



分裂内含肽技术的创新发展与前沿应用*

甘金秋^{1,2)} 邓祥宇²⁾ 王歆燕²⁾ 李佳斌^{1)**}⁽¹⁾ 苏州大学药学院, 苏州 215123; ⁽²⁾ 清华大学化学系, 北京 100084)

摘要 内含肽是一类能够进行自我剪接的蛋白质插入序列, 通过发生自我切除介导其两侧的外显肽以酰胺键连接。其中, 分裂内含肽是内含肽的独特类型, 其自剪接序列由来源于两个独立基因的片段组成, 通过非共价方式组装成活性结构而介导外显肽连接, 从而实现蛋白质的反式剪接。分裂内含肽介导的蛋白质反式剪接具有高度特异性与不可逆性, 已成为实现蛋白质无痕连接的强大工具。近年来, 针对分裂内含肽介导蛋白质剪接的技术革新, 不仅克服了该技术的一些固有局限, 而且进一步拓展了其应用范围, 已被应用于基因递送、蛋白质内部序列编辑、蛋白质相互作用检测、生物逻辑门控以及作用机制解析等方面。为此, 本文总结了近10年来有关分裂内含肽的技术创新, 并综述了分裂内含肽剪接技术在蛋白质前沿领域的创新应用。

关键词 分裂内含肽, 蛋白质反式剪接, 蛋白质连接工具, 化学生物学

中图分类号 Q71, Q812, Q816

DOI: 10.3724/j.pibb.2026.0043

CSTR: 32369.14.pibb.20260043

内含肽 (intervening proteins, inteins) 是一类具有自催化活性的特殊蛋白质序列, 可从前体多肽中自我切除, 并介导两侧外显肽通过天然酰胺键连接。作为 HINT (Hedgehog/INTein) 超家族的重要成员, 内含肽与刺猬蛋白的 C 端结构域以及细菌类内含肽 (bacterial intein-like, BIL) 共享保守的核心结构域^[1-2], 且具备自催化活性。该剪接反应无需任何辅因子、伴侣或能量源, 完全由其自身独特的氨基酸序列介导完成。自首个内含肽 Sce VMA1 的基因于 1990 年在酿酒酵母 (*Saccharomyces cerevisiae*) 编码液泡 H⁺-ATP 酶 α 亚基 1 (vacuolar membrane H⁺-adenosine triphosphatase, VMA1) 的基因中被发现以来^[3-4], 科研人员已在古菌、细菌、真核生物及病毒基因组中发现上千种编码内含肽的基因^[5], 例如异核枝杆菌的 DNA 促旋酶亚基 A 基因 (*Mycobacterium xenopi* DNA gyrase subunit A gene, GyrA) 编码的内含肽 Mxe GyrA^[6], 以及古细菌热球菌属 (*Thermococcus litoralis*) 的 DNA 聚合酶基因编码的内含肽 I-TliI^[7] 等。

基于结构与剪接方式的差异, 内含肽可被划分为全长型 (full-length, FL)、微小型 (mini) 及分

裂型 (split) 3 种形式 (图 1a)。其中, 全长内含肽与微小内含肽的核心区别在于是否包含归巢核酸内切酶 (homing endonuclease, HEN) 结构域, 二者均属于顺式剪接类型^[8], 通过单一前体多肽链内的自催化反应完成剪接 (图 1b)。与两者不同, 分裂内含肽进行反式剪接, 其自剪接序列由两个不同的基因独立翻译合成 (N 端部分缩写为 IntN, C 端部分为 IntC), IntN 和 IntC 通过高亲和力的蛋白质-蛋白质相互作用 (protein-protein interaction, PPI)^[5] 而自发组装成活性结构, 进而发生自剪切反应并介导其两侧外显肽的精准连接 (图 1c)。相较于全长型与微小型的整体内含肽, 分裂内含肽具备更强的可控性、灵活性以及更出色的功能表现^[9-10], 因此受到广泛关注。分裂内含肽早期主要通过人工断裂整体内含肽进行理性设计, 在 HEN 结构域内部或其经典插入位点 (C35 位点) 断裂获得^[11]。然而, 此类人工分裂内含肽由于暴露疏水

* 国家重点研发计划 (2022YFA0913200), 国家自然科学基金 (22477092) 和江苏省基础研究计划 (BK20250150) 资助项目。

** 通讯联系人。

Tel: 0512-65882089, E-mail: lijiajin@suda.edu.cn

收稿日期: 2026-01-26, 接受日期: 2026-04-22

斑块, 往往存在折叠效率低以及剪接产率差等问题^[12]。针对此类内含肽的性能优化主要集中在通过截短片段改善化学合成可及性^[13-14], 但也未能有效解决上述短板。1998年Liu等^[15]在蓝藻(*Synechocystis* sp. PCC6803)的DNA聚合酶III催化亚基中首次鉴定出天然分裂内含肽Ssp DnaE, 避免了人工分裂内含肽使用时需要复性的局限性^[16]。此后, Npu DnaE等超快剪接速率的天然分裂内含肽相继被鉴定^[17], 如GP41-1^[18]是目前已鉴定的具有最快剪接速率的分裂内含肽, 能够在数秒内完成剪接反应, AceL-TerL^[19]则能够在低温(8℃)下进行高效剪接, $t_{1/2}$ 为7 min。分裂内含肽介导的蛋白质反式剪接(protein trans-splicing, PTS)已

在蛋白质半合成^[11, 20]、蛋白质和多肽环化^[21-22]以及条件蛋白质剪接(conditional protein splicing, CPS)^[23-25]等领域展现出强大的应用潜力。另一方面, PTS技术的持续应用拓展也使得其自身存在的问题(如: 对外显肽序列的依赖性、较差的溶解度以及相互之间的交叉反应等)逐渐出现^[11, 16], 并进一步推动了对分裂内含肽的技术创新和性能拓展。本文聚焦近10年来该技术在工程化改造中的创新发展, 阐述其在基础生物学研究、生物技术开发及疾病治疗等领域的新应用模式, 旨在展现分裂内含肽技术的发展全貌, 为其后续的技术创新与应用拓展提供参考。

