

DOI: 10.13376/j.cblls/2025077

文章编号: 1004-0374(2025)07-0764-13



张余, 中国科学院分子植物科学卓越创新中心研究员。2004年毕业于复旦大学, 获学士学位; 2009年毕业于中国科学院上海药物研究所, 获博士学位; 2009—2015年, 在美国罗格斯大学从事博士后研究; 2015年8月起在现单位担任研究员。张余博士从结构生物学角度出发, 探索基因转录的工作机制和调控机制, 率领团队解析了 mRNA 的转录终止机制, 提出多种细菌转录调控模型, 解析了植物功能特化的 RNA 聚合酶的结构和工作机制。研究成果以通讯作者(含共同)在 *Cell*、*Science*、*Nature* 等杂志发表论文 20 余篇。

## RNA聚合酶的结构和转录机制

尤琳琳<sup>1,2#</sup>, 章洪伟<sup>1#</sup>, 张余<sup>1\*</sup>

(1 中国科学院分子植物科学卓越创新中心, 植物性状形成与塑造全国重点实验室, 上海 200032; 2 美国罗格斯大学瓦克斯曼微生物研究所, 新泽西州 皮斯卡塔维 08854)

**摘要:** 转录是 RNA 聚合酶将遗传信息从 DNA 传递到 RNA 的过程, 是中心法则的重要环节。转录主要分为起始、延伸和终止三个阶段。在转录起始阶段, RNA 聚合酶和转录起始因子协同识别启动子 DNA 序列或者染色质状态, 结合到基因组特定区域起始 RNA 合成。在转录延伸阶段, RNA 聚合酶在转录延伸因子的帮助下, 以 DNA 为模板合成 RNA。在转录终止阶段, RNA 聚合酶停止 RNA 合成, 解离 RNA 和 DNA。在转录的三个阶段中, RNA 聚合酶均受到多重控制从而精确控制基因表达。本文综述了模式生物的基因转录分子机制及其调控机制, 以及高等植物中功能特化的 RNA 聚合酶的特殊工作和调控机制。

**关键词:** 基因转录; RNA 聚合酶; 转录调控

**中图分类号:** Q753 **文献标志码:** A

## Structure, transcription mechanism, and regulation of RNA polymerases

YOU Lin-Lin<sup>1,2#</sup>, ZHANG Hong-Wei<sup>1#</sup>, ZHANG Yu<sup>1\*</sup>

(1 State Key Laboratory of Plant Trait Design, CAS Center for Excellence in Molecular Plant Sciences, Chinese Academy of Sciences, Shanghai 200032, China; 2 Waksman Institute of Microbiology, Rutgers University, Piscataway, NJ 08854, USA)

**Abstract:** Transcription is the process by which an RNA polymerase (RNAP) transcribes genetic information from DNA to RNA. The transcription process can be divided into three stages: initiation, elongation, and termination. During transcription initiation, RNAPs and transcription initiation factors cooperatively recognize promoter DNA sequences or chromatin states and initiate RNA synthesis at certain genomic regions. In the elongation stage, RNAPs extend RNA using the genomic DNA as template, with assistance from various elongation factors. In the

收稿日期: 2025-06-20

基金项目: 国家自然科学基金杰出青年基金项目(32425006); 上海市基础研究特区计划项目(JCYJ-SHFFY-2022-012)

<sup>#</sup>共同第一作者

\*通信作者: E-mail: yzhang@cemps.ac.cn

termination stage, RNAPs stop RNA synthesis, release the transcripts, and dissociate from genomic DNA. Across all three stages, RNAPs are subject to multiple layers of regulation for precisely controlled gene expression. This review summarizes the molecular mechanisms of gene transcription in model organisms, with a particular focus on the specialized functions and regulatory mechanisms of RNA polymerases in higher plants.

**Key words:** gene transcription; RNA polymerase; transcription regulation

细胞内执行基因转录的RNA聚合酶(RNA polymerase, RNAP)根据亚基组成与功能特性,可分为单亚基RNAP和多亚基RNAP两大类。单亚基RNAP属于噬菌体类型RNAP,主要在真核细胞的细胞器(动植物细胞线粒体和植物前质体)中负责基因转录,这些单亚基RNAP由细胞核基因组编码,在细胞质中翻译后被运输到细胞器中发挥功能。多亚基RNAP包括细菌RNAP、古菌RNAP,真核生物的RNA Pol I、II、III、IV、V<sup>[1-3]</sup>(其中RNA Pol IV/V为植物所特有),以及植物质体基因组编码的RNAP(plastid-encoded polymerase, PEP)<sup>[3-5]</sup>。多亚基RNAP具有相对保守的催化核心,而古菌和真核生物的RNAP具有高度特化的多个外周亚基,赋予了它们不同的功能和调控机制,以适应其基因表达调控的复杂需求。下文将详细介绍多亚基RNAP结构的共性与特性,以及其转录与调控机制。

## 1 多亚基RNAP的三维结构

### 1.1 细菌RNAP三维结构

细菌RNAP于20世纪60年代被发现,是研究最为深入的多亚基RNAP<sup>[6]</sup>。大肠杆菌RNAP核心酶(RNAP core enzyme)由五个亚基组成,包括两个 $\alpha$ 亚基( $\alpha_2$ )、一个 $\beta$ 亚基、一个 $\beta'$ 亚基和一个 $\omega$ 亚基(图1和表1)。两个 $\alpha$ 亚基为装配亚基,帮助核心酶组装,并参与转录调控; $\beta$ 与 $\beta'$ 为催化亚基,形成催化中心,负责RNA合成; $\omega$ 亚基为辅助亚基,主要结合在 $\beta'$ 亚基,稳定RNAP构象,并在转录过程中发挥调控作用<sup>[7]</sup>。细菌RNAP的整体形状像一个蟹爪, $\beta$ 和 $\beta'$ 组成蟹爪的两个钳子,内部形成活性中心DNA通道,用来加载DNA。除了活性中心通道外,RNAP还含有两个额外通道,即NTP进入活性中心的二级通道(secondary channel),和新生RNA链退出的RNA退出通道(RNA exit channel)<sup>[7]</sup>。尽管RNAP核心酶具备合成RNA的催化活性,但其本身无法在基因特定位点起始转录。只有当RNAP核心酶与 $\sigma$ 因子( $\sigma$  factor)结合形成RNAP全酶(RNAP holoenzyme)后,才能特异性识别启动子DNA,并正确起始转录<sup>[8-10]</sup>。

尽管RNAP的核心亚基高度保守,但在不同种属中,RNAP的亚基仍存在一些差异。在进化上较为古老的蓝细菌中,RNAP的最大亚基 $\beta'$ 被拆分为两个亚基。蓝细菌RNAP结构表明,这种拆分并不改变其三维结构,拆分之后的两个亚基仍能正确组装成一个完整的 $\beta'$ 亚基发挥功能<sup>[11]</sup>。此外,蓝细菌RNAP具有约630个氨基酸的插入序列(sequence insertion 3, SI3),约占全酶氨基酸总数的五分之一。该结构域自NTP通道延伸至DNA通道,在转录起始及延伸阶段发挥关键调控作用<sup>[11-12]</sup>。在革兰氏阳性菌枯草芽孢杆菌中,除核心的五个保守亚基( $\alpha_2\beta\beta'\omega$ )外,RNAP还含有两个额外小亚基,分别为 $\delta$ 和 $\epsilon$ 。 $\delta$ 亚基在促进RNAP转录循环等方面发挥功能<sup>[13-14]</sup>,而 $\epsilon$ 亚基可能在促进RNAP装配中发挥功能<sup>[15]</sup>。

### 1.2 叶绿体PEP的三维结构

叶绿体是植物和藻类执行光合作用的关键细胞器,其起源可追溯至约15亿年前的蓝细菌内共生事件<sup>[16]</sup>。叶绿体基因组在内共生进化过程中大规模向核基因组转移或丢失,现今大多数陆地植物的叶绿体基因组仅保留110~130个基因,主要编码基因转录、蛋白质翻译、蛋白质降解、光合作用的蛋白质机器的核心组分<sup>[17]</sup>。叶绿体PEP在叶绿体从原质体向成熟叶绿体的分化过程以及分化成熟的叶绿体中承担主要的基因转录任务<sup>[4]</sup>。尽管叶绿体基因组在进化过程中大幅缩减,但其转录机器PEP却趋于复杂化。早前研究暗示PEP是由叶绿体基因组编码的核心酶亚基与十余种核基因组编码的PEP相关蛋白(PEP-associated proteins, PAPs)共同组成的分子量约为1 MDa的超大复合物。近期本实验室和其他实验室相继报道了PEP的高分辨率冷冻电镜结构<sup>[5, 18-20]</sup>,结构表明PEP的核心由蓝细菌RNAP来源的亚基构成,形成“催化模块”,负责RNA的合成;而15种PAPs围绕核心模块组装成多个功能性模块,包括“支架模块”“保护模块”“DNA/RNA模块”“调控模块”,形成自然界已知最复杂的基因转录机器<sup>[5]</sup>(图1)。

