



动物细胞基因组的 DNA 顺序结构和表达

十年以前人们对于生物体基因表达的认识主要来自对简单的原核生物如细菌的研究,在这些生物中,基因的表达和调节过程相对地比较简单。但是,真核生物尤其是高等生物的基因表达和调节远比细菌复杂得多。一个病毒的基因组通常有 $2-100 \times 10^6$ 道尔顿的 DNA;细菌有约 2×10^9 道尔顿的 DNA;而一个哺乳动物细胞基因组有 2×10^{12} 道尔顿的 DNA,比细菌大 1000 倍。近年来,生物学家在阐明真核生物基因表达奥秘的努力中正开始取得越来越多的进展。高等生物基因组 DNA 的变性-复性实验说明在这些基因组中存在着和单拷贝 DNA 顺序间隔排列的重复的 DNA 顺序。真核细胞结构基因内插入顺序的存在提示了真核生物基因的不连续性。真核生物基因组中 DNA 顺序结构的特点和基因转录-转译的特异模式是生物体发育、分化的分子基础。

DNA 顺序排列的两种型式

真核细胞中重复顺序一般是和包括结构基因在内的单拷贝 DNA 相间隔的。研究证明许多不同种属的生物显然有着十分相似的 DNA 顺序结构型式,这种型式首先在爪蟾中被证明,故称为“爪蟾型”,其特点是比较短的单拷贝顺序和重复顺序相间排列,在爪蟾中,它的重复顺序平均长度 350 核苷酸,单拷贝 DNA 约 700-900 核苷酸。爪蟾内 70% 的单拷贝 DNA 按这种型式组成,其余单拷贝 DNA 组成较长的顺序。而重复顺序的 75% 具有较短的顺序,其余的至少有 2000 核苷酸。

在果蝇基因组中存在着另一种 DNA 顺序结构型式,其重复顺序平均长度 5600 核苷酸,而单拷贝顺序平均至少 13000 核苷酸,甚至可能多到 30000 核苷酸。该种型式称为“果蝇型”。

大部分生物具有爪蟾型的基因组结构,只有果蝇、摇蚊、蜜蜂三种昆虫属于果蝇型的结构型式。最近另有报告,鸡的基因组介于这两种结构型式之间。真核生物基因组中 DNA 顺序结构的特点包含着至今尚未被人们完全认识的重要意义。

单拷贝顺序信息 RNA 群和优势分类

就分化的动物细胞中发现的 mRNA 群之结构而论,把它们分为三种信息类别可能是有帮助的。**复杂类** mRNA, 在每个细胞中出现的水平是一到几个拷贝;**中等优势** mRNA, 在每个细胞中出现高到几百个拷贝;**超优势** mRNA, 在每个细胞中出现超过 10^4 拷贝。表 1 收录了包括很广范围的生物体和细胞的测定结果。就象表 1 中清楚地说明的那样,动物细胞的 mRNA 浓度通常是从每个细胞几个拷贝到几百个拷贝。复杂类 mRNA 群有多种多样的顺序,足以编码大于 10^4 个不同的蛋白质。名目繁多的结构基因,大约有十分之一确实是表现在中等优势类 mRNA 中的,计算双向凝胶电泳分析中观察到的蛋白质数目,也可获得各种中等优势 mRNA 顺序数目的最小估计值。这些凝胶至少分离开 500 种新合成的蛋白质,这和用互补 DNA(cDNA)杂交测定中等优势类 mRNA 顺序的复杂性近似相符。

超优势 mRNA 存在于某些高度分化细胞中,并且一直是集中研究的目标。在一些特例中,细胞内全部 mRNA 的一大部分可能由一种或几种信息所组成。一个动物细胞含有大于 10^5 的 mRNA 分子,每个细胞的超优势 mRNA 数目常常比典型的中等优势 mRNA 大一百倍以上。例如,下蛋母鸡的输卵管,每个细胞含有 $1-1.5 \times 10^5$ 的卵清蛋白 mRNA;小鼠或小鸡的网织细胞,各有 4×10^4 和 1.5×10^5 的珠蛋白 mRNA 分子。超优势 mRNA 显然是对特定分化状态专一的,而且,在最终分化过程中,它们在核内的合成速度有显著的改变。但是,正像表 1 所指出的,超优势 mRNA 并不总是很多的,有时只占 mRNA 总量的一小部分。虽然超优势 mRNA 对特定类的细胞是极为重要的,但它只代表了生物体所需要的各种结构基因的一小部分,带有超优势 mRNA 的细胞也利用许许多多属于其它优势类别的 mRNA (表 1)。

