



真核生物RNA聚合酶II转录前起始复合物的结构研究进展

陈曦子*, 徐彦辉*

复旦大学生物医学研究院, 复旦大学附属肿瘤医院, 复旦大学生命科学学院, 上海 200032

* 联系人, E-mail: xuyh@fudan.edu.cn; chenxz@fudan.edu.cn

2025-01-02 收稿, 2025-03-08 修回, 2025-04-21 接受, 2025-04-21 网络版发表

国家重点研发计划(2021YFA1300100, 2022YFA1303200)和国家自然科学基金(32271242)资助

摘要 真核生物基因转录是基因表达的关键环节, 决定了细胞的生命过程和发育、分化、衰老等组织和个体层面的生命活动, 影响众多生理病理过程. 发生在基因启动子区的转录起始是基因表达调控的核心, 是细胞内最为重要的生命过程之一, 受到来自染色质结构、表观遗传修饰以及转录因子、转录共激活因子等多重因素的复杂调控, 涉及上百个蛋白质的超大复合物和数千个调控蛋白, 转录起始失调将导致包括肿瘤在内等众多疾病的发生. 本文围绕哺乳动物RNA聚合酶II所介导的转录起始过程, 总结了包括转录前起始复合物的启动子识别、分步组装、转录起始、转录激活等多个过程的最新结构研究进展. 这些研究为深入理解多细胞生物复杂的转录调控机制提供了重要的理论基础, 对理解基因表达调控及相关疾病的发生发展具有重要意义.

关键词 转录前起始复合物, 中介体, 转录起始, 转录调控

1958年克里克首次提出了分子生物学“中心法则”, 描述了遗传信息从DNA传递给RNA, 再从RNA传递给蛋白质的过程. “中心法则”是生物体活动最基本、最重要的规律, 指导遗传密码破译成能够行使生物学功能的生物大分子. 基因表达包括转录和翻译两个过程, 其中转录是以DNA为模板产生RNA的过程, 作为遗传信息传递的首要环节, 其调控直接决定了基因的差异性表达和细胞的命运, 影响生物个体的生长、发育和衰老等几乎所有生命活动. 转录失调将会导致众多疾病的发生, 一半以上的癌症发生都与转录紊乱密切相关.

早期对转录机制的研究主要集中在细菌和噬菌体上, 如乳糖操纵子的发现, 揭示了正负调控机制和乳糖阻遏蛋白、CAP蛋白等关键因子的作用, 为基因表达调控的理解奠定了基础. 由于真核生物复杂的基因组

环境, 直到1969年Roeder等人^[1]才得以成功从海胆中分离出真核生物RNA聚合酶并发现不同RNA聚合酶分别负责合成不同类型的RNA. 真核生物中, 细胞核基因的转录由3种高度保守的RNA聚合酶: RNA聚合酶I(RNA polymerase I, Pol I)、RNA聚合酶II(RNA polymerase II, Pol II)和RNA聚合酶III(RNA polymerase III, Pol III)分工完成, 此外陆生植物还具有两种特有的RNA聚合酶IV(RNA polymerase IV, Pol IV)和RNA聚合酶V(RNA polymerase V, Pol V), 负责植物细胞核基因组部分特定区域的转录. 其中, Pol I主要合成核糖体RNA(rRNA), 包括5.8S rRNA、18S rRNA和28S rRNA; Pol II负责生成信使RNA(mRNA)、小核仁RNA(snoRNA)、核内小RNA(snRNA)、长非编码RNA(lncRNA)和微小RNA(microRNA)等; Pol III则主要参与细胞内短片段RNA的合成, 如转运RNA(tRNA)、5S rRNA、U6 snRNA以及

引用格式: 陈曦子, 徐彦辉. 真核生物RNA聚合酶II转录前起始复合物的结构研究进展. 科学通报, 2025, 70: 2333–2343

Chen X, Xu Y. Structural insights into RNA polymerase II transcription preinitiation complex (in Chinese). Chin Sci Bull, 2025, 70: 2333–2343, doi: 10.1360/TB-2025-0011

7SL RNA等. 细胞内, 由Pol I合成的rRNA装配形成核糖体, 与Pol III生成的tRNA共同作用, 以Pol II转录的mRNA为模板, 合成蛋白质, 实现基因的表达.

在真核生物中, 基因组DNA以染色质形式存在, 染色质由基因组DNA缠绕组蛋白形成的核小体再进一步折叠压缩形成. 压缩紧密的染色质结构一定程度上阻碍了RNA聚合酶和转录因子等转录相关蛋白的结合, 避免了随机转录造成异常基因的表达, 防止细胞功能紊乱和疾病的发生. 因此, 相比于原核生物, 真核生物的基因通常是处于非活化的状态, 需要经历复杂的转录起始过程来调控基因的表达. 研究显示, 转录起始受到来自染色质结构、DNA序列、转录因子、表观遗传等多个层面涉及数千种蛋白的调控. 多层次复杂的转录起始调控为高等生物体利用同一套基因组遗传信息, 表达不同的蛋白, 分化数百种不同的细胞类型, 完成复杂的生长发育过程提供了保证, 因此研究其调控机制对理解基因表达的调控以及相关疾病的发生发展具有非常重要的意义. 本文主要聚焦于RNA聚合酶II所介导的转录起始过程, 总结最新研究进展, 并进一步讨论对转录调控领域的未来研究展望.

1 RNA聚合酶II介导的转录起始

Pol II所介导的转录包括起始、延伸和终止三个主要阶段, 其中转录起始作为首要步骤, 是基因表达的关键调控点. Pol II的转录起始发生在几乎所有编码基因和大部分非编码基因的核心启动子区. 核心启动子是基因上涵盖转录起始位点上游50 bp到下游50 bp, 包含转录起始位点(transcription start site, TSS)的一段启动子区域, 是能够引导Pol II起始转录的最短DNA序列^[2]. Pol II的核心启动子主要分为“集中型”(focused, sharp)

启动子和“分散型”(dispersed, broad)两种类型. 集中型启动子具有单一明确的转录起始位点, 常见于高度调控的细胞特异性基因, 其核心启动子区一般包含一些特定的DNA基序(motif)——核心启动子元件(如TATA框等), 帮助Pol II精确定位转录起始位点. 而分散型启动子则拥有多个相邻且使用频率相似的起始位点, 主要存在于管家基因(house-keeping genes)及调控形态发生和模式形成的发育关键基因中. 由于分散型启动子缺乏明确的核心启动子元件, 其转录起始位点的选择可能更多依赖启动子区域的染色质特性, 如核小体的表观遗传修饰和定位等多个层面的调控^[3].

