

真核转录的结构生物学 ——2006年诺贝尔化学奖简介

胡永林*

(中国科学院生物物理研究所, 北京 100101)

摘要 用 X 射线晶体学方法测定的一系列 RNA 聚合酶 II 复合物结构揭示了真核转录的分子机制。

关键词 真核转录, RNA 聚合酶 II, 晶体结构

学科分类号 Q7

瑞典皇家科学院宣布, 将 2006 年诺贝尔化学奖授予美国斯坦福大学医学院结构生物学教授罗杰·科恩伯格(Roger D. Kornberg), 以表彰他对真核转录分子基础研究所作的杰出贡献。

科恩伯格和他的研究组在真核转录这个领域进行了长时间的研究工作. 他们在 2001 年 6 月 8 日出版的《科学》杂志(Science)上发表的两篇论文, 标志着这方面研究的重大突破. 这两篇论文分别描述了 0.28 nm 分辨率的 RNA 聚合酶 II 晶体结构^[1]和 0.33 nm 分辨率的 RNA 聚合酶 II 与 DNA 和 mRNA 复合物^[2]的晶体结构. 从 2001 年至今, 科恩伯格研究组已经发表了十余篇关于 RNA 聚合酶 II 分子结构的研究论文, 从分子水平上阐述了转录过程中的各个步骤, 如聚合酶在 DNA 分子上的移位(translocation), 选择正确的核苷酸延伸 RNA 链, 以及新合成的 mRNA 与 DNA 模板链的分离(strand separation), 等等. 这些晶体结构在原子分辨率水平上向我们展示了在真核细胞中从双链 DNA 到 mRNA 的转录是如何进行的。

1 原核细胞转录的研究

由于原核转录的机制相对简单, 对其研究开展得比较早. 1965 年, 三位科学家, François Jacob, André Lwoff 和 Jacques Monod 因为对细菌基因表达转录调控的研究而共享诺贝尔生理学或医学奖. 至今, 已经有数个细菌来源的 RNA 聚合酶的晶体结构被解析. 结构和生物化学的研究显示, 细菌的 RNA 聚合酶全酶(holoenzyme)由 5 个亚基($\alpha_2\beta\beta'\sigma$)

组成, 其中 σ 亚基被称为起始因子(initiation factor), 不包含这个亚基的 $\alpha_2\beta\beta'\sigma$ 复合物被称为核心酶(core enzyme). 核心酶可以和一般 DNA 链结合, 不能识别特定序列; σ 亚基结合到核心酶后极大降低了全酶与一般 DNA 序列的结合能力, 同时 σ 亚基与 -35 区和 -10 区的启动子序列作用, 并起始 RNA 合成; 合成起始后, σ 亚基脱离, 核心酶延 DNA 链移动来延伸新合成的 RNA 链; 当核心酶遇到终止子时, 在终止因子的作用下会从 DNA 链脱落, 转录终止。

2 真核细胞转录机制的研究

2.1 早期的研究

真核转录相对于原核转录, 其分子机制要复杂得多. 一方面是因为在真核细胞中, 染色体 DNA 首先是与组蛋白结合形成核小体, 多个核小体又紧密排列形成染色质. 另一方面, 真核细胞的 RNA 聚合酶比细菌的 RNA 聚合酶要复杂得多: 它含有更多的亚基, 也提供更精密的转录调控。

1959 年, Weiss 和 Gladstone 首次报道了在小鼠肝细胞核中有 RNA 聚合酶活性. 随后, 其他科学家的实验显示, 在真核细胞中主要存在 3 种不同的 RNA 聚合酶, 通过 α -鹅膏蕈碱的抑制作用可将此 3 类 RNA 聚合酶区别开: RNA 聚合酶 I 对 α -

* 通讯联系人. Tel: 010-64888546, E-mail: yonglin@ibp.ac.cn

收稿日期: 2006-10-13, 接受日期: 2006-10-13

鹅膏蕈碱不敏感, RNA 聚合酶 II 可被低浓度 α -鹅膏蕈碱($10^{-9} \sim 10^{-8}$ mol/L)抑制, 而 RNA 聚合酶 III 只被高浓度 α -鹅膏蕈碱($10^{-5} \sim 10^{-4}$ mol/L)抑制. 其中, RNA 聚合酶 I 转录 45 S rRNA 前体, RNA 聚合酶 II 转录所有 mRNA 前体和大多数核内小 RNA (snRNA), RNA 聚合酶 III 转录 tRNA、5S rRNA、U6 snRNA 等小分子转录物. 对应于 3 种 RNA 酶, 真核生物的启动子也被分为 3 类(Class I, II, III): 类别 I 启动子只控制 rRNA 前提基因的转录; 类别 II 启动子涉及到众多蛋白质基因的表达调控; 类别 III 启动子涉及到一些小分子 RNA 的转录. 其中, 类别 II 启动子含有 4 类控制元件: 基本启动子(basal promoter)、起始子(initiator)、上游元件(upstream element)和应答元件(response element). 基本启动子序列为中心在 -25 至 -30 左右的 7 bp 保守区, 又被称为 TATA 框(TATA box)或者 Goldberg-Hogness 框. 研究证明, TATA box 是 RNA 聚合酶 II 和下面提到的通用转录因子形成前起始复合物的主要装配点.