比较各种组织的复杂类聚核糖体 mRNA,显示

表1 动物细胞 mRNA 群的优势性分类

组织或细胞 类型	信 息 类 别					
	复 杂 (每个细胞 1-15 拷贝 的顺序)		中 等 优 势 (每个细胞 15-300 拷贝 的顺序)		超 优 势 (每个细胞 10^4-10^5 拷贝的顺序)	
	SCC (kb)	量 (%)	SCC (kb)	量 (%)	SCC (kb)	量 (%)
小鼠肾	2×10^4	45	0.1×10^3	45	7	10
小鼠肝	1×10^4	40	1×10^3	35	10	25
小鼠脑	1.1×10^5	47	1.5×10^3	37	20	14
小鼠 Friend 细胞	0.5×10^4	15	1×10^3	75	2	10
鸡 肝	2×10^4	45	1×10^3	40	2 ^[1]	15
鸡输卵管	3×10^4	35	NA ^[2]	NA ^[2]	2 ^[3]	50
鸡肌纤维	3.2×10^4 ^[4]	50	0.3×10^3	30	12	20
海胆原肠胚	1.7×10^4 ^[4]	20 ^[5]	2×10^3	80	N.O. ^[7]	
爪蟾蝌蚪	3×10^4	40	1×10^3 ^[6]	60	N.O. ^[7]	

[1] 可能是白蛋白 mRNA。

[2] 作者们报告另外一个小的(15%)mRNA 成分, 由每个细胞内存在 4000 次的顺序组成。这个成分的复杂性是 15kb。

[3] 卵白蛋白 mRNA。

[4] 这些测定从单拷贝 DNA 饱和方法得到, 其它数据系用 cDNA 测定。

[5] 在所引的参考资料中, 估计复杂类是 10%, 但新的关于高盐浓度在 RNA 与过量 RNA 反应中的动力学效应的数据说明比较确当的估计是 20%。

[6] 作者们报告两种间隔的中等优势类 mRNA, 由每个细胞内存在 110 和 630 次的顺序组成。这些类别的数据已被合并起来。

[7] 没有测定。

它们是受到严格调节的。在海胆中, 胚胎的复杂类 mRNA 顺序群里, 有不到 20% 的顺序是普遍存在的, 而在卵母细胞、胚胎和成年组织的 mRNA 顺序里, 有相当于成千个结构基因量的单拷贝顺序是不相同的。在某个聚核糖体 mRNA 样品中, 被记为“缺失”的 mRNA 顺序可能是在每个细胞中存在小于 0.05 个拷贝, 而不是通常的一到几个拷贝。最近对编码海胆卵母体 mRNA 顺序的无性繁殖系研究中, 有两个特定复杂类结构基因调节的例子是很有意思的, 这些无性繁殖的顺序以一般复杂类信息的水平出现在早期胚胎聚核糖体中, 但是它们的转录物以后基本上从细胞质里消失。在成年小肠细胞的 mRNA 中也发现一个这类的特别信息。用哺乳类和鸟类材料所作的一些研究也表明分化细胞的复杂类 mRNA 顺序群的调节有变化。复杂类 cDNA 探针在同小鼠肝和肾 mRNA 反应时, 显示 10—20% 的差异。在鸟类肝及输卵管单拷贝 mRNA 顺序群之间, 有同样的相当于几

千个基因的顺序之差别。同样, 小鼠脑 mRNA 的总 cDNA 和 L 细胞 mRNA 交叉反应仅是 45%, 和肾 mRNA 交叉反应是 75% (见下面)。在这些例子中, 不发生交叉反应的, 大部分显然属于复杂类 mRNA。

在不同组织和发育阶段中, 特异的复杂类 mRNA 群的存在表明这些信息的翻译和细胞分化的状态有关。也有例子说明, 肝脏专一的酶是由复杂类 mRNA 编码的。虽然结论决没有定局, 但我们主张这样的观点: 成千上万个复杂类结构基因的调节是动物细胞功能和分化的基本分子活动。

用双向凝胶电泳比较哺乳动物和海胆胚胎发生时合成的蛋白质, 表明中等优势类信息至少在定量水平上被调节, 也就是, 发育时某些特定的蛋白质出现了, 而另一些消失了。“消失”意味着 mRNA 顺序浓度降低到复杂类信息的水平; “出现”则反之。确实, 在 cDNA 杂交研究中看到几个明显的例子, 表明 mRNA 顺序浓度有很大的变化。

核 RNA 顺序群

出现在任何一类细胞或组织 nRNA 中的单拷贝顺序, 仅仅一小部分也出现在它的 mRNA 中。在海胆胚胎中(取决于阶段), nRNA 顺序复杂性的 10—20% 组成胚胎的 mRNA 顺序; 在大鼠肝中, 这个数

