

文章编号: 1000-1336(2008)06-0686-05

## 真核生物转录的延伸机理

艾南平 陈瑞川 刘敏

(厦门大学生命科学学院教育部肿瘤细胞工程重点实验室, 厦门 361005)

**摘要:** 由于对真核生物基因转录延伸期研究的忽视, 现有的真核生物基因转录理论存在种种的偏颇。然而, 新近的研究发现却使人们逐渐认识到, 真核基因转录延伸阶段其实是真核细胞调节基因转录的一个高度有序、而又极为复杂的调控平台。本文尝试对近年来真核生物基因转录延伸领域的研究进展, 包括转录延伸与 mRNA 加工的偶联、基因转录延伸调控的分子机理以及转录延伸调控对发育和应激反应产生的影响, 作概括性的综述。

**关键词:** 真核生物; 基因转录延伸期; 转录延伸调控

**中图分类号:** Q291

真核生物中RNA聚合酶II(RNA polymerase II, Pol II)负责转录编码蛋白质的基因、大多数核内小RNA(snRNA)及微RNA(miRNA)前体, 其过程大致可分为五个紧密相连的阶段: 起始前、起始、启动子区清扫、延伸和终止期<sup>[1,2]</sup>。过去二十年来, 研究人员的主要精力集中于揭示起始前复合体(pre-initiation complex, PIC)在启动子区的组装、转录起始及启动子区清扫的精细过程, 并籍此发现了大量转录前期所必须的转录因子和调控元件。相反, Pol II长时间发挥功能的延伸期却很少受到关注。事实上, 长期以来人们一直错误地认为此过程是不受调控的、机械地添加碱基直至全长的mRNA生成; 直到最近几年才逐渐认识到转录延伸阶段其实是细胞用来调节基因转录的一个极为复杂、而又高度有序的调控平台, 其调控机理甚至比转录起始更严格, 并与mRNA加帽、剪辑和加尾等过程相契合。

### 1. 真核基因转录延伸与mRNA成熟加工紧密偶联

由于对基因转录延伸期研究的忽视, 以往人们误以为mRNA合成与加工过程是相互割裂、顺次发生的, 时至今日这种模式仍出现在许多教科书中。

而其实际过程是<sup>[3]</sup>: 在转录启动时 Pol II大亚基Rpb1的C-末端结构域(CTD)第5位丝氨酸(Ser5)被磷酸化, PIC部分解体, 形成早期转录延伸复合体(early transcript elongation complex, eTEC)并进行早期转录延伸。但在新生RNA延伸至20~30 nt时, DSIF(DRB-sensitivity inducing factor)和NELF(negative elongation factor)协同将eTEC阻滞于靠近启动子区的模板上, 以便提供充足的时间为新生的RNA链5'-端加帽。此时, Ser5磷酸化的CTD可结合RNA加帽相关的3个酶, 依次对新生RNA进行加帽反应。当加帽完成后, 停滞的转录延伸需要重新启动, 于是转录延伸调控中最重要的正性转录延伸因子 $\beta$ (positive transcription elongation factor  $\beta$ , P-TEF $\beta$ )被募集上来。它先分别磷酸化DSIF的Spt5亚基和NELF的RD亚基, 执行解抑制的功能; 该磷酸化的发生使NELF从eTEC上解离下来, 并将DSIF逆转为促进转录延伸的因子。同时, P-TEF $\beta$ 进一步磷酸化CTD第2位丝氨酸(Ser2), 刺激Pol II的转录延伸活性, 重新启动转录延伸, 直至全长mRNA产物的形成<sup>[3]</sup>(图1)。

除与加帽过程契合外, RNA的剪辑、3'-末端切除及加polyA尾也均与真核基因转录延伸过程相偶联。一方面, 基因的转录延伸活性会促进新生mRNA的加工和成熟过程; 另一方面, 这类加工信号也可刺激Pol II的转录延伸活性。

严格上来讲, RNA剪辑时靠近5'-端的内含子在

收稿日期: 2008-06-20

作者简介: 艾南平(1984-), 男, 博士生, E-mail: ainanping@163.com; 陈瑞川(1963-), 男, 博士, 副教授, E-mail: chenrc@xmu.edu.cn; 刘敏(1980-), 男, 博士, 博士后, 联系作者, E-mail: jacqueslm2001@hotmail.com

