

2015年度“中国生命科学领域十大进展”:剪接体的三维结构以及RNA剪接的分子结构基础研究



◀ 牵头科学家施一公,中国科学院院士,美国科学院院士、美国艺术与科学院外籍院士,教育部“长江学者”讲座教授,国家杰出青年科学基金获得者,“千人计划”首批国家特聘专家。现任清华大学副校长。曾获舍尔学者奖及雷拓爱伦学者奖、鄂文西格青年科学家奖、国际赛克勒生物物理学奖、香港求是基金会杰出科学家奖、谈家桢生命科学终身成就奖、瑞典皇家科学院爱明诺夫奖、吴阶平-保罗·杨森医学药理学奖。

✎ 中国科协生命科学学会联合体:在真核细胞中,蕴藏在基因组DNA序列中的遗传信息先传递给信使RNA,转录的信使RNA前体需经剪接体剪接成熟之后再翻译成蛋白质,执行生物学功能。剪接体是一个巨大而又复杂的动态分子机器。清华大学施一公课题组利用酵母细胞内源性蛋白提取获得了性质良好的样品,利用单颗粒冷冻电子显微镜技术,解析了酵母剪接体近原子水平的高分辨率三维结构,阐述了剪接体对信使RNA前体执行剪接的工作机理。这一研究成果于2015年9月在《Science》杂志以2篇“背靠背”的长文发表。

剪接体的三维结构及基因剪接的分子机理研究

杭婧,施一公

清华大学结构生物学高精尖创新中心,北京 100084

1 基因剪接是生命活动的重要过程

从花草鱼虫到飞禽走兽,大自然物种的丰富多样令人惊叹。虽然外表和形态千变万化,但是组成这些生命体的最基本单元——细胞有着最基本的共性:遗传和代谢。生命的延续本质是遗传信息的代代相传。基因表达保证了这些信息被翻译成千变万化的细胞活动,从而呈现出五彩斑斓的生命世界(图1)。执行细胞活动最主要的物质是蛋白质。基因表达,就是把储存在DNA中的遗传信息经过转录和翻译,转化为执行生物功能的蛋白质的过程,这是最基本的生命活动之一,因此被称为分子生物学的“中心法则”。

遗传信息的载体是遗传物质。绝大部分生命体,复杂如人类,简单如细菌,其遗传物质都是脱氧核糖核酸(DNA)。由数量、顺序大不同的ATCG组成的DNA序列就是蕴藏遗传信息的密码。有趣的是,人类有大概23000个基因,这个数量居然与低等的果蝇相近。这是为什么呢?答案在于基因剪接。

从DNA的遗传信息传递到蛋白质,中间要经过核糖核酸RNA。“转录”的过程

就是把蕴藏在基因组DNA中的遗传信息传递到RNA中。在基因表达过程中,RNA扮演了一个“信息使者”的角色。当RNA聚合酶以DNA为模板合成

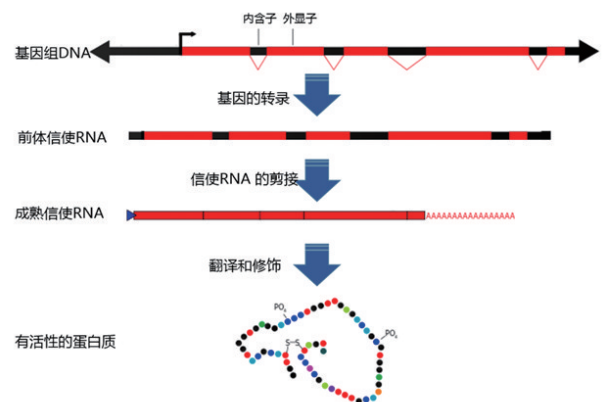


图1 基因表达图示

引用格式:杭婧,施一公.剪接体的三维结构及基因剪接的分子机理研究[J].科技导报,2016,34(13):15-19;doi:10.3981/j.issn.1000-7857.2016.13.002