众所周知, mRNA 转录直接影响了后续的由 mRNA 到蛋白质的翻译过程, 进一步影响蛋白质功能的行使, 因此, RNA 聚合酶 II 在真核生物体内的作用尤为突出.

对 RNA 聚合酶 II 的研究表明, 这个复合物分子包含 12 个亚基, 但并不存在一个亚基有对应于细菌 RNA 聚合酶中 σ 亚基的功能. 而且, 纯化的 RNA 聚合酶 II 并不能对 DNA 进行转录. 进一步的实验又发现了 5 个蛋白质因子, 只有将它们加入到提纯的 RNA 聚合酶 II 中, 后者才具有转录的功能. 因为这 5 个蛋白质因子被所有的转录活动所需要, 它们的存在是维持基础水平的转录(basal transcription)的充要条件, 它们因而被通称为通用转录因子(general transcription factors, 以 TF II X 表示, X 按发现先后次序用英文字母命名), 分别是 TF II B, D, E, F, H. 而其中 TF II D 是一个寡聚蛋白, 含有可以识别上面提到的 TATA box 的 TATA 结合蛋白(TATA-binding protein, TBP).

2.2 科恩伯格的研究工作

实验表明, 从小鼠肝细胞或其他哺乳动物细胞中纯化 RNA 聚合酶 II 非常困难. 科恩伯格从开始就集中在酿酒酵母(*Saccharomyces cerevisiae*) RNA 聚合酶 II 上. 他的研究显示, 除了上面提到的 5 个通用转录因子之外, RNA 聚合酶 II 还需要另一个蛋白质复合物. 这个蛋白质复合物的功能是向 RNA

聚合酶 II 和通用转录因子传递各种调控信号. 这个复合物由近 20 个蛋白质组成^[3]. 因为它的功能, 它被命名为 Mediator (中介者).

通过科恩伯格和其他科学家的生物化学和分子生物学研究, 到 20 世纪末, 我们对真核转录的机制, 特别是这个过程所涉及的蛋白质复合物以及它们之间的相互作用有了很广泛的了解(图 1). 而最近几年一系列出自科恩伯格研究组的关于 RNA 聚合酶 II 晶体结构的论文, 让我们从一个完全不同的角度去了解真核转录这个极其复杂的过程是如何一步一步在分子水平上实现的, 也给科恩伯格带来诺贝尔奖的殊荣.

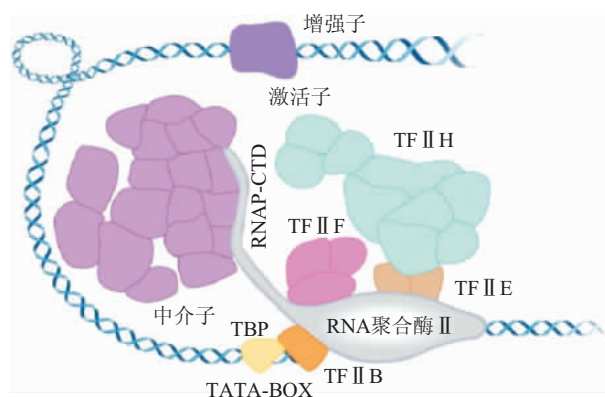


Fig. 1 A cartoon of the elements involved in eukaryotic transcription, including RNA polymerase II, mediator, the general transcription factors TF II B, D, E, F, and H, and DNA

图 1 参与真核转录的各个要素的示意图

其中包括: RNA 聚合酶 II, Mediator, 通用转录因子 TF II B, D, E, F, H, 以及被转录的 DNA(copyright: www.nobelprize.org).

2.3 RNA 聚合酶 II 的分子结构

科恩伯格组的第一篇关于 RNA 聚合酶 II 结构的论文发表于 1991 年. 这是一个 1.6 nm 分辨率的二维晶体电子显微镜结构^[4]. 虽然不能观测到分子的结构细节, 但从电子密度图上可以确定单链和双链核酸的结合位点.