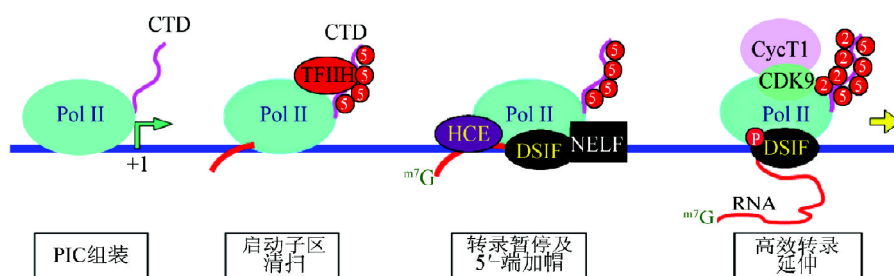


图1 RNA Pol II调控的真核基因转录模式图

转录延伸的过程中就已经被切除了,而靠近3'-末端的内含子的切除则可能发生在全长的mRNA产生以后<sup>[4]</sup>。目前已知的转录延伸和剪辑过程的偶联存在三种模式:其一,启动子或转录增强子可通过结合转录因子再募集剪辑复合体;其二,磷酸化的CTD直接结合具转录和剪辑双重功能的蛋白质,通常这类蛋白质含有WW和/或FF结构域,如酵母的Prp40、Ess1、人源的CA150、SCAF;其三,高度磷酸化的Pol II本身与某些转录或延伸因子以及U snRNPs形成巨大的多亚基复合体,如PSF和TLS,以及P-TEF $\beta$ 和Tat-SF1<sup>[4-6]</sup>。正是由于这些具转录和剪辑双重活性的蛋白质复合体的参与,使得真核基因的转录延伸和剪辑过程相伴而生、同步进行。此外,磷酸化的CTD还能通过直接结合应答3'-末端切除的CPSF、CstF、CF1和CF2以及应答加polyA尾的PAP蛋白<sup>[7]</sup>,或者通过P-TEF $\beta$ 间接偶联<sup>[6]</sup>,将转录延伸与mRNA加工和成熟整合为一体。

## 2. 真核基因转录延伸期的调控机制

完整的真核基因转录调控过程的实质是围绕Pol II活性调控而进行的,并通过Pol II上特殊结构域结合特定蛋白质执行相应的功能而实现的,转录延伸期的调控机制也不例外。

### 2.1 Pol II CTD的结构完整性及其磷酸化循环

酵母的Pol II是由12个亚基组成的复合体,其大亚基Rpb1 C-末端比Pol I、Pol III多一个进化上高度保守的CTD结构域,它由七肽重复序列(Tyr-Ser-Pro-Thr-Ser-Pro-Ser)串连而成。不同物种之间CTD结构域的差别仅限于这七肽重复序列的拷贝数不同,如啤酒酵母的CTD含26个拷贝、线虫为32个、果蝇有42个,而哺乳动物则有52个。Pol II要具备完整的功能,拷贝数有最低限度,例如人类细胞能存活至少需28个拷贝,酵母为9个;而拷贝数在9~20之间的酵母突变体对

低温敏感且多数基因转录有缺陷;仅有39个拷贝的老鼠生长退化或新生儿死亡率增加。由此可见,CTD所含七肽重复序列拷贝数的完整性是至关重要的<sup>[8]</sup>。

在转录过程中,CTD还会经历一个磷酸化和去磷酸化的循环<sup>[8]</sup>。在PIC中的Pol II处于低磷酸化状态,转录一经开始,真核细胞中3个高度保守的激酶CDK7、CDK8和CDK9就开始顺次并特异性地磷酸化CTD。CDK8功能尚不明确,只知是中介复合物组分,作为其它转录活化因子的桥联蛋白,能磷酸化CTD。CDK7(酵母同源物为M015或Kin28)是TFIIH的组分之一,当它与周期蛋白H(cyclin H)、Mat1形成CAK复合体时,可磷酸化并激活CDK1、CDK2、CDK4和CDK6,调控细胞周期;而CDK7与TFIIH核心亚基结合后便只能特异性地磷酸化CTD上七肽重复序列的Ser5<sup>[8]</sup>。Ser5磷酸化后,引发Pol II构象变化,使Pol II从起始复合体中脱出,进入启动子区清扫,转录开始;但在转录20~30 nt后转录暂停并进行新生mRNA加帽。加帽完成后,CDK9(酵母同源物为Bur1或Ctk1)负责磷酸化CTD上七肽重复序列的Ser2,使暂停的转录重新启动并偶联RNA成熟和加工机制。转录结束后,Pol II从DNA模板上解离下来,CTD特异性蛋白质磷酸酶FCP1将已经高度磷酸化的CTD去磷酸化,使Pol II能迅速进入第二轮的转录。除FCP1外,Ssu72和PP1也能部分执行去磷酸化的作用<sup>[6,8]</sup>。因此,CTD除长短被严格控制外,在转录过程中其磷酸化水平及磷酸化位点的调节也是调控真核基因转录的重要一环。