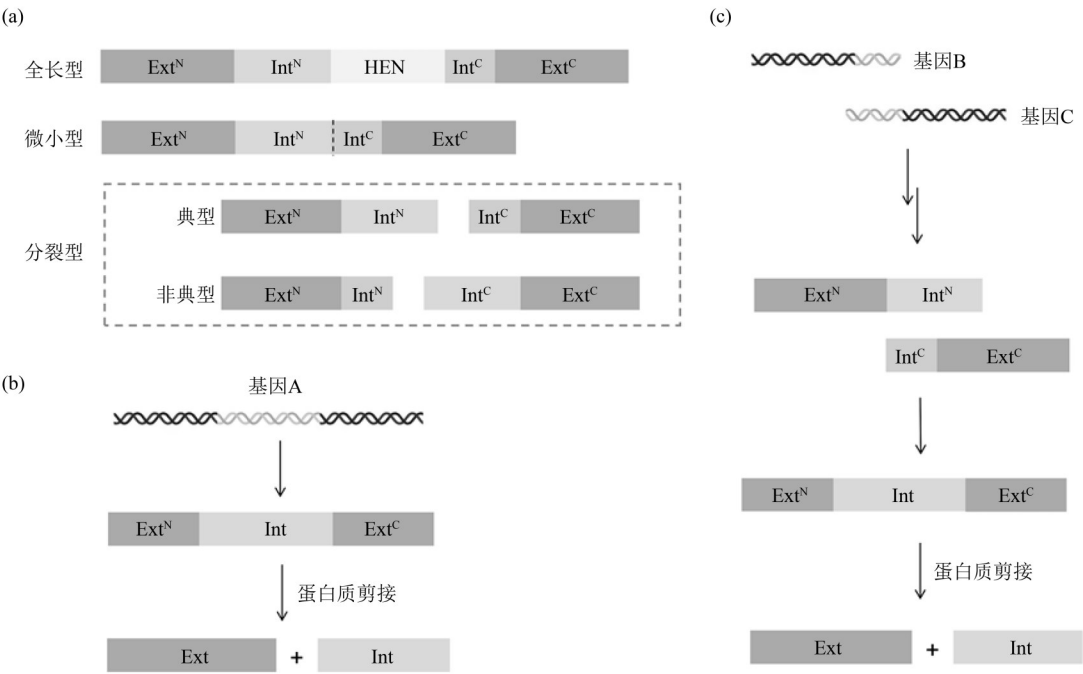


Fig. 1 The types of inteins and their splicing mechanisms

图1 内含肽的种类与剪接方式

(a) 根据结构和剪接方式的差异, 内含肽可分为全长型、微型型和分裂型。全长型和微型型内含肽的主要区别在于是否包含归巢核酸内切酶(HEN)结构域。(b) 顺式剪接过程, 外显肽(Ext^N、Ext^C)和内含肽(Int)由单一前体基因转录翻译而来, 随后通过自剪切反应拼接外显肽(Ext)。(c) 反式剪接反应, 涉及两个独立基因的翻译产物, 自发组装后发生自剪切反应, 实现两侧外显肽连接。

1 分裂内含肽的作用机制与改性基础

1.1 分裂内含肽的结构特征

分裂内含肽是整体内含肽的分裂形式, 即自剪接序列被分裂成Int^N和Int^C两个结构域。两个结构域单独存在时不具备剪接活性, 但能通过分子识别自发组装成具有活性的完整结构, 即Hedgehog/

INTein (HINT) 褶皱, 该马蹄状结构是催化活性的结构基础^[26-27]。该结构在各种内含肽之间高度相似, 主要由β链、环(loop)和两个短螺旋组成, 具有伪二轴对称性(图2)。一些来自古菌的内含肽, 如: 詹氏甲烷球菌转录因子IIB (*Methanocaldococcus jannaschii* transcription factor IIB, *Mja*TFIIB)^[28]、深渊热球菌细胞分裂控制蛋白

21-1 (*Pyrococcus abyssi* cell division control protein 21-1, PabCDC21-1)^[29]等, 其HINT域的边缘区域通常存在额外的 β 链, 能够在极端环境(如: 高

温)下发挥稳定HINT域的作用^[26]。HINT域通常具有动态性, 其局部构象能随剪接反应的进行而发生持续变化^[30]。

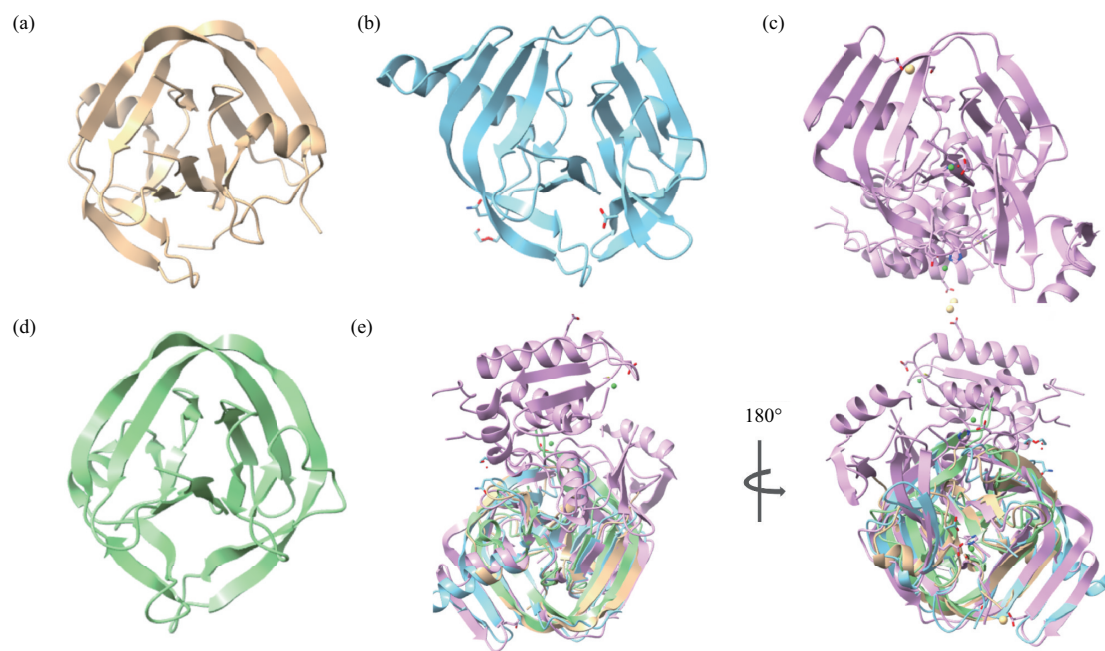


Fig. 2 Structural features of naturally occurring inteins

图2 天然内含肽的结构特征

(a) Npu DnaE典型分裂内含肽的结构(PDB ID: 4LX3)^[31]; (b)来自古菌的PabCDC21-1内含肽结构(PDB ID: 6RPP)^[29]; (c) PhoVMA全长内含肽的结构(PDB ID: 7QST)^[32]; (d) Spl DnaX微小内含肽的结构(PDB ID: 7CFV)^[33]; (e)上述4种内含肽的HINT域的结构叠加显示其HINT域高度相似。

根据分裂位点的不同, 分裂内含肽可以分为典型与非典型两类(图1a)。前者的分裂位点靠近C端, 如GP41-1的天然分裂位点位于HINT褶皱中伪二轴对称相关的环区(C35位点附近)^[34]; 后者的分裂位点则靠近N端, 如AceL-TerL内含肽的天然分裂位点位于N端第24、25位氨基酸对应的表面环区^[12]。此外, 所有分裂内含肽的活性中心均含有高度保守的氨基酸基序, A、B、F、G域。A域第1位(A:1)一般是半胱氨酸(cysteine, Cys)或丝氨酸(serine, Ser); B域含保守的TXXH基序, 其中苏氨酸(threonine, Thr)和组氨酸(histidine, His)残基分别位于B:7和B:10位; F域的核心区域存在一段由数个酸性残基及疏水性残基组成的混合片段, 该片段的两端各分布有一个芳香族残基^[35]; G域通常以组氨酸-天冬酰胺(asparagine, Asn)结尾, 少数情况下倒数第2位残基为甘氨酸(glycine, Gly)或苯丙氨酸(phenylalanine, Phe), G域下游第1位外显肽(+1