与原核生物基因转录不同, Pol II不能直接识别靶基因启动子, 并且不具有打开DNA双链的能力, 研究显示, Pol II需与6个II类通用转录因子(general transcription factor, GTF): TFIIA、TFIIB、TFIID、TFIIE、TFIIF和TFIIH共同组装形成转录前起始复合物(preinitiation complex, PIC)才能起始转录(表1). 在组装过程中, TFIID起始PIC形成, TFIIH完成组装并通过CDK7激酶磷酸化Pol II的羧基端结构域(C terminal domain, CTD), 以及XPB转位酶打开DNA双链, 这两步是转录起始的关键事件, 完整PIC的形成是转录的必要条件. 完整PIC具备基本的转录起始能力, 但细胞中高表达基因的转录往往还需要序列特异性转录因子和转录共激活因子(如中介体Mediator)共同作用来激活转录起始^[4].

转录起始是一个高度复杂且普遍的过程, 涉及染色质结构、转录因子、DNA序列、表观遗传等多个层面包含上百种蛋白、十余个复合物以及超过3000种调控因子的精确协作与动态转换. 上千种序列各异的转录因子可以针对不同基因进行特异性调控, 来自上游

表 1 II类通用转录因子的组成和功能

Table 1 Composition and functions of class II general transcription factors

因子	亚基组成	分子量	功能
TFIID	TBP, TAF1-TAF13	1200 kD	识别并结合启动子DNA、H3K4me3以及乙酰化修饰的核小体, 起始PIC组装
TFIIA	TFIIA α , TFIIA β , TFIIA γ	54 kD	帮助并稳定TFIID与启动子的结合
TFIIB	TFIIB	35 kD	稳定TFIID与启动子的结合, 招募Pol II-TFIIF, 转录起始位点的选择
TFIIF	TFIIF α (RAP74), TFIIF β (RAP30),	87 kD	结合Pol II并促进其被招募至启动子, 招募TFIIE, 起始位点的选择
TFIIE	TFIIE α (p56), TFIIE β (p34),	82 kD	招募TFIIH至启动子
TFIIH	XPB, XPD, p62, p52, p44, p34, p8, CDK7, CyclinH, MAT1	490 kD	转位酶活性: 打开启动子DNA双链, 形成转录泡; 激酶活性: 磷酸化Pol II, 促进转录起始

的各种调控信号最终都将汇聚于PIC, 决定PIC是否能够正确组装, 是否启动转录. 因此, PIC的组装及起始转录可以说是转录调控的关键环节, 过去几十年间, 研究者们采用生物化学、单分子生物物理、超分辨成像和结构生物学等技术, 对这一过程及其调控机制展开了深入的探索.

2 转录前起始复合物PIC的组装

20世纪90年代, 研究发现, PIC的组装是逐步发生: TFIID结合TFIIA识别启动子后, 依次招募Pol II及TFIIB、TFIIF、TFIIE, 最终TFIIH加入, 磷酸化Pol II CTD并打开DNA双链, 起始转录^[5-8]. 完整PIC由50个蛋白组成, 具有高度的复杂性和动态性. 其中, TFIID作为最关键的起始因子之一, 由TATA框结合蛋白(TATA box-binding protein, TBP)和13个TBP-相关蛋白(TBP-associated factor, TAF1-13)所组成, 包含20条多肽链,

分子量1.3 MDa.

早期研究发现, 在体外, 以TATA框元件启动子(TATA box promoter)为模板的转录活性反应体系中, 单独的TBP可以替代TFIID起始转录^[9,10]. 因此, 在过去几十年中, 研究人员采用简化的实验模型, 使用TBP蛋白替代TFIID复合物, 围绕TBP与含TATA框启动子开展了一系列结构和功能研究, 解析了基于TBP的PIC结构(图1(a)): 在PIC中, TBP蛋白结合并弯曲核心启动子上的TATA框元件; 通用转录因子TFIIA结合TBP, 稳定TBP与启动子的结合; 通用转录因子TFIIB的锌带结构域(zinc ribbon)结合Pol II的船坞结构域(dock), 核心结构域结合TBP和启动子, 稳定Pol II、TBP和启动子之间的相互作用; 通用转录因子TFIIF的二聚化结构域(dimerization domain)结合Pol II的叶结构域(lobe), 翼螺旋结构域(winged helix domain, WH)结合在TATA框下游的启动子区域, 进一步稳定Pol II和启动子之间的

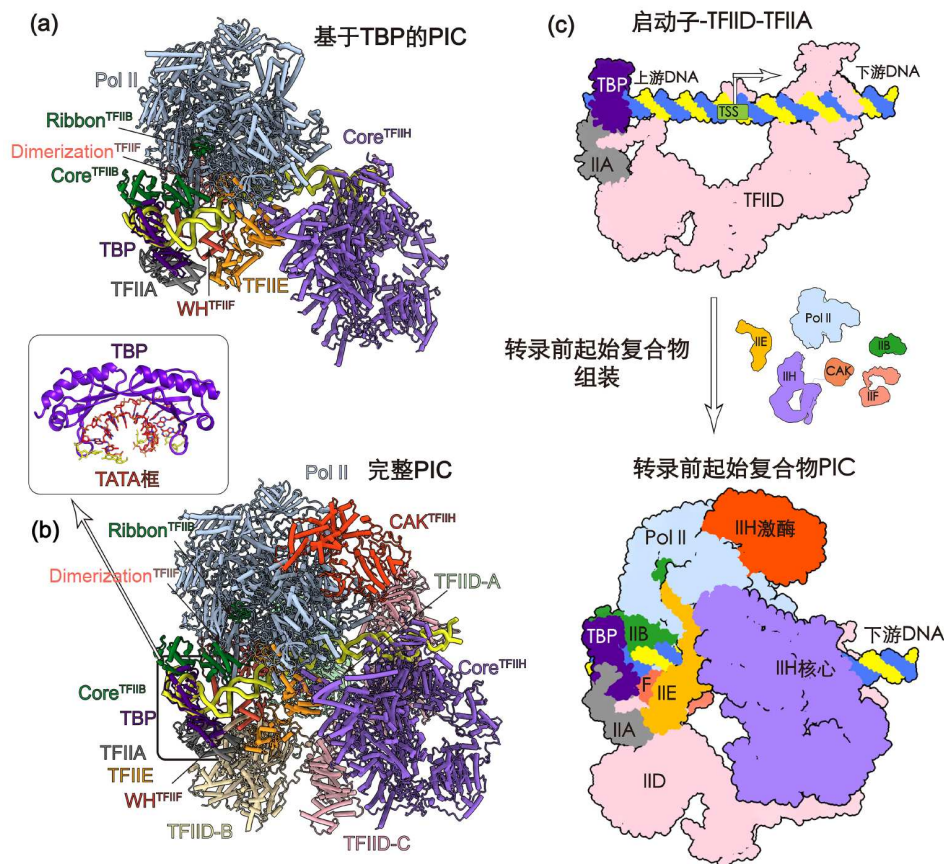


图1 转录前起始复合物PIC的结构模型及组装模式图. (a) 基于TBP的转录前起始复合物结构(PDB:7NVY); (b) 包含TFIID的完整转录前起始复合物结构(PDB:7EGB); (c) 转录前起始复合物的组装模式图. 结构模型由ChimeraX软件绘制

