

转录机器：绳上舞者

王耀来 刘锋

Transcription apparatus: A dancer on a rope

Wang Yaolai Liu Feng

引用信息 Citation: *Acta Physica Sinica*, 69, 248702 (2020) DOI: 10.7498/aps.69.20201631

在线阅读 View online: <https://doi.org/10.7498/aps.69.20201631>

当期内容 View table of contents: <http://wulixb.iphy.ac.cn>

您可能感兴趣的其他文章

Articles you may be interested in

生物分子结合水的结构与动力学研究进展

Research progress of molecular structure and dynamics of biological water

物理学报. 2019, 68(1): 013101 <https://doi.org/10.7498/aps.68.20181273>

外电场作用下MoS₂的分子结构和电子光谱

Molecular structure and electronic spectrum of MoS₂ under external electric field

物理学报. 2019, 68(17): 173101 <https://doi.org/10.7498/aps.68.20190781>

外场作用下C₁₂H₄Cl₄O₂的分子结构和电子光谱研究

Molecular structure and electronic spectrum of C₁₂H₄Cl₄O₂ under external electric field

物理学报. 2018, 67(22): 223101 <https://doi.org/10.7498/aps.67.20181454>

分子结拉伸与界面识别: 破解4, 4'-二吡啶分子结拉伸过程中高低电导之谜

Molecular junction stretching and interface recognition: Decode the mystery of high/low conductance switching in stretching process of 4, 4'-bipyridine molecular junction

物理学报. 2020, 69(20): 208502 <https://doi.org/10.7498/aps.69.20201297>

基于自适应随机共振理论的太赫兹雷达信号检测方法

Adaptive stochastic resonance system in terahertz radar signal detection

物理学报. 2018, 67(16): 160502 <https://doi.org/10.7498/aps.67.20172367>

钾钡共掺菲分子结构和电子特性的第一性原理研究

First principles studies on molecular structure and electronic properties of K- and Ba-codoped phenanthrene

物理学报. 2017, 66(23): 237401 <https://doi.org/10.7498/aps.66.237401>

特邀综述

转录机器: 绳上舞者*

王耀来^{1)†} 刘锋^{2)‡}

1) (江南大学理学院, 无锡 214122)

2) (南京大学物理系, 南京 210093)

(2020年9月30日收到; 2020年11月11日收到修改稿)

转录机器是响应细胞信号并启动基因转录的分子机器. 它控制着基因在恰当的时间以适当的速率表达, 演奏着生命之曲的底层旋律. 为了揭示转录机器的工作原理, 研究人员从其结构、动力学、信号转导等不同的维度进行探索. 研究进程在跌宕起伏中螺旋式上升, 各大研究领域趋于成熟却也趋于独立. 故此, 本文着眼全局, 扼要介绍真核生物转录机器在多领域的重要进展, 着重呈现看似大相径庭的新发现. 结构研究揭示了转录机器的基本架构, 核心争议在于媒介子(也称中介体复合物)的信号传递方式——“变构效应”抑或“招募导引”. 超级增强子的可能工作机制包括令人遐想的“液相分离”和“精确协作”. “转录钟”研究揭示了转录动力学的特征. 转录爆发现象的发现, 带来了转录调控机制的“调频”与“调幅”之争——前者得到了更多的支持. 此外, 本文也概述了转录研究的技术困难和理论模型, 提供了基于系综统计方法的研究策略, 并介绍了转录机器的理想模型. 转录机器在DNA上的运转方式源自进化的长河, 似乎是对物理学定律的完美运用与平衡.

关键词: 分子结构, 信号转导, 随机过程, 生命现象**PACS:** 87.10.-e, 02.50.Cw, 05.10.-a, 31.15.at**DOI:** 10.7498/aps.69.20201631

1 引言

物质的运动, 遵循着简单而又统一的规律, 体现为物理学公式的简洁、完备和优美. 生命, 作为物质存在的一种形态, 也应遵循同样的物理规律. 然而, 面对生命的机理问题, 我们经常手足无措, 望而却步. 似乎, 生命体过于复杂, 超出了物理学的美学追求. 然而, 换一个角度来看, 生命分子机器的构造方式和运转机理, 是源自数十亿年的进化优化, 是对物理学规律的完美运用. 其精其妙, 非“鬼斧神工”可拟.

对理解生命的困难, 物理学家薛定谔在《生命是什么》一书中, 用一个穿越故事做了很好的诠释. 一位蒸汽机时代的工程师, 来到了电气化时代,

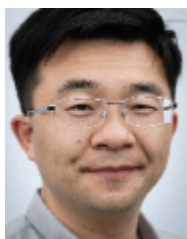
对电动机充满了好奇. 他惊奇地发现, 铜被拉成了很长的细丝、并被塞在了铁里面. 经过仔细探究, 这位工程师确定, 铜正是他制造汽锅的铜, 铁也是他制造气缸、曲轴、飞轮的铁, 铜和铁的每种属性都遵循着本来的物理规律. 他最终认定, 电动机旋转不是因为精灵鬼怪, 而是构造方式的不同导致了截然不同的运作方式——只不过, 他尚不知晓其中的物理规律或者已知规律的应用方式. 时至今日, 我们在理解生命机理的时候, 还是面临类似的境地.

本文将真核细胞中执行基因转录起始的分子机器——转录机器为例, 论述其结构与运转机制. 转录机器响应细胞信号, 控制着基因在恰当的时间以合适的速率转录, 是生命运作的基础^[1-6]. 这是一个广为人知的领域, 却也是一个困惑重重的领域. 在分子生物学教材中, 大约一半的内容是阐

* 国家自然科学基金(批准号: 11804123, 11831015, 11874209)资助的课题.

† 通信作者. E-mail: yaolaiwang@jiangnan.edu.cn

‡ 通信作者. E-mail: fliu@nju.edu.cn



王耀来, 江南大学理学院特聘副教授, 理学博士, 毕业于南京大学物理学院. 主要从事物理学、数学与生命科学的交叉课题研究与教学. 研究兴趣在于细胞与分子生物学的基础性问题, 目前主要研究转录机器的结构、功能、运转动力学及其信号转导机制. 论文发表在 *Biol. Rev.*, *Nucleic Acids Res.*, *Biology* 等期刊上.



刘锋, 南京大学物理学院教授. 1998 年于南京大学获得理学博士学位. 1999 年至今一直在南京大学物理系(学院)工作. 从事理论生物物理的研究, 探究基因表达、细胞信号转导和认知行为的动力学及其调控机制, 当前主要研究基因转录爆发、p53 信号转导网络、空间导航的神经环路机制等. 论文发表在 *Sci. Adv.*, *PNAS*, *J. Neurosci.* 和 *Phys. Rev. Lett.* 等期刊上.

述转录及其调控的. 然而, 在面对转录机器对细胞信号的动态响应时, 我们却常常手足无措, 所知甚少. 另一方面, 我们也习惯了不同研究者的结论相互冲突的情形. 事实上, 多数研究成果并没有错, 恰似“盲人摸象”, 只是由于研究者自身的局限性而导致了结论的片面性. 确知研究现状和待解决的问题, 有助于判断未来的方向、明确我们的使命, 本文的立意即在此.

自孟德尔遗传学说广泛传播以后, 人们就好奇遗传信息的物质载体是什么, 什么样的分子携带可传递的信息. 1953 年, 生物学家沃森和物理学家克里克基于 X 射线衍射数据, 提出了 DNA 的双螺旋结构. 据此, 克里克于 1957 年初步提出了著名的“中心法则”. 其核心内容是: 遗传信息储存在 DNA 序列中, DNA 可以自我复制, DNA 转录为信使 RNA, 信使 RNA 翻译为生理功能的主要执行者——蛋白质. 迄今为止, 中心法则是现代生物学的理论基石, 也似乎是生命王国里唯一的定律^[7-10]. 中心法则提出以后, 大家需要找到一部分子机器, 它能把 DNA 转录为 RNA. 1960 年前后, 生物化学家赫尔维茨等发现了构成转录机器的关键组件——RNA 聚合酶. 随后, RNA 聚合酶 II (RNA polymerase II, Pol II) 被证实是真核生物中转录信使 RNA 的核心组件^[11,12].