位)残基则多为Cys、Ser或Thr^[35]。基于保守基序的组成差异, 分裂内含肽可进一步分为3类。第1类内含肽基本含有这些保守基序, 第2类、第3类通常缺少A:1位置的Cys或Ser, 但第3类额外含有非连续的色氨酸-半胱氨酸-苏氨酸(Trp-Cys-Thr, WCT)保守基序。基序中的3个残基分别定位于B:12、F:4及G:5位置^[36](图3)。值得注意的是, 这些序列和结构特征与分裂内含肽的剪接反应路径直接相关。

1.2 分裂内含肽的蛋白质剪接机制

1.2.1 分裂内含肽的分子识别过程

分裂内含肽的剪接过程大致可以分为分子识别与蛋白质剪接两个主要过程。在识别过程中, 分裂内含肽的IntN和IntC结构域相互结合, 通过缠绕使得催化残基与末端靠近, 构成多个由保守序列构成的类蛋白酶活性位点, 负责剪接过程中各步化学反应^[30]。典型分裂内含肽(以Npu DnaE为例)的IntN包含两个结构不同的区域, 一个是具有明确结

	A域	B域	F域	G域	
第1类	C/S	TXXH		H/G/F N	C/S/T
第2类		TXXH		H/G/F N	C/S/T
第3类		TXXH W	C	T H/G/F N	C/S/T
...-2 -1 A:1		B:7 B:10 B:12	F:4	G:5 G:6 G:7 +1 +2 ...	

Fig. 3 Sequence features of the three types of inteins

图3 三类内含肽的序列特征

图中展示了三类内含肽的保守基序及其在A、B、F、G域中的分布。-1表示N端外显肽的最后一个残基，+1表示C端外显肽的第一个残基，A : 1表示A域的第一个残基。

构的有序区域，另一个是呈现负电荷的无序区域，采用“Capture and Collapse”的机制识别IntC并进行组装^[37]。首先，IntN中的负电无序区域通过静电相互作用识别并结合带正电的IntC；这种初始结合形成的复合物会进一步转化为具有部分天然Npu结构特征的中间体；随后，中间体经历坍塌形成HINT域，该过程由IntC与IntN有序部分之间的疏水作用稳定（图4a）。

相比之下，非典型分裂内含肽，如共识非典型内含肽（consensus atypical, Cat），不经历典型分裂内含肽的相互缠绕过程，其HINT域的形成需要ExtN从IntC中穿行^[38]，且因原本带电的无序N端结构域被转移至C端片段，因此该重组过程主要依赖疏水作用。非典型分裂内含肽（如Cat）的C端结构域由3个交替排列的结构区域组成：无序的靠近N端区域、局部有序的中央核心区域以及无序的

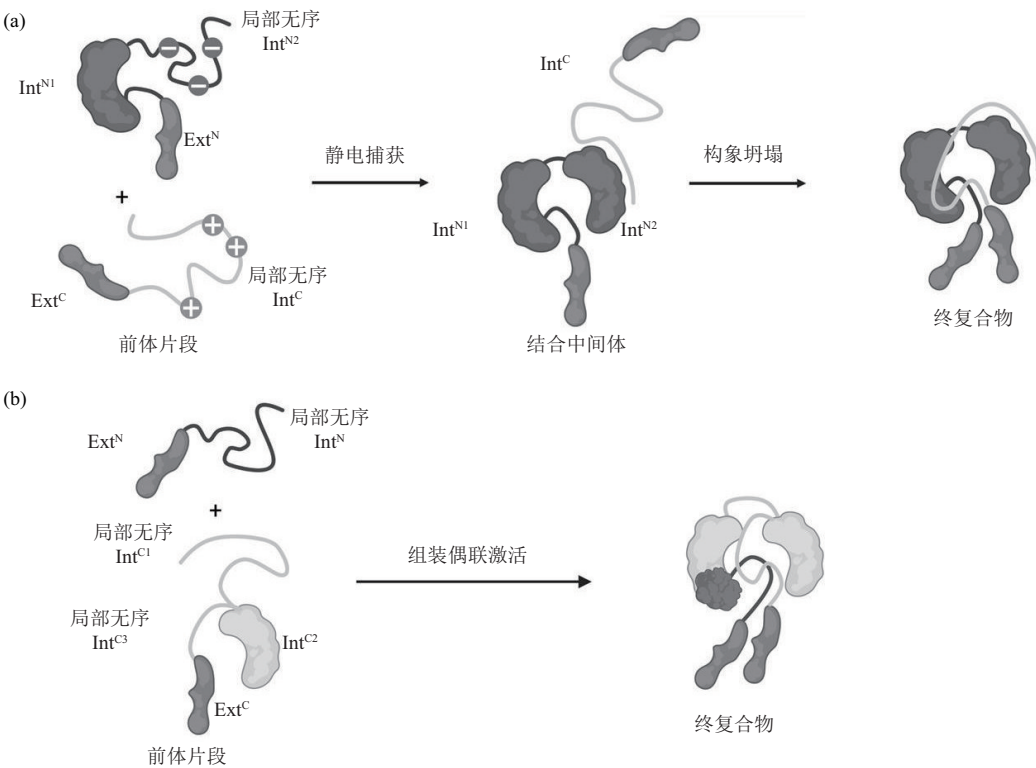


Fig. 4 Recognition mechanism between the two fragments of the intein

图4 内含肽识别机制

(a) 典型分裂内含肽“Capture and Collapse”识别机制。带正电的分裂内含肽C端结构域（Int^C）通过静电相互作用识别N端结构域（Int^N）中的负电无序区域并结合，形成具有部分天然Npu结构特征及分子间相互作用的中间体，随后该中间体经历坍塌形成HINT域。(b) 非典型分裂内含肽识别机制，该过程主要依赖疏水作用。图中Ext^N、Ext^C为N端与C端外显肽。

靠近C端区域(图4b),这一结构确保了与目标肽片段(如CatN)结合时提供更强的亲和力和稳定性^[38]。

两类内含肽在结合前通过调整自身局部结构来减少副反应并促进分离片段之间相互作用,并在结合后通过结构重排形成HINT结构域,从而激活蛋白质剪接反应。

1.2.2 分裂内含肽的剪接反应路径

在分子识别的基础上,分裂内含肽的剪接过程主要存在3类反应路径(图5)。第1类是典型的剪接反应。首先A:1的Cys/Ser与-1位的羰基碳发生N-S/O的酰基迁移反应,在N端剪接位点形成硫酯/氧酯键,完成酰基重排。第二步是+1位的Cys/Ser/Thr进攻酯键,将N端外显肽转移到C端剪接