Figure 1 Structure and assembly model of the transcription pre-initiation complex. (a) Structure of TBP-based PIC structure (PDB: 7NVY). (b) Structure of TFIID-based PIC (PDB: 7EGB). (c) Schematic model of PIC assembly. Generated using UCSF ChimeraX

相互作用;通用转录因子TFIIE连接Pol II的茎杆结构域(stalk)和TFIIF翼螺旋结构域,与Pol II一起将TFIIF结合位点下游的一小段启动子DNA“包围”起来;转录起始位点TSS位于Pol II的催化活性位点上方区域;TSS下游启动子与Pol II的颌(jaw)结构域结合;通用转录因子TFIIH的核心模块中的XPB-MAT1与Pol II茎杆结构域和TFIIE结合,模块中的XPB转位酶结合在TSS下游的启动子区域,帮助Pol II打开启动子双链。在PIC中,Pol II与各通用转录因子紧密结合,在通用转录因子的帮助下,Pol II稳定结合在核心启动子上,为后续转录起始做好了准备^[11~26]。

然而,随着研究的深入,研究人员发现细胞中的TBP并非以游离形式存在,而是作为TFIID复合物的核心亚基发挥作用。值得注意的是,人类基因组中超过85%的基因启动子属于不含TATA框的类型(TATA-less promoters),仅TBP蛋白并不能启动这类基因的转录。几乎所有由Pol II介导的真核基因转录都必须依赖完整的TFIID复合物才能起始^[27~29]。因此基于TATA框启动子和TBP的研究结果并不能完整地揭示在人类细胞中所有基因的起始调控机制。

突破性成果来自2021年,研究人员通过在体外组装,利用冷冻电子显微镜方法,解析了结合不同启动子类型以及不同组装阶段的包含TFIID的多个PIC结构,揭示了PIC在不同启动子上的动态组装过程^[8](图1(b, c))。研究发现,TFIID由三个模块(TFIID-A、TFIID-B和TFIID-C)所组成,具有多个DNA结合区域,并且对DNA序列的保守度不高,能够识别并结合细胞内复杂的各种不同类型的启动子。TFIID结合启动子后,会进一步招募Pol II和其他通用转录因子,并将启动子推至Pol II的催化中心附近,最终组装成完整的PIC。完整的PIC复合物中,Pol II被成功装载进PIC,启动子位于Pol II的催化中心附近,TFIIH的CAK激酶模块结合在Pol II CTD附近,促进CTD磷酸化这一转录起始的关键调控步骤的发生;TFIIH的XPB转位酶结合在启动子下游,为转录起始时打开DNA双链这一转录起始关键步骤提供支持。

研究发现,包含TFIID的完整PIC中,无论启动子是否含有TATA框,TBP都能够稳定结合并弯曲上游启动子,将PIC稳稳地锚定在启动子上,避免转录时DNA双链打开以及转位时PIC在启动子上发生滑移和旋转,从而保证转录能够顺利进行。因此,TBP对于所有基因的转录是重要,TFIID复合物中的TAFs作为识别和支架

蛋白,帮助TBP正确地结合到启动子上游,TBP结合后弯折启动子,将PIC稳定在启动子上,保证转录正常起始^[8]。

关于PIC的组装顺序,目前仍存在一定的争议。20世纪90年代,研究人员通过直接从酵母和哺乳动物细胞中分离复合物,发现了一种包含除TFIID外其他5个通用转录因子、中介体及Pol II的复合体,基于这一发现,他们提出了PIC的全酶组装模型,即PIC的组装过程与原核生物的聚合酶复合体类似,TFIID类似细菌的“ σ 因子”,首先与启动子结合,而其他因子则与Pol II形成复合体后共同被招募至启动子区域,与TFIID协同组装,形成完整的PIC^[31]。近期,研究人员利用单分子荧光显微镜技术,发现细胞内还存在一种激活子依赖的组装方式,即Pol II、TFIIF、TFIIE以及中介体提前被招募至增强子区组装,随后一同被招募到基因启动子区与TFIID及其他几个通用转录因子组装形成PIC^[32]。究竟细胞中PIC是以哪种方式组装仍有待进一步研究。

3 转录起始到早期延伸的转换过程

当PIC在启动子上完成组装后,TFIIH将启动子DNA双链打开形成转录泡,模板链被送进Pol II的催化中心,随后Pol II利用NTP,开始RNA链的合成。从20世纪90年代开始,研究者们利用生化方法围绕Pol II的起始过程开展了大量的研究,研究发现Pol II转录的起始阶段会发生显著的蛋白组成和功能的变化。当生成的RNA链达到2~3个核苷酸(nucleotide, nt)时,由于RNA与Pol II的结合不够稳定,首先会发生“流产式转录”(abortive transcription);随后,当RNA链合成至4 nt时,此时RNA与Pol II结合相对稳定,逐步进入“逃逸承诺”阶段(escape commitment);接着,随着RNA链不断增长,当RNA链达到7~10 nt后,转录泡扩大到了17~18 nt,转录泡会突然“崩解”(bubble collapse),Pol II与通用转录因子分开,完成启动子逃逸(promoter escape),此时Pol II与RNA和模板DNA形成稳定的转录复合物,顺利进入早期延伸阶段^[33~47]。2006年,Ebright等人^[48]通过体外转录和单分子实验发现,细菌RNA聚合酶在转录起始阶段,其转录泡中的模板单链会被持续压缩(scrunching)至聚合酶内以驱动聚合酶从启动子区逃逸进入快速延伸阶段,即“蜷缩”(scrunching)模型。

近期,研究人员解析了16个连续(RNA长度2~17 nt)的转录复合物结构,揭示了Pol II的转录起始过程(图2):PIC组装完成后,TFIIH通过水解ATP产生的能量打开