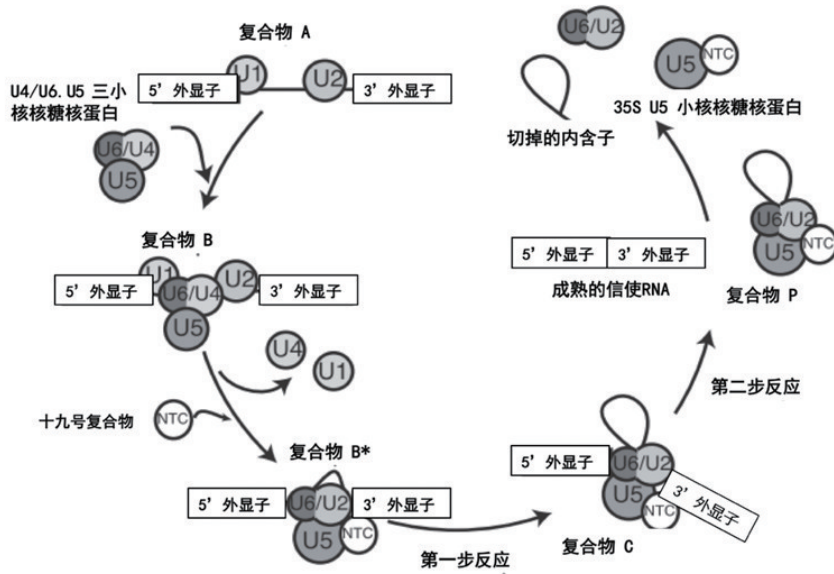


图2 基因剪接的基本流程

与之互补配对的RNA时,遗传信息就传递给了RNA。但是,在真核生物里,这样的RNA被称为前体信使RNA。当它把信息继续向下传递时,可不仅仅是简单的一一对应,而是如同布料被裁减为衣服,一条信使RNA可以变身成不同长度、不同序列的成熟信使RNA(mRNA)。这些mRNA最终被核糖体翻译成具有不同功能的蛋白质分子。从前体信使RNA到mRNA这个过程便是剪接(pre-mRNA splicing)(图2)。用学术语言来说,最终能够被翻译成蛋白质的序列叫编码序列,否则叫非编码序列。在真核生物里,编码序列和非编码序列交替存在,当信息传递到前体信使RNA上时,这些非编码区域(内含子)需要被剔除出去,编码区域(外显子)需要重新排列连接,形成成熟的信使RNA,被转运出细胞核,到细胞质中被核糖体识别并翻译成蛋白质^[1]。

大多数前体信使RNA有十几甚至几十个外显子,剪接过程并非简单地剔除内含子,而是把外显子进行复杂的排列重组,结果便是每个基因能够产生大约10个不同的蛋白。这一现象被称为可变剪接(alternative splicing)。这大约也就是为什么人类和果蝇的基因数目相差无几但是生命活动复杂得多的

原因。

由于前体信使RNA的剪接是真核生物基因表达的关键一步,其功能异常会导致严重后果。许多人类疾病都归咎于基因的错误剪接或是针对剪接体的错误调控^[2]。人类35%的遗传紊乱是由于基因突变导致单个基因的可变剪接引起的,如:单个剪接位点的增加或缺失可能引起 α -或 β -地中海贫血症;可变剪接平衡紊乱导致的某些外显子不正常表达可能导致额颞骨痴呆症。还有一些疾病的起因是剪接体蛋白的突变影响了许多转录本的剪接,比如一些剪接体关键蛋白的突变可能引起常染色体显性视网膜色素变性、慢性淋巴