RNA 聚合酶 II 的 0.28 nm 分辨率晶体结构^[5]包含 10 个亚基, 分子质量超过 500 ku. 在这个结构中, 结合双链 DNA 的钳状结构处于张开构象, 以利于 DNA 启动子进入而开始转录过程. 而钳状结构附近的 3 个柔性区可能负责转录之后 RNA 从 DNA 上解离和 DNA 双链的重新生成. 从结构中也确定了催化转录反应的活性区以及重要的残基. 与

这篇同时发表的另一篇论文, 是 0.33 nm 分辨率的 RNA 聚合酶 II 与 DNA 和 mRNA 的复合物结构^[2]. 在这个结构中, 我们如同看到了双链 DNA 进入 RNA 聚合酶 II, 在离催化活性区 3 个碱基对的地方, 双链被拆开, 转录在模板链上进行(图 2). 一段含有 9 个核苷酸的 mRNA 已经被合成并与 DNA 模板链形成双链, 其 3' 端正处于 RNA 聚合酶 II 催化核苷酸聚合的位点, 而附近的一个通道可让核苷酸分子进入. 活性位点附近的一个天冬酰胺残基确保核苷酸而不是脱氧核苷酸被聚合到 mRNA 分子上. 当脱氧核苷酸或者错误的核苷酸被引入时, 结构就会变得不稳定, 迫使 mRNA 分子后退而将错误的碱基送回酶的活性位点, 继而被切除.

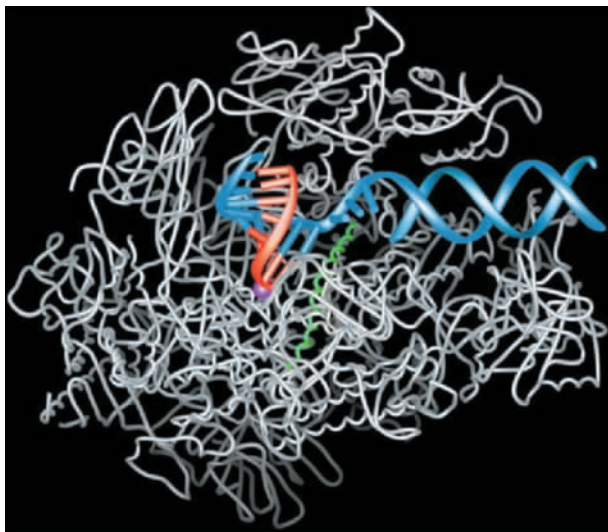


Fig. 2 Structure of an RNA polymerase II transcribing complex

图 2 正在转录的 RNA 聚合酶 II

连接 Rbp1 和 Rbp2 的 α 螺旋显示为绿色, 活性中心的一个 Mg 离子为洋红色, DNA 为蓝色, 合成的 mRNA 为红色. DNA 双链在活性中心之前大约 3 个残基处被解旋, 其模板链在结构中可见, 而正义链在结构中无序而不能被观察到 (copyright: www.nobelprize.org).

如果说上述两篇论文确立了 RNA 聚合酶 II 最重要的结构特征, 以后发表的一系列论文就更详细地描述了转录过程中的一些步骤.

移位: RNA 聚合酶 II 结构中最大的两个亚基, Rbp1 和 Rbp2, 位于复合物的中心, 各从 DNA 分子的一面与其结合. 一个来自于 Rbp1 的 α 螺旋横穿这 2 个亚基并且穿过聚合酶的催化活性部位, 通过 2 个残基侧链与双链解开后被暴露的 DNA 模板链上的碱基生成氢键作用. 这个 α 螺旋可以有直或

者弯两种构型, 它每发生一轮构型的变化, 就会将核酸分子移动 0.3 nm——一个碱基对的长度^[1,5].

核苷酸的选择: 在 RNA 聚合酶 II 的活性中心附近, 有一个漏斗状的通道直通分子外部环境. 核苷酸分子就是从这个通道扩散进入聚合酶的活性中心. 当将 RNA 聚合酶 II 晶体在含有核苷酸的溶液中浸泡后再测结构时, 意外地发现, 如果 RNA 聚合酶 II 中含有 DNA-RNA 杂交分子而且核苷酸分子与 DNA 模板链的 +1 位的碱基相配, 核苷酸分子就会以正确的方向进入, 其碱基与模板链的 +1 位碱基形成氢键, 其 α 磷酸根靠近 mRNA 的 3'-OH 基团以利于与这个基团发生亲核取代反应而延伸 mRNA 链. 当核苷酸分子与 DNA 模板链的碱基不相配, 或者 RNA 聚合酶 II 中不含 DNA-RNA 杂交分子时, 晶体结构中的核苷酸将取相反的方向: 其碱基指向聚合酶分子表面. 而且核苷酸分子在通道中的位置也更靠外. 此时, 核苷酸不可能与 mRNA 的 3'-OH 基团发生反应^[6]. 基于这些晶体结构, Westover 等提出了一个核苷酸选择的“旋转”机制: 核苷酸分子以碱基朝外的方式进入通道, 在与聚合酶分子产生非共价结合后, 核苷酸分子在通道里旋转. 如果它的碱基与 DNA 模板链 +1 位的碱基相配, 它与 mRNA 分子间的磷酸二酯键就会形成. 如果不相配, 由于复合物中的 DNA-RNA 杂交分子本来已经欠旋 (underwound), 引入不相配的核苷酸会使复合物结构更不稳定, 从而阻止这个核苷酸分子与 mRNA 的聚合.