此后的数十年至今, 人们一直在寻找和纯化转录机器的组件, 并解析其结构^[5,13]. 随着单细胞和单分子测量技术的进步, 对转录机器组件之间相互

作用动力学的研究提上了日程. 单分子标记追踪技术导致了转录爆发现象的发现, 颠覆了传统认知. 近年来, 研究进展突飞猛进, 冲突的成果层出不穷. 不同的研究领域在趋于成熟的同时, 也趋于细化和独立化. 遗憾的是, 虽然转录的综述文章众多, 但大都是围绕某一子领域展开的. 本文试图概述近 20 年的重要研究成果, 并介绍关键争议, 旨在促进共同思考. 需要特别提及的是, 转录研究的核心目标是阐明转录机器的运转机制, 明确转录机器对细胞信号的编码和译码机制.

2 结构解析: 转录的物质基础

2.1 转录机器的基本架构

结构是功能的基础. 研究转录机器, 首先要明确转录机器的组件、组件的结构及其空间架构形式^[14-19]. 为解析转录机器的结构, 先采用生化技术, 对细胞核提取物分离纯化、得到转录所需的物质. 这些物质大都是蛋白质或者蛋白质复合物, 以及少量 RNA. 再运用 X 射线衍射、核磁共振、低温冷冻电子显微镜等方法, 研究这些物质的结构, 分析其功能.

研究表明, 转录机器涉及种类繁多的蛋白质, 大致可分为三类^[5,19]. 第一类只包含单个组件, 即 Pol II^[20-23]. Pol II 的功能是读取 DNA 序列信息, 并合成 RNA. 然而, Pol II 自身并不能准确定位转录起始位点, 也不能打开 DNA 双链. 第二类为通用转录因子 (general transcription factors, GTFs). GTFs 与启动子结合, 辅助 Pol II 定位并打开 DNA 双链, 帮助 Pol II 进入转录延伸阶段^[24-29]. 依据当初对核提取物的分离分组方式, GTFs 被命名为 TFII-A, TFII-B, TFII-D, TFII-E, TFII-F, TFII-H, TFII-S 等. GTFs 大都含有多个亚基. 后来, 又发现了一种新的通用转录因子, 被命名为媒介子 (Mediator; 也有人把 Mediator 归类为 GTF, 视作 cofactor)^[30-32]. 媒介子是包含 20 个以上亚基的“庞然大物”, 将激活子携带的调控信号传递给 Pol II, 是调控转录的关键所在. 第三类为转录激活子 (transcriptional activator) 和种类繁多的转录辅助因子 (cofactor). 这些因子参与转录与否, 与特定基因的调控区域是否含有相关序列有关^[19]. 转录激活子是细胞信号 (cellular signaling) 的携带者, 通过媒介子控制转录水平. 细胞核内激活子的

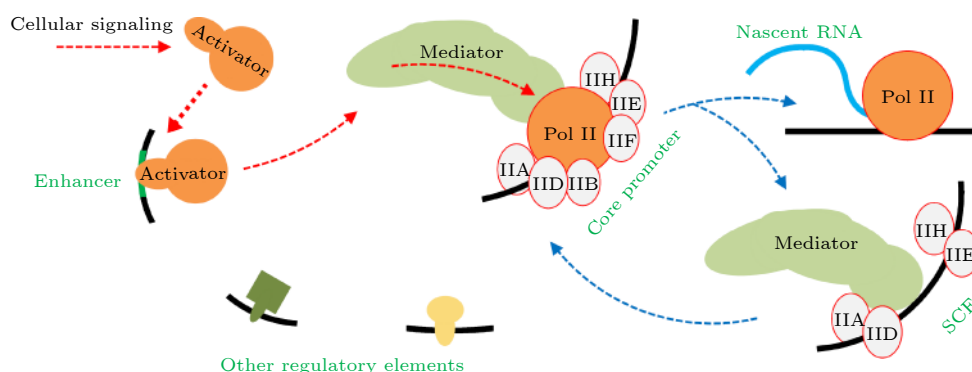


图 1 转录机器的基本架构和信号转导

Fig. 1. Configuration and signaling of the transcriptional apparatus.

浓度随着细胞信号的变化而变化; 一般来讲, 浓度越高, 基因转录水平越高, 直至饱和。

图 1 为转录机器的基本架构方式. 激活子与通常处于启动子上游的增强子 (Enhancer) 结合 [33]. 增强子与核心启动子之间的距离, 可达数千碱基对 (bp) 以上. 在 DNA 环化后, 增强子与启动子靠近. GTFs 将 Pol II 定位于转录起始位点附近 [25,34]. 增强子上的激活子与媒介子相互作用, 媒介子同时与 GTFs 和 Pol II 相互作用. GTFs 和 Pol II 依据来自媒介子的“指令”启动转录 [15]. 在 Pol II 进入延伸态以后, 还有大量 GTFs 滞留在启动子上, 它们被统称为脚手架复合物 (scaffold complex, SCF). 其他的辅助性因子, 直接或间接地与 DNA 上的特定序列结合, 通过改变 DNA 环化弯曲的刚性等方式, 影响转录.

目前, 人们只得到了转录机器基本架构和部分组件复合物的结构, 结构解析任务远未完成. 主要的挑战包括但不限于揭示蛋白质柔性结构域的结构和功能, 变构发生的条件、方式和效应, 不稳定蛋白质的结构和功能等. 假设在未来的某一天, 知晓了全部的组件结构, 这些碎片能拼凑出完整的转录机器吗? 如果能, 还能据此推断出转录机器的工作原理吗? 拼凑过程中我们会犯多大的错误?

2.2 媒介子的工作原理模型

媒介子, 是转录激活子和 Pol II 之间的信息桥梁, 故而是理解转录机器工作机制的关键所在 [30–32,35–42]. 待阐明的问题包括: 媒介子如何与激活子相互作用, 如何与其他 GTFs 相互作用, 如何与 Pol II 相互作用, 如何传递调控信息.

最新的研究得到了一个高达 250 万道尔顿的

媒介子-Pol II 复合物结构 [35]. 其结构提示, 媒介子与 Pol II 的 CTD 结构域 (即 Pol II 大亚基的 C 端结构域 carboxy-terminal domain, 是一条长达 182 个氨基酸残基的柔性链) 具有很高的亲和性. CTD 结构域是 Pol II 的一条柔性“长尾巴”. 安装在启动子区域的媒介子, 可以通过“黏连”、“拉拽” CTD 的方式, 招募 Pol II (图 2(a)). Pol II 定位到

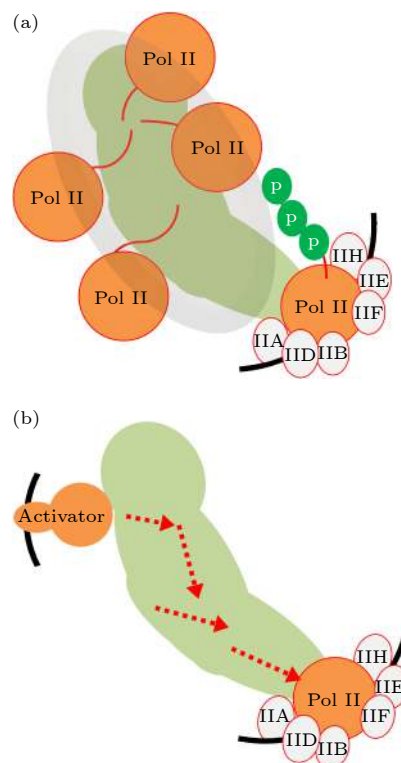


图 2 媒介子工作原理的两大模型 (a) 相分离模型; (b) 变构模型

Fig. 2. Two models of how the Mediator complex operates: (a) The Mediator acts to nucleate the flexible CTDs of Pol IIs, with the efficiency of Pol II assembly elevated; (b) an enhancer-bound activator induces allostery in the Mediator, resulting in a facilitated circumstance for transcription initiation.