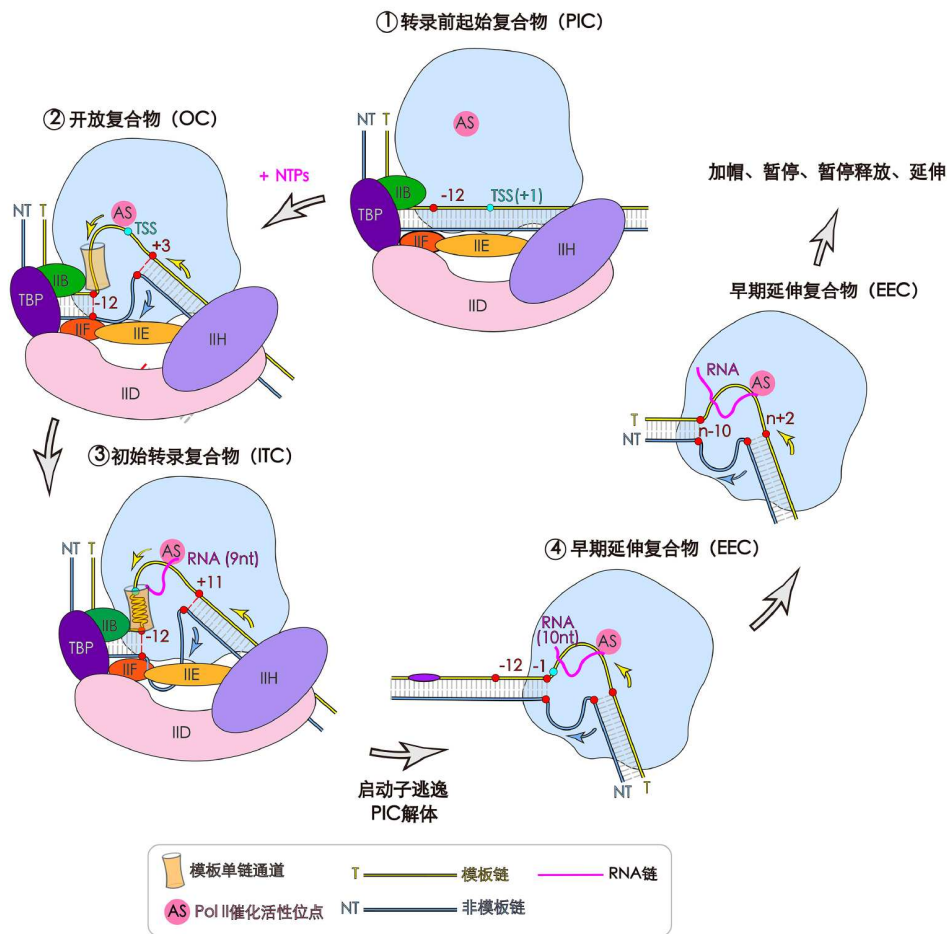


图2 Pol II介导的转录起始到早期延伸转换模式图

Figure 2 Schematic model of transcription initiation-to-early elongation transition

启动子区部分DNA双链, 模板链进入Pol II催化活性中心, 在转录起始位点附近产生13 nt的转录泡, 形成开放复合物(open complex, OC). 随着NTPs的加入, Pol II以模板链为模板开始合成RNA, 开放复合物转变为初始转录复合物(initial transcription complex, ITC), 在此阶段, 通用转录因子与Pol II及启动子保持紧密结合, 构象基本与PIC一致. 上游启动子因TBP的作用而稳定弯折, 下游启动子则在TFIIH转位酶的作用下持续进入Pol II的催化活性中心, 确保转录泡持续开放, 避免RNA合成早期因转录泡不稳定而失败. 随着RNA链从2 nt延长至9 nt, DNA-RNA杂交链逐渐稳定, 当RNA链延长至10 nt时, Pol II内堆积的模板链与RNA-DNA杂交链前移共同“胀破”模板链通道, 参与组成模板链通道的TFIIB及相关的通用转录因子与Pol II解离, ITC转变为早期延伸复合物(early elongation complex, EEC). 这一过程

中, 通用转录因子从启动子区域解离, 上游的DNA重新配对形成双链, 转录泡从21 nt缩小至11 nt, Pol II也脱离最初结合的启动子区域. 形成EEC后, 随着RNA链逐步延长, 转录延伸进入稳定重复过程, Pol II随着合成的核苷酸向前推进, 同时新生成的RNA链通过RNA通道伸出^[49].

为了能够精确起始转录, Pol II的转录起始需要依赖通用转录因子. 通用转录因子、Pol II与启动子之间紧密的相互作用形成稳定的PIC, 为实现高效且稳定的转录起始提供了保证. 然而, 这种稳定作用在起始完成后, Pol II离启动子时却成了障碍, 因此PIC在转录起始完成后需要解体, 通用转录因子与Pol II分离, 解除对启动子的束缚, 确保Pol II顺利进入延伸阶段. 通用转录因子与转录泡的协同作用不仅支持Pol II在初始RNA合成阶段结合并打启动子, 还为Pol II成功起始转录

后摆脱束缚顺利进入稳定快速高效的RNA延伸合成阶段提供了支持.

4 中介体激活转录

完整的PIC能够完成基本的转录起始, 具备基础的转录活性. 在细胞中, 活跃基因的转录水平通常显著高于这种基础水平, 这些基因的高效转录需要序列特异性转录因子(sequence-specific transcription factors, TF)的调控. 1990年, Soutourina^[4]和Flanagan等人^[50]发现一个重要的转录共激活因子——中介体(Mediator), 当它与转录因子共存时, 能够显著提高转录水平. 人体中绝大多数高活跃基因的表达都需要中介体的参与, 中介体可以将不同信号通路的转录激活信号, 传递到PIC上激活转录.

中介体复合物由30个蛋白组成, 其核心模块参与了PIC的招募和组装, 主要由头部、中部和尾部三个亚模块构成, 包含26种蛋白质. 近期PIC-中介体复合物的结构研究显示, PIC-中介体复合物包含76个蛋白, 总分子量4.1 MDa, 当中介体结合PIC后, 其头部和中部模块会发生显著构象变化, 夹住Pol II的CTD, 形成“三明治结构”, 促进暴露于CDK7激酶附近的Pol II CTD被磷酸化. 头部和中部之间的动态相互作用形成“CTD磷酸化门控”, 通过“门”的开关调节Pol II CTD的释放, 确保其能够被CDK7激酶持续有效地磷酸化, 当CTD被完全磷酸化后, Pol II与中介体解离, 进入转录延伸阶段. 在

PIC-中介体复合物中，中介体与TFIID分别结合于TFIIH的上下两侧，共同稳定TFIIH在复合物中的定位，更好地支持TFIIH功能的发挥^[51](图3(a)). 因此，对于细胞内高活跃基因的表达，中介体通过接收来自上游转录因子的激活信号，促进启动子区PIC的组装及CDK7激酶活性，增强Pol II CTD磷酸化，提升转录效率。

5 染色质上的转录起始

真核生物通过调控基因启动子区域的染色质结构实现基因转录的调控, 这些染色质结构包括DNA序列、转录因子、表观遗传修饰、染色质重塑复合物及高级染色质结构等^[3,52-55]. 转录因子结合基因的增强子区和近端启动子区, 促进转录前起始复合物的招募和组装; 而表观遗传修饰则通过改变基因组稳定性, 招募转录调控相关复合物, 从而调节基因的转录活性.