链分离: 合成的 mRNA 分子必须从 DNA 模板链上分离. 这个功能, 是由 Rbp1 和 Rbp2 两个亚基中的 3 个柔性区执行^[7]. 首先, 来自 Rbp2 的残基 Lys471 和 Arg476 与 RNA 的 -5, -6, -7 位磷酸根形成氢键, 它们的功能可能是保证 DNA 和 RNA 双链不会过早 (在 -8 位之前) 分开. 其次, 来自于 Rbp1 的残基 Ser318 和 Arg320 与 -9 和 -10 位 DNA 的核糖和磷酸根作用, 从而降低 DNA-RNA 之间的作用. 最后, 来自于 Rbp1 的 Phe252 及其附近的几个残基完全插入到 DNA 和 RNA 中间, 迫使两条单链从不同的出口离去.

因为篇幅所限, 很多结果不能在这里介绍. 同时, 我们也可以肯定, 更多重要的结构将会陆续发表, 例如 RNA 聚合酶 II 与 Mediator 和通用转录因子的复合物结构. 这些结构将帮助我们认识真核转录的另一个重要方面, 即这个过程是如何被调控的.

3 话题外的花絮

罗杰·科恩伯格的获奖, 也是历史上第六次父子同获诺贝尔奖。罗杰·科恩伯格的父亲, 阿瑟·科恩伯格(Arthur Kornberg), 因为对 RNA 和 DNA 的生物合成机制的研究有重大贡献而与 Severo Ochoa 共享 1959 年的诺贝尔生理学或医学奖, 他得奖时也是斯坦福大学医学院的教授。老科恩伯格仍然健在。

除了对 RNA 聚合酶 II 的结构有深入的研究, 罗杰·科恩伯格也对另一重要的生物学研究领域作出了卓越的贡献: 1974 年, 他以“Chromatin structure: a repeating unit of histones and DNA”为题在《科学》杂志(Science)上发表论文, 首次提出核小体的结构是以组蛋白八聚体为其核心而外面缠绕长约 200 个碱基对的 DNA 链。这个结构后来被实验完全证实。科恩伯格的这些研究工作, 是他在英国剑桥大学 MRC 分子生物学实验室作博士后时进行的。同一时间在这个实验室作博士后研究的另一位科学家, 安德鲁·菲尔(Andrew Z. Fire), 也正巧

是 2006 年的诺贝尔生理学或医学奖获得者之一。

参 考 文 献

- 1 Cramer P, Bushnell D A, Kornberg R D. Structural basis of transcription: RNA Polymerase II at 2.8 Angstrom Resolution. *Science*, 2001, **292**: 1863~1876
- 2 Gnatt A L, Cramer P, Fu J, *et al.* Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution. *Science*, 2001, **292**: 1876~1882
- 3 Myers L C, Kornberg R D. Mediator of transcriptional regulation. *Annu Rev Biochem*, 1990, **69**: 729~749
- 4 Darst S A, Edwards A M, Kubalek E W, *et al.* Three-dimensional structure of yeast RNA polymerase II at 16 Å resolution. *Cell*, 1991, **66**: 121~128
- 5 Bushnell D A, Cramer P, Kornberg R D. Structural basis of transcription: α -Amanitin-RNA polymerase II cocrystal at 2.8 Å resolution. *Proc Nat Acad Sci USA*, 2002, **99**: 1218~1222
- 6 Westover K D, Bushnell D A, Kornberg R D. Structural basis of transcription: nucleotide selection by rotation in the RNA polymerase II active center. *Cell*, 2004, **119**: 481~489
- 7 Westover K D, Bushnell D A, Kornberg R D. Structural basis of transcription: separation of RNA from DNA by RNA polymerase II. *Science*, 2004, **303**: 1014~1016

Structural Biology of Eukaryotic Transcription: a Brief Introduction to The Nobel Prize in Chemistry 2006

HU Yong-Lin*

(Institute of Biophysics, The Chinese Academy of Sciences, Beijing 100101, China)

Abstract A brief introduction to the X-ray crystallographic studies on RNA polymerase II complexes and the enzymatic mechanisms revealed by the crystal structures.

Key words crystal structure, RNA polymerase II, enzymatic mechanism

*Corresponding author. Tel: 86-10-64888546, E-mail: yonglin@ibp.ac.cn

Received: October 13, 2006 Accepted: October 13, 2006