“支架模块”由PAP1、PAP3、PAP5、PAP7、



表1 多亚基RNAP的组成

| 细菌RNAP          | 叶绿体PEP <sup>#</sup> | 古菌RNAP      | Pol I | Pol II | Pol III | Pol IV | Pol V |
|-----------------|---------------------|-------------|-------|--------|---------|--------|-------|
| α <sup>I</sup>  | α <sup>I</sup>      | Rpo3 (D)    | RPAC1 | RPB3   | RPAC1   | RPB3   | RPB3  |
| α <sup>II</sup> | α <sup>II</sup>     | Rpo11 (L)   | RPAC2 | RPB11  | RPAC2   | RPB11  | RPB11 |
|                 |                     | Rpo10 (N)   | RPB10 | RPB10  | RPB10   | RPB10  | RPB10 |
|                 |                     | Rpo12 (P)   | RPB12 | RPB12  | RPB12   | RPB12  | RPB12 |
| β′              | β′1+β′2             | Rpo1′+Rpo1″ | A194  | RPB1   | RPC1    | NRPD1  | NRPE1 |
| β               | β                   | Rpo2        | A135  | RPB2   | RPC2    | NRPD2  | NRPD2 |
| ω               | ω(PAP12)            | Rpo6        | RPB6  | RPB6   | RPB6    | RPB6   | RPB6  |
|                 | PAP1                | Rpo5 (H)    | RPB5  | RPB5   | RPB5    | RPB5   | NRPE5 |
|                 | PAP2                | Rpo8 (G)*   | RPB8  | RPB8   | RPB8    | RPB8   | RPB8  |
|                 | PAP3                |             | RPA12 | RPB9   | RPC10   | NRPD9  | RPB9  |
|                 | PAP4                | Rpo4 (F)    | RPA43 | RPB4   | RPC9    | NRPD4  | NRPD4 |
|                 | PAP5                | Rpo7 (E)    |       | RPB7   | RPC8    | NRPD7  | NRPE7 |
|                 | PAP6                |             | RPA49 |        | RPC5    |        |       |
|                 | PAP7                |             | RPA34 |        | RPC4    |        |       |
|                 | PAP8                |             |       |        | RPC3    |        |       |
|                 | PAP9                |             |       |        | RPC6    |        |       |
|                 | PAP10 <sup>I</sup>  |             |       |        | RPC7    |        |       |
|                 | PAP10 <sup>II</sup> |             |       |        |         | RDR2   |       |
|                 | PAP11               | Rpo13*      |       |        |         |        |       |
|                 | PAP13               |             |       |        |         |        |       |
|                 | PAP14               |             |       |        |         |        |       |
|                 | PAP15               |             |       |        |         |        |       |

注：\*，表示只存在某些特殊物种中；<sup>#</sup>，表示PEP的PAP1-15 (除了PAP12外)和其他RNA聚合酶亚基不具有同源性。

和 RPB12)，在 RNAP 组装中发挥核心作用；Rpo1′、Rpo1″、Rpo2 共同构成催化模块（对应于 Pol II 的 RPB1 和 RPB2），负责 RNA 合成；Rpo4 和 Rpo7 形成 Stalk 结构域（对应 Pol II 的 RPB4 和 RPB7）；Rpo5 和 Rpo6（对应 Pol II 的 RPB5 和 RPB6）以及在某些物种中存在的 Rpo8 和 Rpo13 作为辅助亚基，参与 RNAP 的组装及活性调控<sup>[25]</sup>。由于古菌基因转录系统兼具真核 RNAP 结构特征和细菌基因调控特点，其独特性不仅使古菌成为研究基因转录调控的重要模型，也为深入理解真核生物的起源提供了重要的研究对象。

1.4 真核生物Pol I~V的三维结构

细菌和古菌中只存在单一的 RNAP，而大多数的真核生物细胞核中存在着三种 RNAP：Pol I、Pol II 和 Pol III<sup>[25-27]</sup>。Pol I 负责合成核糖体 RNA (ribosome RNA, rRNA) 前体，并进一步加工为 28S、5.8S 和 18S rRNA；Pol II 负责合成信使 RNA 前体 (messenger RNA, mRNA) 以及非编码 RNA (non-coding RNA, ncRNA)，如 miRNA、snRNA 和小 snoRNA 等；Pol III 则负责合成短链非编码 RNA，如转运 RNA (transfer RNA, tRNA)、5S rRNA 以及 U6 小核 RNA 等。

Pol I、II 和 III 均展现出保守的蟹爪型核心结构<sup>[27]</sup>，并围绕这一核心结构搭载了数量不等的特异性亚基（图 1 和表 1）。其中，人类细胞的 Pol I 由 13 个亚基组成，Pol II 由 12 个亚基组成，Pol III 由 17 个亚基组成<sup>[2]</sup>。三种 RNAP 的核心结构均由 10 个与古菌 RNAP 同源的亚基构成<sup>[27]</sup>，包括两个最大的催化亚基 (Pol I 中为 RPA194/RPA135，Pol II 中为 RPB1/RPB2，Pol III 中为 RPC1/RPC2)、一个参与校对的亚基 (Pol I 中为 RPA12，Pol II 中为 RPB9，Pol III 中为 RPC10)、五个共有亚基 (RPAC1/2/3/4/5 或 RPB5/6/8/10/12)，以及 Pol I/III 之间共用的 RPAC1/RPAC2 (Pol II 中为 RPB3/RPB11)<sup>[28-30]</sup>。值得注意的是，三种 RNAP 的最大亚基的 C 末端结构域 (C-terminal domain, CTD) 具有较大区别。Pol II 最大亚基 RPB1 的 CTD 由保守的七肽重复序列 Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7 组成<sup>[31]</sup>，重复数量从酵母的 26 个到人源的 52 个不等<sup>[32]</sup>。七肽序列上的氨基酸在转录过程中发生频繁的磷酸化与去磷酸化，调控 RNAP 的转录活性<sup>[33-34]</sup>。除核心亚基外，其他亚基对于相应的 RNAP 功能也非常重要。Pol II 的 stalk 模块包含 RPB4 和 RPB7 两个亚基（对应

Pol I 的 RPA43 亚基和 Pol III 的 RPC9/RPC8 亚基), 其作为转录起始、延伸和终止因子的重要结合位点在整个转录过程中发挥关键作用<sup>[27, 30]</sup>。Pol I 和 III 的部分亚基具有 Pol II 通用转录因子类似的结构域, 如 Pol I 的 RPA49/RPA34 亚基 (Pol III 中为 RPC5/RPC4), 与 Pol II 的转录因子 TFIIF ( $\alpha/\beta$  二聚体) 结构相似<sup>[35-36]</sup>, Pol III 的专属亚基 RPC3/RPC6 与 Pol II 的转录因子 TFIIE $\alpha/\beta$  结构相似。Pol I 的 RPA12 亚基和 Pol III 的 RPC10 亚基高度同源, 由 Pol II 的 RPB9 亚基与其转录延伸因子 TFIIS 融合而成, 在转录过程中具有校正功能, 确保了延伸的高效和准确。以上结果暗示, Pol I、Pol III 在演化过程中将 Pol II 的通用转录因子整合为稳定亚基, 实现其功能特化。

与其他真核生物不同的是, 在高等植物中还存在着两类 RNAP, 分别为 Pol IV 和 Pol V<sup>[37]</sup>, 其主要功能是合成非编码 RNA, 通过 RNA 介导的 DNA 甲基化 (RNA-directed DNA methylation, RdDM) 途径建立和维持基因组 DNA 甲基化, 沉默基因组的转座子活性。高等植物基因组含有大量的重复序列和转座子, 如玉米和小麦基因组中的转座子含量高达 80%<sup>[38]</sup>, 通过 RdDM 途径建立的 DNA 甲基化表观遗传修饰使转座子处于沉默状态, 维持基因组的稳定<sup>[39]</sup>。RdDM 通路主要分为两个步骤: (1) Pol IV 和 RNA 依赖型聚合酶 2 (RNA-dependent RNA polymerase 2, RDR2) 协作在基因组特定区域 (如转座子) 合成双链 RNA, 随后该双链 RNA 被 DCL3 和 HEN1 加工修饰生成成熟的 24 nt 小干扰 RNA (small interfering RNA, siRNA), 并被 AGONAUTE4/6 (AGO4/6) 进一步切割保留单链; (2) Pol V 在相同区域转录 RNA, 而携带 24 nt siRNA 的 AGO4/6 蛋白通过两层相互作用定位到 Pol V 的转录区域, 第一层相互作用是 AGO4/6 蛋白和 Pol V 之间的蛋白质相互作用, 第二层相互作用是 AGO4/6 蛋白携带的 24 nt siRNA 与 Pol V 的转录本发生碱基配对。两层相互作用的建立能够招募甲基转移酶 DRM2 对该基因组区域进行甲基化<sup>[3, 39]</sup>。