启动子上以后, TFIID 磷酸化 CTD, 导致 CTD 与媒介子脱离、允许 Pol II 进入延伸态. 也就是说, 媒介子与 Pol II 的作用, 是通过 CTD 的磷酸化来控制凝聚态的相分离^[43]. 或许是因为分辨率不足, 该研究未发现媒介子的构象变化, 也难以解释激活子对媒介子的影响.

然而, 稍早的研究提示, 在媒介子和激活子结合以后, 媒介子会发生构象变化; 在激活子与媒介子脱离接触以后, 媒介子构象复原^[44,45]. 只有当媒介子处于变构态时, Pol II 才能迅速且顺利地安装并迅速启动转录(图 2(b)). 该观点解释了激活子与媒介子的相互作用, 也直接阐释了信号传递机制. 然而, 具体细节尚待解答: 激活子是如何引发变构的? 不同激活子的引发机制是否相同? 媒介子的变构是如何影响脚手架结构的? 是如何促进 Pol II 启动转录的?

其实, 前述两种观点并不绝对排斥, 可以被整合. 但遗憾的是, 结构解析未能证实变构效应.

2.3 增强子的功能

增强子属于顺式调控元件(即 DNA 上实施基因调控功能的片段), 其跨度通常为 1—2 个 DNA 大沟(单个大沟匹配激活子的单个 DNA 结合结构域), 长约 11—22 bp^[46]. 增强子的空间尺度, 相比于细胞核是非常渺小的; 这意味着, 为数不多的激活子要在巨大、拥挤的细胞核中准确找到增强子, 并不容易. 单纯以自由扩散的方式, 激活子在短时间内很难找到增强子, 也就无法完成信号的传递. 研究表明, 激活子能够在 DNA 染色质上滑动、跳跃、同时扫描增强子, 以加快速度^[47–49]. 激活子在搜寻增强子的过程中, 并不消耗自由能——就像麦克斯韦妖一样, 受到热力学第二定律的严格限制.

近年来, 一个新观点挑战了传统的“搜寻”概念. 在增强子团簇——大量间距很小的增强子, 也称超级增强子——上, 结合有大量的激活子^[50,51]. 这些激活子通过其无序结构域, 彼此相互结合, 形成了一个有别于核液的“液滴”^[52–54]. 此“液-液”相分离, 类似于油与醋的相分离. 在激活子液滴内部, 还包含有 GTFs、媒介子等蛋白质. 相比单个激活子, 激活子液滴一旦与启动子靠近, 就可以快捷地激活转录(图 3(a)). 不过, 具体的物理机制尚待阐明.

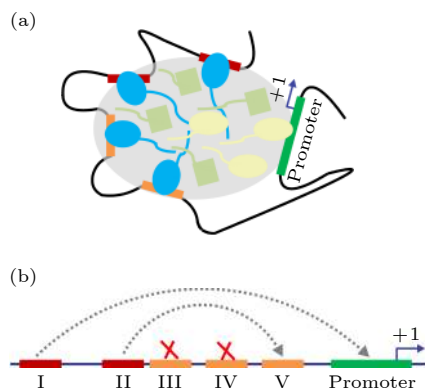


图 3 增强子工作原理的两种模型 (a) 液滴模型; (b) 分工协作模型

Fig. 3. Two models of how the enhancers function: (a) In the phase separation model, enhancers recruit transcriptional activators that further recruit various coactivators and the transcriptional apparatus via low-affinity disordered regions; (b) every enhancer plays a unique role and different enhancers cooperate to orchestrate transcription regulation. Shown is an example of regulatory mode at the *glnAp2* promoter.

“液滴说”可能仅仅代表了一种表面认知. 对大肠杆菌 *glnAp2* 基因的增强子团簇的研究表明, 液滴内部存在着严格的秩序. *glnAp2* 基因是一个原核基因, 但其调控模式与真核生物相似——结合在增强子上的激活子需要通过 DNA 环化才能与启动子接触^[55–58]. 从进化的角度来看, *glnAp2* 基因处于原核生物调控机制向真核生物调控机制的过渡位置, 具有特殊的地位. 如图 3(b) 所示, 在 *glnAp2* 基因的上游, 有 5 个相邻的增强子, 其中两个对激活子具有强亲和性(图中红色标记), 3 个具有弱亲和性(橙色). 按传统观点或者“液滴说”, 这 5 个增强子能招募大量的激活子, 形成“液滴”. 然而, 液滴模型无法解释复杂的转录动力学行为. 基于结构解析的理论研究表明, 每个增强子都有不可或缺的独特功能^[59]. 增强子与启动子之间的距离、彼此的相对距离、每个增强子的螺旋朝向和亲和性等, 都是被进化所精密选择的. 基于这些空间和拓扑因素, II 号增强子可与 V 号增强子协作, 实现 DNA 环化. 在 II-V 环化形成以后, 由 I 号增强子上的激活子激活转录. III 和 IV 的主要功能是抑制转录, 在激活子浓度很高时抑制 DNA 环化. 对真核生物增强子的研究, 目前很难达到精细的程度. 尽管如此, 酵母菌 *GAL1/GAL10* 基因启动子上的研究也表明, 增强子存在明确的分工^[60,61].

3 动力学测量：“转录钟”的启发

研究转录机器的终极目标,是揭示其动态运转的机制.也就是说,需要揭示哪个蛋白质在哪个时刻与启动子(包括启动子上的蛋白质)发生相互作用,以及如何发生作用.这是一个随机过程,研究的目的是获得各种动力学常数.

对转录动力学的探索,最初是基于一种定量生化技术——染色质免疫沉淀技术(chromatin immunoprecipitation assays, ChIP).其原理是:用甲醛浸泡细胞,导致蛋白质变性,使得结合在DNA上的蛋白质与DNA发生胶连.继而,整个细胞内的DNA被超声波打碎,而受到黏连蛋白质保护的DNA片段会保留下来.接着,用免疫抗体抓取感兴趣的蛋白质,并对该蛋白质结合的DNA进行测序.如果在实验开始前,用鹅膏蕈碱将所有细胞的转录进程同步化,在转录进程的不同阶段提取部分细胞,实施甲醛胶连并定量测量,就可以获得蛋白质-DNA的结合随时间的变化关系.

ChIP测量结果^[62–68]表明,转录进程具有周期性.在每个周期的初始阶段,各类致使核小体从核心启动子区域脱离的染色质异构酶发挥作用.接下来,转录机器在启动子上完成组装,Pol II启动转录.随之而来的是转录机器解散,核小体重新占据启动子.这个周期为数十分钟.最典型的是人*pS2*基因,其周期约为40分钟.测量结果还显示,每种蛋白质与启动子的结合都是稳定的,可达数十分钟.

对DNA-蛋白质相互作用的另一种测量方法,是光致光漂白荧光恢复(flourescence recovery after photobleaching technology, FRAP)技术.其原理是:在带有荧光基团的蛋白质与DNA密集结合以后,结合区域在显微镜下呈现为一个亮斑.用激光瞬时照射该亮斑,使其失去发光能力.如果蛋白质与DNA的结合是稳固的,被打黑的区域会长时间保持暗态.反过来,如果结合不稳定,随着已结合蛋白质被新来的蛋白质替换,暗区就会重新亮起来.结果表明,亮度很快恢复,说明蛋白质与DNA的结合时间在几秒到几十秒,最长的也不过两分钟^[62–64,69–73].

两种测量技术给出了令人困惑的结果,蛋白质与DNA的结合既呈现了长达数十分钟的长周期,

也呈现了数秒的短周期.此现象后来被称为“转录钟”^[74,75].理论研究将微观的分子间相互作用特性与宏观可测量的量联系起来,完美解释了转录钟^[76–78].在转录进程中,蛋白质与DNA的结合是非常短暂而又频繁的,并不存在稳定的结合.虽然FRAP技术无法分辨蛋白质与DNA的相互作用是简单的无功能碰撞还是有功能的相互作用,但给出了蛋白质与DNA相互作用的时间尺度.不同种类的蛋白质,在不同的转录进程阶段与启动子短暂而又频繁地相互作用.这种特质在ChIP测量中表现为长时间的稳定结合——这是测量技术和转录进程的阶段性共同造成的假象.