位点处,形成支链酯键中间体。第三步是G域最后一位Asn的环化,使内含肽释放。第四步是酰胺键的形成,即硫酯/氧酯中间体再次发生S/O-N酰基转移形成酰胺键。在第2类剪接反应中,直接由+1位Cys的巯基进攻N端剪接位点的酰胺键,典型的代表是KlbA家族的内含肽^[39]。第3类剪接反应则由WCT基序中位于F域的Cys发挥第一类中A:1位置的Cys/Ser的类似作用,对N端剪接位点酰胺键发起进攻。值得注意的是,A:1位置是否存在Cys/Ser并不能作为第一类内含肽的判断标准,如分枝杆菌噬菌体Catera中的Gp206内含肽(mycobacteriophage Catera Gp206 intein, MP-Catera Gp206)^[36],在A:1位置存在Ser,但采用第2类剪接反应。

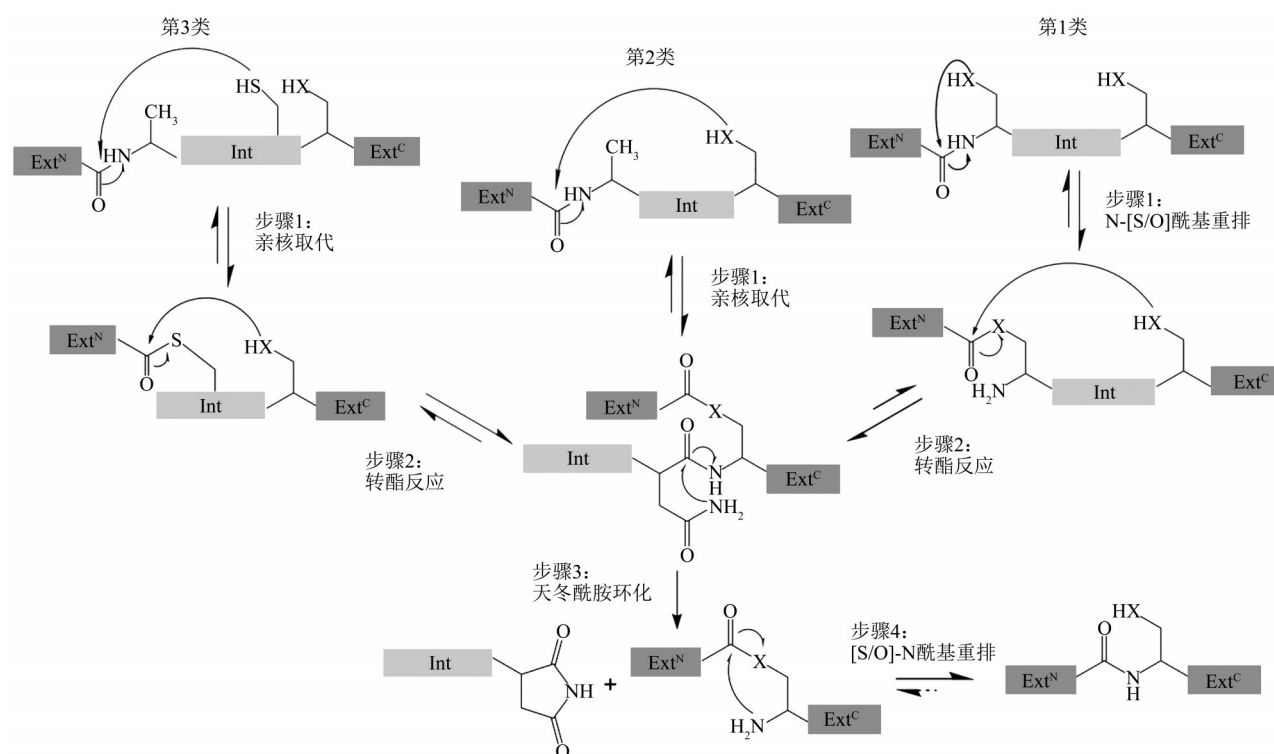


Fig. 5 Schematic of the three types of splicing reactions

图5 三类剪接反应示意图

图中展示了3类不同的剪接反应路径,其中Ext^N、Ext^C为外显肽的N端与C端,Int为内含肽。

1.3 非常规剪接反应

除发生外显肽连接的常规剪接反应外,分裂内含肽在特定条件下也可能发生副反应,即仅发生N端或C端断裂,单独释放某一端的外显肽。这种N端或C端的断裂过程表现出对温度^[40]、pH^[41]的敏感性。外显肽残基通过影响两个剪接位点的精确

对齐以及反应基团的空间取向,从而干扰高度协调的剪接过程。对外显肽特定的残基进行点突变,可以选择性调控某一端剪切反应的速率。N端切割主要受-1位外显肽残基的影响^[42-43],而C端切割普遍受+1位外显肽残基的影响^[44]。目前常用的诱导C端裂解的方式是将A:1的Cys突变为丙氨酸

(alanine, Ala), 当接触 IntN 后可介导 C 端外显肽自动断裂, 但也存在 IntC 无需与 IntN 接触而独立发生 C 端断裂的情况^[45]。

2 分裂内含肽的工程化新发展

早期获取分裂内含肽的核心策略集中于对整体内含肽的人工分割, 常见方式包括: a. 通过在核酸内切酶结构域进行分割^[46]; b. 根据计算机预测结果或已经获得的晶体结构为参考进行分割^[13, 47]; c. 借鉴已知分裂内含肽的保守分裂位点。应用这些方案成功获得了 Ssp DnaB M86^[48]、Mtu RecA^[49] 以及 Psp Pol-1^[50] 等分裂内含肽, 但由于这类分子未经过长期自然进化的筛选, 存在溶解性差、副反应占主导等问题, 在活性方面被认为劣于天然的分裂内含肽^[26, 51]。

目前聚焦于在已有分裂内含肽的基础上开展优化设计, 主要分为定向进化和基于构效关系的理性设计两种思路。定向进化策略通过随机诱变与功能筛选的迭代循环, 以结果为导向的设计过程使其展现出较高的定制化程度。通过易错聚合酶链式反应 (polymerase chain reaction, PCR) 获得突变文库, 结合卡那霉素抗性筛选与体外噬菌体展示技术, 获得了一系列对外显肽序列容忍度更高、剪接速率更快、pH 温度敏感的特异工程化分裂内含肽变体^[48, 52-53]。随着对分裂内含肽的结构特征、保守基序功能及剪接反应动力学机制认知的深化, 基于构效关系的理性设计近年来逐渐成为主要的技术路径, 通过挖掘天然分裂内含肽的序列与结构信息, 实现了对其核心性能的精准调控。