值约为 11%; 小鼠脑中大约 18%; 培养的果蝇细胞约 4—6%。最近在海胆和小鼠两个系统的研究中都取得一个惊人的结果: 某类细胞的聚核糖体 mRNA 顺序群, 在其它类细胞 mRNA 中虽然是完全缺失的, 但却出现在它们的 nRNA 中(表 2)。同样地, 上面提到过的无性繁殖的海胆母体 mRNA 顺序, 虽

表 2 海胆和小鼠不同组织 mRNA 及 nRNA 中结构基因顺序群的比较

探 针	和亲本 mRNA 反应 ^[1]		和其它 mRNA 反应		和 nRNA 反应	
	mRNA	%	mRNA	%	nRNA	%
与海胆原肠胚 mRNA 互补的单拷贝 DNA	原肠胚	100	小肠 胚胎细胞	12 <13	小肠 胚胎细胞	97 101
与小鼠脑 mRNA 互补的总 cDNA	脑	100	肾	78	肾	102
与小鼠脑稀有信息互补的 cDNA ^[2]	脑	100	肾	56	肾	100

[1] 海胆探针和亲本 mRNA 的反应为 78%, 以 100% 表示, 其他 RNA 样品的反应以相对百分率表示

[2] 此 cDNA 探针是一个复杂类信息探针, 制备方法: 从脑聚核糖体 poly(A)-RNA 转录得到 cDNA, 并和亲本 RNA 反应到 Cot 20, 不反应部分(总 cDNA 的 38%) 收集起来即是, 它和脑 mRNA 的反应力是 90%。

然只是在早期胚胎的聚核糖体才有, 但却是也存在于晚期胚胎的 nRNA 中, 并达到和其它单拷贝顺序转录物同样的水平。另有报告, 在不生成红血球的组织之 RNA 中存在着低水平的珠蛋白 mRNA 顺序; 在脾脏和肝脏 RNA 中出现卵清蛋白 mRNA 顺序, 虽然相反的结果也有报道。上述资料的解释是, 每个分化了的细胞核不仅包含着生物体曾经用过的所有基因, 而且还有着几乎所有这些基因的转录物。

虽然表现为信息的那些结构基因顺序的转录物确是普遍都有的, 因此每组 nRNA 都是全套的, 但这些可能只是总的 nRNA 复杂性的一小部分。在海胆中, 估计这一部分最高约 25%。不同类细胞的 nRNA 并不包含有一模一样的顺序群, 比较海胆的成年和胚胎期 nRNA, 证明它们共同有一个单拷贝顺序核心, 它占了总 nRNA 顺序复杂性的约 80%, 我们预测这共有的顺序核心包括了结构基因转录物中的单拷贝插入顺序。然而, 成年海胆小肠 nRNA 的 20% 并不出现在胚胎 nRNA 中, 海胆其它 nRNA 顺序群的差别更小些。哺乳类各组织的 nRNA 也具有一个很大的共同的单拷贝顺序群。如果不包括脑 nRNA, 其数量达到总 nRNA 复杂性的 40%; 如果包括脑, 则为约 20%。因此, 比起海胆来哺乳类的每一类细胞有较多 nRNA 的转录是细胞专一的。小

鼠脑 nRNA 中相当多的单拷贝转录物显示不含有结构基因顺序。概括起来, 从现在所引用的实验结果得出一个令人惊奇的结论: 单拷贝结构基因顺序可能是普遍地存在于各种 nRNA 中的, 而至少某些非结构基因顺序却是被专一地转录的。

海胆胚胎的不均一 nRNA, 大部分以约 20 分钟的半寿期进行转换, 在原肠胚期, 大部分的快速转换 nRNA 是由每个核内存在的一个的单拷贝顺序转录物组成的, 海胆 nRNA 中的复杂类结构基因顺序转录物之稳定态浓度和总的单拷贝顺序转录物浓度是一样的, 不论这些转录物中是不是有那些要被输送到细胞质里去的。将 nRNA 单拷贝顺序的平均合成速度(例如, 结构基因转录物平均)和 mRNA 在细胞质聚核糖体中出现的速度相比较, 在海胆胚胎中, 复杂类和中等优势类 mRNA 的转换速度常数差不多是相等的(半寿期, $t_{1/2}$, 约 5 小时), 所以, 正如以前所指出的, 它们的优势性的差别只是和 mRNA 在细胞质内出现的速度相平行。中等优势类 mRNA 出现的速度是典型的单拷贝 nRNA 转录物合成速度的两倍之内, 而细胞质内复杂类 mRNA 出现的速度比每个顺序 nRNA 合成速度远远低得多。所以, 在海胆胚胎中, nRNA 结构基因转录可能有一个近于一致的速度, 并且取决于要被加工和输出的核内前身物部分, 使中等