位于转录起始位点上下游的第一个核小体分别为-1核小体和+1核小体,与经典核小体所不同,启动子区域的核小体包括特有的组蛋白变体(如H2A.Z和H3.3)以及表观遗传修饰(如组蛋白乙酰化和H3K4me3)^[52-58]。这些特征在表观遗传修饰酶和染色质重塑复合物的动态调控下形成,影响核心启动子区域核小体的状态及转录前起始复合物的组装。

在人类活跃转录的基因中, +1核小体通常位于转录起始位点(TSS)下游约40 bp处^[59-61]。传统观点认为, +1核小体与PIC之间没有直接关联, Pol II只有在转录

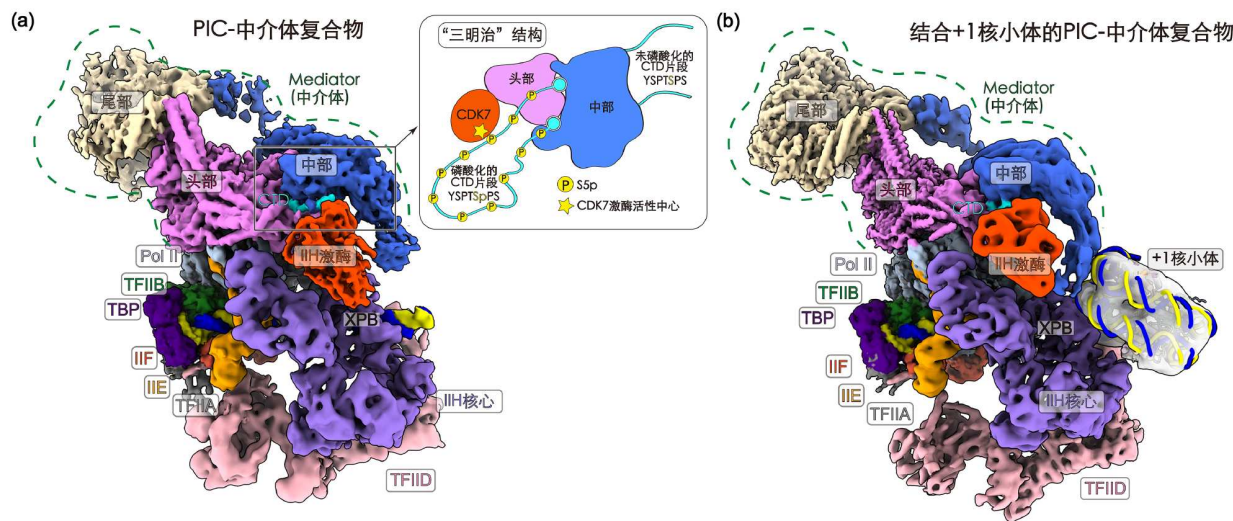


图 3 转录起始超级复合物结构. (a) PIC-中介体复合物的冷冻电子显微镜密度图(EMD-31207)及“磷酸化门控模式图”; (b) 结合+1核小体的PIC-中介体复合物的冷冻电子显微镜密度图(EMD-34359). 由ChimeraX软件绘制

Figure 3 Structure of PIC-Mediator complex. (a) Cryo-EM structure of the PIC-Mediator complex and the proposed CTD-gating model. (b) Cryo-EM structure of +1 nucleosome-bound PIC-Mediator complex. Generated using UCSF ChimeraX

进入延伸阶段后才会遇到+1核小体,并且+1核小体的存在会成为延伸过程中的“障碍物”抑制基因的转录。然而,近期的研究发现,+1核小体能通过非序列依赖的静电相互作用与PIC结合,促进PIC招募和组装并激活转录起始,当+1核小体上的组蛋白变体H2A.Z替换成H2A,乙酰化组蛋白替换成未修饰组蛋白时,均会减弱其与PIC的相互作用,影响转录起始活性(图3(b))^[62-64]。因此,转录调控可能直接通过+1核小体上的组蛋白变体(如H2A.Z和H3.3)及组蛋白修饰(包括H3K4三甲基化与组蛋白乙酰化)实现。

TSS及其上下游的核小体在转录调控中扮演了关键的角色,核心启动子区域的核小体直接影响了转录起始效率,当核小体占据核心启动子区域时,其产生的空间位阻效应会抑制PIC的组装,从而导致基因沉默;对于此类基因的激活,需要通过染色质重塑复合物介导的核小体移除或重新定位来实现。在活跃转录的基因中,精确定位的+1核小体与PIC结合,促进PIC组装、稳定PIC,激活转录起始。对于无TATA框启动子,+1核小体可能参与PIC的招募,使其按+1核小体的位置组装,进而在复合物的框架内决定了转录的方向和转录起始位点(图4)。相比于原核生物主要依赖DNA序列来

调控基因转录,真核生物通过染色质结构的复杂和动态的调控方式,既满足了多细胞生物对基因表达多样化的需求,也实现了对基因表达的精确控制。

6 展望

真核生物的转录起始是一个高度动态的多步骤过程,包括转录前起始复合物PIC的组装、RNA生成、启动子逃逸、RNA加帽、暂停与释放、转录延伸复合物形成以及Pol II克服+1核小体的结构障碍进入有效延伸阶段等。在转录起始位点到+1核小体之间约40 bp的范围内,Pol II需要经历多次复合物的组装与解离,涉及上百种转录调控蛋白的相互协同作用,这些因子是如何动态调控的机制目前还未得到阐明,仍有待进一步探索。

染色质高级结构(如3D基因组中增强子-启动子环化)、转录因子与转录共激活因子的动态招募,以及蛋白-DNA相互作用网络与表观遗传修饰之间的级联调控等复杂因素如何共同调控转录起始,特别是在复杂染色质微环境中,多重调控元件(包括DNA序列、染色质结构及表观遗传修饰)如何时空协调转录起始,目前仍未被阐明。核心启动子区域的核小体富集多种表观