图4示意了最典型的人*pS2*基因转录钟.其转录进程长周期 T 约为40分钟.图中给出了四种蛋白质ER α , Pol II, p/CIP, Brg1参与转录进程的情况.一个点表示蛋白质与启动子的一次相互作用.单次相互作用的时间,即短周期,约为数秒到几分钟.点越密集的区域,该蛋白质与启动子发生相互作用的概率越高.每种蛋白质都在特定的阶段参与转录,但每次参与的时间都很短.

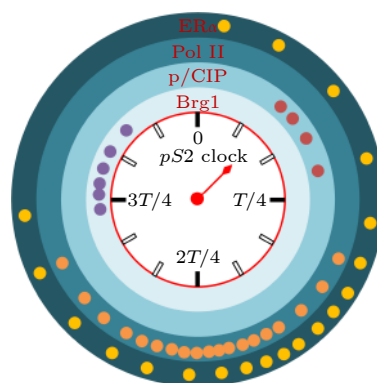


图4 人*pS2*基因启动子的“转录钟”
Fig. 4. Transcriptional clock of the human *pS2* promoter.

“转录钟”表明,转录机器是一架高度动态演化的机器.蛋白质在特定的转录阶段参与转录进程,作用时间很短但频率很高.推测起来,这种高动态允许转录机器对环境信号做出快速的响应.需要说明的是,传统测量技术并不能揭示蛋白质是如何参与转录的,也不能给出具体的转录动力学常数.单分子荧光光谱共定位技术(colocalization single-molecule spectroscopy method)为直接测量蛋白质-DNA相互作用动力学提供了可能,但远不足以深入研究转录机器的运转机制^[79].

4 单分子追踪: 颠覆传统的转录爆发

研究转录机器的另一个方法,是把它视作一个黑盒子.只研究其信号输入输出关系,即研究转录激活子的水平与转录起始速率的关系,而不考虑具体的分子机制.传统上,转录起始事件被认为是一个泊松过程;相邻转录起始事件之间的时间间隔服从指数分布^[80,81].为了研究单细胞的转录起始动力学,名为 MS2 或 PP7 的序列被插到待研究基因^[82–85].这些序列被转录以后,会折叠成发夹结构,进而结合荧光蛋白质.从而,新生成的信使 RNA 在显微镜下呈现为一个亮斑,可以被及时观测到.这种单分子跟踪技术的结果表明,转录起始事件的发生,非但不稳恒,而是断断续续的“爆发事件”^[84,86–94].如图 5 所示,基因多数时间处于

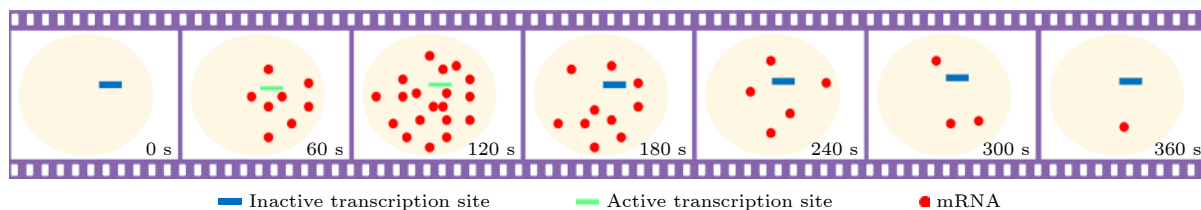


图 5 转录爆发^[95]

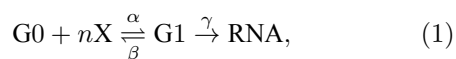
Fig. 5. Transcriptional burst^[95].

5 理论建模: 转录机器的唯象模型

如前所述,现有的转录研究成果具有碎片化、甚至相冲突的特征,客观上亟需大量的理论研究介入,以进行分析整合并为进一步的探索梳理思路.

5.1 希尔函数与两态模型

早期,转录被简化为酶促反应:



其中, G_0 和 G_1 分别表示处于未激活态和激活态的基因; X 为激活子; n 为激活转录所需要的激活子个数 (n 通常被视作等效值,可以是非整数); α , β , γ 为化学反应速率常数.据此, RNA 的平均生成速率 v 用希尔函数来描述,

$$v = \gamma \frac{\alpha[X]^n}{\alpha[X]^n + \beta}. \quad (2)$$

希尔函数在系统生物学中有着广泛的应用^[112].需要说明的是,希尔函数是细胞群体层次描述性模

型,不与具体的转录机制相对应;实验上观察到的单细胞的行为特征并不与“平均细胞”一致.

型,没有转录行为,但会突然进入爆发期.爆发期很短,通常不超过几分钟,但会有多达数十个的 Pol II 排队进入延伸阶段;紧接着,基因进入沉默态,没有转录起始事件的发生,直到下一次爆发^[95].从简单的原核生物到高等哺乳动物和人类,无论是看家型基因还是应激型基因,转录爆发普遍存在^[84,86–94].转录爆发的发现和证实,颠覆了传统认知,相关的综述文献多达几十篇^[96–105].

面对普遍存在的转录爆发现象,人们需要回答一个难题,即断断续续的爆发过程是如何传递调控信号的^[6,106–109].目前主流的观点^[90,93,105,106,107,110,111]有如下三种:一是频率调制,转录爆发的频率与激活子的浓度正相关,即转录响应的信号转导是数字信号;二是幅度调制,即单次转录爆发的信使 RNA 生成量与激活子的浓度正相关;三是认为前二者兼而有之.

型,不与具体的转录机制相对应;实验上观察到的单细胞的行为特征并不与“平均细胞”一致.

在进行随机模拟时,酶促反应模型等价于著名的“两态模型”,也称为“开关模型”或者“电报模型”^[113–115].两态模型中,基因的启动子被简化为处于“开”或“关”两个状态,转录起始事件只在“开”状态发生.通常,“开”和“关”两个状态的寿命,以及转录起始事件之间的时间间隔,均被视作服从指数分布.两态模型是迄今为止应用最广泛的模型,被用来解释和拟合包括转录爆发在内的实验现象和数据.因两态模型的简单易用性,研究人员对其低精度性有很高的容忍度,尽管理论上已经证明,它不宜用于分析具体的分子机制或随机原理^[116].

5.2 多尺度模型与连续性模型

更为精细的模型,可统称为多态模型.这类模型,依其建立的方式,分为“自上而下”和“自下而上”两类.前者是基于特定的考虑来设定模型,进而参照实验结果验证其可靠性,如下文将提到的理想模型.后者是通过实验“直接测量”转录机器状态

转换的动力学常数,从而建立转录机器的运转模型,最典型的例子是多尺度模型和连续性模型^[86,88].

转录机器的状态,依据转录钟研究的成果,具有数目庞大且转换迅速的特点.现有技术难以直接探测这些状态.因此,实验上测量转录动力学的策略,是诉诸于寻找限速步骤.具体来说,以两态模型为参照,检验“开”和“关”两个状态的寿命是否服从指数分布,检验“开”状态时相邻转录起始事件之间的间隔是否服从指数分布.实验结果促成了两个截然不同的模型,即“多尺度模型”和“连续性模型”.多尺度模型表明,在媒介子和 TBP(TATA 结合蛋白质,是 SCF 的关键组件)位于启动子上的前提下,启动子状态转换服从两态模型^[86].此条件下,“关”状态(记作 OFF1)的寿命服从期望为 100 秒的指数分布.若 TBP 不在启动子上,则两个 OFF1 状态之间包含了多个期望不同的 OFF2 状态.连续性模型则表明,“开”状态时相邻转录起始事件之间的间隔并不服从单个指数分布,而是一系列指数分布——其期望值在一个很宽的范围内连续变化^[88].

多尺度模型和连续性模型属于唯象模型,背后的分子机制停留在猜测阶段.它们源自对特定基因转录动力学的直接测量,普适性尚待进一步的实验验证.更多的转录模型,可参阅综述文献^[95, 96].