2.1 降低聚集倾向

许多在核酸内切酶位点被人工切割的内含肽在细胞中共表达时显示出良好的剪接活性, 但当分别表达和纯化两个片段以进行体外组装时, IntN 结构域无序区易形成富含 β 折叠的高分子质量的可溶聚集物, 导致具有活性的可溶单体量急剧下降, 显著降低剪接效率, 已成为大多数典型分裂内含肽应用的主要障碍^[51]。

在非典型分裂内含肽中, 原本来自 IntN 的无序且带相反电荷的区域被整合到较长的 IntC 片段时, 能够通过“分子内饱和”抑制聚集^[38]。Mootz 等^[54] 通过移动分裂位点, 在典型分裂内含肽上实现了类似的减少聚集效果。Aes123 PolB1 的分裂位点位于第 120 位和第 121 位氨基酸之间, 所得 N 端片段 (Aes^N, 120 个氨基酸) 和 C 端片段 (Aes^C,

39 个氨基酸) 的剪接反应效率并不理想, 可能由于易错误折叠而导致 Aes^N 的转化率不超过 40%。为提升反式剪接的效率, 他们将分裂位点移动至第 26 位氨基酸处, 并删除可能导致其错误折叠的第 27~33 位氨基酸序列, 从而获得该内含肽的无半胱氨酸 (cysteine-less, CL) 变体以及制备对应的 N 端和 C 端片段 (CL^N 和 CL^C)。这种工程化改造将天然 Aes 内含肽转化为能高效折叠的变体 CL, 其 N 端与 C 端片段混合后, 可几乎完全自发结合形成活性构象, 而无需经历变性-复性步骤^[50], 且缩短后的 CL^N 可通过固相合成轻松制备。然而, 该策略的代价是导致 CL 的 PTS 速率大幅降低。

近期, Mootz 等^[51] 深入探究 Aes123 PolB1 的分子机制, 通过结构解析与计算模拟发现, 其 N 端片段 (Aes^N) 中存在的无序结构区域是驱动 β 折叠介导聚集的关键。基于生物信息学网络工具 AMYLPRED2^[55] 的序列分析, Mootz 等预测出在第 60 到 120 位氨基酸之间存在 3 个聚集倾向区域 (aggregation-prone regions, APRs), 随后将 APRs 中易形成 β 折叠和疏水相互作用的残基替换为极性 or 带正电的氨基酸, 构建了将 3 个氨基酸 (T69K/F75H/M118N) 进行替换的突变体 CL_m。该突变体的 N 端片段 (CL_m^N) 纯化后单体比例 >95%, 显著改善了野生型 Aes^N 的聚集倾向 (聚集率 ~71%), 同时保留了超快速剪接特性 ($t_{1/2}$ = 16.3 s), 其剪接产率可达 89%。这种基于序列信息预测 APRs 的策略具有普适性, 仅需输入内含肽的氨基酸序列, 即可借助 AMYLPRED2 等工具定位 APRs, 后续结合定点突变实现对聚集的抑制。研究团队通过该方法对人工分裂的 Ssp DnaB 内含肽、天然分裂的 Npu DnaE 内含肽进行类似改造, 均成功降低了其体外聚集率^[51]。

2.2 提升剪切效率

Muir 等^[56] 基于共识设计 (同源蛋白质序列的进化信息预测稳定突变) 开发出高性能的共识快速 (consensus fast, Cfa) DnaE 分裂内含肽, 在不同温度下均展现出优于传统快剪接内含肽 Npu DnaE 的性能。他们首先通过比较天然“快” (Npu DnaE) 和“慢” (Ssp DnaE) 内含肽的序列特征, 结合嵌合体构建和批量诱变技术, 鉴定出位于活性位点附近的 4 个关键“加速”残基 (K₇₀、M₇₅、M₈₁、S₁₃₆), 基于此在美国联合基因组研究所 (Joint Genome Institute, JGI) 和美国国家生物技术信息中心 (National Center for Biotechnology

Information, NCBI) 数据库中筛选出 73 个天然存在且预测为“快”的 DnaE 内含肽序列, 经过多序列比对 (multiple sequence alignment, MSA) 选取每个位点出现频率最高的氨基酸构建出共识序列 Cfa DnaE。Cfa DnaE 展现出超快的剪接速率, 30°C 下的 $t_{1/2}$ 仅为 20 s, 并且能在高浓度变性剂 (4 mol/L 盐酸胍或 8 mol/L 尿素) 下高效工作, 显著拓展了蛋白质反式剪接技术的适用范围。利用该内含肽, Liu 等^[57] 在 2023 年提出了骨架安装的分裂内含肽辅助连接技术 (backbone-installed split intein-assisted ligation, BISIAL), 在 8 mol/L 尿素的强变性条件下实现了难溶性 D 型膜蛋白片段在低浓度 ($\mu\text{mol/L}$) 下的高效连接。

另一个通过共识设计提高剪切效率的案例是非典型分裂内含肽 Cat-TerL^[38]。通过对宏基因组数据库中使用基本局部比对搜索工具 (basic local alignment search tool, BLAST) 发现的 TerL 内含肽进行 MSA, 得到的 Cat-TerL 表现出超快剪接活性 (30°C 时 $t_{1/2}=59$ s), 并在温度及变性剂耐受方面显著优于天然同源物变体 AceL*。

2.3 降低外显肽序列依赖性

目前广泛使用的内含肽普遍对剪切位点侧翼的 3~5 个氨基酸残基 (尤其 +2 和 +3 位置^[5]) 存在序列依赖性, 即需要特定的序列才有较高的剪接效率^[58]。例如, DnaE 型内含肽的剪接活性高度依赖于 +2 位的疏水氨基酸残基^[59-60]。这种序列依赖性导致在应用内含肽时需引入特定的连接位点或对目标蛋白质的氨基酸序列进行突变, 而此类操作可能会影响目标蛋白质的功能。近期结果表明, 通过理性设计降低外显肽在内含肽剪接过程中的参与度, 可增强其对外显肽序列变化的耐受性 (表 1)。

研究表明, Npu DnaE 的 His₁₂₅ 残基可催化最后一步天冬酰胺环化过程, His₁₂₅ 与 +2 位的芳香族氨基酸存在 Π - Π 堆积相互作用, 而位于 +2 位的小位

