

度保守性; DNA 序列比较表明: 基因组中各种基因序列也有较高的或至少是中等程度的保守性。而 mt DNA 的非编码区部分无论在长度上或序列上都是进化分歧最大的区域。至于其它的基因间序列及重叠序列有的亦因动物种类而有所差别,例如: 人 COII 基因与 tRNA^{Leu} 基因之间有 25 个核苷酸间隔,牛及小鼠者均仅 3 个核苷酸;人及牛 tRNA^{Leu} 基因与 URF 1 之间均只有 2 个核苷酸,而小鼠者则有 9 个核苷酸。基因组中有三对相邻蛋白基因,人、牛及小鼠的 URFA6L 与 ATPase 6 基因分别有 46、40 及 43 个核苷酸重叠;而 URF4L 与 URF 4 间则均为 7 个核苷酸重叠; URF 5 与 URF 6 系相反两条链编码,人基因组中者为紧挨连接,而牛及小鼠者则分别有 17 及 14 个核苷酸重叠。有的相邻 tRNA 基因序列也有重叠现象^[1-3]。

三、D 环区域与 H 链复制起点

哺乳类 mtDNA 的非编码区位于 tRNA^{Pro} 基因与 tRNA^{Phe} 基因之间,这个区域除含 H 链复制起点外,人 HeLa 细胞研究结果表明在 H 链复制起点附近还含有 H 链及 L 链的转录起始位点^[5]。由于在分离到的 mt DNA 一部分分子中这个区域含有与 DNA 复制有关的 D 环 (Displacement loop, 置换环),因而通常称为 D 环区域。人的 D 环区域长约 1,100 bp,小鼠者较短为 879 bp。D 环由三条 DNA 链组成,其中一条被称为 D 环链的短链(即 7SDNA),是由 H 链复制起点开始合成,并置换 H 链而与 L 链互补;被置换的那段 H 链则以单链形式存在。D 环链一般长约 520—700 核苷酸。同一动物来源的 D 环链往往因长度上的差别而有好多种。这种差别可以在 3' 端,也可以在 5' 端,不同动物可有所不同^[6]。

哺乳类 mt DNA 的复制是由 H 链合成开始,虽然一般认为 H 链是由 D 环中的 D 环链沿 H 链复制方向扩展而产生的,但是还没有确切的实验证据证明 D 环链直接被用作引物,并扩展成新生的 H 链。D 环链的转化速度要比 mtDNA 的复制速度高得多,这表明 D 环链即使

用作引物也只有很小部分可扩展成 H 链。D 环链的高转化速度及多型性说明 D 环的功能比原来预期的可能要复杂得多。除 D 环链可能作为引物外, D 环还与基因组 DNA 复制及表达的调控、以及膜依附等过程有关^[6]。

D 环区域是哺乳类 mt DNA 无论在序列上及长度上最容易发生变异的区域。不同种类动物 mt DNA 分子大小的差别主要是由于这个区域的长度不同所造成的;但 D 环区域也有比较保守的部分。比较人、牛、小鼠及大鼠序列可以看到, D 环的中央区域是 D 环区域中最保守的部分,它含有若干同源区段;而 D 环的两端区域则均缺乏同源序列^[2,3,6]。这与以前有人用异双链技术对绵羊和山羊 mt DNA D 环的观察结果是一致的^[7]。据推测, D 环中央区域可能对基因表达负有重要作用。人的 D 环序列测定结果表明, D 环序列还存在种属内差异,最容易发生变异的部分也是在 D 环的两端区域^[8]。tRNA^{Phe} 基因与 D 环之间区域是保守性最低的区域,不同动物 D 环区域长度不同主要来自这个区域的差别。这个区域也含有保守序列。人、小鼠及大鼠中在靠近 D 环链 5' 端处有三个短的同源区段,它们可能对 H 链合成的起始或终止负有作用。D 环与 tRNA^{Pro} 基因之间区域也是高度进化分歧的区域,这个区域中靠近 D 环链 3' 端处所含的短的保守序列(或重复序列)可能与 D 环链合成的终止有关^[6]。

四、L 链复制起点

当 H 链复制沿基因组进行至分子全长的大约三分之二处的 L 链复制起点时, L 链才开始复制。与 H 链复制起点不同, L 链复制起点在结构上有较高的保守性,它是一个由三十几个核苷酸组成的反向重复序列,在基因组中位于 5 个 tRNA 基因的中间(图 1)。H 链合成进行至 L 链复制起点时,上述反向重复序列暴露而形成发卡型结构而起始 L 链合成^[9]。人和牛者上述发卡型结构的双链区(茎部)结构相同,靠近 5' 端的半个环均由 T(胸苷酸)组成,小鼠者茎部与人及牛者也几乎相同,环部分则几乎

