 个碱基 5'-TG-GCA GA-3' 更保守。在脊椎动物、昆虫和甲壳纲

中这一保守序列位于 16S rRNA 基因下游 tRNA-Leu(UUR) 基因的 8—14 位上, 与其相结合的蛋白因子也已被提纯(13)。

4. 非编码区

在线粒体基因组中主要存在两段非编码区, 一段是控制区(control region), 另一段是 L-链复制起始区。

控制区又称 D-环区(displacement-loop region), 位于 tRNA-Pro 和 tRNA-Phe 基因之

间, 是整个线粒体基因组序列和长度变异最大的区域, 但其中也包含有保守片段。从总体上看此区包含有 H-链复制起始区  $O_H$ , 保守序列节段(conserved sequence blocks, CSB I, II, III), L-链启动子(L-strand promoter, LSP), H-链启动子(H-strand promoter, HSP)及终止结合序列(termination associated sequences, TAS)(见图 1)。

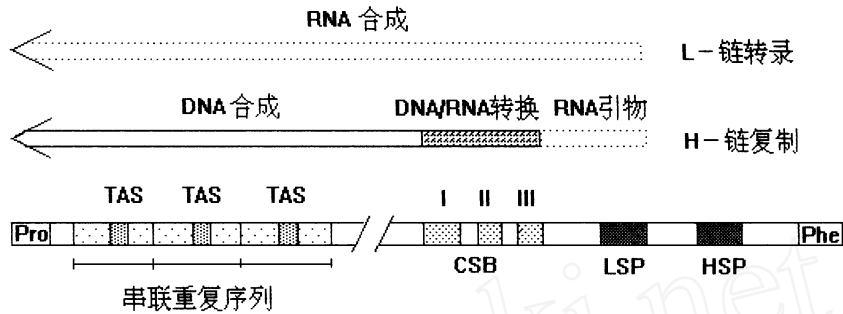


图 1 动物线粒体 D-环区结构示意图

脊椎动物 D-环 5 端通常含有串联重复序列, 拷贝数在 1—8 个之间, 是线粒体基因组长度变异的主要原因。拷贝数不仅种间有差异, 种内个体间也存在差异, 只是小于种间。在有些脊椎动物中, 个体内也存在着多种重复序列数目不同的线粒体基因组, 称为异质性(Heteroplasm y), 例如一种鲟鱼一个个体最多含 5 种不同长度的线粒体基因组。重复序列能够形成发卡结构, 发卡结构引起的高频率回复突变可能是异质性形成的原因, 也不排除父本 m tDNA 的渗漏造成异质性(14, 15)。一般每个重复序列中都含有一个保守的 TAS, TAS 可能是与 H-链复制终止有关的信号, 和 TAS 结合的反式作用因子已在牛的线粒体中发现。由于 D-环包含  $O_H$ , HSP, LSP 等重要元件, 所以 D-环长度的变异势必会影响线粒体基因组的复制和转录, 从而影响到整个生物体的代谢速率。

线粒体 D-环区 3 端一般都存在 3 个 CSB, 即 CSB I, II, III。CSB I 是最保守的成分, 几乎所有的脊椎动物都含有 CSB I, 肺鱼例外, 它只有

CSB II, III(7)。CSB II, III 保守性比 CSB I 低, 在牛、海豚、羊、鲸中 CSB III 完全缺失。CSB 是 H-链复制起始时 RNA 引物的加工信号, 对线粒体基因组的复制有重要作用(见下文)。

L-链复制起始区( $O_L$ )长约 30—50bp, 位于 L-链 tRNA-A sn 和 tRNA-Cys 基因间, 与这两个基因常有重叠,  $O_L$  可以折叠成茎环结构, 在脊椎动物中茎的保守性很强, 茎长相同, 序列间有高度同源性, 并富含 G-C 对, 而环的碱基组成、长度变异比较大。已测序的哺乳动物及果蝇的  $O_L$  环上都含有一富含 T 的序列, 虹鳟和鳕鱼的  $O_L$  环上有连续的 C 组成的序列, L-链复制起始时, RNA 引物的合成就是从这段富含嘧啶的序列开始的(6)。5'-GCCGG-3' 是  $O_L$  中 RNA 引物与 DNA 合成的转换区, 位于 H-链  $O_L$  茎区 5-端, 其序列组成和位置从人、海豹、果蝇, 到鲤鱼、泥鳅、鳕鱼、虹鳟 Bichir 都是十分保守的(6, 9)。