研究表明 Pol IV 和 Pol V 由 Pol II 演化而来, 它们与 Pol II 的亚基组成相似, 与 Pol II 共享部分亚基 (RPB3/6/8/9/10/11/12)<sup>[40-41]</sup>。但是, Pol IV 和 Pol V 具有自己专属的亚基 (Pol IV 的专属亚基是 NRPD1/5/7 和 RDR2; Pol V 的专属亚基是 NRPE1/5/7) (图 1 和表 1), 这些亚基由 Pol II 相应亚基的编码基因复制进化而来<sup>[42-45]</sup>, 赋予了 Pol IV 和 Pol V 在细胞内的

特殊功能。

## 2 细菌、古菌转录起始和终止机制

### 2.1 细菌RNAP转录起始机制

转录起始是转录的第一步, 也是最复杂的步骤。细菌 RNAP 核心酶需要与  $\sigma$  因子形成全酶才能精确起始基因转录<sup>[8-10]</sup>。细菌含有多种  $\sigma$  因子, 不同类型的  $\sigma$  因子识别不同启动子序列, 负责各自基因转录。根据结构和机制差异, 细菌  $\sigma$  因子可分为两大家族:  $\sigma^{70}$  家族和  $\sigma^{54}$  家族。 $\sigma^{70}$  家族成员较多, 其包含细菌看家  $\sigma$  因子 (如大肠杆菌  $\sigma^{70}$ ), 负责大部分必需基因转录, 还包含数量最多的 ECF  $\sigma$  因子 (extra-cytoplasmic  $\sigma$  factor), 响应特定环境调控基因表达<sup>[46-47]</sup>。 $\sigma^{54}$  家族成员较少, 大多数细菌只编码 1 个  $\sigma^{54}$  因子, 其主要负责氮代谢、运动、生物膜形成等特定生理过程的基因转录。

RNAP 与  $\sigma$  因子形成 RNAP- $\sigma$  全酶才具备识别基因启动子的能力。RNAP- $\sigma^{70}$  识别的启动子包含两个核心元件: -35 区 (5'-TTGACA-3') 和 -10 区 (5'-TATAAT-3'), -35 和 -10 区之间由 16~19 bp 的间隔区连接, 其中最佳间隔为 17 bp<sup>[8]</sup>。此外, 部分启动子还包含额外调控元件, 如 UP 元件 (UP element: -40~-60 区)、-10 延长区 (extended-10 element, EXT)、鉴别区 (discriminator element, DISC) 以及核心酶识别区 (core recognition element, CRE), 这些序列在不同程度上影响转录效率<sup>[8, 10]</sup>。RNAP- $\sigma^{54}$  识别的启动子核心元件则包括 -24 区 (5'-TGGCAC-3') 和 -12 区 (5'-TTGCT-3')<sup>[48-49]</sup>。

RNAP- $\sigma^{70}$  的转录起始过程主要分为以下几个步骤: (1) RNAP-启动子 DNA 闭合复合物 (RNAP-promoter DNA close complex, RPC): RNAP- $\sigma^{70}$  识别 UP 元件或 -35 区与启动子区形成闭合复合物<sup>[50]</sup> (图 2A); (2) RNAP-启动子 DNA 中间复合物 (RNAP-promoter DNA intermediate complexes, RPi):  $\sigma^{70}$  结构域特异性识别启动子 -10 区, 并逐步解链该区域, 其中非模板链的 -11A 和 -7T 结合在相应的蛋白口袋中<sup>[50-51]</sup>; (3) RNAP-启动子 DNA 开放复合物 (RNAP-promoter open complex, RPo): RNAP 将转录泡完全打开, 将单链模板链 DNA 置于催化中心, 形成稳定的开链复合物<sup>[50-53]</sup> (图 2B); (4) RNAP-启动子 DNA 转录起始复合物 (RNAP-promoter DNA initial transcription complexes, RPitc): RNAP 起始 RNA 合成, 产生 productive product 或 abortive product; (5) 启动子逃逸 (promoter escape): 当新生 RNA 延长至

一定长度(10~15 nt)后, RNAP 解离启动子 DNA 往前行进, 进入转录延伸阶段<sup>[54-55]</sup>。RNAP- $\sigma^{\text{ECF}}$  全酶与 RNAP- $\sigma^{70}$  全酶结构类似, 且 RNAP- $\sigma^{\text{ECF}}$  采用与 RNAP- $\sigma^{70}$  类似的方式识别启动子 DNA, 但是  $\sigma^{70}$  和  $\sigma^{\text{ECF}}$  识别 -35 和 -10 区的氨基酸各不相同, 因此二者能够识别不同序列的启动子 DNA<sup>[56-58]</sup>。

$\sigma^{54}$  与  $\sigma^{70}$  在蛋白质序列上没有相似性。 $\sigma^{54}$  由 3 个区段组成, 区段 I 由两个 helix 组成, 参与 DNA 解链和转录泡形成; 区段 II 同样由两个 helix 组成, 主要结合在 RNA 聚合酶内部; 区段 III 包含 CBD (core binding domain)、ELH-HTH 和 RpoN 三个结构域, 其中 ELH-HTH 和 RpoN 分别负责识别 -12 区和 -24 区<sup>[48-49, 59]</sup>。不同于 RNAP- $\sigma^{70}$  依赖型转录起始, RNAP- $\sigma^{54}$  结合启动子 DNA 形成闭合复合物后, 无法自发形成开放复合物 (RPO), 而是需要远离启动子位点的 ATP 依赖型激活蛋白, 也称为细菌增强子结合蛋白 (bEBPs), 解开双链 DNA 起始 RNA 合成<sup>[49, 60-61]</sup>。

## 2.2 细菌RNAP转录起始调控

细菌主要通过  $\sigma$  因子和转录因子两个层面来调控转录起始。平均每种细菌编码大约 10 个  $\sigma$  因子, 不同  $\sigma$  因子识别各异的启动子 DNA 核心序列调控基因表达。每种细菌还编码大量转录因子, 例如大肠杆菌编码约 300 个转录因子 (占总基因数量 6% 左右)。大多数转录因子含有 DNA 结合结构域 (DNA-binding domain, DBD) 和调控结构域 (regulatory domain, RD), 其通过特异性结合启动子, 从而抑制或者促进 RNA 聚合酶与启动子 DNA 结合, 实现转录抑制或者激活<sup>[62]</sup>。

细菌转录因子的激活方式复杂多样。根据作用机制, 其可分为经典转录激活机制和非经典转录激活机制。经典转录激活机制按照转录因子结合在 DNA 的位置以及与 RNAP- $\sigma$  的相互作用方式可分为两类<sup>[63-65]</sup>。第一类 (Class I) 转录因子结合在启动子 DNA-35 区上游, 通过与 RNAP- $\alpha$ CTD (RNAP- $\alpha$  C-terminal domain) 相互作用招募 RNAP, 促进转录起始<sup>[66]</sup>。第二类 (Class II) 转录因子结合位置与 -35 区部分重叠, 其通过与 RNAP- $\alpha$ CTD 和  $\sigma^{70}$  共同相互作用, 促进转录起始<sup>[67]</sup>。

除了经典转录激活机制外, 近年来多种非经典转录激活机制被报道。(1) 改变 DNA 构象的转录激活。MerR 家族转录因子在结合配体后, 以二聚体形式结合在 -35 区与 -10 区之间, 并通过弯曲和扭转 DNA, 将原本约 19 bp 的间隔优化为最佳的 17 bp, 从而促进 RNAP 的转录<sup>[15, 68-70]</sup>。(2) 不依赖于

DNA 相互作用的转录激活。这类转录激活因子包括 Crl、SutA、DksA、TraR 和 MglA-SspA 等<sup>[65, 71]</sup>。上述转录因子通过稳定 RNAP 转录各阶段结构, 激活基因转录。调控  $\sigma^{54}$  家族的转录因子也称为细菌增强子结合蛋白, 其结合在核心启动子上游区域, 在宿主整合因子 (integration host factor, IHF) 帮助下, 利用 ATP 水解酶活性, 重塑  $\sigma^{54}$ 、打开转录泡等<sup>[49, 60-61]</sup>。

## 2.3 细菌RNAP转录终止

转录终止是 RNAP 停止 RNA 延伸, 释放 RNA, 并从 DNA 上解离的过程<sup>[72]</sup>。细菌基因通过两种方式实现转录终止<sup>[72-74]</sup>, 一种是固有转录终止 (intrinsic termination) 或自发转录终止, 也称不依赖  $\rho$  因子的转录终止; 另一种是依赖于  $\rho$  因子的转录终止 ( $\rho$ -dependent termination) (图 2C 和 D)。除此之外, 还存在 Mfd 蛋白介导的转录终止以及近期发现的 RNAP 间碰撞介导的转录终止等其他形式<sup>[72-73, 75-78]</sup>。