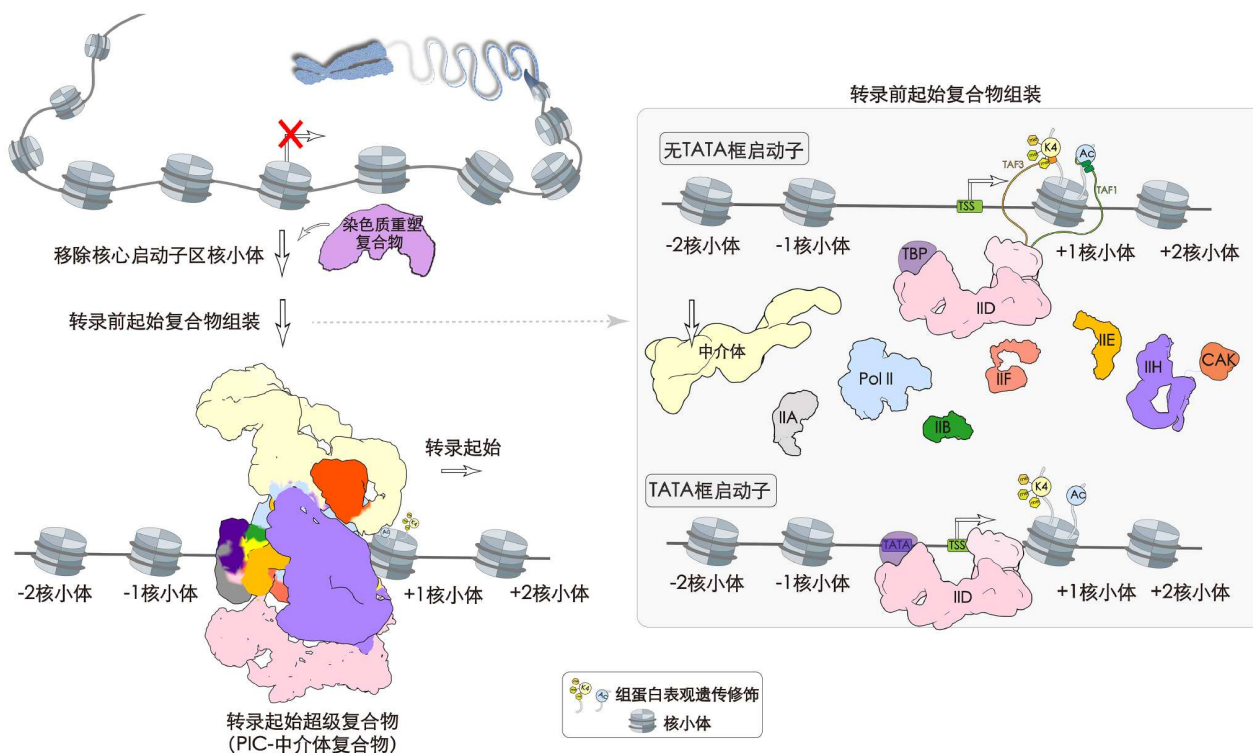


图4 转录前起始复合物在染色质上的组装模式图

Figure 4 Schematic model of transcription pre-initiation complex assembly on chromatin

遗传修饰特征(如组蛋白H3K4me3、乙酰化修饰以及组蛋白变体H2A.Z和H3.3),这些修饰作为修饰酶与染色质重塑复合物的关键结合位点,通过调控核小体稳定性或介导转录复合物招募实现对转录的精细调控。当前转录起始的结构研究主要依赖于相对简化的研究体系(如使用裸露DNA或含强核小体定位序列Widom 601的核小体模板),然而这些体系无法模拟染色质上复杂的表观遗传修饰及基因组结构。未来研究还需构建更接近生理状态的染色质模板,深入系统地研究染色质高级结构、转录因子、转录共激活因子、表观遗传修饰、染色质重塑复合物以及转录机器等的多层次相互作用网络,阐明其协同调控转录起始的分子机制。

转录起始发生的时间尺度(转录前起始复合物组装到进入转录延伸)在分钟级别,空间尺度(转录复合物在基因组上的位移)在约14纳米之内(40个核苷酸的长度),由于这些时空维度的特殊性,对该过程的研究具有极大的挑战。过去几十年来,生物化学、分子生物学(测序等)、结构生物学、单分子和光学成像等方法对这一体系的阐释起到了极大的推动作用。这些技术手段各具优势的同时也存在一定的局限:例如,结构生物学虽然能提供静态的高分辨率复合物结构信息,却无法

捕获转录起始前复合物组装的中间态以及转录起始的动态过程,因此不能完全模拟出细胞内真实的转录起始环境;单细胞多组学虽然可以检测细胞内真实的RNA产物、转录复合物结合或覆盖的基因组区域等,但属于终点检测,无法描述分子水平上发生的动态变化过程。值得注意的是,近年来发展的活细胞成像技术能够实时观测转录因子的招募以及RNA生成等,这为动态研究带来了新希望,但目前的技术分辨率仍存在瓶颈,尚不足以动态跟踪单基因上转录复合物分子的动态转录起始事件。因此,未来对于转录起始的研究需要综合多种新的技术方法,例如,利用冷冻电子断层成像(Cryo-ET)技术直接研究细胞内原位的转录复合物结构,开发更高分辨率的测序技术方法和细胞内超分辨率成像技术直接检测细胞内单分子的转录事件等。同时,通过在体外构建更接近生理条件的完整研究体系,利用生化活性分析、单分子以及冷冻电镜单颗粒结构解析技术,综合细胞内功能分析,在分子水平上全面、精确地解析转录起始的动态过程。这些研究的推进将更加全面地理解转录起始调控机制,揭示转录调控的基本原理,从而从根本上深化对生命机制和疾病发生过程的理解,为相关人类疾病的治疗提供重要的理论依据。

参考文献

- 1 Roeder R G, Rutter W J. Specific nucleolar and nucleoplasmic RNA polymerases. *Proc Natl Acad Sci USA*, 1970, 65: 675–682
- 2 Juven-Gershon T, Kadonaga J T. Regulation of gene expression via the core promoter and the basal transcriptional machinery. *Dev Biol*, 2010, 339: 225–229
- 3 Haberle V, Stark A. Eukaryotic core promoters and the functional basis of transcription initiation. *Nat Rev Mol Cell Biol*, 2018, 19: 621–637
- 4 Soutourina J. Transcription regulation by the Mediator complex. *Nat Rev Mol Cell Biol*, 2018, 19: 262–274
- 5 Buratowski S, Hahn S, Guarente L, et al. Five intermediate complexes in transcription initiation by RNA polymerase II. *Cell*, 1989, 56: 549–561
- 6 Van Dyke M W, Roeder R G, Sawadogo M. Physical analysis of transcription preinitiation complex assembly on a class II gene promoter. *Science*, 1988, 241: 1335–1338
- 7 Zawel L, Reinberg D. Initiation of transcription by RNA polymerase II: a multi-step process. *Prog Nucleic Acid Res Mol Biol*, 1993, 44: 67–108
- 8 Chen X, Qi Y, Wu Z, et al. Structural insights into preinitiation complex assembly on core promoters. *Science*, 2021, 372: eaba8490
- 9 Peterson M G, Tanese N, Pugh B F, et al. Functional domains and upstream activation properties of cloned human TATA binding protein. *Science*, 1990, 248: 1625–1630
- 10 Verrijzer C P, Chen J L, Yokomori K, et al. Binding of TAFs to core elements directs promoter selectivity by RNA polymerase II. *Cell*, 1995, 81: 1115–1125
- 11 Nogales E, Louder R K, He Y. Structural insights into the eukaryotic transcription initiation machinery. *Annu Rev Biophys*, 2017, 46: 59–83
- 12 Liu X, Bushnell D A, Wang D, et al. Structure of an RNA polymerase II–TFIIB complex and the transcription initiation mechanism. *Science*, 2010, 327: 206–209
- 13 Kostrewa D, Zeller M E, Armache K J, et al. RNA polymerase II–TFIIB structure and mechanism of transcription initiation. *Nature*, 2009, 462: 323–330
- 14 Robinson P J, Trnka M J, Bushnell D A, et al. Structure of a complete mediator-RNA polymerase II pre-initiation complex. *Cell*, 2016, 166: 1411–