5. 潜在的开放阅读框

两栖类和哺乳类 HSP 的转录起始位点

于 tRNA-Phe 基因上游 35nt 处,而鸡的却在 tRNA-Phe 基因上游 156nt 处,这一段间隔区中存在一个潜在的开放阅读框(ORF),编码一个 26 个氨基酸的多肽,相应的 RNA 长 155nt,包含起始密码子 ATG 和一个线粒体通用的终止密码子。在人的 HeLa 细胞线粒体中发现了相似的 RNA (7sRNA),也含有一个线粒体通用的终止密码子及 Poly A 尾,是 HeLa 细胞中含 Poly A 尾的最多的 mtRNA,其间也有一个潜在的 ORF,编码 23 或 24 个氨基酸的多肽。7sRNA 的一部分被发现与线粒体核糖体有联系,但还不知它是否被翻译(16)。

## 二、mtDNA 的复制和表达调控

### 1. mtDNA 的复制

动物 mtDNA 的复制不受细胞周期的限制,复制与转录相关联(见图 1)。首先从位于 D-环的 LSP 转录出一段 RNA,由 RNA 加工酶(mitochondrial RNA processing, RNase MRP)将此 RNA 在 CSB 区内的某处切断,形成 H-链复制的引物,由 DNA 聚合酶  $\gamma$  (DNA polymerase  $\gamma$ , DNA pol  $\gamma$ ) 完成 H-链的复制。H-链的复制通常终止在 D-环 5'-端的 TAS 处,产生一段短的 7sDNA,7sDNA 代替相应的 H-链与 L-链配对,形成一个三链结构,这就是 D-环名称的来历。在鼠中只有 5% 的复制事件能够超越 D-环得以完成。当 H-链的复制叉移动到 OL 时,L-链才开始复制。OH 与 OL 一般相距 5000bp 以上。

RNase MRP 由蛋白质和 RNA 两种成分组成,都为核基因编码。小鼠和人的 RNA 组分分别长 275bp、265bp,同源性达 84%,与小鼠相比,人的 RNA 组分基因似乎存在局部的缺失。RNase MRP 为位点特异性的核酸内切酶,主要定位于细胞核中,但其在核中的功能还不十分清楚,有研究表明它是酵母核内 5.8s rRNA 加工所必需的,并可与核 tRNA 的加工酶 RNase P 作用于共同的底物。只有一小部分 RNase MRP 定位于线粒体中,其功能是形成 H-链复制起始的 RNA 引物。RNase MRP 以保

守的方式作用于线粒体的 D-环,识别由 LSP 转录出的 RNA 中保守的 CSB 序列,如酵母的 RNase MRP 可与人的 RNase MRP 在相同的位点切割人的 D-环 mtDNA。在鼠体外实验中 RNase MRP 的切割位点主要位于 CSB II, CSB III 也是 RNase MRP 识别所必需的,而与 CSB I 无关。但大鼠中切割位点主要位于 CSB I,包括 CSB I 在内的区域可折叠成热力学上稳定的三叶草样二级结构,这在进化上是保守的, RNase MRP 的切点就位于可折叠成稳定二级结构的序列中(17)。因此 RNase MRP 的识别信号也许不是某段保守序列,而是特殊的二级结构,这还需进一步的实验证明。