固有转录终止是 RNAP 转录至固有终止子后自发终止转录的方式, 是在所有细菌中均保守的转录终止方式<sup>[79]</sup>。固有转录终止子由一段富含 G/C 的发卡序列和紧跟其后的 polyT 序列组成。RNAP 在 polyT 的第 7 或第 8 位发生终止<sup>[80-81]</sup>。固有转录终止分为 4 个步骤<sup>[82]</sup>。(1) polyU 诱导转录暂停: RNAP 转录至终止位点时, RNA 链的 polyU 序列诱导 RNA-DNA 杂合双链在活性中心处于半移位非活性状态 (即 RNA 链发生移位, 而模板 DNA 链未发生移位), 抑制 NTP 添加, 从而暂停 RNA 延伸; (2) RNA 发卡折叠: 转录暂停为 RNA 发夹折叠提供时间窗口, RNA 发卡折叠进入 RNA 退出通道, 诱导 RNA 聚合酶构象变化, 诱发转录泡闭合; (3) RNA-DNA 杂合双链被破坏: RNA 发卡进一步折叠, 破坏 RNAP 催化中心 RNA-DNA 杂合双链, 诱导 RNA 解离; (4) RNAP 解离 DNA: RNAP 首先解离 RNA, 再从基因组 DNA 上脱离。部分细菌编码促转录终止因子帮助 RNAP 从基因组解离<sup>[82-84]</sup>。

依赖  $\rho$  因子的转录终止是 RNAP 在  $\rho$  因子的帮助下终止转录的方式。 $\rho$  因子通过结合富含嘧啶的 rut (rho utilization, rut) RNA, 并利用其 ATP 依赖的 5'→3' RNA 转位酶活性, 促使 RNAP 终止转录。 $\rho$  因子通过六聚体发挥功能<sup>[85]</sup>。 $\rho$  因子的转录终止过程大概分为 4 步<sup>[86]</sup>: (1) 处于开环状态的  $\rho$  因子识别并结合富含嘧啶的 rut RNA; (2)  $\rho$  因子与 rut RNA 和 ATP 的结合诱导  $\rho$  闭环, 激活  $\rho$  催化活性<sup>[87]</sup>; (3)  $\rho$  因子水解 ATP, 沿 5'→3' 方向向 RNAP 移动; (4)  $\rho$  因子利用其转位酶活性, 破坏 RNAP 活性中

心 RNA-DNA 杂合双链,使 RNA 从 RNAP 中解离。在上述过程中,转录延伸因子 NusG 通过 N 端与 RNAP 结合、C 端与  $\rho$  因子结合发挥桥梁作用,连接 RNAP 和  $\rho$  因子。在上述经典观点之外,一些研究提出  $\rho$  因子以转录延伸复合物为中心进行转录终止的新观点:在转录延伸过程中, $\rho$  因子与转录延伸因子 NusA 和 NusG 协同作用,采取一种“搭便车”的方式先与 RNAP 结合,再捕获 RNA,进而诱导 RNAP 发生一系列构象变化,进入失活状态,最终终止转录<sup>[13, 88-90]</sup>。

#### 2.4 噬菌体蛋白抑制细菌RNAP转录终止

噬菌体依赖宿主细菌来繁殖后代,其通常劫持宿主的 RNAP 转录自身基因,并抑制宿主基因转录<sup>[91]</sup>。此外,部分噬菌体编码抗终止蛋白,拮抗宿主 RNAP 转录终止,帮助噬菌体在合适的时间切换其基因组的早期基因到晚期基因的表达。

$\lambda$  噬菌体 N 蛋白是第一个被报道的抗转录终止因子,开启噬菌体早期基因到中期基因的表达<sup>[92]</sup>。N 蛋白识别 RNA 上的 nut (N-utilization, nut) 位点并结合转录延伸复合物,其活性需借助宿主 Nus 因子 (NusA、NusB、NusE 和 NusG)。N 蛋白与 Nus 因子形成复合物结合 RNAP,阻止 RNA 发卡在 RNA 退出通道中形成,并稳定 RNA-DNA 杂合双链,抑制固有转录终止。NusE 与 NusG 的 C 端结构域结合,阻断  $\rho$  因子与 RNAP 结合,从而抑制依赖  $\rho$  因子的转录终止<sup>[92]</sup>。 $\lambda$  和 21 噬菌体的 Q 蛋白开启从中期基因向后期基因的转录。Q 蛋白和 RNAP 的复合物结构表明,Q 蛋白在 RNA 出口通道形成环状结构,阻止 RNA 发卡生成,抑制终止<sup>[93-94]</sup>。Xp10 噬菌体的 P7 蛋白是一种双功能调控因子,既可抑制宿主转录,又可促进噬菌体后期基因表达<sup>[95]</sup>。P7 与 RNAP 的复合物结构表明,P7 结合在 RNAP 的 RNA 退出通道,抑制 RNA 发卡在 RNA 退出通道形成,从而发挥抗终止作用<sup>[96]</sup>。

综上,这些噬菌体抗转录终止蛋白通过保守的机制拮抗转录终止,即结合在 RNAP 的 RNA 退出通道附近,阻止 RNA 发卡在 RNA 退出通道折叠,或者阻止终止蛋白  $\rho$  因子结合 RNAP,从而抑制转录终止,开启噬菌体中后期基因的表达。

#### 2.5 古菌RNAP转录起始机制

细菌的 RNA 聚合酶启动转录时需要依赖  $\sigma$  因子,而古菌的 RNAP 则需要借助三种通用转录起始因子:TBP (TATA-binding protein, TBP)、TFB (transcription factor B, TFB) 和 TFE (transcription factor E, TFE),这

三种因子分别与真核生物中的 TBP、TFIIB 和 TFIIE 具有同源性。古菌的启动子主要包括 TATA box 和 BRE 元件 (B recognition element),其中 TATA box 通常位于转录起始位点上游 -25 bp 位置,是 TBP 的特异性结合位点,而 BRE 元件是 TFB 特异性结合位点。在转录起始过程中,TBP 首先识别并结合 TATA box,使局部 DNA 发生弯曲,随后 TFB 结合 BRE 元件,与 TBP 形成稳定的复合物,并招募 RNAP,形成最小的 DNA-TBP-TFB-RNAP 转录起始前复合物。该复合物随后招募 TFE,促进双链 DNA 的解链,形成转录起始开放复合物,起始 RNA 合成<sup>[22, 97-98]</sup>。由于缺乏完整的转录起始复合物三维结构,古菌的转录起始机制尚未阐明。

#### 2.6 古菌RNAP转录终止机制

古菌与细菌类似,主要存在两种转录终止途径:固有转录终止和 FttA 依赖型转录终止,此外还有 Eta 蛋白介导的转录终止。固有转录终止子的特征是非模板链中富含 5~10 个 T 的序列<sup>[98]</sup>。古菌转录终止因子 FttA (factor terminates transcription in Archaea) 也称作 aCPSF1 (archaeal cleavage and polyadenylation specificity factor 1),负责古菌中大多数转录单元的转录终止。FttA 特异性识别固有转录终止子的 polyU 序列<sup>[99]</sup>,此外 FttA 具有核糖核酸内切酶活性和 5'→3' 核糖核酸外切酶活性<sup>[100-101]</sup>。转录延伸因子 Spt5 能够促进 FttA 的转录终止活性<sup>[100]</sup>。FttA 转录终止复合物显示:FttA 结合在 RNA 退出通道外侧,RNA 从 RNAP 的 RNA 退出通道进入 FttA 的活性切割中心;Spt5 的 N 端结构域结合 RNAP,C 端结构域结合 FttA,在 RNAP 和 FttA 之间发挥桥梁的作用。FttA 首先利用其核糖核酸内切酶活性切割 RNA,然后利用 5'→3' 核糖核酸外切酶活性进行 RNA 5'→3' 切割和转位,最终破坏转录延伸复合物,引发转录终止<sup>[102]</sup>。

### 3 叶绿体PEP的转录起始与终止机制

PEP 的转录起始依赖于保守的启动子序列,典型启动子通常包含位于 -35 和 -10 区域的两个顺式元件,由 PEP- $\sigma$  全酶识别。植物细胞核通常编码多个  $\sigma$  因子,例如拟南芥编码 6 个  $\sigma$  因子 (SIG1~6),负责在不同的发育阶段或者响应不同环境信号调控 PEP 转录活性。当前,PEP 转录起始的具体分子机制仍需深入研究<sup>[4]</sup>。

PEP 的转录终止机制也尚未阐明。部分叶绿体基因 3' 端存在可形成发卡结构的反向重复序列,但

其主要功能可能在于稳定转录本、避免 RNA 降解, 而非直接介导转录终止<sup>[103]</sup>。PEP 是否通过类似细菌的固有转录终止机制终止转录有待进一步阐明。在拟南芥中, 有研究表明 RHON1 参与 *rbcl* 与其下游基因 *accD* 之间的转录终止, 然而具体机制未知<sup>[104]</sup>。