全由T组成。可见L链复制起点的茎部、环的大小及大部分一级结构均具有高度保守性^[3]。

五、rRNA 基因

哺乳类线粒体含有两种 rRNA, 即大亚单位 rRNA 及小亚单位 rRNA, 一般前者为 16 S, 后者为 12 S, 均由线粒体 DNA 编码; 但哺乳类线粒体没有 5 S 或 5.8 S rRNA, 也没有发现有其基因存在。

哺乳类线粒体 rRNA 基因具有较高保守性。人与牛的 12 S 及 16 S rRNA 基因 DNA 序列的同源率分别为 77% 及 76%, 人与小鼠者分别为 75% 及 74%。rRNA 分子中含有若干不同长度的同源序列。上述线粒体 rRNA 与非线粒体系统 rRNA 在一级结构上则只有很低的同源性。人、牛、小鼠及大鼠 DNA 序列比较表明, 线粒体 16 S rRNA 中在靠近 3' 端处有一个长二百核苷酸以上的同源区域, 这是该分子中最保守的部分, 也是线粒体 16 S rRNA 与大肠杆菌 23 S rRNA 之间同源性最高的区域。这个区域具有大肠杆菌 23 S rRNA 3' 端区域的二级结构特征。有证据表明氨酰基 tRNA 3' 端与核糖体的结合和这个区域有关^[2,3]。哺乳类线粒体 12 S rRNA 与大肠杆菌 16 S rRNA 在一级结构上只有很低的同源性, 但前者能折叠成一种与后者二级结构模型大致相类似的结构, 它很像大肠杆菌 16 S rRNA 二级结构的一种简缩形式。在原核中普遍存在的 29 个寡核苷酸, 虽然在人线粒体 12 S rRNA 中只有 11 个, 在牛中仅 9 个; 但人及牛线粒体 12 S rRNA 保留着原核小亚单位 rRNA 二级结构中所有 9 处碱基配对特征^[3,10]。

六、tRNA 基因

所有序列已知的哺乳类 mtDNA 均只含有 22 种 tRNA 基因。由于哺乳类线粒体 tRNA 除有的与密码子第三位碱基可 G:U 配对辨认两种密码子外, 有的反密码子第一位为 U 的 tRNA 还可 U:N 配对而辨认同族四种密码子, 因此其种类比细胞质中者少, 只有 22 种编码 tRNA 的

基因即可满足需要。

哺乳类线粒体 tRNA 在结构上也与其它系统中的有明显差别; 除个别例外, 绝大多数都能形成“三叶草”型二级结构, 但却往往缺少(一部分或全部)在原核及非线粒体真核系统 tRNA 中普遍存在的某些结构特征(这些特征是: 都含有一个 T- ϕ -C-R-A 序列和由 7 个核苷酸组成的“T ϕ C”环, 以及在“DHU”环及其它部位的某些共同碱基)。一个极端的例子是 tRNA^{Sec}_{Gly}, 缺少整个“DHU”臂, 人、牛及小鼠中者均如此。此外, 与其它系统 tRNA 相比, 线粒体 tRNA 二级结构中含有较多的非标准碱基对, 主要是 A-C 和 G-U。线粒体 tRNA 的另一个显著特点是 A + U 比率比非线粒体中者要高得多。所有上述这些结构特点不可避免地会导致线粒体 tRNA 三级结构弱化^[3]。

与其它系统相比, 哺乳类线粒体 tRNA 不仅结构不正规, 各种 tRNA 之间有较大差别, 而且还有比较明显的种属差异。人、牛及小鼠线粒体 tRNA 基因比较表明, tRNA 最保守的部分是反密码子环, 其次为“DHU”臂茎部及氨基酸受体茎顶端等处; 而“DHU”环, 尤其是“T ϕ C”环则无论在序列上或长度上都是最容易发生变异的区域^[2,3], 例如人 tRNA^{Thr} 的“T ϕ C”环为 CCT 三个核苷酸, 而牛者则由 ACAACTAA 8 个核苷酸组成。

虽然线粒体 tRNA 与其它系统的相比, 其种属差异非常显著, 但在哺乳类线粒体基因组中 tRNA 基因仍不失为保守性最高的一类基因, 例如人与小鼠的线粒体 tRNA 基因同源率平均值为 80%, 其中 tRNA^{Gly} 最低, 为 68%; 最高的是 tRNA^{Met} 达 90%。