6 系综统计: 转录机器的理想模型

如前所述,界定转录机器的状态并测定状态转换动力学存在着重重困难.但是,理论研究可以在一定程度上弥补实验技术的不足,甚至给出可检验的机理模型.巧的是,多尺度模型和连续性模型,被早期的理想模型所解释和囊括^[44,95].该理想模型建立在转录机器的基本架构方式之上,能兼容已有模型并很好地解释实验现象,是结构架构和动力学研究之间的一座桥梁.该模型无论被进一步的研究证实或证伪,皆可作为进一步研究的出发点.因此,下文先概述研究转录机器的系综统计方法,继而介绍理想模型.

尽管转录存在基因特异性,但也存在普遍性.这种普遍性,体现为图 1 归纳的转录机器的基本架构.故此,我们可以设想一个理想转录机器,它只具有这些普遍性特征,且在启动子的附近有一个增强子.理想转录机器的状态,取决于蛋白质与 DNA、蛋白质与蛋白质之间的相互结合状态.由于分子运动的随机属性,所有的结合与解离反应都是随机事

件.设想一个系综,它包含了大量的理想转录机器,且这些转录机器处在完全相同的环境下,在“稳定”地启动转录.故而,在该系综内,处于某特定状态的转录机器的数量,不随时间变化.换言之,随机抽取一个转录机器,其处于某特定状态的概率也是确定的.据此,可以定义一系列关键状态: A 为增强子被激活子结合; S 为核心启动子处于被脚手架结合的状态; M 为信使 RNA 处于孕育阶段,即 Pol II 处于从 PIC (转录预起始复合物) 到 OPC (开放复合物; 此状态下, DNA 双链已打开,形成了转录泡),再到进入延伸态之前的状态; J 为媒介子同时连接了增强子上的激活子和核心启动子上的结构包括 SCF, PIC, OPC 等.根据转录机器的结构架构,可知这些状态间的关系,进而可得到信使 RNA 生成和转录机器状态之间的概率方程:

$$P(M) = P(S)P(A_J|S)P(J|A_J S)P(M|J) + q, \quad (3)$$

其中, A_J 是分离变量过程中引入的一个随机事件,是 A 的子集,表示的是与媒介子相互作用的激活子; q 与 A_J 无关,代表基底转录.

在此概率方程中,只有 $P(A_J|S)$ 的值与激活子浓度直接相关,内在包含了激活子与增强子的相互作用编码激活子浓度的机制.考虑一段时间内,属于 A_J 的激活子与增强子的结合-解离相互作用发生了 m 次.令 τ_{on}^i 和 τ_{off}^i 分别表示第 i 次相互作用的结合时间和处于分离态的时间,令 R 表示增强子被激活子结合的时间占比,则

$$P(A_J|S) = \lim_{m \rightarrow \infty} R = \lim_{m \rightarrow \infty} \frac{\sum_{i=1}^m \tau_{\text{on}}^i}{\sum_{i=1}^m (\tau_{\text{on}}^i + \tau_{\text{off}}^i)}, \quad (4)$$

其中, $\tau_{\text{on}}^i = \frac{1}{a_{\text{off}}} \ln\left(\frac{1}{r_g}\right)$, $\tau_{\text{off}}^i = \frac{1}{a_{\text{on}}} \ln\left(\frac{1}{r_k}\right)$, a_{on} 和 a_{off} 分别表示结合和解离的倾向性函数(propensity function), a_{on} 与激活子的浓度正相关, a_{off} 为常数, r_g 和 r_k 均服从 $(0, 1]$ 区间的均匀分布且相互独立.则,

$$\begin{aligned} P(A_J|S) &= \lim_{m \rightarrow \infty} \frac{\sum_{i=1}^m \tau_{\text{on}}^i}{\sum_{i=1}^m (\tau_{\text{on}}^i + \tau_{\text{off}}^i)} \\ &= \frac{\int_0^1 \frac{1}{a_{\text{off}}} \ln\left(\frac{1}{r_g}\right) dr_g}{\int_0^1 \frac{1}{a_{\text{off}}} \ln\left(\frac{1}{r_g}\right) dr_g + \int_0^1 \frac{1}{a_{\text{on}}} \ln\left(\frac{1}{r_k}\right) dr_k} \\ &= \frac{a_{\text{on}}}{a_{\text{on}} + a_{\text{off}}}. \end{aligned} \quad (5)$$

生理条件下, 要实现激活子浓度的编码, 就必须在激活子浓度未发生明显变化的短时间内, R 快速收敛. 模拟表明, 10 次左右的结合-解离就能使 R 较好地收敛 (如图 6(a) 所示, 图中 m 为结合解离的次数, 横轴为与激活子浓度正相关的无量纲量 a_{on}/a_{off}). 这也意味着, 在短时间内, 激活子必须与增强子发生多次结合解离, 即激活子与增强子的相互作用是短暂且极为快速的. 生理条件下, 有专门的分子伴侣将处于结合态的激活子剥离 [72-74,117,118]. 据此推断, 激活子与增强子之间的快速而又短暂的相互作用是激活子浓度的编码方式.

如果激活子的编码机制如上所述, 那么如何确保该信号能高保真地转换为信使 RNA 的生成量呢? 转录机器分子运动的随机性, 很容易淹没这种编码信号. 要不被淹没, 则 $P(S), P(J|A_jS), P(M|J)$ 这 3 个因式必须足够大, 尽可能接近 1. 这 3 个因式接近 1 的方式, 就是转录机器的运转方式. 要使 $P(S)$ 大, 唯一的实现方式, 就是脚手架要在启动子状态许可的情况下, 尽快完成安装, 且要尽可能保持稳定. 要使 $P(J|A_jS)$ 大, 即在增强子上存在激活子、且核心启动子上存在脚手架的条件下, 媒介子必须迅即地连接二者. 考虑到激活子的快速结合解离编码了激活子浓度, 媒介子通过自由扩散等方式去连接是来不及的. 可行的方案是, 媒介子自身作

为脚手架结构的一部分, 等待激活子的到来. $P(M|J)$ 大, 就是在媒介子连接激活子和脚手架的情况下, Pol II 必须尽可能频繁地启动转录. 这里的关键问题是, Pol II 是如何知道媒介子处于激活子结合状态的? 生理上可能的答案似乎只有一种, 那就是激活子与媒介子的结合会导致媒介子的结构发生变化, 该变化使得 Pol II 能被快速地招募、能快速地启动转录 [45].

把上述结论归结起来, 即可得到理想转录机器的工作原理 (图 6(b)). 脚手架结构较为稳定, 媒介子是脚手架结构的一部分. 在脚手架与激活子之间存在钳形空间 (clamp-like space), 激活子快速地出入该空间. 在钳形空间被激活子占据的时间段内, 媒介子通过变构传递信息, Pol II 一个接一个地迅速启动转录, 形成一次转录爆发. 也就是说, 转录起始必然以爆发的形式发生.

脚手架结构是由媒介子、TBP 等大量结合稳定性各不相同的蛋白质架构而成, 因此, 脚手架架构程度的不同及启动子被核小体占据程度的不同, 等同于将转录爆发之间的时间间隔进行了多尺度划分——此即多尺度模型描述的动力学特征. 在转录爆发期, 转录起始事件的发生极为频繁, 此时转录起始所需的 Pol II、TFIIF 等的浓度不可再视为恒量——这种涨落会导致起始速率谱呈现连续性特征, 即连续性模型所描述的动力学特征.