细胞性白血病以及脊髓性肌肉萎缩。

除遗传性疾病外,癌细胞比正常细胞需要更加大量的剪接这一现象引起了研究者的高度关注。相比正常细胞,癌细胞的生长速度更快,因此需要更多的能量来源和更迅速的生理周期。想办法特异性地把癌细胞中剪接体的识别和剪接能力降低甚至消除是一些科学家正在努力的治疗癌症的方向^[3]。但是,和广泛使用的放疗、化疗一样,靶向剪接体的癌症治疗药物也有类似的副作用,即正常细胞也会受到一定程度的损伤,“杀敌一千,自损八百”,因此找到特异性更高、副作用更小的药物是科学家们致力的关键。

2 剪接的发现和前期研究

要完成对前体RNA如此复杂的剪接需要一个巨大且高度动态的分子机器——剪接体(spliceosome)。为了完成基因剪接这一使命,剪接体需要由总计逾百种蛋白质和RNA分子共同协作。其中包括5个亚复合物,称为小核核糖核蛋白(small nuclear ribonucleoproteins),简称U snRNPs。每个snRNP以1条小核RNA(U snRNA)为结合骨架,搭配十几个蛋白质,完成特定的功能。此外还有十九号复合物NTC、十九号复合物相关蛋白NTR等2个亚组分(图3^[5])。整个剪接过程是一个高度动态而又异常精确的过程,期间,剪接体内各组分以前体信使RNA分子为中心,各个亚复合物、多

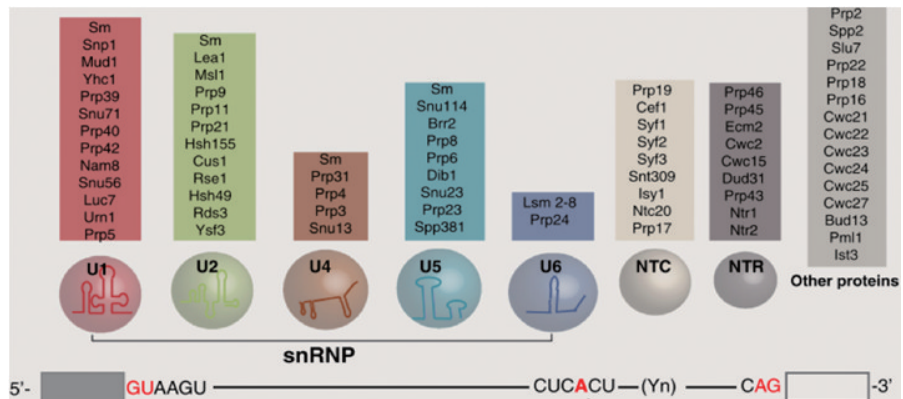


图3 剪接体主要由snRNP、NTC、NTR等组分组成

种因子按照特定顺序不断进行聚散,并发生大规模的结构重组,从而正确识别内含子与外显子的交接位点以及催化内含子的切除与外显子的拼接^[4]。

剪接体的发现始于20世纪70年代。自1953年遗传物质DNA双螺旋结构解析以来,研究者们试图解释中心法则,即基因表达的来龙去脉。人们发现,在低等单细胞生物——细菌中,DNA转录到RNA再翻译到蛋白,是一一对应的。根据物种进化的保守性,科学家们推断其他物种也是如此,哺乳动物之所以复杂是因为其拥有更庞大的DNA库。

这一论断被剪接现象的发现彻底推翻。1977年,在利用电子显微镜观察DNA和与之互补配对RNA时,2位科学家, Richard Roberts 和 Phillip Sharp 独立地发现了DNA要远远比与之对应的RNA长得多;这些长出来的DNA间歇性的以环状形式延伸出来^[6,7](图4(a)^[8])。此时2位科学家还没有相互了解到彼此的工作。不久,他们同时参加了美国冷泉港的一个会议,分别展示自己工作后才发现,他们讲的居然是同一件事。这次会议改变了一切,2位科学家也因此共同分享了1993年的诺贝尔生理学或医学奖。