七、蛋白质基因

1. 蛋白质基因的保守性及“静止取代”

前已述及, 哺乳类线粒体基因组含有 13 种蛋白质编码序列。当将人、牛及小鼠的读码排列并进行比较时, 所看到的密码变化的一个很大部分是核苷酸改变而不引起氨基酸改变, 即所谓“静止取代”(silent substitution)。这种保守

性密码变化主要发生在密码子第三位上。哺乳类线粒体蛋白基因密码子第三位上导致核苷酸变化的碱基取代的速度是很高的^[3]。

哺乳类线粒体的已知蛋白基因的保守性高于未知蛋白基因者。人与小鼠线粒体已知蛋白基因的核苷酸同源率为71% (CO II) 至77% (CO I), 平均74%; 而其氨基酸同源率则为71% (CO II) 至90% (CO I), 平均82%。除CO II基因外, 所有其余已知蛋白基因的核苷酸同源率均低于氨基酸同源率, 这种情况反映了线粒体遗传密码的衍变。在核苷酸序列歧化中, “静止取代”所占百分率为43% (CO II) 至74% (CO I), 此项数值与氨基酸同源率值成正相关。人与小鼠线粒体各种URF的核苷酸同源率在61%至71%之间, 平均67%; 其氨基酸同源率则为46%至78%, 平均64%; URF中“静止取代”所占百分率一般要比已知蛋白基因中者低得多^[2]。人与牛线粒体的蛋白基因同源性与上述情况颇相类似, 唯其同源率值均略高于人与小鼠者。

哺乳类mtDNA的进化速度与核编码的一些单拷贝基因相比是比较快的, 可以认为“静止取代”是线粒体系统在进化速度比较迅速情况下维持其蛋白质的结构与功能的一种方式。

2. 关于未知蛋白基因(URF)

虽然8种URF的保守性与已知蛋白基因相比均较低, 对它们编码的蛋白质的功能也一无所知, 但URF的转录产物在结构及代谢性质上都具有mRNA的特征^[11-13], 也就是说, 它们应是编码蛋白质的基因。最近有人根据人线粒体DNA序列用化学方法合成URFA6L编码的C端九肽及N端八肽, 然后用这两种多肽产生的抗体分别鉴定HeLa细胞线粒体转译产物, 结果两种抗体均能与转译产物中的一种蛋白组分产生特异性免疫沉淀。经电泳“指纹”分析证明, 此蛋白组分为URFA6L的表达产物, 进一步证明了URF确实是线粒体系统所合成蛋白质的基因^[14]。

在8种URF中, 以与ATPase 6基因序列有重叠的URFA6L的序列为最短, 保守性也最

低。由URFA6L序列推出的蛋白质其氨基酸组成极不寻常, 亮、脯、苏、赖四种氨基酸占氨基酸残基总数一半以上。其N端为非极性。由N端起一连有二十多个中性氨基酸; C端则富含赖氨酸, 呈罕见的碱性。值得注意的是大肠杆菌的unc操纵子中的ATPase 6基因前面的一个基因其所编码的蛋白质有与上述情况相类似的电荷分布^[3]。

3. 遗传密码系统中的一些特殊情况

哺乳类线粒体蛋白基因在遗传密码使用上有与原核和非线粒体真核的通用密码不同的某些特殊之处。在哺乳类线粒体系统中UGA编码色氨酸而不用为终止密码子。虽然大部分基因用AUG作为起译密码子, 但牛线粒体基因有的用AUA, 人者有的用AUA及AUU, 小鼠中起译密码子为AUN, 即需用四种密码子。AUA也为甲硫氨酸编码, 但不编码异亮氨酸。在人及牛线粒体基因中, AGA及AGG被用作终止密码子, 而不编码精氨酸, UAA亦被用作终止密码子, 小鼠中则只用UAA作为终止密码子。在线粒体13种蛋白基因中只有一部分基因DNA中含有终止密码子, 还有一部分基因虽然其DNA序列中没有终止密码子, 但其DNA序列表明其转录产物3'端为U或UA, 由于线粒体mRNA 3'端含有多聚腺苷酸(poly(A))尾巴, 因此上述转录产物在转录后加工过程中加接多聚腺苷酸便形成UAA终止密码子^[1-3]。

八、基因表达的特点

与真核细胞核基因相比, 哺乳类线粒体基因比较简单, 内部没有不表达的间隙序列(intron, 内含子)。哺乳类线粒体mRNA也与一般真核mRNA不同, 其5'端没有帽子结构, 并且没有具有一定长度的前导序列^[12], 3'端没有不编码的拖尾区序列, 但有多聚腺苷酸尾巴^[15]。以上特点及基因组中所特有的基因排列方式说明其基因表达与非线粒体系统有所不同。现将一种已为人们接受的哺乳类线粒体DNA H链转录模型^[15]简述如下: H链是作为