DNA pol  $\gamma$  是线粒体中存在的唯一的 DNA 聚合酶,含有两个亚基:一个是 125—140KD 的大亚基,具有 DNA 结合、5'-3'聚合和 3'-5'外切功能;另一个是 35KD 左右的小亚基,可能有维持 DNA pol  $\gamma$  结构的作用。DNA pol  $\gamma$  负责 mtDNA 的复制和修复。Pinz 等研究了 DNA pol  $\gamma$  对线粒体中两种常见的损伤——脱碱基损伤和氧化损伤(发生率在线粒体中是核中的 10 倍)——的作用,发现 DNA pol  $\gamma$  的 3'-5'外切酶活性不能十分有效地校正这两种损伤造成的不正常配对,所以这两种损伤与 DNA pol  $\gamma$  的功能可能是 mtDNA 突变率高的一个原因(18)。

在 DNA 的复制过程中,单链结合蛋白(SSBP)结合于 D-环中被置换出的单链上,以防止单链自身形成二级结构而阻碍 DNA pol  $\gamma$  的作用。SSBP 的突变会改变 mtDNA 复制的忠实性,在序列中积累错误。人的线粒体 SSBP 是由 4 个含 132 个氨基酸的蛋白亚基组成的同型四聚体,其基因已被克隆,并在染色体上定位(19)。

### 2. mtDNA 的转录与转录后加工

mtDNA 的转录与原核生物相似,两条链分别转录出多顺反子,再经过转录后加工过程,生成各种成熟的 RNA。

脊椎动物的 mtDNA 启动子都位于 D-环内。酵母的线粒体启动子与脊椎动物的不同,共

鉴定出 13 个, 散在分布于整个线粒体基因组中, 为一个保守的 9bp 序列 (5'-A T A T A A G-T A -3')。爪蟾和鸡的线粒体启动子可以进行双向转录, 爪蟾中发现两个双向启动子, 鸡中只有一个。爪蟾的每个启动子含有两个保守的 8nt 序列 (A C P u T T A T A), 分别位于两条 DNA 链上, 相互有重叠, 这种对称结构导致了启动子的双向性 (20)。鸡的双向启动子含有一个倒转重复序列, 19nt 中有 16nt 可配对折叠成一个十字型结构, H-链和 L-链的转录起始位点就位于十字型结构形成的环内, 在两个起始位点旁侧也各含有一 8nt 序列 (H-链 5'-A C A T T A T T -3', L-链 5'-A C G A T A A A -3'), 与爪蟾启动子的 8nt 保守序列各有 6nt 相同, 相似的序列在小鼠 H-链启动子区也发现过 (16)。小鼠、牛、人等哺乳动物含有两个线粒体启动子, 一般认为是非双向的, 分别负责两条链的转录, 转录方向相反。但人中发现有比例很小的反方向转录事件

发生, 且影响正方向转录的突变也影响反方向转录, 说明人线粒体启动子也具有一定的双向转录性质。从以上研究结果可以得出这样的结论: 双向转录启动子可能是代表了从酵母的单个单向启动子向人类两个基本的非双向启动子进化的中间阶段 (20) (见图 2)。

m tDNA 的转录需要核基因编码的 RNA 聚合酶和转录因子参加。人的 m tRNA 聚合酶为非特异的 RNA 聚合酶, 转录因子 h m t T F A 可结合于 H S P 和 L S P 转录起始位点的上游元件上, 是高水平特异转录的必要因素, h m t T F A 与非特异 N A 序列有高亲合性, 使 DNA 解旋、弯曲、被包裹。核苷酸序列分析表明, h m t T F A 与小的核 DNA 结合蛋白 (高迁移率族 high mobility group, H M G) 有同源性。H M G 蛋白包含一 80 个氨基酸的串联重复区, 称 H M G 盒 (H M G box), h m t T F A 中含有两个串联的 H M G 盒 (20)。