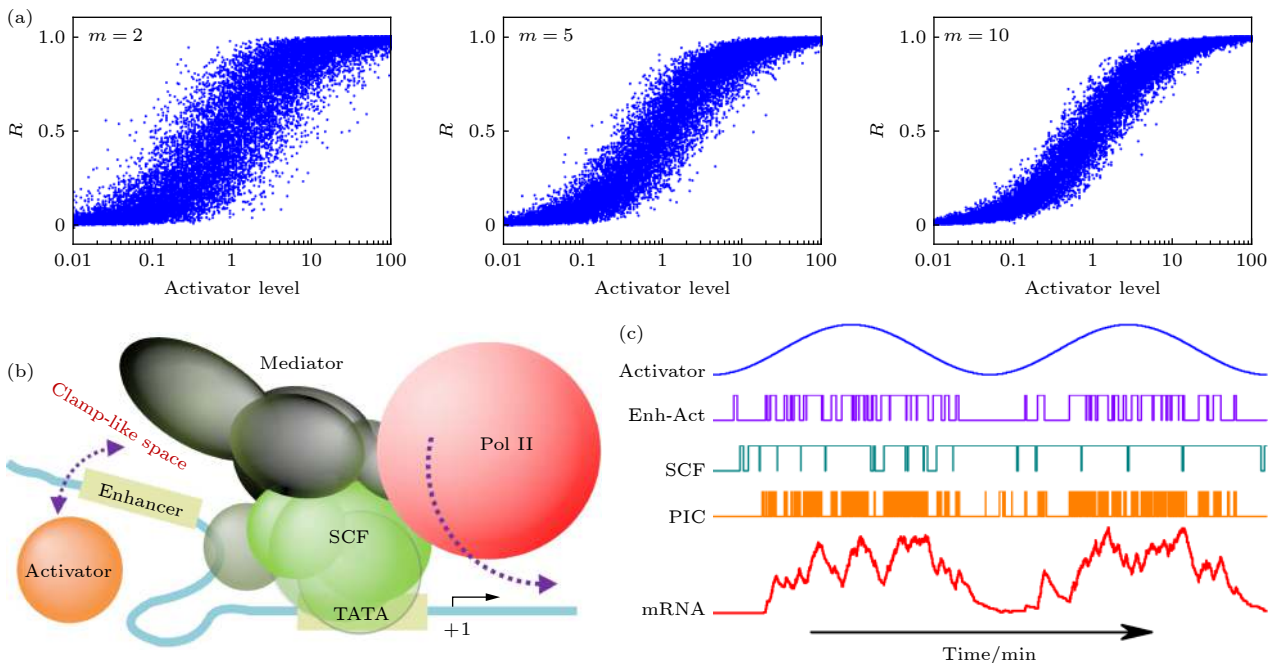


图 6 转录机器的理想化模型 [44]

Fig. 6. Ideal model for how the transcriptional apparatus operates [44].

图 6(c) 模拟了转录机器对调控信号的响应. 伴随着激活子浓度随时间变化, 激活子频繁而又短暂地与增强子相互作用 (图中标示为 Enh-Act), 增强子处于结合态的概率与激活子的浓度正相关. 脚手架结构 (标示为 SCF) 相对稳定, 平均寿命远大于激活子结合时间. 在脚手架结构和媒介子存在、且增强子被激活子结合的时间段内, 转录起始事件频繁地发生 (标示为 PIC). RNA 的生成速率, 或者 RNA 寿命不是很长的条件下的细胞内 RNA 数目 (具有翻译活性的 RNA 寿命大都很短), 随激活子浓度的变化而变化.

转录机器的理想化模型, 支持“频率调制”的观点. 转录起始事件的发生, 总是以转录爆发为单位的——在激活子居于钳形空间内部时, 大量 Pol II 进入延伸阶段, 造成一次爆发. 激活子浓度越高, 爆发发生的概率越大, 但单次爆发的平均持续时间并不改变. 也就是说, 转录调控的信号转导方式, 是一种“数字”信号. 需要说明的是, 实验上记录到的单个“脉冲”或者“单峰”, 不一定对应着一次爆发, 也可能是多个相邻爆发的叠加^[119].

7 结 语

细胞中, 每时每刻都有大量的转录机器驾驭着随机运动的分子, 在秩序井然地运转着. 它们响应调控信号, 适时适量地启动基因表达, 将遗传信息向生理功能转换, 奏响了生命之曲的基层旋律. 如今, 转录机器的结构架构和运转机制若隐若现, 然揭秘之旅似乎山重水复. 诚然, 实验技术的进步是第一位的, 但目前更迫切需要理论研究来整合已有知识, 并洞察具体机理.

转录机器的庞大和复杂, 使得结构解析和动力学测量均面临着不小的挑战. 理论研究方面, 我们渴望拥有行之有效的研究策略甚至是理论体系. 下面是我们的一些思考, 供读者参考. 生命是开放系统, 能自发利用能量对抗熵增, 从而实现自身的有序架构和动态运转. 生命生存压力的首要来源, 就是从生态圈食物链中获取能量, 是故, 能量对生命而言是宝贵的. 在能量和熵的博弈中, 生命在转录机器上的开销是巨大的: 组件众多、架构复杂、运转有序. 因此, 信息转导所要求的有序性, 或许是研究转录机器的可行切入角度.

在漫长的进化过程中, 转录机器具有保守性,

同时趋于复杂化. 对进化上具有过渡意义的 *glnAp2* 基因转录机器的研究启示: 转录机器的运转, 无论从哪个维度来看, 都是最优解. 这些维度包括顺式元件的长度、方向性及相互间隔, 反式元件的结构和稳定性, 转录机器整体的空间架构和拓扑, 动力学上的亲和性等. 液相分离, 使得相关的蛋白质在空间上靠近从而方便了彼此间的协作, 可视作更高层级的组织优化. 转录机器理想模型的得出, 也是基于“其工作机制是对信号响应的优化”思想. 故此, 本文题名为“转录机器: 绳上舞者”, 旨在唤起读者思考: 转录机器的运转机制是否源自对各类物理规律的平衡和优化运用?

参考文献

- [1] Brivanlou A H, Darnell Jr J E 2002 *Science* **295** 813
- [2] Blake W J, Kaern M, Cantor C R, Collins J J 2003 *Nature* **422** 633
- [3] Gregor T, Tank D W, Wieschaus E F, Bialek W 2007 *Cell* **130** 153
- [4] Weake V M, Workman J L 2010 *Nat. Rev. Genet.* **11** 426
- [5] Fuda N J, Ardehali M B, Lis J T 2009 *Nature* **461** 186
- [6] Senecal A, Munsky B, Proux F, Ly N, Braye F E, Zimmer C, Mueller F, Darzacq X 2014 *Cell Rep.* **8** 75
- [7] Koonin E V 2015 *Biol. Direct.* **10** 52
- [8] Bussard A E 2005 *EMBO Rep.* **6** 691
- [9] Crick F 1970 *Nature* **227** 561
- [10] Crick F H 1958 *Symp. Soc. Exp. Biol.* **12** 138
- [11] Hurwitz J 2005 *J. Biol. Chem.* **280** 42477
- [12] Young R A 1991 *Annu. Rev. Biochem.* **60** 689
- [13] Lis J T 2019 *Nat. Struct. Mol. Biol.* **26** 777
- [14] Krishnamurthy S, Hampsey M 2009 *Curr. Biol.* **19** R153
- [15] Kornberg R D 2007 *Proc. Natl. Acad. Sci. U.S.A.* **104** 12955
- [16] Wang D, Bushnell D A, Westover K D, Kaplan C D, Kornberg R D 2006 *Cell* **127** 941
- [17] Westover K D, Bushnell D A, Kornberg R D 2004 *Science* **303** 1014
- [18] Hahn S 2004 *Nat. Struct. Mol. Biol.* **11** 394
- [19] Thomas M C, Chiang C M 2006 *Crit. Rev. Biochem. Mol. Biol.* **41** 105
- [20] Cramer P, Bushnell D A, Kornberg R D 2001 *Science* **292** 1863
- [21] Cramer P, Bushnell D A, Fu J, Gnatt A L, Maier-Davis B, Thompson N E, Burgess R R, Edwards A M, David P R, Kornberg R D 2000 *Science* **288** 640
- [22] Fu J, Gnatt A L, Bushnell D A, Jensen G J, Thompson N E, Burgess R R, David P R, Kornberg R D 1999 *Cell* **98** 799
- [23] Meredith G D, Chang W H, Li Y, Bushnell D A, Darst S A, Kornberg R D 1996 *J. Mol. Biol.* **258** 413
- [24] Murakami K, Tsai K L, Kalisman N, Bushnell D A, Asturias F J, Kornberg R D 2015 *Proc. Natl. Acad. Sci. U.S.A.* **112** 13543
- [25] Fazal F M, Meng C A, Murakami K, Kornberg R D, Block S M 2015 *Nature* **525** 274
- [26] Schweikhard V, Meng C, Murakami K, Kaplan C D, Kornberg R D, Block S M 2014 *Proc. Natl. Acad. Sci. U.S.A.* **111** 6642