### 2.2 多种转录延伸因子及辅助因子参与调控Pol II催化活性

为实现真核基因转录延伸过程的精确调控,Pol II蛋白需要结合多种转录延伸因子和辅助因子,其特殊的CTD结构域以及CTD磷酸化位点和水平的变化为此提供了一个适宜的蛋白质相互作用的结构

平台,并借此协调了高度动态、严格受控的基因转录延伸过程,从而克服了数个转录延伸的瓶颈,以整合转录延伸及诸多上下游事件。这些结合上来的转录延伸相关因子包括直接调控PoI II酶活性的蛋白质,也有通过结合PoI II蛋白调控TEC依时序性组装的蛋白质。不同蛋白质的结合和解离使得所形成的特定的TEC功能也不尽相同,有的引起基因转录的停滞,有的参与转录的重新启动,有的监控新生RNA的加工修饰和转运(表1)。因此,贯穿整个转录延伸过程的TEC多态性是确保基因转录完整性和保真性的重要因素。

### 3. 真核基因转录延伸期调控与发育和细胞应激反应的新认识

过去的半个世纪人们普遍认为,真核基因转录水平的高低取决于对基因转录起始步骤的调控,即PIC在启动子区的组装被认为是至关重要的限速步骤;尤其当首轮转录完成后,由于DAB三元复合物(TFIID/TFIIA/TFIIB)仍滞留在启动子区,第二轮及随后mRNA的大量转录不再受PIC组装的限制。然而最近这种理论却面临了严峻的挑战<sup>[9-11]</sup>,它同样来自于对真核基因转录延伸期的进一步研究。

2007年3个实验室采用不同的实验模型分别证实,正常的发育过程和应激反应时,细胞全基因组中会出现3类基因。第一类基因,启动子区没有组装PoI II,基因处于失活状态,未进行任何转录。第二类基因,PoI II组装、位置分布正常,基因正常表达。第三类基因,尽管PoI II组装正常,染色质空间结构已经舒展,组蛋白乙酰化标示转录业已启动,但PoI II却停在了靠近启动子区的位置,基因仅具潜在的活性、但不进行转录;这类基因包括同源异型基因以及应答发育和环境信号的基因(如Wnt、Notch和TGF- $\beta$ ),总数约占全基因的10%,因此这绝不可能是一个偶然的现象<sup>[9-11]</sup>。

显然,在发育或者应激条件下,真核生物细胞为快速应对外界刺激,在转录起始调控的基础上又

增添了一道天然保障,即通过调控基因转录延伸快速重启来影响基因的表达水平;这种调控模式与上述提到的PoI II CTD的磷酸化状态调控转录延伸重启的模式相吻合,我们新近报道的P-TEF $\beta$ 活性调控的分子模式也与上述发现吻合<sup>[12]</sup>。我们的研究显示在胞外信号刺激条件下Ca<sup>2+</sup>信号和PP1 $\alpha$ 信号途径被分别激活,PP2B协同PP1 $\alpha$ 去磷酸化CDK9(P-TEF $\beta$ 的催化亚基)第186位苏氨酸(T186),使P-TEF $\beta$ 从抑制态的大复合体中释放出来,但由于T186处于去磷酸化状态,使得被组装到启动子区的P-TEF $\beta$ 仍处于无活性的状态,此时它无法激活被阻滞在靠近启动子区的PoI II<sup>[12]</sup>。细胞等待进一步的信号以确定是否开启相应基因的转录,这也许恰当地解释了为什么上述与组织发育或细胞应激相关的第三类基因处于转录暂停状态的原因。因此,P-TEF $\beta$ 在基因启动子区的组装可能还牵涉到P-TEF $\beta$ 与基因特异转录因子之间的相互作用<sup>[6]</sup>。

随着真核基因转录延伸领域研究的深入,人们已经无法预期它将会从多大程度上改变现有的真核基因转录调控的基本理论。总之,这是一个重要的契机,为全面揭示正常的生理发育活动以及相关的病理条件下,基因转录水平的调控提供了一个新的思路。