20世纪70年代末、80年代初,剪接的普遍适用和剪接体的广泛存在通过体外实验^[8](图4(b)^[11])建立之后被大众所熟知。1983年,耶鲁大学科学家 Joan Steitz 首次分离出了剪接体组分 snRNP^[9]。所谓眼见为实,揭示剪接体组装和催化机理最直接有效的方法就是直观地揭示组装和催化过程中每一步所形成的复合物的结构,阐释剪接体复合物内部生物大分子的相互作用以及在不同催化进程下剪接体复合物的构象变化。2002年,德国马克斯·普朗克研究所的 Reinhard Lührmann 在《Science》杂志上发表了里程碑式的工作,展示了剪接体结构的高度柔性^[10]。

可正是这样的不稳定性、巨大的分子量和高度动态的化学成分和形态结构,给剪接体的结构生物学研究带来了巨大的挑战。长期以来,科学家们能够

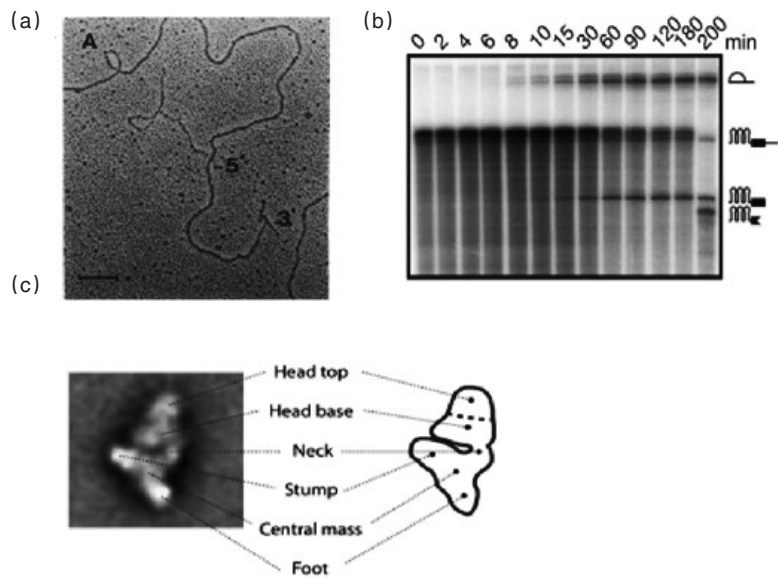


图4 电镜下的DNA:RNA配对链(a)、剪接反应的体外建立(b)及剪接体的低分辨率电镜结构(c)

获得的剪接体结构都维持在一个很低的分辨率范围内,如同百米之外观察一个人一样,只能大致推测其轮廓,无法分析精细结构(图4(c)),也无法看出这个人是在眨眼还是在张口。高分辨率结构的缺失严重限制了对于它们工作机制的理解。因此,剪接体的近原子分辨率研究被公认为结构生物学最难的课题之一,也是剪接领域最重要的挑战。

3 剪接体的原子结构

清华大学施一公2008年便致力于针对剪接体结构与工作机制的攻关。在2014年初首次获得了剪接体复合物中重要组成蛋白Lsm七聚体及其在RNA结合状态下的晶体结构。这一成果发表于《Nature》杂志^[12]。之后,研究组便聚焦于最富挑战性的攻坚课题:完整剪接体的结构生物学研究。在不断地探索和尝试中,最终通过酵母内源性提取,并利用先进的单颗粒冷冻电子显微镜技术,在世界上首次解析了近原子水平分辨率的剪接体三维结构^[13,14]。

为了获得性质良好的剪接体,研究组对传统的串联亲和和层析纯化方法进行了步骤简化和改良,省略了繁琐而又耗时的透析步骤,完成了裂殖酵母剪接体复合物的内源性提取(图5)。再通过合适的手段,提高复合物的浓度,使之能够适用于后续的冷冻电镜样品制备。