## 4 Pol II 转录起始和终止机制

### 4.1 Pol II 转录起始机制

真核生物的转录起始尤为复杂, Pol II 需要在通用转录起始因子的帮助下才能识别基因启动子区域, 并组装成转录起始前复合物 (preinitiation complex, PIC) 起始转录<sup>[105-108]</sup>。真核生物基因的核心启动子是一段以转录起始位点 (transcription start site, TSS) 为中心, 上下游各长约 50 bp 的 DNA 序列。通用转录起始因子包括 TFIID、TFIIA、TFIIB、TFIIF、TFIIE 和 TFIIH<sup>[109]</sup>。通用转录起始因子与启动子 DNA 和 Pol II 的结合存在时顺性, 首先 TFIID 识别并结合至核心启动子, 随后招募 TFIIA、TFIIB、TFIIF 与 Pol II 形成转录前核心起始复合物 (core preinitiation complex, cPIC), 接着招募 TFIIE 形成转录前中间起始复合物 (intermediate preinitiation complex, mPIC), 最后招募 TFIIH 形成转录前完整起始复合物 (holo preinitiation complex, hPIC) (图 2E)。Pol II 转录起始活性能够被共激活因子 Mediator 进一步激活。Mediator 促进 PIC 复合物组装的同时能够激活 TFIIH 的 CDK7 激酶活性<sup>[110]</sup>, 对 Pol II CTD 的 Ser5 进行磷酸化修饰, 促进转录起始向转录延伸阶段转换。当组装成 hPIC 后, TFIIH 利用 DNA 移位酶的活性将 DNA 泵入 Pol II 催化中心, 并同时促进其解链形成转录泡。Pol II 起始 RNA 合成之后, 随着 RNA 延长, 不断增大转录泡的单链 DNA 以“蜷缩” (Scrunch) 机制在催化中心积累挤压, TFIID 的 DNA 移位酶活性进一步将下游 DNA 输入催化中心。当 RNA 长度到达 10 nt 之后, 上述“蜷缩”过程积累的能量帮助 Pol II 挣脱转录起始因子, 进入转录延伸阶段<sup>[107]</sup>。

### 4.2 Pol II 转录终止机制

Pol II 针对不同类型的 RNA 采用不同的转录终止机制<sup>[111]</sup>: Pol II 利用切割和多聚腺苷酸因子 (cleavage and polyadenylation factor, CPF) 和切割因子 (cleavage factor, CF) 途径 (CPF-CF 途径) 终止 mRNA 转录; 在单细胞真核生物酿酒酵母中, Pol II 利用 Nrd1-Nab3-Sen1 途径 (NNS 途径) 终止 ncRNA 转录<sup>[2]</sup>, 在后生动物细胞中, Pol II 利用 Integrator 复

合物终止 ncRNA 转录 (图 2F~H)。

CPF-CF 复合物识别富含腺嘌呤的多聚腺嘌呤位点 (polyA site), 并在下游区域切割、释放新生成的 mRNA 前体, 随后 mRNA 前体添加 polyA 尾巴后被输送至细胞质中。Pol II 向前行进一小段距离后停止转录, Pol II 从基因组上解离, 标志转录终止。目前关于 Pol II 终止 mRNA 转录的机制存在着三种模型: 鱼雷模型 (torpedo model)、异构模型 (allosteric model) 和整合模型 (sitting-duck torpedo model)<sup>[2]</sup>。鱼雷模型认为 CPF-CF 复合物切割 mRNA 后, 残留 RNA 的 5' 端是 5'→3' 核酸外切酶 (酿酒酵母 Rat1、人 XRN2、拟南芥 XRN3) 的识别和结合位点。当核酸外切酶结合 RNA 后, 在切割 RNA 的同时追赶正在转录的 Pol II 延伸复合物。当核酸外切酶追上 Pol II 后, 诱导 Pol II 解离 DNA 和 RNA<sup>[112]</sup>。异构模型认为, 当 Pol II 进入转录终止阶段后, 转录延伸因子解离诱导 Pol II 构象改变, 促使 Pol II 自发停止 RNA 合成, 并解离 DNA 和 RNA<sup>[113]</sup>。此外, 近年来的整合模型认为 Pol II 转录终止兼具异构模型和鱼雷模型的特征, 在该模型中, CPF-CF 切割 mRNA 后 Pol II 构象发生变化, 导致转录延伸速率下降, 使得核酸外切酶能够追上行进中的 Pol II 复合物, 而 Pol II 解离 DNA 和 RNA 的过程仍然依赖核酸外切酶<sup>[114]</sup>。

核酸外切酶介导的转录终止复合物三维结构显示, 核酸外切酶稳定结合在 Pol II RNA 退出通道外侧, 其结合位置和转录延伸因子重叠, 揭示核酸外切酶在转录终止阶段替换延伸因子, 削弱 Pol II 和 RNA/DNA 杂合双链的亲合力, 减缓 Pol II 转录速率。核酸外切酶直接将 Pol II 转录的 RNA 5' 端引导至其催化中心, 利用自身的 RNA 外切酶活性缩短 RNA, 最终诱导 RNA 从 Pol II 活性中心解离。上述结构生物学证据支持 Pol II mRNA 终止的整合模型。

酵母细胞 ncRNA 转录终止的 NNS 复合物由 RNA 结合蛋白 Nrd1 和 Nab3 以及核酸解旋酶 Sen1 组成。Nrd1 和 Nab3 分别识别 RNA 上的特异转录终止序列 GUAA/G 和 UCUUG 以及 Pol II CTD 丝氨酸磷酸化 (Ser5P) 信号, 随后通过蛋白质间的相互作用招募 Sen1 蛋白至转录延伸复合物<sup>[115-116]</sup>。Sen1 结合在 Pol II RNA 退出通道附近, 与 RPB3 亚基存在直接作用。Sen1 终止机制与细菌 ρ 因子类似, 利用自身 5'→3' 解旋酶功能释放聚合酶活性中心的 RNA<sup>[116]</sup>。

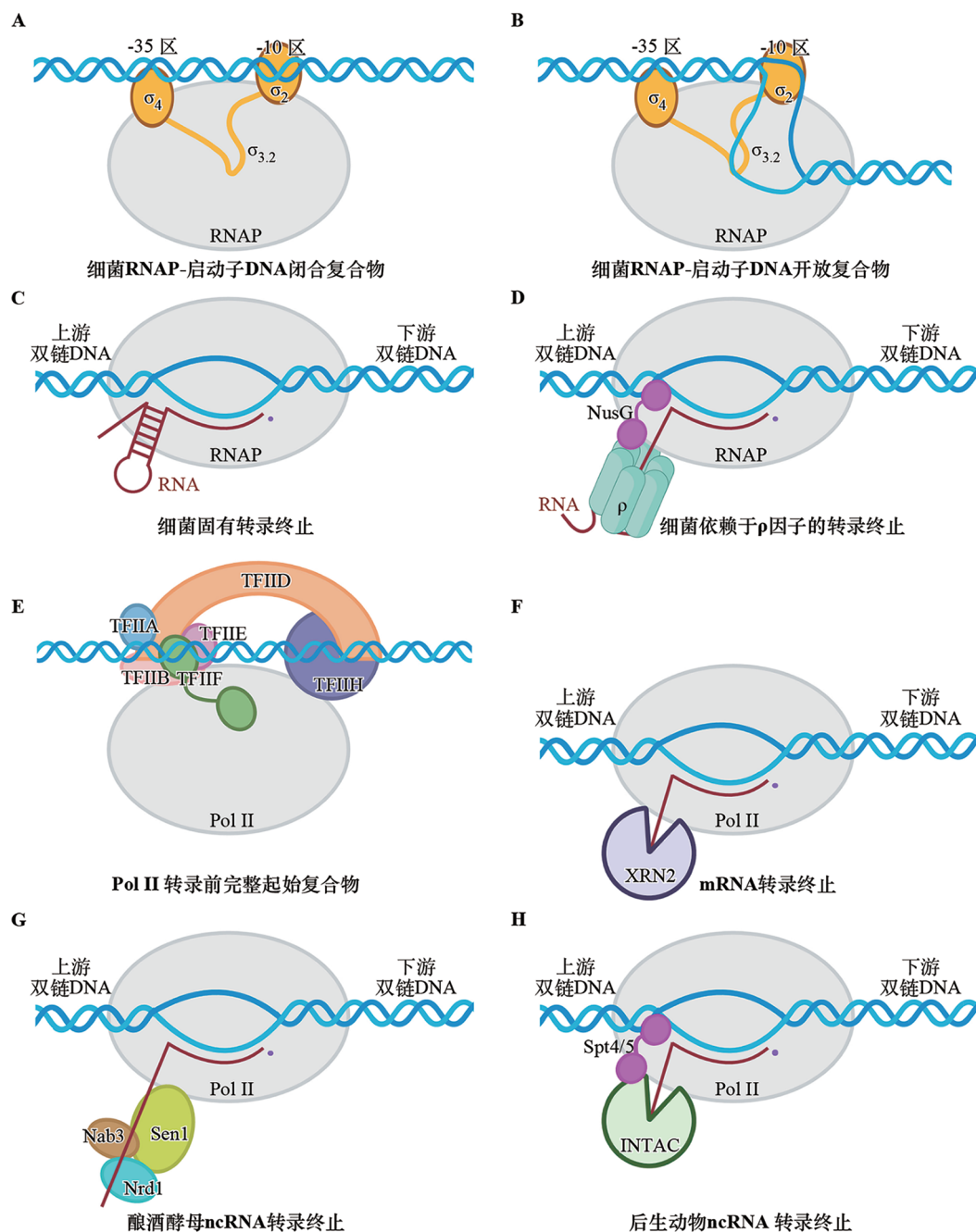
在后生动物细胞中, 转录起始后 Pol II 通常会

在 TSS 下游 20~200 bp 发生停顿,称为启动子近端暂停 (promoter-proximal pausing),是转录延伸过程中重要的限速步骤和检查点<sup>[117]</sup>。不能继续延伸的 Pol II 在 INTAC (integrator-containing PP2A-AC) 复合物的作用下,从启动子附近直接解离,提前终止转录<sup>[118]</sup>。INTAC 复合物由两个模块组成,包括 Integrator 整合复合物和 PP2A 磷酸酶复合物。在 INTAC 介导的转录终止过程中,Integrator 模块切割新生 RNA 的同时破坏活性中心中的 RNA-DNA