阻氨基酸 (如: Ala) 会导致 His₁₂₅ 的侧链更具动态性, 表现出更多不利于天冬酰胺环化的构象^[61]。基于该机制分析, Muir 等^[62] 对 Npu DnaE 的 His₁₂₅ 所在环 (loop) 进行饱和诱变, 以改变 His₁₂₅ 的构象动力学, 并在 +2 位存在不利天冬酰胺环化的甘氨酸的情况下进行卡那霉素筛选。通过重构氨基糖苷磷酸转移酶 (卡那霉素抗性), 将剪接活性与大肠杆菌的抗生素耐药性耦合, 获得了一种外显肽残基依赖性更小的 Npu 变体 (Npu_{GEP}), 该变体在环区 (残基 122~124) 的甘氨酸-谷氨酸-脯氨酸基序突变显著增强了一系列 +2 位外显肽为小位阻或极性氨基酸的剪接效果, 如: Ala、Gly、精氨酸 (arginine, Arg)、天冬氨酸 (aspartate, Asp) 等。相同的环区突变也降低了 DnaE 家族中另一成员 Cfa 的外显肽依赖性 (Cfa_{GEP})。随后, Muir 等^[63] 将该策略应用至内含肽 Cat, 获得对外显肽在 -1 位的氨基酸残基有着增强耐受性的 Cat* (含 S70T 和 S79P 突变), 对该位置的多种类型氨基酸 (如: Ala、Gly、Ser 和 Pro 等) 均有较高的剪接效率。随后, 他们通过分子动力学模拟发现, Cat* 可通过增强 His₇₈ 邻近环区 (残基 78~90) 的柔性, 补偿天然外显肽缺失的空间约束, 推动 His₇₈ 自发翻转至维持催化所需的“近端构象”。

与 Npu DnaE 和 Cat 的外显肽依赖性源于外显肽的氨基酸侧链对 His 残基的稳定作用不同, GP41-1 通过其内部 THN 基序 (残基 123~125) 独特的“偏转构象”, 使 Thr₁₂₃ 的羟基与 His₁₂₄ 的主链形成稳定氢键, 从而将催化天冬酰胺环化的关键氨基酸 His₁₂₄ “锚定”在活性中心附近, 同时使 THN 基序整体向外偏转, 解除 Asp₁₀₇ 对 C 端切割的速率限制 (Asp₁₀₇ 可与 Asn₁₂₅ 形成氢键而抑制后者环化)^[58], 此时外显肽与活性中心的距离过近反而会干扰 His₁₂₄ 与 Thr₁₂₃ 的相互作用。Yang 等^[58] 通过在 GP41-1 引入 D107G 突变 (GP41-1_{D07G}) 增强了 F 域

Table 1 Extein-dependent improvement of split inteins

表1 分裂内含肽的外显肽依赖性改善表

内含肽种类	外显肽位置	初始可耐受的氨基酸 ¹⁾	拓展后可耐受的氨基酸	参考文献
Npu _{GEP}	+2	Phe ²⁾	Phe、Ala、Asp、Thr、Cys、Val	[62]
Cfa _{GEP} ^c	+2	Phe ^b 、Glu、Ser、Val、Thr、Ala	Phe、Arg、Glu、Ser、Val、Thr、Ala、Cys、Lys	[62]
Cat*	-1	Glu ^b 、Ser	Glu、Gly、Arg、Ala、Ser	[63]
GP41-1 _{D107G} ³⁾	+1	除 Asp/Glu/Pro	除 Pro	[58]

¹⁾ 表中可耐受的氨基酸指: 当外显肽该位置为此氨基酸时, 内含肽的剪接速率常数不低于该位置天然氨基酸剪接速率常数的一个数量级, 或产率不少于该位置天然氨基酸产率的 50%。例如, 当天然外显肽的剪接速率常数为 10^{-2} 时, 能够耐受的外显肽氨基酸的剪接速率常数至少应为 10^{-3} 。²⁾ 对应未改造前的分裂内含肽在该位置的天然氨基酸。³⁾ 通过产率评估的外显肽耐受性。

的柔性, 让 Asn₁₂₅ 更易环化形成琥珀酰亚胺, 大幅提升了+1位外显肽残基的耐受性, 如: 野生型 GP41-1 在+1位为 Asp/Glu/Pro 时剪切效率很低, 而突变后则仅不耐受 Pro, 且在融合绿色荧光蛋白 (green fluorescent protein, GFP)、谷胱甘肽 S-转移酶 (glutathione S-transferase, GST) 以及麦芽糖结合蛋白 (maltose-binding protein, MBP) 等不同目标蛋白质时均保持高效剪接效率。

2.4 正交分裂内含肽的开发

分裂内含肽的正交性, 指其仅能与自身对应互补片段配对并发生剪接反应, 而与其他分裂内含肽的片段无交叉反应, 这是应用分裂内含肽实现多个蛋白结构域间选择性拼接的技术基础。然而, 部分剪切活性优异的内含肽 (如: 蓝藻属的 Oli、Tvu 和 Nsp 内含肽) 之间存在显著的交叉反应^[64], 难以实现多个外显肽间的选择性拼接。通过改变分裂内含肽两片段间的电荷网络能够构建正交分裂内含肽^[65]。2020年, Iwai 等^[34]通过引入酸性残基与碱性残基, 在 Npu DnaB 上针对特定局部区域 (模拟 GP41-1 的 $\beta 3/\beta 6$ 电荷网络) 增强或部分增强静电作用, 分别获得 CI-Npu DnaB 与 Oth-Npu DnaB 两种变体。随后的交叉剪接活性测试显示, CI-NpuDnaB^N 与 CI-NpuDnaB^C 能够剪接, 而与 Oth-Npu DnaB^C 间无反应活性, 实现了 Npu DnaB 不同突变体间的部分正交。

同年, Wang 等^[66]构建了迄今规模最大的正交分裂内含肽库。他们首先筛选出具有不同天然外显肽且低序列同源性的 34 个候选内含肽, 其中 14 个来自文献、20 个来自内含肽数据库 (InBase), 涵盖此前从全球海洋采样 (Global Ocean Sampling, GOS) 筛选出的 GP41-1、GP41-8、NrdJ-1 与 IMDPH^[67] 等内含肽, 并以基于分裂单体樱桃红荧光蛋白 (monomeric Cherry red fluorescent protein, mCherry) 的报告平台评估所有分裂内含肽的功能与正交性。结果显示, 24 对分裂内含肽在体内有反式剪接活性, 其中 15 对可在体内互斥正交, 10 对可在体外同时使用。随后 Wang 等^[66]利用所筛选的正交分裂内含肽在体外成功合成超大分子质量 (最高可达 377 ku) 的重复蛋白——金黄色葡萄球菌表面蛋白 G (*Staphylococcus aureus* surface protein G, SasG), 解决传统重组表达重复蛋白时遗传不稳定、宿主毒性等难题^[68], 展示了正交分裂内含肽优异的蛋白质合成能力。

近期, Mootz 等^[69]对无半胱氨酸型 (序列中

不含 Cys) 正交内含肽做出了重要补充。他们从弯曲杆菌噬菌体 CP21 (*Campylobacter* phage CP21) 的 B 型 DNA 聚合酶 (DNA polymerase B, PolB) 及海洋沉积物宏基因组数据中, 鉴定出 3 个未表征的无半胱氨酸候选分裂内含肽 (CP21 PolB1、LCGC14 PolB2、GOM-OPS-1 PolB2), 并基于 AMYLPRED2 工具的序列分析, 筛选出 IntN 片段聚集倾向低的 LCGC14 PolB2 内含肽, 该内含肽与已知高效无半胱氨酸内含肽 CL_m^[51] 无交叉剪接反应。利用这一正交对, 研究团队建立了新型串联反式剪接策略, 成功实现了中间片段携带标记的非核糖体肽合成酶 (non-ribosomal peptide synthetase, NRPS) 的三片段拼接。