一个多顺反子从其转录起始位点进行转录的,原初转录产物通过一种加工过程产生各种 rRNA、tRNA 及 mRNA 分子。由于从 H 链转录产物或转录复合物中分离不到太大的 RNA,因此转录产物可能是依附在 mtDNA 上,以复合物形式边转录边加工的。加工切割位点是在 tRNA 序列前后;tRNA 的二级结构可能是加工酶的主要识别信号。有个别加工位点没有 tRNA 序列,在这种情况下,加工酶可能是通过识别与 tRNA 二级结构相类似的某种茎一环结构(发卡型结构)而加工的。切割所产生的 mRNA 随即在 3' 端进行腺苷酸化作用加接多聚腺苷酸尾巴。线粒体 mRNA 3' 端的腺苷酸化作用可能与核者不同,因为核 mRNA 在加接多聚腺苷酸位点前面有一作为腺苷酸化信号的 AAUAAA 序列,而大多数线粒体 mRNA 3' 端则没有这种序列,而且也没有其它共同性序列^[2,3]。除 mRNA 外,哺乳类线粒体 rRNA 在加工过程中可能也进行腺苷酸化。据报道,仓鼠及人 HeLa 细胞线粒体大及小亚单位 rRNA 绝大多数分子的 3' 端有不同长度的寡聚腺苷酸尾巴,而且还发现 rRNA 3' 端寡聚腺苷酸前面的序列也是参差不齐的^[16,17]。L 链转录时也像 H 链一样是由其转录起始位点开始进行全长转录。L 链信息量低,但其转录速度却比 H 链至少快两三倍,而且大部分转录产物转化速度非常高^[9],这是出乎一般意料之外的。这意味着除转录 L 链编码的 8 种 tRNA 基因及 URF 6 外,L 链转录可能还有其它作用。

由于哺乳类线粒体 mRNA 5' 端缺乏前导序列,其 5' 端编码序列也没有什么共同特征,因此,在转译过程中,线粒体 mRNA 如何与核糖体相互识别并结合是一个有待研究的问题。

九、基因序列进化的一些情况

哺乳类线粒体基因组的结构具有高度保守性,基因组中的基因也有较高的或至少是中等的保守性,但与核的某些基因相比,mtDNA 序列的进化速度是比较高的,大约每核苷酸位置每年有 5×10^{-9} — 10×10^{-9} 碱基取代变化^[18,19]。

在此种情况下,线粒体各种基因在序列上均呈现出一定的进化分歧,但所有基因的序列变化都不是随机的。基因中都存在着某些高度保守的部分,即使进化分歧最大的 D 环区域也含有比较保守的部分。正如人们所预期的,已知的一些例子表明:结构中的保守部分是那些负有重要功能或者在维系生物大分子的结构和功能上具有重要作用的部分。此外,应该提一下的是:在进化过程中哺乳类线粒体蛋白基因密码子三个位置的碱基取代速度有显著差别;第三位碱基取代速度最高,其次为第一位碱基,最低的是第二位碱基。密码子第一、二、三位碱基取代速度比为 1.9:1.0:6.0^[3]。这种碱基取代方式及前面介绍的反密码子与密码子的 U:N 配对方式可使“静止取代”在整个碱基取代中占有相当大的比率,从而在相当程度上维持了氨基酸序列的保守性。以上所述均反映了进化过程中功能对结构的制约作用,同时也说明基因序列的进化比较对深入研究其结构与功能关系也有重要意义。由于动物 mtDNA 进化较迅速,比较其序列的同源性还可作为测定相近动物种属的亲缘关系的一种有效手段。

十、关于线粒体起源问题

哺乳类线粒体基因组与非线粒体真核或原核的遗传系统之间都有很大差别;在某些方面即使有同源性也是很低的。在非线粒体真核和原核遗传系统的某些有共同性的方面,例如遗传密码及 tRNA 结构特征,线粒体系统中者亦与之有所不同,更显示了线粒体基因组是一种独特的遗传系统。哺乳类线粒体与其它线粒体例如酵母线粒体的基因组也有很大区别。虽然二者的几种已知蛋白基因之间有约 50% 同源性,但二者的 URF 则几乎无任何同源性。酵母线粒体的遗传密码既与其它系统有所不同,又与哺乳类线粒体者有某些差别;它的 mRNA 有 5' 端前导序列和 3' 端拖尾区非编码序列;而且酵母线粒体基因组中基因之间有较大间隔,有的基因内部有间隙序列^[4,20]。以上不仅说明不同线粒体系统之间存在着一些重大差别,而