- [27] Murakami K, Elmlund H, Kalisman N, Bushnell D A, Adams C M, Azubel M, Elmlund D, Levi-Kalisman Y, Liu X, Gibbons B J, Levitt M, Kornberg R D 2013 *Science* **342** 1238724
- [28] Liu X, Bushnell D A, Kornberg R D 2013 *Biochim. Biophys. Acta.* **1829** 2
- [29] Liu X, Bushnell D A, Wang D, Calero G, Kornberg R D 2010 *Science* **327** 206
- [30] Kelleher R J, 3 rd, Flanagan P M, Kornberg R D 1990 *Cell* **61** 1209
- [31] Flanagan P M, Kelleher R J, 3 rd, Sayre M H, Tschochner H, Kornberg R D 1991 *Nature* **350** 436
- [32] Kim Y J, Bjorklund S, Li Y, Sayre M H, Kornberg R D 1994 *Cell* **77** 599
- [33] Kim T K, Shiekhhattar R 2015 *Cell* **162** 948
- [34] Haberle V, Lenhard B 2016 *Semin. Cell Dev. Biol.* **57** 11
- [35] Robinson P J, Trnka M J, Bushnell D A, Davis R E, Mattei P J, Burlingame A L, Kornberg R D 2016 *Cell* **166** 1411
- [36] Robinson P J, Trnka M J, Pellarin R, Greenberg C H, Bushnell D A, Davis R, Burlingame A L, Sali A, Kornberg R D 2015 *eLife* **4** e08719
- [37] Plaschka C, Lariviere L, Wenzek L, Seizl M, Hemann M, Tegunov D, Petrotchenko E V, Borchers C H, Baumeister W, Herzog F, Villa E, Cramer P 2015 *Nature* **518** 376
- [38] Poss Z C, Ebmeier C C, Taatjes D J 2013 *Crit. Rev. Biochem. Mol. Biol.* **48** 575
- [39] Casamassimi A, Napoli C 2007 *Biochimie* **89** 1439
- [40] Takagi Y, Kornberg R D 2006 *J. Biol. Chem.* **281** 80
- [41] Kornberg R D 2005 *Trends Biochem. Sci.* **30** 235
- [42] Kornberg R D 2005 *Trends Biochem. Sci.* **30** 221
- [43] Guo Y E, Manteiga J C, Henninger J E, Sabari B R, Dall'Agnesa A, Hannett N M, Spille J H, Afeyan L K, Zamudio A V, Shrinivas K, Abraham B J, Boija A, Decker T M, Rimel J K, Fant C B, Lee T I, Cisse I I, Sharp P A, Taatjes D J, Young R A 2019 *Nature* **572** 543
- [44] Wang Y, Liu F, Wang W 2012 *Sci. Rep.* **2** 422
- [45] Meyer K D, Lin S C, Bernecky C, Gao Y, Taatjes D J 2010 *Nat. Struct. Mol. Biol.* **17** 753
- [46] Natoli G, Saccani S, Bosisio D, Marazzi I 2005 *Nat. Immunol.* **6** 439
- [47] Levine M, Cattoglio C, Tjian R 2014 *Cell* **157** 13
- [48] Wang Y M, Austin R H, Cox E C 2006 *Phys. Rev. Lett.* **97** 048302
- [49] Elf J, Li G W, Xie X S 2007 *Science* **316** 1191
- [50] Whyte W A, Orlando D A, Hnisz D, Abraham B J, Lin C Y, Kagey M H, Rahl P B, Lee T I, Young R A 2013 *Cell* **153** 307
- [51] Spitz F, Furlong E E M 2012 *Nat. Rev. Genet.* **13** 613
- [52] Hahn S 2018 *Cell* **175** 1723
- [53] Boija A, Klein I A, Sabari B R, Dall'Agnesa A, Coffey E L, Zamudio A V, Li C H, Shrinivas K, Manteiga J C, Hannett N M, Abraham B J, Afeyan L K, Guo Y E, Rimel J K, Fant C B, Schuijers J, Lee T I, Taatjes D J, Young R A 2018 *Cell* **175** 1842
- [54] Sabari B R, Dall'Agnesa A, Boija A, Klein I A, Coffey E L, Shrinivas K, Abraham B J, Hannett N M, Zamudio A V, Manteiga J C, Li C H, Guo Y E, Day D S, Schuijers J, Vasile E, Malik S, Hnisz D, Lee T I, Cisse I I, Roeder R G, Sharp P A, Chakraborty A K, Young R A 2018 *Science* **361** eaar3958
- [55] Rippe K 2000 *Biochemistry* **39** 2131
- [56] Atkinson M R, Pattaramanon N, Ninfa A J 2002 *Mol. Microbiol.* **46** 1247
- [57] Lilja A E, Jenssen J R, Kahn J D 2004 *J. Mol. Biol.* **342** 467
- [58] Huo Y X, Tian Z X, Rappas M, Wen J, Chen Y C, You C H, Zhang X, Buck M, Wang Y P, Kolb A 2006 *Mol. Microbiol.* **59** 168
- [59] Wang Y, Liu F, Wang W 2016 *Nucleic Acids Res.* **44** 10530
- [60] Elison G L, Xue Y, Song R, Acar M 2018 *Cell Rep.* **25** 737
- [61] Donovan B T, Huynh A, Ball D A, Patel H P, Poirier M G, Larson D R, Ferguson M L, Lenstra T L 2019 *EMBO J.* **38** e100809
- [62] Karpova T S, Kim M J, Spriet C, Nalley K, Stasevich T J, Kherrouche Z, Heliot L, McNally J G 2008 *Science* **319** 466
- [63] Métivier R, Reid G, Gannon F 2006 *EMBO Rep.* **7** 161
- [64] Métivier R, Penot G, Hubner M R, Reid G, Brand H, Kos M, Gannon F 2003 *Cell* **115** 751
- [65] Kang Z, Pirskanen A, Janne O A, Palvimo J J 2002 *J. Biol. Chem.* **277** 48366
- [66] Liu Y, Xia X, Fondell J D, Yen P M 2006 *Mol. Endocrinol.* **20** 483
- [67] Shang Y, Hu X, DiRenzo J, Lazar M A, Brown M 2000 *Cell* **103** 843
- [68] Becker M, Baumann C, John S, Walker D A, Vigneron M, McNally J G, Hager G L 2002 *EMBO Rep.* **3** 1188
- [69] Darzacq X, Shav-Tal Y, de Turris V, Brody Y, Shenoy S M, Phair R D, Singer R H 2007 *Nat. Struct. Mol. Biol.* **14** 796
- [70] Johnson T A, Elbi C, Parekh B S, Hager G L, John S 2008 *Mol. Biol. Cell* **19** 3308
- [71] Catez F, Ueda T, Bustin M 2006 *Nat. Struct. Mol. Biol.* **13** 305
- [72] Bosisio D, Marazzi I, Agresti A, Shimizu N, Bianchi M E, Natoli G 2006 *EMBO J.* **25** 798
- [73] Reid G, Hubner M R, Métivier R, Brand H, Denger S, Manu D, Beaudouin J, Ellenberg J, Gannon F 2003 *Mol. Cell* **11** 695
- [74] Hager G L, McNally J G, Misteli T 2009 *Mol. Cell* **35** 741
- [75] Reid G, Gallais R, Métivier R 2009 *Int. J. Biochem. Cell Biol.* **41** 155
- [76] Wang Y, Liu F, Li J, Wang W 2014 *J. R. Soc. Interface* **11** 20140253
- [77] Lemaire V, Lee C F, Lei J, Métivier R, Glass L 2006 *Phys. Rev. Lett.* **96** 198102
- [78] Krasnov A N, Mazina M Y, Nikolenko J V, Vorobyeva N E 2016 *Cell Biosci.* **6** 15
- [79] Gourse R L, Landick R 2012 *Cell* **148** 635
- [80] Lenstra T L, Rodriguez J, Chen H, Larson D R 2016 *Annu. Rev. Biophys.* **45** 25
- [81] Raj A, van Oudenaarden A 2008 *Cell* **135** 216
- [82] Hocine S, Raymond P, Zenklusen D, Chao J A, Singer R H 2013 *Nat. Methods* **10** 119
- [83] Lim F, Peabody D S 2002 *Nucleic Acids Res.* **30** 4138
- [84] Chubb J R, Treck T, Shenoy S M, Singer R H 2006 *Curr. Biol.* **16** 1018
- [85] Bertrand E, Chartrand P, Schaefer M, Shenoy S M, Singer R H, Long R M 1998 *Mol. Cell* **2** 437
- [86] Tantale K, Mueller F, Kozulic-Pirher A, Lesne A, Victor J M, Robert M C, Capozzi S, Chouaib R, Backer V, Mateos-Langerak J, Darzacq X, Zimmer C, Basyuk E, Bertrand E 2016 *Nat. Commun.* **7** 12248
- [87] Fukaya T, Lim B, Levine M 2016 *Cell* **166** 358
- [88] Corrigan A M, Tunnacliffe E, Cannon D, Chubb J R 2016 *eLife* **5** e13051
- [89] Chong S, Chen C, Ge H, Xie X S 2014 *Cell* **158** 314
- [90] Suter D M, Molina N, Gatfield D, Schneider K, Schibler U, Naef F 2011 *Science* **332** 472