杂合双链,促进 Pol II 从 DNA 上解离<sup>[119]</sup>。

#### 4.3 Pol IV和Pol V的特殊转录机制

虽然 Pol IV/V 由 Pol II 进化而来,然而漫长的进化使 Pol IV/V 的催化活性中心和外部结构单元与 Pol II 具有显著区别。一方面,Pol IV/V 不与 Pol II 的转录起始和延伸因子相互作用(如 TFIIB、TFII E、TFIIF 和 TFIIS),从而保证 Pol IV/V 和 Pol II 转录互不干扰<sup>[120, 121]</sup>;另一方面,Pol IV/V 催化中心的核心元件发生变化,导致 Pol IV/V 的 RNA 延伸过



图片在<https://BioRender.com>绘制生成。

图2 细菌和Pol II转录起始与终止示意图

程容易发生停滞和倒退<sup>[122-123]</sup>。

Pol IV 和 RDR2 形成稳定的双 RNA 聚合酶复合物, 在该复合物中两个 RNA 聚合酶的催化中心由一个内部的通道相连接。Pol IV 首先被 CLSYs、ZMP、SHH1、REM 等蛋白以未知的机制招募到特定基因组区域, 起始单链 RNA 的合成。Pol IV 起始 RNA 合成之后, 由于 Pol IV 具有较弱的 RNA 延伸能力, 其转录过程容易发生暂停和倒退, 在其倒退的过程中, 单链 RNA 从 Pol IV-RDR2 的内部通道进入 RDR2 催化中心, 作为模板开始第二条 RNA 的合成, RDR2 在不断合成双链 RNA 的同时会诱导 Pol IV 进一步倒退, 促使 RNA 从 Pol IV 催化中心完全转移到 RDR2, 最终 RDR2 完成双链 RNA 合成并将其释放<sup>[120, 122]</sup>。

Pol V 被 DDR (DRD1-DMS3-RDM1) 复合物招募到基因组特定区域起始转录。在转录延伸过程中, Pol V 与 DNA 和 RNA 能够形成特殊的相互作用, 导致其转录易发生停顿和回溯。这种特殊的转录特征使 Pol V 在转录延伸过程中, 能够提供 AGO4/6 的结合时间窗口, 从而进一步招募 DRM2, 在基因组特定位置上沉积甲基化修饰。目前关于 Pol V 转录起始和转录偶联 DNA 甲基化的分子机制仍不清楚。

## 5 展望

近年来, 随着单颗粒冷冻电镜和单分子磁镊、单分子荧光等技术的发展, 细菌、古菌、真核细胞 RNA 聚合酶的高分辨率三维结构已得到全面解析, 然而转录过程中复杂中间状态的高分辨率三维结构尚未得到全面解析, 上述过程的分子机制尚未阐明。细菌基因转录与 DNA 修复、蛋白质翻译等过程直接偶联, 真核细胞 Pol II mRNA 转录与内含子剪接、表观遗传修饰等过程直接偶联, 上述过程的高分辨率三维结构尚待全面解析。此外, 细胞内基因转录过程的原位结构解析才初步开展, 解析 RNA 聚合酶在细胞内的原位结构将揭示基因转录在生理状态下的工作机制和调控机制, 未来冷冻电镜断层扫描技术的发展能够为基因转录的结构生物学研究带来新的突破。

**致谢:** 感谢国家蛋白质科学研究(上海)设施在冷冻电镜样品制备、样品筛选和数据收集等方面提供的大力支持。感谢国家自然科学基金委杰出青年基金(32425006)、上海市基础研究特区计划(JCYJ-

SHFY-2022-012)对相关研究的资助。

## [参 考 文 献]

- [1] Yang DL, Huang K, Deng D, et al. DNA-dependent RNA polymerases in plants. *Plant Cell*, 2023, 35: 3641-61
- [2] Girbig M, Misiaszek AD, Müller CW. Structural insights into nuclear transcription by eukaryotic DNA-dependent RNA polymerases. *Nat Rev Mol Cell Biol*, 2022, 23: 603-22
- [3] Xie G, Du X, Hu H, et al. Molecular mechanisms underlying the establishment, maintenance, and removal of DNA methylation in plants. *Annu Rev Plant Biol*, 2025, 76: 143-70
- [4] Flynn N, Chen X, Chen M. Plastid transcription: a major regulatory point in chloroplast biogenesis [M]//Burch-Smith TM. *Chloroplast gene expression: regulation, stress signaling and biotechnology*. Cham: Springer Nature Switzerland, 2024: 1-38
- [5] Wu XX, Mu WH, Li F, et al. Cryo-EM structures of the plant plastid-encoded RNA polymerase. *Cell*, 2024, 187: 1127-44
- [6] Chamberlin M, Berg P. Deoxyribonucleic acid-directed synthesis of ribonucleic acid by an enzyme from *Escherichia coli*. *Proc Natl Acad Sci U S A*, 1962, 48: 81-94
- [7] Zhang GY, Campbell EA, Minakhin L, et al. Crystal structure of core RNA polymerase at 3.3 Å resolution. *Cell*, 1999, 98: 811-24
- [8] Mazumder A, Kapanidis AN. Recent advances in understanding  $\sigma$ 70-dependent transcription initiation mechanisms. *J Mol Biol*, 2019, 431: 3947-59
- [9] Lee DJ, Minchin SD, Busby SJ. Activating transcription in bacteria. *Annu Rev Microbiol*, 2012, 66: 125-52
- [10] Chen J, Boyaci H, Campbell EA. Diverse and unified mechanisms of transcription initiation in bacteria. *Nat Rev Microbiol*, 2021, 19: 95-109
- [11] Shen L, Lai G, You L, et al. An SI3-sigma arch stabilizes cyanobacteria transcription initiation complex. *Proc Natl Acad Sci U S A*, 2023, 120: e2219290120
- [12] Qayyum MZ, Imashimizu M, Leanca M, et al. Structure and function of the SI3 insertion integrated into the trigger loop/helix of cyanobacterial RNA polymerase. *Proc Natl Acad Sci U S A*, 2024, 121: e2311480121
- [13] Tewary A, Prajapati RK, Mukhopadhyay J. Mechanism of delta mediated transcription activation in *Bacillus subtilis*: interaction with  $\alpha$  CTD of RNA polymerase stabilizes  $\delta$  and successively facilitates the open complex formation. *J Mol Biol*, 2023, 435: 168366
- [14] Pei HH, Hilal T, Chen ZA, et al. The  $\delta$  subunit and NTPase HelD institute a two-pronged mechanism for RNA polymerase recycling. *Nat Commun*, 2020, 11: 6418
- [15] Fang C, Li L, Zhao Y, et al. The bacterial multidrug resistance regulator BmrR distorts promoter DNA to activate transcription. *Nat Commun*, 2020, 11: 6284
- [16] Bock R. Witnessing genome evolution: experimental reconstruction of endosymbiotic and horizontal gene transfer. *Annu Rev Genet*, 2017, 51: 1-22