1422.e16

- 15 Plaschka C, Hantsche M, Dienemann C, et al. Transcription initiation complex structures elucidate DNA opening. *Nature*, 2016, 533: 353–358
- 16 He Y, Yan C, Fang J, et al. Near-atomic resolution visualization of human transcription promoter opening. *Nature*, 2016, 533: 359–365
- 17 Plaschka C, Larivière L, Wenzek L, et al. Architecture of the RNA polymerase II–mediator core initiation complex. *Nature*, 2015, 518: 376–380
- 18 Murakami K, Tsai K L, Kalisman N, et al. Structure of an RNA polymerase II preinitiation complex. *Proc Natl Acad Sci USA*, 2015, 112: 13543–13548
- 19 Murakami K, Elmlund H, Kalisman N, et al. Architecture of an RNA polymerase II transcription pre-initiation complex. *Science*, 2013, 342: 1238724
- 20 He Y, Fang J, Taatjes D J, et al. Structural visualization of key steps in human transcription initiation. *Nature*, 2013, 495: 481–486
- 21 Schilbach S, Aibara S, Dienemann C, et al. Structure of RNA polymerase II pre-initiation complex at 2.9 Å defines initial DNA opening. *Cell*, 2021, 184: 4064–4072.e28
- 22 Aibara S, Schilbach S, Cramer P. Structures of mammalian RNA polymerase II pre-initiation complexes. *Nature*, 2021, 594: 124–128
- 23 Yang C, Fujiwara R, Kim H J, et al. Structural visualization of *de novo* transcription initiation by *Saccharomyces cerevisiae* RNA polymerase II. *Mol Cell*, 2022, 82: 660–676.e9
- 24 Rengachari S, Schilbach S, Aibara S, et al. Structure of the human Mediator–RNA polymerase II pre-initiation complex. *Nature*, 2021, 594: 129–133
- 25 Abdella R, Talyzina A, Chen S, et al. Structure of the human Mediator-bound transcription preinitiation complex. *Science*, 2021, 372: 52–56
- 26 Schilbach S, Hantsche M, Tegunov D, et al. Structures of transcription pre-initiation complex with TFIID and Mediator. *Nature*, 2017, 551: 204–209
- 27 Donczew R, Hahn S. Mechanistic differences in transcription initiation at TATA-less and TATA-containing promoters. *Mol Cell Biol*, 2018, 38: e00448-17
- 28 Donczew R, Warfield L, Pacheco D, et al. Two roles for the yeast transcription coactivator SAGA and a set of genes redundantly regulated by TFIID and SAGA. *eLife*, 2020, 9: e50109
- 29 Taatjes D J. The continuing SAGA of TFIID and RNA polymerase II transcription. *Mol Cell*, 2017, 68: 1–2
- 30 Grünberg S, Warfield L, Hahn S. Architecture of the RNA polymerase II preinitiation complex and mechanism of ATP-dependent promoter opening. *Nat Struct Mol Biol*, 2012, 19: 788–796
- 31 Wilson C J, Chao D M, Imbalzano A N, et al. RNA polymerase II holoenzyme contains SWI/SNF regulators involved in chromatin remodeling. *Cell*, 1996, 84: 235–244
- 32 Baek I, Friedman L J, Gelles J, et al. Single-molecule studies reveal branched pathways for activator-dependent assembly of RNA polymerase II pre-initiation complexes. *Mol Cell*, 2021, 81: 3576–3588.e6
- 33 Holstege F C, van der Vliet P C, Timmers H T. Opening of an RNA polymerase II promoter occurs in two distinct steps and requires the basal transcription factors IIE and IIH. *EMBO J*, 1996, 15: 1666–1677
- 34 Cai H, Luse D S. Transcription initiation by RNA polymerase II *in vitro*. Properties of preinitiation, initiation, and elongation complexes. *J Biol Chem*, 1987, 262: 298–304
- 35 Luse D S, Jacob G A. Abortive initiation by RNA polymerase II *in vitro* at the adenovirus 2 major late promoter. *J Biol Chem*, 1987, 262: 14990–14997
- 36 Luse D S, Kochel T, Kuempel E D, et al. Transcription initiation by RNA polymerase II *in vitro*. At least two nucleotides must be added to form a stable ternary complex. *J Biol Chem*, 1987, 262: 289–297
- 37 Ferguson H A, Kugel J F, Goodrich J A. Kinetic and mechanistic analysis of the RNA polymerase II transcription reaction at the human interleukin-2 promoter. *J Mol Biol*, 2001, 314: 993–1006
- 38 Kugel J F, Goodrich J A. Promoter escape limits the rate of RNA polymerase II transcription and is enhanced by TFIIE, TFIIH, and ATP on negatively supercoiled DNA. *Proc Natl Acad Sci USA*, 1998, 95: 9232–9237
- 39 Kugel J F, Goodrich J A. Translocation after synthesis of a four-nucleotide RNA commits RNA polymerase II to promoter escape. *Mol Cell Biol*, 2002, 22: 762–773
- 40 Kugel J F, Goodrich J A. A kinetic model for the early steps of RNA synthesis by human RNA polymerase II. *J Biol Chem*, 2000, 275: 40483–40491
- 41 Holstege F C P. Three transitions in the RNA polymerase II transcription complex during initiation. *EMBO J*, 1997, 16: 7468–7480
- 42 Hieb A R, Baran S, Goodrich J A, et al. An 8 nt RNA triggers a rate-limiting shift of RNA polymerase II complexes into elongation. *EMBO J*, 2006, 25: 3100–3109
- 43 Pal M, Ponticelli A S, Luse D S. The role of the transcription bubble and TFIIB in promoter clearance by RNA polymerase II. *Mol Cell*, 2005, 19: 101–110

