



基因表达的调节：重复顺序可能的作用

E. H. Davidson 和 R. J. Britten

动物基因组的协调的控制系统是编码在重复顺序关系的网络之中的设想，至今已有十年之久。我们和其他一些工作者提出过这样一个概念：在基因组中，基因可能受重复顺序上发生的专一相互作用的调节。在本文中，我们设想了一个以转录和转录后所发生的控制过程为基础的动物细胞基因调控模式系统，并提出核 RNA (nRNA) 包含着基因组结构基因区连续合成的 RNA 拷贝，而且在这些“拷贝”和互补的重复顺序转录物之间所发生的相互调节作用是借助于 RNA-RNA 二联体的形成。

本文涉及对细胞质中复杂类和中等优势类 mRNA 可能进行定性和定量调节所采取的方式。和产生中等优势类和复杂类信息的结构基因相比，产生超优势类信息的结构基因是如此之少，所以很自然地会设想由特异的、直接的板机来控制每一个该种基因的转录起始速度。这种板机可能包括专一的激素应答系统。但是在每类细胞中，成千上万个基因的调节可能需要更加多种多样的涉及顺序专一性相互作用的控制系统。

一些研究者曾经提出，nRNA 双链区可能在 mRNA 前身物的加工中起一些功能性作用，然而大多数设想都是求助于链内的二联体结构，对 nRNA 分子间的二联体却很少注意。Boncinelli 提出，将结构基因区域分隔开的分子间 nRNA 二联体会被核酸内切酶切除，随后两边的结构基因区相连接，生成成熟的信息分子。Federoff 等人在电镜中看到 HeLa 细胞的分子间 nRNA 二联体，并表明这种二联体是由重复顺序转录物形成的。但是他们谨慎地指出，分子间 nRNA 二联体可能只是体外情况下存在于纯化的 nRNA 制品中，至于这种结构是否也在动物细胞核的环境中形成，则还不知道。在下面的讨论中，我们假定是这样的。

调节模型的要点

在这一节中我们对动物细胞基因表达的调节提出一个融会贯通的解释，这是一个和当前的知识相协调的调节系统模型。这个模型假定几乎所有的结构基因是位于基因组中，以每类细胞特定的基本速度持续转

录的一定区域之内的。基因组的这些区域被称作构造转录单位 (Constitutive Transcription Units) (CTU)。我们设想，各个 CTU 转录为 RNA 就产生一个构造转录物 (Constitutive Transcript) (CT)，像图 1 所示，它包含一个结构基因编码区 (插入顺序和领头顺序即使有，也是很少的) 和短而分散的控制顺序。CT 可能还包括非编码的单拷贝 DNA 区的转录物，而不是插入顺序。因为 CT 是存在于所有 nRNA 中的，我们设想 CT 将构成某个生物体不同 nRNA 之间共有的大多数单拷贝顺序群。海胆中这种复杂性约为 1.6×10^8 核苷酸 (约为总 nRNA 复杂性的 80%)，而在啮齿类中则至少是 2×10^8 核苷酸 (约为脑 nRNA 复杂性的 20—30%)，就象上面讨论过的，产生出超优势 mRNA 的基因似乎具有控制转录速度的机制，其转录物可能并不象复杂类和中等优势类信息那样是构造性的。但是超优势 mRNA 前身物所受到的加工反应可能和其他类信息受到的相同。

我们推想基因组中有一些区域是以细胞特有的型式转录的，而且不包含结构基因的编码区。这些区域称作集成调节转录单位 (Integrating Regulatory Transcription Units) (IRTU) (图 1)。IRTU 转录产生的 RNA 起控制结构基因表达的功能。IRTU 由散布着的重复顺序和单拷贝顺序或成群的重复顺序所组成。来自 IRTU 的 RNA 称作集成调节转录物 (Integrating Regulatory Transcript) (IRT)。这些区域被称作“集成”，是因为在一个 IRTU 中的一群不同的重复顺序可能共同参与对许多不同结构基因表达的控制。各个 IRTU 中每套重复顺序群的联合转录是调节系统协调并综合成一个整体的重要部分。在调节系统中，这些 IRTU 和我们以前模型中推测的集成基因 (integrator gene) 具有同样的逻辑上必然的功能。因此 IRTU 的转录受对内部或外界信号分子起反应的感应器件 (Sensor elements) 控制。这些 IRT 应组成细胞专一的 nRNA 顺序群。一小部分 IRTU 可能也是到处都有表达，因此它们的转录物是包含在共有的 nRNA 顺序核心之内。我们设想，IRT 和 CT 群合在一起构成了除了超优势信息前身物以外的核内总的非均一 RNA。