- [91] Tripathi T, Chowdhury D 2008 *EPL* **84** 68004
- [92] Golding I, Paulsson J, Zawilski S M, Cox E C 2005 *Cell* **123** 1025
- [93] Raj A, Peskin C S, Tranchina D, Vargas D Y, Tyagi S 2006 *PLoS Biol.* **4** e309
- [94] Sanchez A, Golding I 2013 *Science* **342** 1188
- [95] Wang Y, Ni T, Wang W, Liu F 2019 *Biol. Rev.* **94** 248
- [96] Tunnaclyffe E, Chubb J R 2020 *Trends Genet.* **36** 288
- [97] Yao J 2017 *J. Mol. Biol.* **429** 14
- [98] Nicolas D, Phillips N E, Naef F 2017 *Mol. Biosyst.* **13** 1280
- [99] Hnisz D, Shrinivas K, Young R A, Chakraborty A K, Sharp P A 2017 *Cell* **169** 13
- [100] Chubb J R 2017 *Wiley Interdiscip. Rev.-Dev. Biol.* **6** e284
- [101] Bressloff P C 2017 *J. Phys. A-Math. Theor.* **50** 133001
- [102] Reinius B, Sandberg R 2015 *Nat. Rev. Genet.* **16** 653
- [103] Munsky B, Neuert G 2015 *Phys. Biol.* **12** 045004
- [104] Munsky B, Fox Z, Neuert G 2015 *Methods* **85** 12
- [105] Boeger H, Shelansky R, Patel H, Brown C R 2015 *Genes* **6** 469
- [106] Skupsky R, Burnett J C, Foley J E, Schaffer D V, Arkin A P 2010 *PLoS Comput. Biol.* **6** e1000952
- [107] Dar R D, Razooky B S, Singh A, Trimeloni T V, McCollum J M, Cox C D, Simpson M L, Weinberger L S 2012 *Proc. Natl. Acad. Sci. U.S.A.* **109** 17454
- [108] Molina N, Suter D M, Camavo R, Zoller B, Gotic I, Naef F 2013 *Proc. Natl. Acad. Sci. U.S.A.* **110** 20563
- [109] Corrigan A M, Chubb J R 2014 *Curr. Biol.* **24** 205
- [110] Giri R, Papadopoulos D K, Posadas D M, Potluri H K, Tomancak P, Mani M, Carthew R W 2020 *eLife* **9** e53638
- [111] Lammers N C, Galstyan V, Reimer A, Medin S A, Wiggins C H, Garcia H G 2020 *Proc. Natl. Acad. Sci. U.S.A.* **117** 836
- [112] Tian X, Huang B, Zhang X P, Lu M, Liu F, Onuchic J N, Wang W 2017 *Proc. Natl. Acad. Sci. U.S.A.* **114** 5337
- [113] Ko M S 1992 *Bioessays* **14** 341
- [114] Sanchez A, Choubey S, Kondev J 2013 *Methods* **62** 13
- [115] Peccoud J, Ycart B 1995 *Theor. Popul.* **48** 222
- [116] Pedraza J M, Paulsson J 2008 *Science* **319** 339
- [117] Freeman B C, Yamamoto K R 2002 *Science* **296** 2232
- [118] Stavreva D A, Muller W G, Hager G L, Smith C L, McNally J G 2004 *Mol. Cell. Biol.* **24** 2682
- [119] Wang Y, Qi J, Shao J, Tang X Q 2020 *Biology* **9** 339

INVITED REVIEW

Transcription apparatus: A dancer on a rope^{*}

Wang Yaolai^{1)†} Liu Feng^{2)‡}¹⁾ (*School of Science, Jiangnan University, Wuxi 214122, China*)²⁾ (*Department of Physics, Nanjing University, Nanjing 210093, China*)

(Received 30 September 2020; revised manuscript received 11 November 2020)

Abstract

Laws of physics govern all forms of matter movement. However, lives, which are composed of chemical elements which everyone is familiar with, are largely beyond physical description available. This is because the construction of life is not the same as that of general matters, rendering it unknown how physics laws are utilized. In this paper, we present our thinking on the transcriptional apparatus (TA). The TA is a huge molecular machine acting to sense regulatory signals and initiate transcripts at right time and with right rate. The operation of the TA is fundamental to almost all forms of lives. Although great progress has been made in recent years, one often has to face contradictory conclusions from different studies. Additionally, the studies of transcription are divided into several fields, and different fields are increasingly separate and independent. Focusing on eukaryotic transcription, in this review we briefly describe major advances in various fields and present new conflicting view points. Although the structural studies have revealed the main components and architecture of the TA, it is still unclear how the Mediator complex transmits signals from activators to the core transcriptional machinery at the promoter. It is believed that the Mediator functions to recruit RNA polymerase II onto the promoter and promote the entry into transcriptional elongation, which fails to explain how the signal transduction is achieved. On the other hand, the allosteric effect of the Mediator allows for signal transmission but is not supported by structural study. It is reported that enhancers, especially super enhancers, act to recruit activators via forming a so-called liquid drop and phase separation. By contrast, it is suggested that enhancers should cooperate delicately to orchestrate transcription. Results on the kinetics of protein-promoter interaction also contrast with each other, leading to a paradox called “transcriptional clock”. It is then concluded that proteins interact frequently and transiently with promoters and different proteins interact with the promoter at different stages of transcriptional progression. The phenomenon of transcriptional burst questions how the cellular signaling is achieved through such a noisy manner. While the burst frequency or size, or both are potentially modulated by transcriptional activators, more evidence supports the mode of frequency modulation. The technical difficulties in investigating the mechanism of transcription include 1) structural characterization of flexible and/or unstable proteins or protein complexes, 2) measurement of intermolecular kinetics, 3) tracking of single molecule movement, and 4) lack of methodology in theoretical research. We further propose a research strategy based on the ensemble statistical method, and introduce a model for how the TA dynamically operates. The model may act as a benchmark for further investigations. The operating mechanism of the TA should reflect an optimal use of physics laws as a result of long-term biological evolution.

Keywords: molecular structure, signal transduction, stochastic process, biological phenomena**PACS:** 87.10.-e, 02.50.Cw, 05.10.-a, 31.15.at**DOI:** 10.7498/aps.69.20201631

^{*} Project supported by the National Natural Science Foundation of China (Grant Nos. 11804123, 11831015, 11874209).

[†] Corresponding author. E-mail: yaolaiwang@jiangnan.edu.cn

[‡] Corresponding author. E-mail: liu@nju.edu.cn