3 分裂内含肽介导蛋白质反式剪接的创新应用

3.1 基因递送

腺相关病毒 (adeno-associated virus, AAV) 是病毒载体的典型代表, 因具有宿主细胞转染范围广、免疫原性较低及可介导外源基因长期稳定表达等优势, 已成为基因递送的重要技术, 但其包装容量 (约 4.7 kb) 限制了大基因的递送^[70]。将大蛋白质编码序列分割并分别装载于多个 AAV 载体, 是解决装载容量限制的有效策略。然而, 基于同源重组或 DNA 转剪接的基因组层面重组策略, 要求载体基因组以特定方式连接, 存在重组效率低以及易产生副产物等问题^[71]。利用分裂内含肽介导的蛋白质反式剪接, 将大基因拆分为两段并分别与 IntN、IntC 的编码序列融合, 经多 AAV 载体共递送至靶细胞后, 可在翻译后阶段实现靶细胞内蛋白质的自发拼接, 进而形成长全蛋白质, 为突破 AAV 包装容量限制提供了潜在策略。

许多疾病由功能蛋白质缺失或突变引发, 递送功能蛋白质是治疗此类疾病的有效手段。基于 AAV 的疗法被认为是治疗遗传性和遗传性疾病最有效的下一代治疗方法之一^[72], 借助基于分裂内含肽的 AAV 载体输送治疗蛋白质的策略近年来已在 Dravet 综合征 (Dravet syndrome, DS)^[73]、Stargardt 病 1 型 (Stargardt disease 1, STGD1)^[74-75] 等多种疾病的动物模型中实现了对病理表型的改善。例如, Auricchio 等^[76]利用 Npu DnaE 与 Rma DnaB 两对正交分裂内含肽递送约 290 ku 的中心体蛋白 290 (centrosomal protein 290, CEP290), 减缓了 Leber 先天性黑蒙 10 型 (Leber congenital

amaurosis 10, LCA10) 小鼠的视网膜外核层厚度变薄, Holt 等^[77]通过递送约 220 ku 的立体纤毛蛋白 (stereocilin, STRC), 显著改善了 STRC 基因缺陷小鼠的听力功能。

杜兴氏肌营养不良症 (Duchenne muscular dystrophy, DMD) 是一种致命的遗传疾病, 由位于 X 染色体上的 DMD 基因发生突变引起, 可导致活动能力受损、心肌病、呼吸衰竭和早逝^[71, 78-79]。此前 DMD 的基因治疗主要是利用 AAV 递送微型肌营养不良蛋白 (micro-dystrophins, μ Dys)^[80], 但受到载体容量的限制, 所递送 μ Dys 的长度仅为全长的 1/3, 导致动物模型与患者出现不完全的表型改善。针对该问题, Regnier 与 Chamberlain 等^[71]开发了基于分裂内含肽的肌营养不良蛋白 (dystrophins, Dys) 递送新策略。研究团队首先构建两种基因克隆片段以表达中型肌营养不良蛋白 (midi-dystrophin, Midi-Dys): 5'端克隆覆盖肌营养不良蛋白 N 端结构域至血影蛋白重复序列 19 (spectrin-like repeat 19, SR19) 末端 (缺失 SR5-15), 3'端克隆则从铰链区 3 (Hinge3) 延伸至其 C 端结构域。使用 GP41-1 以及精准靶向肌肉的嗜肌 AAV (myotropic AAV, MyoAAV)^[81], 成功实现了 Midi-Dys 的递送。该策略在肌营养不良症小鼠模型中展现出良好的疗效: 在低剂量 (2×10^{13} vg/kg, 每千克体重的病毒载体基因组当量) Midi-Dys 处理下, 可使年轻小鼠肌肉比力恢复至野生型水平, 血清肌酸激酶水平恢复正常, 同时显著减少老年小鼠肌肉纤维化并改善心脏功能, 效果优于 μ Dys。同期另一工作借助 Cfa 与 GP41-1 两对分裂内含肽递送全长肌营养不良蛋白 (full-length dystrophin, FL-Dys), 同样显著改善了 DMD 小鼠模型中的肌肉组织病理学和功能^[78]。

除了直接递送功能蛋白质外, 分裂内含肽也已被应用于基因编辑蛋白质的递送。可编程基因编辑工具的出现革新了生命科学领域, 并在遗传疾病治疗性矫正方面展现出临床前景^[82]。其中, 胞嘧啶碱基编辑器 (cytosine base editor, CBE) 可实现 C: G $\rightarrow$ T: A 的碱基对替换, 腺嘌呤碱基编辑器 (adenine base editor, ABE) 则能精准完成 A: T $\rightarrow$ G: C 的碱基对替换。这类工具能实现精确的单个碱基对替换, 同时避免双链断裂 (double-strand DNA break, DSB) 相关副产物^[83], 在修复人类基因突变方面具有比其他修复方法更高的效率^[84], 然而其较大的分子质量限制了靶向递送的应用。基

于分裂内含肽的 AAV 载体系统为递送该类基因编辑工具提供了可行方案: CBE/ABE 可被拆分后与分裂内含肽序列融合, 分别装载于不同 AAV 载体, 转导至靶细胞后经剪接形成完整功能的碱基编辑器, 这种基因编辑过程能够在肝脏、心脏、肌肉、视网膜和大脑等体细胞组织进行^[85]。目前, 该策略已经在多种疾病治疗研究中取得突破性进展^[83, 85-86]。

钙/钙调蛋白激酶 II δ (calcium/calmodulin-dependent protein kinase II δ , CaMKII δ) 作为多种心脏疾病的核心驱动因子, 其调控域内的 281 位/282 位的甲硫氨酸残基 (methionine, Met) 易被氧化激活, 进而引发钙稳态紊乱、心肌细胞凋亡与纤维化, 最终导致心脏功能障碍^[87]。近期, Olson 等^[88]借助分裂内含肽递送碱基编辑器, 实现 CaMKII δ 调控域上 Met 到缬氨酸 (valine, Val) 的定点突变, 从而彻底消除其氧化激活位点。该策略能有效抑制小鼠缺血/再灌注 (ischemia/reperfusion, IR) 损伤后 CaMKII δ 的异常激活, 从而减少心肌细胞凋亡与纤维化。此外, 在 IR 发生时, 对小鼠进行 CaMKII δ 基因编辑, 能使心脏从原本严重的损伤中恢复功能, 为心脏病提供了新的可能治疗途径。