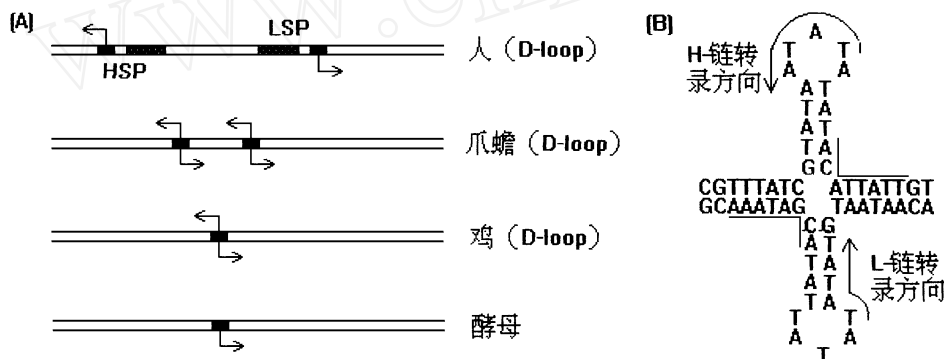


图 2 酵母和三种脊椎动物线粒体启动子的比较 (引自文献 20) (A) 和鸡线粒体启动子的结构 (引自文献 17) (B)

酵母的 m tRNA 聚合酶为核 R P O 4 I 基因编码的 150KD 的核心 m tRNA 聚合酶, 氨基酸序列显示其与噬菌体 T 7、T 3 和 S P 6 的单亚基 RNA 聚合酶有同源性; 转录因子 s c m t T F B 是核 M T F 1 基因编码的 40KD 蛋白, 与细菌的  $\sigma$  因子有一定的相似性。核心 m tRNA 聚合酶具有弱的非特异的 DNA 结合能力, 与 s c m t T F B 一起能和启动子特异结合, 形成转录起始复合物, 这一步为转录起始的限速步, 至少在体外实

验中是这样的。起始复合物的各成分相互作用, 可使 DNA 弯曲, 弯曲程度与不同的启动子突变导致转录活性的改变相关, 证明 DNA 扭曲是转录启动时的重要现象。转录启动后不久, s c m t T F B 就被释放出来, 表明其功能类似于细菌的  $\sigma$  因子, 是形成有活性的 m tRNA 聚合酶全酶的辅助因子。酵母线粒体含有与 h m t T F A 同源的 s c m t T F A, 且丰度很高, 也含有两个串联的 H M G 盒, 具有包裹、解旋、弯曲

DNA 的作用,与  $h\text{mtTFA}$  的不同在于它不是酵母  $\text{mtDNA}$  转录起始必需的特异因子,不能替代  $\text{scmtTFB}$ 。但在有核心  $\text{mtRNA}$  聚合酶和  $\text{scmtTFB}$  存在的体外实验中,加入  $\text{scmtTFA}$  可提高转录效率 3—4 倍,说明  $\text{scmtTFA}$  在  $\text{mtDNA}$  的表达中也有一定的作用(20)。

爪蟾的线粒体转录体系也含有 RNA 聚合酶和两个转录因子  $\text{xltTFA}$ 、 $\text{xltTFB}$ 。爪蟾  $\text{mtRNA}$  聚合酶大小有 140KD 左右,与酵母核心 RNA 聚合酶相似,推测它也与噬菌体 RNA 聚合酶相关。 $\text{xltTFA}$  与  $h\text{mtTFA}$  同源,含有 HMG 盒,与  $\text{mtDNA}$  启动子转录起始位点上游序列结合,刺激转录的起始,也有包裹组装  $\text{mtDNA}$  的作用。酵母、爪蟾、人三种  $\text{mtTFA}$  的氨基酸序列分析显示, $\text{xltTFA}$  和  $h\text{mtTFA}$  都含有一保守的 C-端序列。为刺激转录活性所必需;而  $\text{scmtTFA}$  缺少这段序列,可能是其不能显著刺激酵母  $\text{mtDNA}$  转录,在转录中只能扮演次要角色的原因。 $\text{xltTFB}$  分子量 40KD,是转录起始必需的成分,不能被  $\text{xltTFA}$  替代。 $\text{sc}$ -和  $\text{xltTFB}$  相似的作用可能与酵母和爪蟾线粒体转录起始位点附近都存在富含 AT 的短序列相关(21)。