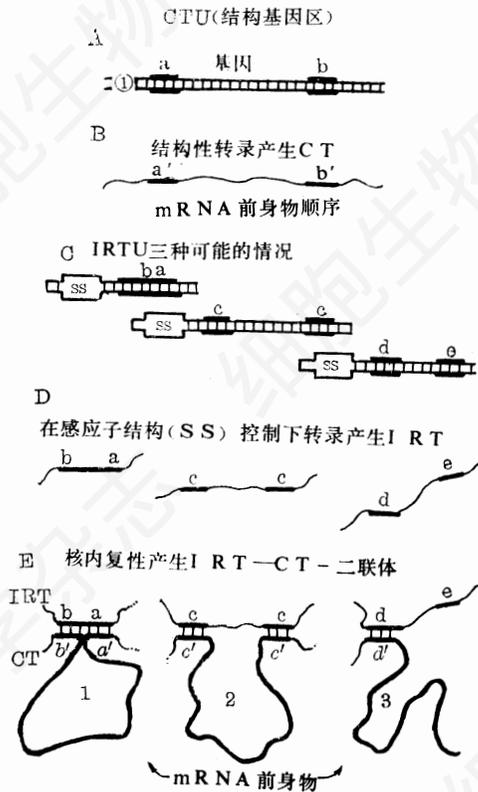


图 1 调节模型的要点

本图指出模型所提示的 DNA 区和 nRNA 转录物的顺序结构。小号字表示短的重复顺序，其它所有区域是单拷贝的。(A) 基因组的包括有一个 CTU 的一段，此 CTU 是在所有种类细胞中都被转录的，并包含了一个结构基因，或许极少的插入顺序以及旁侧片段。“I”代表转录起始位置。(B) 从(A)区段所产生的一个 nRNA 转录物即 CT。(C) 集成调节转录单位 (IRTU) 的三种可能形式，其区别在于散布的重复顺序的排列。推测基因表达的调节是以这些区段转录的控制为基础的。“SS”代表核蛋白“感应器”，它控制着对外界信号应答时 IRTU 的转录。(D) 从(C)中区段所合成的 IRT。(E) 在 IRT 和 CT 重复成分之间，依赖顺序碱基配对所产生的分子间 nRNA 二联体的三种形式。这些例子利用了(D)所表示的三种 IRT 顺序结构。二联体的形成被认为是 mRNA 进一步加工所必需的。

基因表达的调节

在这个模型中，CT 的重复顺序区和相应的 IRT 互补顺序之间所形成的 RNA-RNA 二联体调节基因的表达。我们设想这些二联体是由核内的复性所造成(图 1)，它们是细胞特异 mRNA 群的存在和加工所需要的，从而使信息成功地输送到聚核糖体成为可

能。这种 RNA 二联体可能或不一定能从它们已启动的一系列加工反应中去除掉。

IRT 上特定的重复顺序转录物在核内的浓度将决定和 CT 互补顺序形成二联体的速度，后者的形成会和 IRT 及 CT 的降解发生竞争。我们推想降解是以 nRNA 转换动力学所表明的速度随机地始动的。IRT 上某种重复顺序家族(family)的浓度主要取决于在转录的 IRTU 中该家族成员(family members)的数目。某些 IRTU 始动转录的频率也可能是变化的。

有理由这样设想，当比较大的重复顺序家族被最大限度地利用时，就会产生出较高绝对浓度的重复转录物。我们想象，作为中等优势类信息前身物的 CT 包含着属于大家族的重复顺序。在进行广泛的转录时，这些重复顺序家族在 nRNA 内产生很高浓度的转录物，在 CT 上的大多数互补重复顺序会很快地形成二联体，结果携带在这些 CT 上的几乎所有结构基因转录物都被加工成 mRNA。对比之下带有复杂类 mRNA 顺序的 CT 上的重复成分，可能属于小的重复顺序家族，从这种家族所产生的重复转录物在核内的最高浓度可能相对地比较低一些。这样的低浓度重复转录物也可能起因于较大家族的利用低于最高限度。不过无论是那一种情况，这类 CT 大部分会在二联体生成之前被降解，因此仅其中一小部分潜在的前身物群被加工，并到达聚核糖体。无论是大的还是小的重复家族，当只有很少成员被转录时，IRT 重复转录物的绝对浓度就低到不能促进足够量 IRT-CT 二联体的形成，结果与这些重复顺序相邻的结构基因的表达就不会达到可以被检测的程度。在这个模型中，每个核里全套 CT 上的重复顺序本身是不能促进 mRNA 加工的，只有 IRT-CT 相互作用才导致这一结果。其含意是 IRT-CT 二联体具有特别的性质，这些性质可能取决于包含在二联体内的专一顺序对或由二联体形成所引起的核糖核蛋白体结构之改变。