- 44 Coppola J A, Luse D S. Purification and characterization of ternary complexes containing accurately initiated RNA polymerase II and less than 20 nucleotides of RNA. *J Mol Biol*, 1984, 178: 415–437
- 45 Pal M, Luse D S. The initiation–elongation transition: lateral mobility of RNA in RNA polymerase II complexes is greatly reduced at +8/+9 and absent by +23. *Proc Natl Acad Sci USA*, 2003, 100: 5700–5705
- 46 Ly E, Powell A E, Goodrich J A, et al. Release of human TFIIB from actively transcribing complexes is triggered upon synthesis of 7- and 9-nt RNAs. *J Mol Biol*, 2020, 432: 4049–4060
- 47 Zawel L, Kumar K P, Reinberg D. Recycling of the general transcription factors during RNA polymerase II transcription. *Genes Dev*, 1995, 9: 1479–1490
- 48 Kapanidis A N, Margeat E, Ho S O, et al. Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. *Science*, 2006, 314: 1144–1147
- 49 Chen X, Liu W, Wang Q, et al. Structural visualization of transcription initiation in action. *Science*, 2023, 382: eadi5120
- 50 Flanagan P M, Kelleher Iii R J, Sayre M H, et al. A mediator required for activation of RNA polymerase II transcription *in vitro*. *Nature*, 1991, 350: 436–438
- 51 Chen X, Yin X, Li J, et al. Structures of the human Mediator and Mediator-bound preinitiation complex. *Science*, 2021, 372: eabg0635
- 52 Struhl K, Segal E. Determinants of nucleosome positioning. *Nat Struct Mol Biol*, 2013, 20: 267–273
- 53 Lai W K M, Pugh B F. Understanding nucleosome dynamics and their links to gene expression and DNA replication. *Nat Rev Mol Cell Biol*, 2017, 18: 548–562
- 54 Jiang C, Pugh B F. Nucleosome positioning and gene regulation: advances through genomics. *Nat Rev Genet*, 2009, 10: 161–172
- 55 Kornberg R D, Lorch Y. Primary role of the nucleosome. *Mol Cell*, 2020, 79: 371–375
- 56 Weber C M, Ramachandran S, Henikoff S. Nucleosomes are context-specific, H2A.Z-modulated barriers to RNA polymerase. *Mol Cell*, 2014, 53: 819–830
- 57 Jin C, Zang C, Wei G, et al. H3.3/H2A.Z double variant-containing nucleosomes mark ‘nucleosome-free regions’ of active promoters and other regulatory regions. *Nat Genet*, 2009, 41: 941–945
- 58 Rhee H S, Bataille A R, Zhang L, et al. Subnucleosomal structures and nucleosome asymmetry across a genome. *Cell*, 2014, 159: 1377–1388
- 59 Schones D E, Cui K, Cuddapah S, et al. Dynamic regulation of nucleosome positioning in the human genome. *Cell*, 2008, 132: 887–898
- 60 Mavrich T N, Jiang C, Ioshikhes I P, et al. Nucleosome organization in the *Drosophila* genome. *Nature*, 2008, 453: 358–362
- 61 Tramantano M, Sun L, Au C, et al. Constitutive turnover of histone H2A.Z at yeast promoters requires the preinitiation complex. *eLife*, 2016, 5: e14243
- 62 Wang H, Schilbach S, Ninov M, et al. Structures of transcription preinitiation complex engaged with the +1 nucleosome. *Nat Struct Mol Biol*, 2023, 30: 226–232
- 63 Chen X, Wang X, Liu W, et al. Structures of +1 nucleosome-bound PIC-Mediator complex. *Science*, 2022, 378: 62–68
- 64 Žumer K, Maier K C, Farnung L, et al. Two distinct mechanisms of RNA polymerase II elongation stimulation *in vivo*. *Mol Cell*, 2021, 81: 3096–3109.e8

Summary for “真核生物RNA聚合酶II转录前起始复合物的结构研究进展”

Structural insights into RNA polymerase II transcription preinitiation complex

Xizi Chen^{*} & Yanhui Xu^{*}

Institutes of Biomedical Sciences, Fudan University Shanghai Cancer Center, and School of Life Sciences, Fudan University, Shanghai 200032, China

^{*} Corresponding authors, E-mail: xuyh@fudan.edu.cn; chenxz@fudan.edu.cn

RNA polymerase II (Pol II)-mediated transcription is a tightly regulated process central to gene expression, governing cellular differentiation, development, and disease. This review synthesizes recent structural advances in Pol II-driven transcription initiation, focusing on promoter recognition, stepwise assembly of the preinitiation complex (PIC), transition to elongation, and chromatin-mediated regulation.

Transcription initiation involves the assembly of a megadalton-scale PIC, comprising Pol II, general transcription factors (GTFs: TFIIA, TFIIB, TFIID, TFIIE, TFIIF, TFIIH), and coactivators such as the Mediator complex. Cryo-electron microscopy (cryo-EM) has resolved near-atomic structures of PICs at distinct assembly stages. TFIID, a 20-subunit complex containing TATA-box binding protein (TBP) and TBP-associated factors (TAFs), nucleates PIC assembly by binding diverse promoters, including TATA-less promoters. Structural studies reveal TFIID's modular architecture (TFIID-A/B/C), enabling flexible promoter engagement. Subsequent recruitment of Pol II and GTFs positions the promoter near Pol II's active site, with TFIIH's XPB translocase unwinding DNA and its CDK7 kinase phosphorylating Pol II's C-terminal domain (CTD) to initiate transcription. The transition from initiation to elongation involves dynamic conformational changes. Structural snapshots of transcribing complexes—open complex (OC), initial transcription complex (ITC), and early elongation complex (EEC)—reveal RNA-DNA hybrid stabilization, template-strand scrunching, and TFIIB displacement as critical steps.

Mediator, a 30-subunit coactivator, bridges sequence-specific transcription factors (TFs) and the PIC. Cryo-EM structures of PIC-Mediator demonstrate the Mediator's “CTD-gating” mechanism: its Head and Middle modules stabilize Pol II's CTD, facilitating phosphorylation by TFIIH, while coordinating enhancer-promoter communication. Mediator also stabilizes TFIIH within the PIC, ensuring efficient CTD phosphorylation. Notably, +1 nucleosomes downstream of the transcription start site (TSS) interact electrostatically with PICs, enhancing assembly and activation. Histone variants (e.g., H2A.Z) and modifications (e.g., H3K4me3, acetylation) at promoter-proximal nucleosomes further modulate PIC recruitment and activate transcription.

Unresolved questions persist, including the precise sequence of PIC assembly—whether TFIID initiates recruitment or preassembled Pol II-GTF modules integrate into the complex. Furthermore, how chromatin architecture, 3D genome looping, and epigenetic modifications spatiotemporally regulate initiation *in vivo* requires integration of cryo-electron tomography (cryo-ET), single-molecule imaging, and native chromatin templates. Recent studies using Widom 601 nucleosomes highlight +1 nucleosome-PIC interactions but emphasize the need for physiological models to recapitulate endogenous regulatory complexity.

In summary, structural breakthroughs have elucidated the mechanistic basis of Pol II transcription initiation, from PIC assembly to elongation transition. These insights provide a framework for understanding transcriptional regulation. Future interdisciplinary approaches combining high-resolution structural biology, live-cell imaging, and functional genomics will further unravel the spatiotemporal dynamics of this fundamental process.

transcription pre-initiation complex, mediator, transcription initiation, transcription regulation

doi: [10.1360/TB-2025-0011](https://doi.org/10.1360/TB-2025-0011)