- [17] Daniell H, Jin S, Zhu XG, et al. Green giant--a tiny chloroplast genome with mighty power to produce high-value proteins: history and phylogeny. *Plant Biotechnol J*, 2021, 19: 430-47
- [18] Vergara-Cruces A, Pramanick I, Pearce D, et al. Structure of the plant plastid-encoded RNA polymerase. *Cell*, 2024, 187: 1145-59.e21
- [19] Wang T, Wang GL, Fang Y, et al. Architecture of the spinach plastid-encoded RNA polymerase. *Nat Commun*, 2024, 15: 9838
- [20] do Prado PFV, Ahrens FM, Liebers M, et al. Structure of the multi-subunit chloroplast RNA polymerase. *Mol Cell*, 2024, 84: 910-25.e5
- [21] Jun SH, Hyun J, Cha JS, et al. Direct binding of TFE $\alpha$  opens DNA binding cleft of RNA polymerase. *Nat Commun*, 2020, 11: 6123
- [22] Blombach F, Matelska D, Fouqueau T, et al. Key concepts and challenges in archaeal transcription. *J Mol Biol*, 2019, 431: 4184-201
- [23] Pilotto S, Fouqueau T, Lukyanova N, et al. Structural basis of RNA polymerase inhibition by viral and host factors. *Nat Commun*, 2021, 12: 5523
- [24] Tarau D, Grunberger F, Pils M, et al. Structural basis of archaeal RNA polymerase transcription elongation and Spt4/5 recruitment. *Nucleic Acids Res*, 2024, 52: 6017-35
- [25] Werner F, Grohmann D. Evolution of multisubunit RNA polymerases in the three domains of life. *Nat Rev Microbiol*, 2011, 9: 85-98
- [26] Cramer P. Organization and regulation of gene transcription. *Nature*, 2019, 573: 45-54
- [27] Vannini A, Cramer P. Conservation between the RNA polymerase I, II, and III transcription initiation machineries. *Mol Cell*, 2012, 45: 439-46
- [28] Chedin S, Riva M, Schultz P, et al. The RNA cleavage activity of RNA polymerase III is mediated by an essential TFIIS-like subunit and is important for transcription termination. *Genes Dev*, 1998, 12: 3857-71
- [29] Werner M, Thuriaux P, Soutourina J. Structure-function analysis of RNA polymerases I and III. *Curr Opin Struct Biol*, 2009, 19: 740-5
- [30] Kuhn CD, Geiger SR, Baumli S, et al. Functional architecture of RNA polymerase I. *Cell*, 2007, 131: 1260-72
- [31] Kim YJ, Björklund S, Li Y, et al. A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell*, 1994, 77: 599-608
- [32] Hsin JP, Manley JL. The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes Dev*, 2012, 26: 2119-37
- [33] Zaborowska J, Egloff S, Murphy S. The pol II CTD: new twists in the tail. *Nat Struct Mol Biol*, 2016, 23: 771-7
- [34] Jeronimo C, Collin P, Robert F. The RNA polymerase II CTD: the increasing complexity of a low-complexity protein domain. *J Mol Biol*, 2016, 428: 2607-22
- [35] Carter R, Drouin G. The increase in the number of subunits in eukaryotic RNA polymerase III relative to RNA polymerase II is due to the permanent recruitment of general transcription factors. *Mol Biol Evol*, 2010, 27: 1035-43
- [36] Geiger SR, Lorenzen K, Schreieck A, et al. RNA polymerase I contains a TFIIF-related DNA-binding subcomplex. *Mol Cell*, 2010, 39: 583-94
- [37] The Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*, 2000, 408: 796-815
- [38] Dubin MJ, Mittelsten Scheid O, Becker C. Transposons: a blessing curse. *Curr Opin Plant Biol*, 2018, 42: 23-9
- [39] Matzke MA, Mosher RA. RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nat Rev Genet*, 2014, 15: 394-408
- [40] Huang L, Jones AM, Searle I, et al. An atypical RNA polymerase involved in RNA silencing shares small subunits with RNA polymerase II. *Nat Struct Mol Biol*, 2009, 16: 91-3
- [41] Haag JR, Pikaard CS. Multisubunit RNA polymerases IV and V: purveyors of non-coding RNA for plant gene silencing. *Nat Rev Mol Cell Biol*, 2011, 12: 483-92
- [42] Luo J, Hall BD. A multistep process gave rise to RNA polymerase IV of land plants. *J Mol Evol*, 2007, 64: 101-12
- [43] Ream TS, Haag JR, Wierzbicki AT, et al. Subunit compositions of the RNA-silencing enzymes Pol IV and Pol V reveal their origins as specialized forms of RNA polymerase II. *Mol Cell*, 2009, 33: 192-203
- [44] Tucker SL, Reece J, Ream TS, et al. Evolutionary history of plant multisubunit RNA polymerases IV and V: subunit origins via genome-wide and segmental gene duplications, retrotransposition, and lineage-specific subfunctionalization. *Cold Spring Harb Symp Quant Biol*, 2010, 75: 285-97
- [45] Wang SB, Liang HP, Xu Y, et al. Genome-wide analyses across Viridiplantae reveal the origin and diversification of small RNA pathway-related genes. *Commun Biol*, 2021, 4: 412
- [46] Feklistov A, Sharon BD, Darst SA, et al. Bacterial sigma factors: a historical, structural, and genomic perspective. *Annu Rev Microbiol*, 2014, 68: 357-76
- [47] Mascher T. Signaling diversity and evolution of extracytoplasmic function (ECF) sigma factors. *Curr Opin Microbiol*, 2013, 16: 148-55
- [48] Campbell EA, Kamath S, Rajashankar KR, et al. Crystal structure of *Aquifex aeolicus*  $\sigma^N$  bound to promoter DNA and the structure of  $\sigma^N$ -holoenzyme. *Proc Natl Acad Sci U S A*, 2017, 114: E1805-E14
- [49] Danson AE, Jovanovic M, Buck M, et al. Mechanisms of  $\sigma$ -dependent transcription initiation and regulation. *J Mol Biol*, 2019, 431: 3960-74
- [50] Saecker RM, Mueller AU, Malone B, et al. Early intermediates in bacterial RNA polymerase promoter melting visualized by time-resolved cryo-electron microscopy. *Nat Struct Mol Biol*, 2024, 31: 1778-88
- [51] Chen J, Chiu C, Gopalkrishnan S, et al. Stepwise promoter melting by bacterial RNA polymerase. *Mol Cell*, 2020, 78: 275-88

- [52] Zuo Y, Steitz TA. Crystal structures of the *E. coli* transcription initiation complexes with a complete bubble. *Mol Cell*, 2015, 58: 534-40
- [53] Zhang Y, Feng Y, Chatterjee S, et al. Structural basis of transcription initiation. *Science*, 2012, 338: 1076-80
- [54] Li L, Molodtsov V, Lin W, et al. RNA extension drives a stepwise displacement of an initiation-factor structural module in initial transcription. *Proc Natl Acad Sci U S A*, 2020, 117: 5801-9
- [55] Kapanidis AN, Margeat E, Ho SO, et al. Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. *Science*, 2006, 314: 1144-7
- [56] Li L, Fang C, Zhuang N, et al. Structural basis for transcription initiation by bacterial ECF sigma factors. *Nat Commun*, 2019, 10: 1153
- [57] Lin W, Mandal S, Degen D, et al. Structural basis of ECF-sigma-factor-dependent transcription initiation. *Nat Commun*, 2019, 10: 710
- [58] Fang C, Li L, Shen L, et al. Structures and mechanism of transcription initiation by bacterial ECF factors. *Nucleic Acids Res*, 2019, 47: 7094-104
- [59] Yang Y, Darbari VC, Zhang N, et al. Structures of the RNA polymerase- $\sigma$ 54 reveal new and conserved regulatory strategies. *Science*, 2015, 349: 882-5
- [60] Gao F, Ye F, Buck M, et al. Subunit specialization in AAA+ proteins and substrate unfolding during transcription complex remodeling. *Proc Natl Acad Sci U S A*, 2025, 122: e2425868122
- [61] Ye F, Gao F, Liu X, et al. Mechanisms of DNA opening revealed in AAA+ transcription complex structures. *Sci Adv*, 2022, 8: eadd3479
- [62] Madan Babu M, Teichmann SA. Evolution of transcription factors and the gene regulatory network in *Escherichia coli*. *Nucleic Acids Res*, 2003, 31: 1234-44
- [63] Busby S, Ebright RH. Transcription activation by catabolite activator protein (CAP). *J Mol Biol*, 1999, 293: 199-213
- [64] Lawson CL, Swigon D, Murakami KS, et al. Catabolite activator protein: DNA binding and transcription activation. *Curr Opin Struct Biol*, 2004, 14: 10-20
- [65] Kompaniets D, Wang D, Yang Y, et al. Structure and molecular mechanism of bacterial transcription activation. *Trends Microbiol*, 2024, 32: 379-97
- [66] Liu B, Hong C, Huang RK, et al. Structural basis of bacterial transcription activation. *Science*, 2017, 358: 947-51
- [67] Feng Y, Zhang Y, Ebright RH. Structural basis of transcription activation. *Science*, 2016, 352: 1330-3
- [68] Fang CL, Philips SJ, Wu XX, et al. CueR activates transcription through a DNA distortion mechanism. *Nat Chem Biol*, 2021, 17: 57-64
- [69] Fang CL, Zhang Y. Bacterial MerR family transcription regulators: activation by distortion. *Acta Biochim Biophys Sin*, 2022, 54: 25-36
- [70] Yang Y, Liu C, Zhou W, et al. Structural visualization of transcription activated by a multidrug-sensing MerR family regulator. *Nat Commun*, 2021, 12: 2702
- [71] He D, You L, Wu X, et al. *Pseudomonas aeruginosa* Suta wedges RNAP lobe domain open to facilitate promoter DNA unwinding. *Nat Commun*, 2022, 13: 4204
- [72] Ray-Soni A, Bellecourt MJ, Landick R. Mechanisms of bacterial transcription termination: all good things must end. *Annu Rev Biochem*, 2016, 85: 319-47
- [73] Roberts JW. Mechanisms of bacterial transcription termination. *J Mol Biol*, 2019, 431: 4030-9
- [74] Mitra P, Ghosh G, Hafeezunnisa M, et al. Rho protein: roles and mechanisms. *Annu Rev Microbiol*, 2017, 71: 687-709
- [75] Shi J, Wen A, Zhao M, et al. Structural basis of Mfd-dependent transcription termination. *Nucleic Acids Res*, 2020, 48: 11762-72
- [76] Kang JY, Llewellyn E, Chen J, et al. Structural basis for transcription complex disruption by the Mfd translocase. *Elife*, 2021, 10: e62117
- [77] Wang L, Watters JW, Ju XW, et al. Head-on and co-directional RNA polymerase collisions orchestrate bidirectional transcription termination. *Mol Cell*, 2023, 83: 1153-64
- [78] You L, Zhang Y. Progress on molecular mechanisms of bacterial transcription termination. *Hereditas*, 2024, 46: 982-94
- [79] Peters JM, Mooney RA, Kuan PF, et al. Rho directs widespread termination of intragenic and stable RNA transcription. *Proc Natl Acad Sci U S A*, 2009, 106: 15406-11
- [80] Gusarov I, Nudler E. The mechanism of intrinsic transcription termination. *Mol Cell*, 1999, 3: 495-504
- [81] Yarnell WS, Roberts JW. Mechanism of intrinsic transcription termination and antitermination. *Science*, 1999, 284: 611-5
- [82] You L, Omollo EO, Yu C, et al. Structural basis for intrinsic transcription termination. *Nature*, 2023, 613: 783-9
- [83] Kang W, Ha KS, Uhm H, et al. Transcription reinitiation by recycling RNA polymerase that diffuses on DNA after releasing terminated RNA. *Nat Commun*, 2020, 11: 450
- [84] Harden TT, Herlambang KS, Chamberlain M, et al. Alternative transcription cycle for bacterial RNA polymerase. *Nat Commun*, 2020, 11: 448
- [85] Geiselmann J, Yager TD, Gill SC, et al. Physical properties of the *Escherichia coli* transcription termination factor rho. 1. Association states and geometry of the rho hexamer. *Biochemistry*, 1992, 31: 111-21
- [86] Molodtsov V, Wang CY, Firlar E, et al. Structural basis of Rho-dependent transcription termination. *Nature*, 2023, 614: 367-74
- [87] Skordalakes E, Berger JM. Structure of the Rho transcription terminator: mechanism of mRNA recognition and helicase loading. *Cell*, 2003, 114: 135-46
- [88] Song E, Uhm H, Munasingha PR, et al. Rho-dependent transcription termination proceeds via three routes. *Nat Commun*, 2022, 13: 1663
- [89] Said N, Hilal T, Sunday ND, et al. Steps toward translocation-independent RNA polymerase inactivation by terminator ATPase p. *Science*, 2021, 371: eabd1673