单颗粒冷冻电子显微镜技术指的是利用特殊的仪器把蛋白质包埋在非

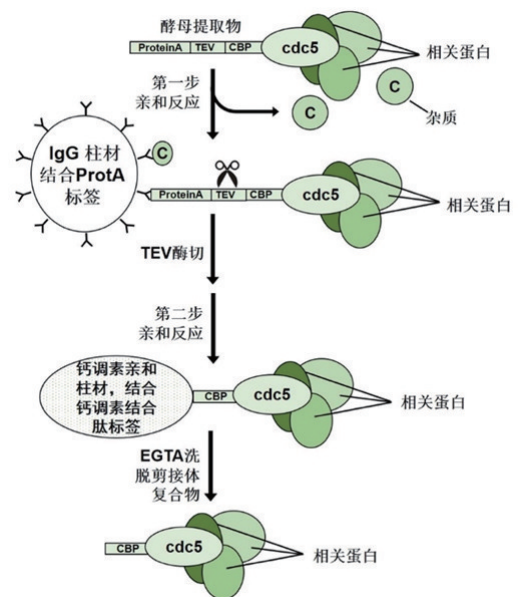


图5 酵母剪接体内源性提取示意

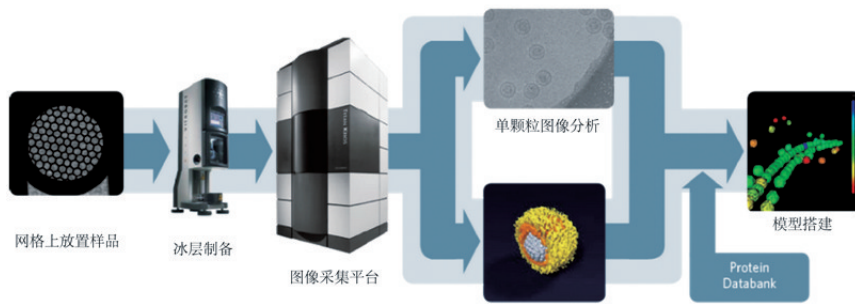


图6 单颗粒冷冻电镜技术工作流程

常薄的冰层内,最大程度保持蛋白的天然状态,再利用高分辨显微镜进行观察和大量的图像采集(图6)。采集后的图像可以用来数据分析和三维重构。但在2013年之前,结构生物学的研究主要集中在X射线晶体学方面,电镜的主要作用是对生物蛋白样品进行一个初步的鉴定和轮廓的观察,并不能够获得原子水平分辨率,因此没有被广泛使用。2013年底,美国加州大学洛杉矶分校(UCSF)科学家程亦凡首次报道了利用冷冻电镜技术解析的近原子水平分辨率的TRPV1蛋白结构。这一研究成果震惊了整个生物学界,这意味着电镜技术终于可以与X射线晶体学在生物大分子的研究中分庭抗礼了。这个革命源于电子直接探测相机的诞生和图像分析软件的更新。

施一公研究组在获得优质的酵母内源性剪接体后,利用最先进的冷冻电镜技术,最终获得了剪接体处在反应后期的高分辨率三维结构(图7),搭建了分辨率为3.6 Å的剪接体原子模型,其中包括37种蛋白质的10574个氨基酸和4条RNA分子,其整体分子量接近130万D。结构包括3个U snRNP、大部分NTC组分和一部分NTR组分。在模型中可以看出,剪接体的外形轮廓十分不对称,为便于描述,我们将其分为手臂I、手臂II、头部区域和中心区域4部分。其中手臂I主要由U2 snRNP构成,手臂II由NTC亚复合物组成。亚复合物之间相对独立但又密切联系,形成

了分子量和体积巨大的复合物。一些大分子量蛋白,如Spp42蛋白,如同“平台”一样,让其他分子依附在上面完成功能;而另一些蛋白则以延伸状形式存在,以prp45蛋白为例,类似“套索”将庞大的剪接体复合物固定成完整的整体。

与此同时,研究组对剪接体的3条U snRNA分子以及内含子的分支位点区域进行了细致的结构分析,搭建了前体信使RNA被剪切、连接的原子模型,揭示了剪接反应的催化中心构成和多个关键分子的作用方式,证明剪接体本质上是一个在蛋白质调控下的核酶(ribozyme),而蛋白质组分的作用在于帮助递送关键的RNA分子使其在合适