从酵母和爪蟾的转录系统分析, $\text{mtTFB}$  可能是转录起始的必需因子,而  $\text{mtTFA}$  则为高水平转录的激活因子。但到目前为止,还没有从人的线粒体中分离出  $\text{mtTFB}$  的同源物。假设的  $h\text{mtTFB}$  也许包含于人的  $\text{mtRNA}$  聚合酶中,因为人的  $\text{mtRNA}$  聚合酶还未纯化到单基因组分;另一种可能是进化过程中, $h\text{mtTFA}$  适应人线粒体启动子新的组织方式,同时具有了  $\text{mtTFA}$  和  $\text{mtTFB}$  两种因子的功能,这还有待于进一步研究证明(20)。

$\text{tRNA}$  基因是线粒体初级转录本加工的“标点”信号。 $\text{tRNA}$  的加工主要包括两步:1. 核酸内切酶识别  $\text{tRNA}$  前体,切去两端序列;2. 在 3'-端加上 CCA 序列。对人  $\text{mt tRNA}$  加工的研究表明,5'-和 3'-端(5'-PO<sub>4</sub>, 3'-OH)是由两种核酸内切酶—— $\text{mt RNase P}$  和前体  $\text{tRNA}$  3'-核酸内切酶——分别作用于  $\text{tRNA}$

前体的特异位点产生的。人的细胞中含有两类  $\text{RNase P}$ :一类是负责核  $\text{tRNA}$  5'-端成熟的核  $\text{RNase P}$ ( $\text{nRNase P}$ ),另一类是负责  $\text{mt tRNA}$  5'-端成熟的  $\text{mt RNase P}$ 。这两类  $\text{RNase P}$  的作用底物不同,但体外实验中  $\text{nRNase P}$  可加工某些  $\text{mt tRNA}$ ,如  $\text{tRNA}^{\text{Leu}}$ (UUR),可能是因为其序列保守,结构与核  $\text{tRNA}$  非常相似。 $\text{mt tRNA}$  3'-端及其随后的  $\text{rRNA}$  或蛋白质基因 5'-端的形成只经过前体  $\text{tRNA}$  3'-内切酶的一次内切作用即可完成,这与真细菌  $\text{tRNA}$  3'-端的产生不同,真细菌  $\text{tRNA}$  前体 3'-端要先由核酸内切酶切断,然后再经过核酸外切酶从 3'-端逐个切去附加序列,进行进一步的修剪(trimming)。内切形成的  $\text{mt tRNA}$  3'-端在 ATP(CTP)- $\text{tRNA}$  特异的核苷酸转移酶催化下加上 CCA 序列,这样才形成了成熟的 3'-端(22)。从初级转录本上切除  $\text{tRNA}$  是动物  $\text{mtRNA}$  加工的主要步骤,但有些  $\text{mtRNA}$  两端的形成不能用  $\text{tRNA}$  的加工解释,如  $\text{CO I}$   $\text{CO III}$  和  $\text{Cyt b mRNA}$  的 5'-端,ATPase6 和 ND5  $\text{mRNA}$  的 3'-端,这还需要更深入的研究。

### 3. 线粒体蛋白质基因的翻译

线粒体蛋白质基因翻译起始的机制还不很清楚,到目前为止,只有一个线粒体翻译起始因子被鉴定,它相当于原核生物的起始因子-2(IF-2),被命名为线粒体翻译起始因子-2(线粒体 IF-2),为一单体蛋白,属于 GTP 酶家族,可在有活性的线粒体 IF-2·GTP 与无活性的线粒体 IF-2·GDP 之间转换,在 GTP 和  $\text{mRNA}$  存在的条件下,促使起始  $\text{tRNA}$  与 28S 核糖体小亚基结合。牛的线粒体 IF-2 已被纯化,人和酵母的基因已被克隆测序,推测的氨基酸序列除 N-端外,与细菌 IF-2 有极高的同源性。通过对牛的线粒体 IF-2 基因在 *E. coli* 的表达产物的分析表明,线粒体 IF-2 分为 N-端区、G-区和 C-端区三部分,其中 G-区和 C-端区是其活性所必需的,C-端区有结合核糖体的位点,但不结合  $\text{Met-tRNA}$ , $\text{mt IF-2}$  是通过先与  $\text{mt}$  核糖体结合,然后再与  $\text{Met-tRNA}$  和 GTP 结合发挥作用的(23)。