且表明不同线粒体系统以及它们与非线粒体真核及原核之间的差别也是错综复杂的。由现有基因组方面资料看来,还很难对线粒体的起源作出明确的推断。

结 束 语

近年在哺乳类线粒体基因组的研究上虽然取得了很大进展,但在基因组 DNA 复制、基因表达以及其调控机制等还有许多问题需要搞清楚;8 种 URF 蛋白的性质和功能更有待研究解决;已知的线粒体基因组全序列也仅限于少数几种动物。为了进一步阐明哺乳类 mt DNA 的结构和功能及探索解决线粒体进化起源问题,除在上述有关方面进行更深入研究外,还需要更多的全序列资料。

参 考 文 献

- [1] Anderson, S. et al.: *Nature*, **290**, 457, 1981.
- [2] Bibb, M. J. et al.: *Cell*, **26**, 167, 1981.
- [3] Anderson, S. et al.: *J. Mol. Biol.*, **156**, 683, 1982.

- [4] 刘蓉:《生物化学与生物物理进展》,1981 年,第 5 期,第 7 页。
- [5] Cantatore, P. et al.: *Nucleic Acids Res.*, **8**, 2605, 1980.
- [6] Walberg, M. W. et al.: *Nucleic Acids Res.*, **9**, 5411, 1981.
- [7] Upholt, W. B. et al.: *Cell*, **11**, 571, 1977.
- [8] Greenberg, B. D. et al.: *Gene*, **21**, 33, 1983.
- [9] Martens, P. A. et al.: *J. Mol. Biol.*, **135**, 327, 1979.
- [10] Anderson, S. et al.: *Mitochondrial Genes* (ed. P. Slonimski et al.), p. 5, 1982.
- [11] Ojala, D. et al.: *Cell*, **22**, 393, 1980.
- [12] Montoya, J. et al.: *Nature*, **290**, 465, 1981.
- [13] Gelfand, R. et al.: *Mol. Cell Biol.*, **1**, 497, 1981.
- [14] Mariottini, P. et al.: *Cell*, **32**, 1269, 1983.
- [15] Ojala, D. et al.: *Nature*, **290**, 470, 1981.
- [16] Dubin, D. T. et al.: *Cell*, **23**, 271, 1981.
- [17] Dubin, D. T. et al.: *Mitochondrial Genes* (ed. P. Slonimski et al.), p. 89, 1982.
- [18] Brown, W. M. et al.: *Proc. Nat. Acad. Sci. U.S.A.*, **76**, 1967, 1979.
- [19] Brown, W. M.: *Proc. Nat. Acad. Sci. USA.*, **77**, 3605, 1980.
- [20] Mobrega, F. G. et al.: *J. Biol. Chem.*, **255**, 9828, 1980.

[本文于 1984 年 6 月 1 日收到]

膜脂—膜蛋白的相互作用(下)

杨福愉 黄 芬 王金凤

(中国科学院生物物理研究所,北京)

四、膜蛋白对膜脂的影响

1. 膜蛋白与膜脂的结合

膜蛋白按其不同类型可以多种方式与脂双层结合,现叙述如下:

(1) 外周蛋白的结合

外周蛋白通过静电力(蛋白带电荷)或范德瓦力(蛋白不带电荷)与膜结合。在蛋白与脂的结合面处含有大量的水。

(a) 静电结合 细胞色素 C 是外周蛋白通过静电力与细胞膜表面结合的一个典型例子(图 3 a)。细胞色素 C 整个分子具有可观的非对称电荷分布,在中性 pH 时,它带有过量的

正电荷。细胞色素 C 与含有酸性和中性磷脂的脂双层静电结合。当带负电荷的磷脂如 DPG 增加时,结合到脂双层的细胞色素 C 的量也增加。早在 1975 年 Van 和 Griffith^[34] 就已通过 ESR 谱观察自旋标记了的细胞色素 C—磷脂系统的相互作用,并得出细胞色素 C 只是位于脂双层表面,对脂双层的特性影响很小的结论。但是,他们也指出,上述结果并不排除细胞色素 C 与脂双层的静电结合对靠近极性“头部”脂双层的扰动以及可能引起的脂的分相。

(b) 范德瓦相互作用 分子中非极性基团具有偶极矩和诱导偶极矩,这些偶极矩与偶极矩,偶极矩与诱导偶极矩之间的吸引力,共同组