时间彼此接近从而实现剪接反应,为阐明剪接反应的分子机制提供了重要基础(图8)。

4 展望

对剪接体研究的突破不仅初步解答了基础生命科学领域长期以来备受关注的核心问题,也为进一步揭示与剪接体相关疾病的发病机理提供了坚实基础,并为后续的药物靶点研究和疾病成因研究提供结构基础和理论指导。

这是一个很好的开始,但离完全阐释剪接体的工作机理还很远,还有大量的问题亟待解决。首先是剪接体的不同阶段构象变化。任何一个机器在行使功能时都需要发生形态上的变化,才能完成底物到产物的转化。剪接体更是如此。在整个剪接过程中,亚基组分的不断被招募和解离是剪接反应得以进行的基础。

其次,研究重心由低等生物向高等生物转化,也是未来的趋势。生物在进化时会保留关键组分的保守性,进化越低等,相关调控会越简单。因而低等的模式生物是研究者们对于一些细胞通路研究的首选。但是,也正因为低等生物简单性,无法完全模拟人类细胞的真实情况,在获得酵母剪接体的基础上,对哺乳动物细胞剪接体的结构生物

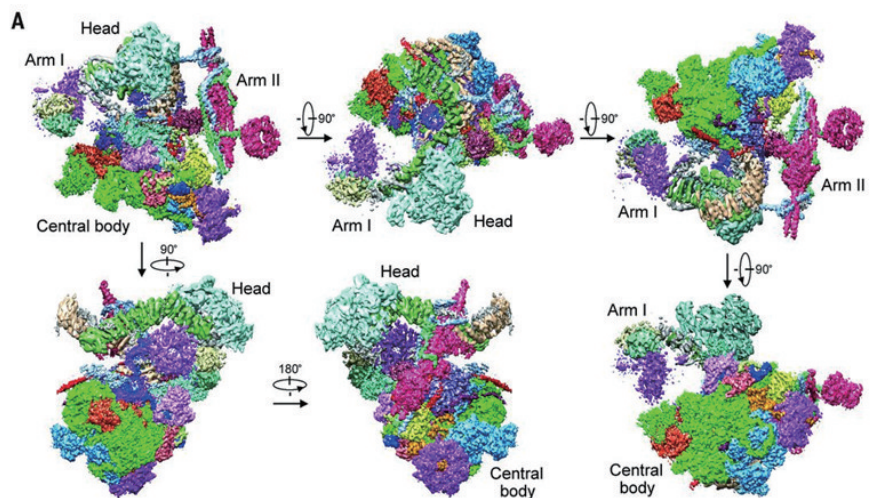


图7 完整剪接体的冷冻电镜结构

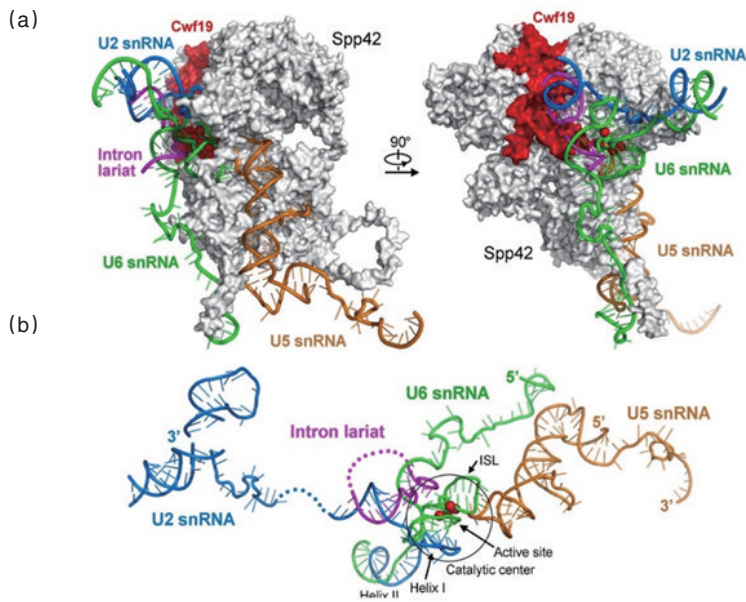


图8 RNA分子在整个剪接体中所在的位置和形状(a)及由RNA组成的反应活性中心(b)