#### 4. 线粒体基因表达产物的定位和功能

一般的看法认为线粒体基因表达产物都定位于线粒体内,mt rRNA 和 mt tRNA 位于线粒基质中,与核基因编码的酶和蛋白因子共同完成线粒体蛋白质的翻译;线粒体基因编码的蛋白质单独或与核基因编码的蛋白质形成复合体,镶嵌在线粒体的内外膜和嵴上,行使生物氧化的功能。除产生ATP这一主要作用外,线粒体还参与了血红素、脂类、氨基酸、嘧啶等的形成及无机离子的代谢调控(24)。更有趣的是近几年研究发现,线粒体基因产物并不完全局限于线粒体中,例如:16s rRNA 可进入细胞质,参与果蝇胚胎发育中极细胞的形成(25);线粒体编码的ND1蛋白作为一种组织相容性抗原,出现在小鼠的细胞表面(26)。

#### 参考文献

- [1] Anderson, S. et al (1981) Nature 290: 457- 465
- [2] Meyer, A. (1994) DNA technology and phylogeny of fish. In: Beaumont, A. R. (ed) Genetics and evolution of aquatic organisms Chapman and Hall, London pp. 219- 249
- [3] 吴乃虎, 王钢锋等(1991)草鱼和鲤鱼线粒体DNA的分离纯化及其COI基因的分子克隆 动物学报 37(4): 375- 382
- [4] Tzeng, C. S. et al (1992) Nucleic Acids Research 20(18): 4853- 4959
- [5] Chang, Y. S. et al (1994) J. Mol. Evol. 38(2): 138- 155
- [6] Zardoya, R. et al (1995) J. Mol. Evol. 41(6): 942- 951
- [7] Zardoya, R. , Meyer, A. (1996) Genetics 142(4): 1249 - 1263
- [8] Lee, W. J. , Kocher, T. D. (1995) Genetics 139(2): 873 - 887
- [9] Noack, K. et al (1996) Genetics 114: 1165- 1180
- [10] Zardoya, R. , Meyer, A. (1997) Genetics 146: 995- 1010
- [11] 米志勇, 吴乃虎等(1996)鲤鱼线粒体 rRNA -Phe 基因结构的特异性 实验生物学报 29(3): 269- 272
- [12] Orti, G. et al (1996) J. Mol. Evol. 42(2): 169- 182
- [13] Valverde, J. R. et al (1994) Proc Natl Acad Sci USA 91(12): 5368- 5371
- [14] Brown, J. R. et al (1996) Genetics 142(2): 525- 535
- [15] Brown, J. R. et al (1992) Genetics 132: 221- 228
- [16] L'Abbe, D. et al (1991) J. Biol. Chem. 266(17): 10844 - 10850
- [17] Tullo, A. et al (1995) Eur. J. Biochem. 227: 657- 662
- [18] Pinz, K. et al (1995) J. Biol. Chem. 270(16): 9202- 9206
- [19] Tiranti, V. et al (1995) Genomics 25(2): 559- 564
- [20] Shadel, G. , Clayton, D. A. (1993) J. Biol. Chem. 268(22): 16083- 16086
- [21] Bogenhagen, D. F. (1996) J. Biol. Chem. 271(20): 12036- 12041
- [22] Rossmannith, W. et al (1995) J. Biol. Chem. 270(21): 12885-12891
- [23] Ma, J. , Spremulli, L. L. (1996) J. Biol. Chem. 271(10): 5805- 5811
- [24] Boore, J. L. (1997) BioEssays 19(9): 751- 753
- [25] Kobayashi, S. et al (1993) Science 260(5113): 1521- 1524
- [26] Loveland, B. et al (1990) Cell 60(6): 971- 980