### 参考文献

- [1] Leclerc V et al. Mol Biol Cell, 1996, 7(4): 505-513
- [2] Sims RJ 3rd et al. Genes Dev, 2004, 18(20): 2437-2468
- [3] Peterlin BM et al. Mol Cell, 2006, 23(3): 297-305
- [4] Kornblihtt AR et al. RNA, 2004, 10(10): 1489-1498
- [5] Fong YW et al. Nature, 2001, 414(6866): 929-933
- [6] Zhou Q et al. Microbiol Mol Biol Rev, 2006, 70(3): 646-659
- [7] Rosonina E et al. Genes Dev, 2006, 20(9): 1050-1056
- [8] Chen R et al. In: Ma J eds. Gene Expression and Regulation, Chapter 14, Beijing: HEPC and Springer Press, 2006, 239-256
- [9] Guenther MG et al. Cell, 2007, 130(1): 77-88
- [10] Muse GW et al. Nat Genet, 2007, 39(12): 1507-1511
- [11] Zeitlinger J et al. Nat Genet, 2007, 39(12): 1512-1516
- [12] Chen R et al. Genes Dev, 2008, 22(10): 1356-1368

表1

延伸因子	亚基	正性/负性	特征
TFIIF	RAP30 RAP74	正性	参与起始前复合体的形成, 重启转录暂停, 提高PoI II转录速率, 调控TFIIS
Elongins	Elongin A Elongin B Elongin C	正性	重启转录暂停, 提高PoI II转录速率, 泛素化相关事件
ELL	EAP20 EAP30 EAP45	正性	重启转录暂停, 提高PoI II转录速率
DSIF	Spt4 Spt5	正性	刺激转录延伸, 抑制转录终止早期, 活化加帽
NELF	NELF-A NELF-B NELF-C NELF-D NELF-E	正性	中断PoI II介导的基因转录, 预留时间进行5'-端加帽和检控点调控
CSB	-	正性	活化转录延伸, 调控TFIIS, DNA损伤和转录相伴的核酸切除修复中援引PoI II起作用
FCP1	-	正性	活化转录延伸, 去磷酸化CTD Ser5以循环利用PoI II, 参与加帽过程
TFIIS	-	正性	激活PoI II介导的新生转录产物的剪切
Spt6	-	正性	活化转录延伸, 调控染色质结构, 组蛋白伴侣
HDAg	-	正性	活化转录延伸, 结合PoI II, 置换NELF, 功能异于TFIIF
19S Proteasome	Rpt4 Sug1 (Rpt6) Others?	正性	通过H2B单泛素化汇集, 参与H3-K4甲基化
P-TEFb <sup>a</sup>	Cdk9 cyclinT	正性	重启NELF介导的转录暂停, 磷酸化CTD Ser2和DSIF (Spt5)
Ssu72	-	负性	去磷酸化CTD Ser5, 参与3'-末端加工过程
SWI/SNF	BRG1 Others	负性	ATP依赖性的染色质重构
Isw1p (Isw1b)	Isw1p Ioc2p Ioc4p	负性	调控Ser5和Ser2的磷酸化, H3-K4和H3-K36甲基化, 募集3'-末端加工因子
Chd1	-	负性	遗传性牢固结合转录延伸因子, 定位在编码区内
FACT	Spt16 SSRP1	负性 <sup>b</sup>	通过染色质便利转录延伸, 调控染色质结构, 组蛋白伴侣活性
Set1	-	负性	甲基化组蛋白H3-K4, 定位于启动子和编码区
Set2	-	负性	甲基化组蛋白H3-K6, 定位于编码区

<sup>a</sup> 推测 Ctk1 和 Bur1 是 P-TEFb 催化亚基 Cdk9 的酵母同源物。<sup>b</sup> 影响染色质模板上 PoI II 转录速率。-, 未发现亚基。

## Highlights on eukaryotic transcription elongation

Nan-Ping AI Rui-Chuan CHEN Min LIU

(Key Laboratory of the Ministry of Education for Cell Biology and Tumor Cell Engineering, School of Life Sciences,

Xiamen University, Xiamen 361005, China)

**Abstract** For many years, elongation stage of eukaryotic transcription has been thought of as repetitive, unregulated pre-mRNA synthesizing step. In addition, the maturation of pre-mRNA has been considered as a post-transcriptional step separable from the transcription process. Only until recently, it has become increasingly clear that the elongation stage of eukaryotic transcription is a highly regulated process that is capable of not only generating full-length RNA transcripts but also coupling transcription with pre-mRNA maturation. Here, we try to summarize the recent progress in studies focusing on elongation stage of eukaryotic transcription, including the coupling mechanisms of elongation with pre-mRNA maturation and the intrinsic links of general transcription to the regulation of cell proliferation and differentiation.

**Keywords** eukaryotic cells; transcriptional elongation step; elongation regulation