学研究将成为下一个挑战。

最后,充分了解细胞对剪接体的调控机理,是未来药物研发的关键。在剪接体循环中,有8个蛋白通过水解ATP提供能量,起到了“发动机”的作用。这些ATP水解酶的突变或者失活会对剪接反应造成直接的影响。此外,还有许许多多调控因子参与到剪接位点的识别、和其他信号通路进行联络等多个环节,需要充分了解这些调节因子的作用机制和调节方法。而生命科学研究的根本目的之一是探索细胞生存的奥秘,从而揭示疾病的致病原因,发现甚至设计靶向药物,以兹为治疗相关疾病提供新线索、新思路、新靶点。

参考文献

- [1] Wahl M C, Will C L, Luhrmann R. The spliceosome: design principles of a dynamic RNP machine[J]. *Cell*, 2009, 136(4): 701-718.
- [2] Cooper T A, Wan L, Dreyfuss G. RNA and disease[J]. *Cell*, 2009, 136(4): 777-793.
- [3] Hsu T Y, Simon L M, Neill N J, et al. The spliceosome is a therapeutic vulnerability in MYC-driven cancer[J]. *Nature*, 2015, 525(7569): 384-388.
- [4] Will C L, Luhrmann R. Spliceosome structure and function[J/OL]. *Cold Spring Harbor Perspectives in Biology*, 2011, 3(7): doi: 10.1101/cshperspect.a003707.
- [5] Chen W, Moore M J. Spliceosomes[J]. *Current Biology*, 2015, 25(5): R181-183.
- [6] Berget S M, Moore C, Sharp P A. Spliced segments at the 5' terminus of adenovirus 2 late mRNA[J]. *PNAS*, 1977, 74(8): 3171-3175.
- [7] Chow L T, Gelinaz R E, Broker T R, et al. An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA[J]. *Cell*, 1977, 12(1): 1-8.
- [8] Padgett R A, Mount S M, Steitz J A, et al. Splicing of messenger RNA precursors is inhibited by antisera to small nuclear ribonucleoprotein[J]. *Cell*, 1983, 35(1): 101-107.
- [9] Hinterberger M, Pettersson I, Steitz J A. Isolation of small nuclear ribonucleoproteins containing U1, U2, U4, U5, and U6 RNAs[J]. *Journal of Biological Chemistry*, 1983, 258(4): 2604-2613.
- [10] Makarov E M, Makarova O V, Urlaub H, et al. Small nuclear ribonucleoprotein remodeling during catalytic activation of the spliceosome[J]. *Science*, 2002, 298(5601): 2205-2208.
- [11] Wolf E, Kastner B, Deckert J, et al. Exon, intron and splice site locations in the spliceosomal B complex[J]. *EMBO Journal*, 2009, 28(15): 2283-2292.
- [12] Zhou L J, Hang J, Zhou Y L, et al. Crystal structures of the Lsm complex bound to the 3' end sequence of U6 small nuclear RNA[J]. *Nature*, 2014, 506(7486): 116-120.
- [13] Hang J, Cai B, Xue P, et al. The joint effects of lifestyle factors and comorbidities on the risk of colorectal cancer: A large Chinese retrospective case-control study[J]. *PLoS One*, 2015, 10(12): e0143696.
- [14] Yan C Y, Hang J, Wan R X, et al. Structure of a yeast spliceosome at 3.6-angstrom resolution[J]. *Science*, 2015, 349(6253): 1182-1191.

(责任编辑 王媛媛)