- [90] Hao Z, Epshtein V, Kim KH, et al. Pre-termination transcription complex: structure and function. *Mol Cell*, 2021, 81: 281-92
- [91] Nechaev S, Severinov K. Bacteriophage-induced modifications of host RNA polymerase. *Annu Rev Microbiol*, 2003, 57: 301-22
- [92] Krupp F, Said N, Huang YH, et al. Structural basis for the action of an all-purpose transcription anti-termination factor. *Mol Cell*, 2019, 74: 143-57
- [93] Shi J, Gao X, Tian T, et al. Structural basis of Q-dependent transcription antitermination. *Nat Commun*, 2019, 10: 2925
- [94] Yin Z, Kaelber JT, Ebricht RH. Structural basis of Q-dependent antitermination. *Proc Natl Acad Sci U S A*, 2019, 116: 18384-90
- [95] Nechaev S, Yuzenkova Y, Niedziela-Majka A, et al. A novel bacteriophage-encoded RNA polymerase binding protein inhibits transcription initiation and abolishes transcription termination by host RNA polymerase. *J Mol Biol*, 2002, 320: 11-22
- [96] You LL, Shi J, Shen LQ, et al. Structural basis for transcription antitermination at bacterial intrinsic terminator. *Nat Commun*, 2019, 10: 3048
- [97] Fouqueau T, Blombach F, Cackett G, et al. The cutting edge of archaeal transcription. *Emerg Top Life Sci*, 2018, 2: 517-33
- [98] Gehring AM, Walker JE, Santangelo TJ. Transcription regulation in Archaea. *J Bacteriol*, 2016, 198: 1906-17
- [99] Li J, Yue L, Li Z, et al. aCPSF1 cooperates with terminator U-tract to dictate archaeal transcription termination efficacy. *Elife*, 2021, 10: e70464
- [100] Sanders TJ, Wenck BR, Selan JN, et al. FttA is a CPSF73 homologue that terminates transcription in Archaea. *Nat Microbiol*, 2020, 5: 545-53
- [101] Yue L, Li J, Zhang B, et al. The conserved ribonuclease aCPSF1 triggers genome-wide transcription termination of Archaea via a 3'-end cleavage mode. *Nucleic Acids Res*, 2020, 48: 9589-605
- [102] You L, Wang C, Molodtsov V, et al. Structural basis of archaeal FttA-dependent transcription termination. *Nature*, 2024, 635: 229-36
- [103] Ji D, Manavski N, Meurer J, et al. Regulated chloroplast transcription termination. *Biochim Biophys Acta Bioenerg*, 2019, 1860: 69-77
- [104] Chi W, He B, Manavski N, et al. RHON1 mediates a Rho-like activity for transcription termination in plastids of *Arabidopsis thaliana*. *Plant Cell*, 2014, 26: 4918-32
- [105] Chen X, Qi Y, Wu Z, et al. Structural insights into preinitiation complex assembly on core promoters. *Science*, 2021, 372: eaba8490
- [106] Chen X, Wang X, Liu W, et al. Structures of +1 nucleosome-bound PIC-Mediator complex. *Science*, 2022, 378: 62-8
- [107] Chen X, Liu W, Wang Q, et al. Structural visualization of transcription initiation in action. *Science*, 2023, 382: eadi5120
- [108] Chen X, Yin X, Li J, et al. Structures of the human Mediator and Mediator-bound preinitiation complex. *Science*, 2021, 372: eabg0635
- [109] Thomas MC, Chiang CM. The general transcription machinery and general cofactors. *Crit Rev Biochem Mol Biol*, 2006, 41: 105-78
- [110] Glover-Cutter K, Larochelle S, Erickson B, et al. TFIIF-associated Cdk7 kinase functions in phosphorylation of C-terminal domain Ser7 residues, promoter-proximal pausing, and termination by RNA polymerase II. *Mol Cell Biol*, 2009, 29: 5455-64
- [111] Porrua O, Libri D. Transcription termination and the control of the transcriptome: why, where and how to stop. *Nat Rev Mol Cell Biol*, 2015, 16: 190-202
- [112] Richard P, Manley JL. Transcription termination by nuclear RNA polymerases. *Genes Dev*, 2009, 23: 1247-69
- [113] Kim M, Vasiljeva L, Rando OJ, et al. Distinct pathways for snoRNA and mRNA termination. *Mol Cell*, 2006, 24: 723-34
- [114] Luo W, Johnson AW, Bentley DL. The role of Rat1 in coupling mRNA 3'-end processing to transcription termination: implications for a unified allosteric-torpedo model. *Genes Dev*, 2006, 20: 954-65
- [115] Chaves-Arquero B, Perez-Canadillas JM. The Nrd1-Nab3-Sen1 transcription termination complex from a structural perspective. *Biochem Soc Trans*, 2023, 51: 1257-69
- [116] Rengachari S, Hainthaler T, Oberthuer C, et al. Mechanism of polyadenylation-independent RNA polymerase II termination. *Nat Struct Mol Biol*, 2025, 32: 339-45
- [117] Dollinger R, Gilmour DS. Regulation of promoter proximal pausing of RNA polymerase II in Metazoans. *J Mol Biol*, 2021, 433: 166897
- [118] Fianu I, Chen Y, Dienemann C, et al. Structural basis of Integrator-mediated transcription regulation. *Science*, 2021, 374: 883-7
- [119] Fianu I, Ochmann M, Walshe JL, et al. Structural basis of Integrator-dependent RNA polymerase II termination. *Nature*, 2024, 629: 219-27
- [120] Huang K, Wu XX, Fang CL, et al. Pol IV and RDR2: a two-RNA-polymerase machine that produces double-stranded RNA. *Science*, 2021, 374: 1579-86
- [121] Zhang HW, Huang K, Gu ZX, et al. A cryo-EM structure of KTF1-bound polymerase V transcription elongation complex. *Nat Commun*, 2023, 14: 3118
- [122] Fang C, Huang K, Wu X, et al. Transcription elongation of the plant RNA polymerase IV is prone to backtracking. *Sci Adv*, 2024, 10: eadq3087
- [123] Xie G, Du X, Hu H, et al. Structure and mechanism of the plant RNA polymerase V. *Science*, 2023, 379: 1209-13