系统的特征

这个模式系统的一个重要特点是基因表达的主要控制取决于基因组水平上 IRTU 转录的调节。某些 IRTU 上的始动转录信号分子可能出现在其他细胞或远离的组织中，也可能出现在特定细胞的胞质中甚至核内。这里面存在着很多反馈关系。某些关系反映了不同组织之间相互作用的复杂性；另一些可能起着把细胞“锁”在它的分化功能途径中的作用。控制着 IRTU 功能的感应器结构也具备和以前模型中一样的

约束机制。按现在的说法, DNA 中的感应器顺序可能为在和一个 IRTU 相邻位置上产生一个核蛋白感应器结构提供识别信息。我们设想这种感应器结构本身是较早时期基因组其它部分中受发育控制的活动的产物, 而且它的存在决定了细胞在以后时期对特定信号分子反应的能力。因此, 约束作用就是由于是否存在一套被细胞系以前历史所决定的适当的感应器结构。

以各种不同重叠方式一起行使功能的结构基因群的存在, 看来是任何基因调节系统必不可少的部分。大多数结构基因在基因组中以单拷贝存在, 它们在特定类型细胞中的表达是和其它特定结构基因群一起发生的。本模型有三个特点, 它们提供了一套丰富多彩的可能组合。某个重复顺序家族的成员可能存在于许多 CTU 结构基因顺序的旁边, 因此当同一个重复家族成员的转录物在很多个 IRT 上出现并达到足够浓度时, 就决定了这整套基因即一个基因串联组的表达。任何一个 CTU 中和结构基因区相连的重复顺序可能不止一个, 这样的话任何一个基因(也许是大多数基因)可能是几个串联组的成员。一个解释是, 从某些 CT 生成的 mRNA 前身物结构可能是不同的, 这取决于形成二联体涉及了哪个重复成分(或哪些成分)。另外, 一个 IRTU 可能包含了许多控制着很多不同串联组的不同的重复顺序。含有同一重复家族(或几个家族)成员的几个 IRTU 可能被激活对一个共同的外来信号作出反应, 而不同感应器也可能对同一个信号分子起应答。这种组合特征的提出是以基因活动的复杂情况为依据的, 同时还有进化过程中可塑性的含意。

RNA-RNA 二联体控制加工的可能机制

根据这个设想, RNA 二联体能保护 CT 免受降解性核酸酶的作用, 或可作为一个特异的加工性核酸内切酶的作用位点, 也就是在重复顺序区域内可能有着对特定核酸酶敏感的顺序。最近对海胆基因组的几个无性繁殖重复顺序一级结构研究也暗示存在着这样一个位点。在海胆重复顺序中, 一个共同有的短顺序是和存在于单拷贝区内插入部分和编码部分之间连接处的交感顺序相同的。在哺乳类和鸟类系统中, 这种短顺序可能被一种 RNA 内切酶所识别, 它在海胆重复 DNA 中的存在提示了可能也有某些相同的功能。

核酸内切酶控制 CT 降解的作用可能通过另一种途径。我们在图 1E 中表明了 RNA-RNA 二联体和 CT 中潜在的 mRNA 顺序之间几种可能的关系。图 1E 例 1 显示一个环状结构, 那里一个 IRT 重复

顺序会在基因区的末端于同样的重复顺序之间形成一个“桥”。这个结构将保护基因顺序不被降解, 而当被切割时, 就产生出前身物的末端以供接上帽子和聚腺苷化, 接着就发生随后的加工步骤。图 1E 例 2 表明了一个不同的情况, 其中同一个重复家族的两个被隔开的成员完成一个相仿的功能。我们注意到, 如果重复成分“C”是在相反的方位上, 那么 IRT 和 CT 就都能够形成已在 nRNA 中所观察到的回摺圈 (fold-back-loop) 结构。例 1 和例 2 中结构的另一种可能的改变是形成紧密相间的串联或成簇的 IRT 重复物。从推理上讲, 如图 1E 例 3 所示, 仅仅在基因转录物的一个末端上有一个二联体区就可能足够了。

当 IRT 重复物的一个家族在核内的浓度十分低时, 互补的 CT 重复物之浓度还是能被观察得到, 因此 CT 重复物可能说明了在 nRNA 中看到的最低重复转录物水平。如上面所提到的, 我们推测即使形成 CT-CT 二联体, 它们的结构在某些情况下对进一步加工也是不合适的; 或者因为链内 CT 结构可能被足够浓度的互补 IRT 所竞争或取代, 这种二联体可能不会形成。