优势类和复杂类 mRNA 具有稳定态的浓度。

显然不必用假定结构基因转录起始速度各不相同来说明海胆胚胎 mRNA 优势程度之差异。从小鼠 L 细胞也能得出这些结论来。L 细胞的中等优势和稀有信息之出现速度与测出的任何 nRNA 单拷贝顺序转录物合成速度相比,或是相同或是低一些。为了进行这一计算,从总的不均一 nRNA 合成速度和由小鼠其它细胞测量值推算的 nRNA 复杂性,来测定每一个这种转录物合成的大概速度。非诱导的小鼠 Friend 细胞提供一个这方面的例子:和在细胞质内优势程度相差大于 30 倍的各种 mRNA 互补的 cDNA,却是以基本上单一成分的动力学特性和 nRNA 起反应,也就是说, nRNA 中所有的结构基因转录物几乎是达到同样的稳定浓度。这个结果意味着,如若它们在核内的半寿期没有很大差别的话,它们是以同样速度转录的。其它一些 cDNA 研究也获得相似的结果。

结构基因表达是否有转录水平的调节?

上面的论述表明,在结构基因转录中,只有超优势类 mRNA 的基因转录才被证明具有细胞专一的变化。最近的资料表明,在分化时这一类数量不多但很特别的结构基因的转录起始速度有很大的提高。还不知道这些基因究竟是从完全“关闭”状态还是从低水平“打开”状态被调高起来的。

没有确切的证据说明动物细胞中大部分结构基因有转录水平的调节。虽然我们从前引用的不太多的资料,推想复杂类和中等优势类结构基因顺序是普遍地存在于所有 nRNA 中的,但是,这一点也并不一定要求这些基因仅仅是在转录后水平上被调节的。普遍存在的 nRNA 基因顺序转录物可能有着另外的核内功能,它们和确实在转录水平上被调节的真正的 mRNA 前身物,在结构上可能是有区别的。但是,目前的资料的最简单解释是带着结构基因顺序的 nRNA 分子全部是 mRNA 的潜在前身物。根据上面的结论,我们认为产生复杂类和中等优势类 mRNA 的单拷贝结构基因是以差不多相同的速度连续地转

录的。我们称这个平均速度为每个顺序 nRNA 合成之“基本”速度,它对每类细胞或生物体是特征性的。直截了当的解释是,细胞质 mRNA 群的组成在性质和数量上都是受转录后控制的,这个控制作用决定了从每个活跃基因产生的 mRNA 前身物有多少百分数被加工和输送到细胞质中去。这个观点将编码超优势类 mRNA 基因的控制机制和所有其它结构基因的控制机制区分开来,依据其转录速度是不是明显超过基本速度。和杂交数据相一致,核内转录复合物的电镜观察表明强烈地转录的区域是非常少的,在任何某一时刻,大多数转录单位仅仅只有一个转录物,或是没有。两栖类卵母细胞的灯刷染色体是一个例外,那里的基本速度显示出最大的提高,因为几乎所有的转录单位都被新生的转录物紧紧地包裹着, nRNA 合成的总速度约为一个典型的体细胞核的 100 倍。

nRNA 中的重复顺序转录物

虽然前面只是谈了 nRNA 里的单拷贝顺序转录物,但大多数快速转换的 nRNA 分子也还包括了重复顺序成分。海胆胚胎、HeLa 细胞和大鼠腹水细胞的 nRNA 都显示出和基因组 DNA 非常相象的间隔的顺序结构。用海胆基因组重复顺序的无性繁殖系进行研究,表明在不同类细胞的 nRNA 中,特定重复顺序转录物的浓度可能相差至少一百倍,因此,在原肠胚 nRNA 中,某些重复顺序家族在 nRNA 转录物中有很高的浓度,而其它重复家族仅有低水平的表达。在成年动物小肠 nRNA 中,不同的重复顺序家族都被高度地表达,而且每个重复家族的两个互补顺序都存在 nRNA 中。从这些发现引出一个重要的假设: nRNA 中互补的重复转录物之间在核内生成的二联体可能对基因表达起一种顺序专一的调节作用。

[中国科学院上海细胞生物学研究所
徐永华编译根据 E. H. Davidson 和 R. J.
Britten, Science, v. 204, № 4397, 1052—
1059, 1979, 等文]