#### Advance in Molecular Biology of Animal Mitochondrial DNA

Zhang Fang Mi Zhiyong

(Institute of Developmental Biology, Chinese Academy of Sciences, Beijing, 100080)

**Abstract** The mitochondrial genome of vertebrates is a double-stranded and circular DNA molecule. The complete sequences of several species have been determined. It contains genes coding for 13 proteins, two rRNAs, 22 tRNAs, and one major noncoding region (D-loop) that contains the origin of H-strand DNA replication ( $O_H$ ) and the promoters of H- and L-strand (HSP,

(下转第6页)

我国已有各类生物技术公司 200 多家, 其中已取得基因工程药物生产文号的不足 30 家, 1997 年只有两家公司(科兴、三生)年销售额超过亿元, 其余各公司销售额在几百万至几千万不等。各种干扰素加起来 1997 年销售约达 5 亿元, 居首位。全国生产基因工程药物公司总销额不及美国或日本一家中等公司的年产值。由此可见, 我国基因工程药物的产业化、规模化与国外仍有很大差距。

### 3. 我国基因工程药物研究回顾与展望

#### 3.1 仿制过多, 创新不够

通过近十几年来, 我国科学家的努力和政府的支持, 与研究化学合成药物一样, 主要集中在仿制上, 国外有什么基因工程药物, 我们就仿制什么药物, 而且仿制速度越来越快。我国已批准上市的基因工程药物仅有 rhu IFN  $\alpha$ 1b 一种是首创的, 其余均是仿制的。从表面看来这样做, 风险较小, 速度较快, 似乎与先进国家的差距缩小。实际上这样的仿制, 可能侵犯了国外的知识产权或专利。不是不可以仿制, 而是反对盲目仿制。要集中力量, 尽快仿制一批到本世纪末或下世纪初专利到期的、疗效明确、应用前景广阔的基因工程药物。

创新的难度大、周期长、投资高、风险亦高, 但不受专利限制, 不会侵犯知识产权。而且有新

可创, 如蛋白工程产品、各种融合蛋白、各种细胞因子突变体和衍生物、小分子功能肽类等等。可通过分子设计、有控制的基因修饰及基因合成, 创造世界上原来没有的, 但生物功能更优越的新型基因工程药物, 进行真正不落人后的开拓性工作。

#### 3.2 重覆过多, 水平不高, 浪费太大

基因工程药物的广阔前景和效益吸引了国内众多科学家、企业家和政府部门, 纷纷投向这一领域的研究与开发。甚至斥巨资给予大力的鼓励和支持, 使这一领域的热度越来越高。这对促进这一领域的迅猛发展、进一步缩小与国外的差距是一难得的机遇。但近年来呈现过热的状况, 表现在重覆研制同一类药物的单位越来越多, 如研制 rhuG-CSF 有 18 家公司, rhuGM-CSF 有 16 家公司, 这两种药物的临床作用基本相同, 若都投入生产, 势必没有市场, 造成极大的浪费。这种盲目无序的发展, 令人忧虑, 极需要治理和引导。实际上没有必要去过多地重覆研制同一种药品, 尚有众多已发现的基因药物有待研究开发, 应充分利用目前的大好时机, 开发一批目前国外处于 I—II 期临床的、有确切疗效苗头的基因工程药物。应广泛参与国际合作, 抓住机遇, 开发更多、更新的基因工程药物。

---

(接第 31 页)

LSP). Now adays scientists pay more attention to the mechanism of replication and expression of mitochondrial genome. The replication and transcription are believed to be strictly related. The RNA primer of H-strand DNA replication is the RNA which is transcribed from LSP and cleaved at OH by RNase MRP. The mitochondria-specified RNA polymerase, transcription factor A and B which are required for initiation of transcription have been isolated. The mitochondrial transcripts are processed by two kinds of RNase at 5'- and 3'-end of tRNAs. A Translation initiation factor, mtIF-2, which coding gene also has been cloned. This paper summarizes these recent advances of mtDNA in detail.

**Key words:** mitochondrial DNA, structure, replication, gene expression