RNA-RNA 二联体决定分子间调节作用的设想, 提示它们可能也涉及分子内的其它加工阶段, 例如插入顺序的去除。在同一转录物的隔开相当距离的几个区段内或在另一些 RNA 分子内可能找到相同的顺序成分。这里我们形象地设想一个和图 1E 中例 1 十分相象的结构, 不过现在这个环包含着插入顺序, 而不是图中的 mRNA 前身物。Berg 设想了一个与此十分相似的结构, 它是在同一个 SV40 转录物的不同部位中生成的。当然, 一个分子内的而不是分子间的链配对复性并不要求除了基因本身以外的顺序同一性。但是从本文所提出的论证来看, 如果插入顺序加工受到分子间链配对复性的影响, 这些相互作用可能起的调节作用是应该考虑到的。虽然许多插入顺序似乎是单拷贝, 参与想象的 RNA-RNA 二联体形成的成分可能免于证明为重复物, 因为它们的长度很短, 存在的频率非常低。

同当前认识的一致性和预见

我们在这里记录一些和这个模型一致的最新实验结果, 并提出几点预见, 以提供一个直接的机会让实验来证明其正确与否。在这一模型里被统一起来的实验事实之简目如下:

(1) 大多数生物体中, DNA 序列结构显示重复

顺序和单拷贝顺序的间隔排列。

(2) 基因组的重复顺序中, 某些家族是很大的, 而另一些是很小的。

(3) nRNA 分子很快地转换。

(4) 在研究过的所有组织中, nRNA 显示共同具有一个很大的单拷贝顺序转录物核心, 并有高度复杂性。

(5) 在不同类细胞中, nRNA 单拷贝顺序群仍然有着很大的差异。

(6) 各类 nRNA 分子都特征性地含有重复顺序转录物。

(7) nRNA 重复转录物中特定家族顺序的浓度随细胞类型而不同。

(8) 大多数重复家族以某个水平出现在 nRNA 中。

(9) 每个重复顺序的两个互补体都存在于 nRNA 中。

(10) 复杂类结构基因在核内产生的转录物只有一部分被输送到细胞质中去。

(11) 在某些种类细胞中存在的结构基因转录物也全部出现在其它类细胞的 nRNA 中, 但后者并不在聚核糖体水平上表达所有这些基因。

这个模型和我们以前的模型都预见, 与某类细胞中表达的结构基因相邻的重复顺序群是全部重复顺序家族的一个亚群。虽然这种预见可能是正确的, 但由于技术上的困难, 企图予以验证, 还未获得肯定的结果。另一个与此密切相关的预见是, 和结构基因相邻的顺序组成反映了协同表达的基因串联组的概念。换句话说, 和功能上相关的几个结构基因群相邻的重复顺序常常属于同一家族。在小鼠 α 和 β 珠蛋白基因的例子中, 已经观察到每个基因旁边约 1.5Kb 的 3' 片段上有着 150—200 个同样的核苷酸顺序。这个模型提示这种同一性可能存在于功能上相关的几个基因的 5' 端、3' 端, 也可能两个末端都有。

另一个预见是, 哪里需要稳定高浓度的某些 IRT

重复顺序, 则该家族成员的极大部分将被转录。在某些种类的细胞中, 较大的重复顺序家族象是和表达为中等优势类 mRNA 的基因相连的。

这个模型假设 IRTU 包括了大部分的基因组, 因此该基因组的顺序结构将充分地反映 IRTU 的功能结构及其进化历史。基因组的很多单拷贝顺序被发现可能就是 IRTU 中将重复转录物隔开的成分。我们原先就讨论过, 很多的 nRNA 单拷贝顺序具有间隔顺序的特征。

如果图 1E 中例 1 的“搭桥”设想是正确的话, 那么在基因组的许多地方, 特定的重复顺序对将是相互很靠近的。象 (a、b) 这样一对顺序在 CTU 中常常被分隔在基因区的两端 (图 1B), 但是在 IRTU 别的位置上却是在一起的 (图 1, C 和 D)。

本模型预见核内 RNA 二联体是存在的, 并且以一定浓度出现在这些二联体中的重复顺序群会是细胞或组织专一的。此外还预见, 在某类细胞中与编码中等优势类 mRNA 的结构基因相邻的重复顺序, 就是以高浓度出现在这类细胞 nRNA 中的顺序。相比之下, 和复杂类 mRNA 之结构基因相邻的重复顺序则是以低浓度出现在 nRNA 之中。

和体细胞核一样, 在成熟的海胆卵细胞质中重复顺序转录物 (包括每个转录物的两个互补成分) 呈现特异的分布。卵细胞质的某些重复顺序转录物可能被隔离在早期胚胎细胞的核内, 这一假设和我们原先所作的一个假设相似。因此, 母体 RNA 中重复顺序转录物的这种特异的分布可能确定了胚胎核内相应的调节方案。这也能说明早期胚胎和卵母细胞结构基因顺序群之间的很大相似性。另外, 在早期卵裂时, 卵细胞质重复顺序转录物可能被局限在某个部位, 因此导致基因表达的早期特定模式。

[中国科学院上海细胞生物学研究所徐永华摘译自 Science, V. 204, No 4397, 1052—1059, 1979]