3.2 蛋白质内部序列编辑

分裂内含肽技术已被应用于蛋白质内部序列的精准编辑 (图 6)。Muir 等^[89]利用两对正交分裂内含肽实现了对折叠蛋白内部柔性区域的序列替换, 避免了传统蛋白质半合成路径中需要经历的多步连接-蛋白质折叠等复杂且困难的步骤。结合 N¹⁴/N¹⁵ 同位素标记与质谱技术, 他们筛选出了 Cat 与 Cfa 两对内含肽, 并通过工程化改造得到 Cfa 变体 (mut-Cfa) 以匹配 Cat 的剪接动力学, 从而减少动力学不匹配时单独剪接步骤中可能出现的半剪接中间体。随后, 动力学匹配的两对正交分裂内含肽 (mut-Cfa/Cat) 被成功应用于对增强型绿色荧光蛋白 (enhanced green fluorescent protein, eGFP)、常染色质组蛋白赖氨酸 N-甲基转移酶 (euchromatic histone-lysine N-methyltransferase, G9a) 等蛋白质进行序列替换, 显著扩展了体外蛋白质半合成的应用范围。

Shalem 和 Burslem 等^[90]将该技术进一步拓展至哺乳动物活细胞内。他们选择 β 肌动蛋白 (β -actin, ACTB) 和 α 辅肌动蛋白 1 (α -actinin-1, ACTN1)、细胞周期检查点激酶 1 (checkpoint

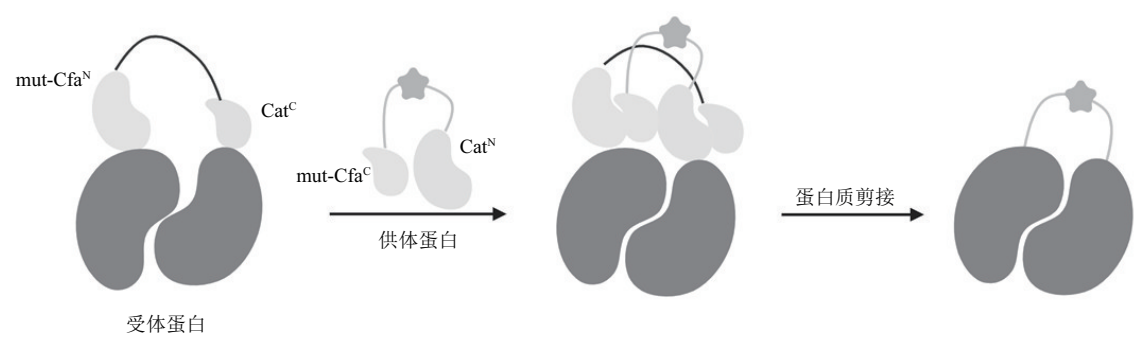


Fig. 6 Principle of protein internal sequence replacement mediated by orthogonal split inteins
图6 正交分裂内含肽介导的蛋白质内部序列替换原理
两对正交分裂内含肽（mut-Cfa以及Cat）协同剪接完成折叠蛋白质的内部序列的替换，其中上标表示分裂内含肽的不同结构域。

kinase 1, Chk1) 以及转录因子细胞骨髓细胞瘤病毒癌基因同源物 (cellular myelocytomatosis viral oncogene homolog, c-Myc) 等作为模型系统, 通过表达含有不同分裂内含肽序列的受体蛋白, 结合电穿孔转染供体, 成功实现对瞬转/稳转以及内源表达的蛋白质的编辑。随后, 他们将该技术应用于活细胞内无环己酰亚胺 (一种翻译延长抑制剂) 的 c-Myc 稳定性与周转研究, 以更加精确的时间分辨率测量其半衰期 ($t_{1/2}=41\text{ min}$)。该技术在电穿孔转染供体后仅需 10 min 即可完成编辑, 能够实现对哺乳动物活细胞内源蛋白质的精准、快速且可控的时空编辑。值得注意的是, 目前该策略的编辑区域仍局限于环区及无序区域。

3.3 蛋白质-蛋白质相互作用探测

生物大分子之间的相互作用, 如 PPI, 是细胞内最基本的分子事件, 其检测在生命科学、医学、药学等基础及应用研究中均具有重要价值^[91], 但传统分析方法通常存在难以检测瞬时/弱相互作用、适用场景有限以及通量低等问题。针对上述局限, Stagljär 等^[92]报道了利用分裂内含肽介导蛋白质剪接技术 (split-intein-mediated protein ligation, SIMPL) 检测瞬时/弱 PPI 的新方法。在该体系中, 分裂内含肽的 N 端和 C 端结构域分别通过 FLAG 与 V5 标签融合于诱饵蛋白 (Bait) 和猎物蛋白 (Prey), 形成 Bait-V5-Int^N 与 Prey-FLAG-Int^C 融合蛋白。当诱饵蛋白和猎物蛋白发生相互作用时, 分裂内含肽的两个结构域得以邻近并触发剪接反应, 使 FLAG 与 V5 标签共价连接, 随后可通过蛋白质印迹法 (Western blot, WB)、酶联免疫吸附分析 (enzyme-linked immunosorbent assay, ELISA) 等方法检测标签信号 (图 7)。应用该技术, Stagljär

等在 HEK293、HeLa 等细胞系中实现了对表皮生长因子受体 (epidermal growth factor receptor, EGFR) /SHC 衔接蛋白 1 (Src homology 2 domain-containing transforming protein 1, SHC1)、Kirsten 大鼠肉瘤病毒癌基因同源物 (Kirsten rat sarcomaviral oncogene homolog, KRAS) /快速进展性纤维肉瘤 1 (rapidly accelerated fibrosarcoma 1, RAF1) 等参与生理或致癌信号转导的经典蛋白相互作用的检测, 并且能捕获激酶与底物之间瞬时且弱的相互作用, 以及线粒体等特定亚细胞结构或秀丽隐杆线虫等多细胞生物模型中的 PPI。此外, 基于 SIMPL 建立的 ELISA 平台可实现对 PPI 的高通量筛选, 不仅能进行量化分析, 而且能检测 EGFR 酪氨酸激酶抑制剂 (tyrosine kinase inhibitor, TKI) AG1478、B 细胞淋巴瘤 2 蛋白 (B-cell lymphoma 2, BCL2) /BCL-2 相关 X 蛋白 (BCL-2-associated X protein, BAX) 的 PPI 抑制剂维奈克拉对相互作用的抑制效果, 为药物研发提供了潜在的分析工具。

3.4 生物逻辑门控

生物分子逻辑门是合成生物学中实现细胞行为精准调控的核心工具, 旨在通过控制细胞内部的分子回路响应输入信号并输出特定的生物学反应。分裂内含肽凭借其“条件性蛋白质剪接”机制, 成为构建生物逻辑门的理想平台。在特定条件下 (输入信号后), 分裂内含肽片段在细胞内重新结合并触发剪接反应, 从而激活细胞内的生物学反应 (如: 基因激活)。以“AND”逻辑门控为例, 分裂的基因调控元件 (如: T7 RNA 聚合酶) 与相应的分裂内含肽片段相连, 当两个输入信号同时存在时, 发生反式剪接而重构出完整的 T7 RNA 聚合酶, 进而通过 T7 启动子转录基因^[93]。“NAND”逻辑门控

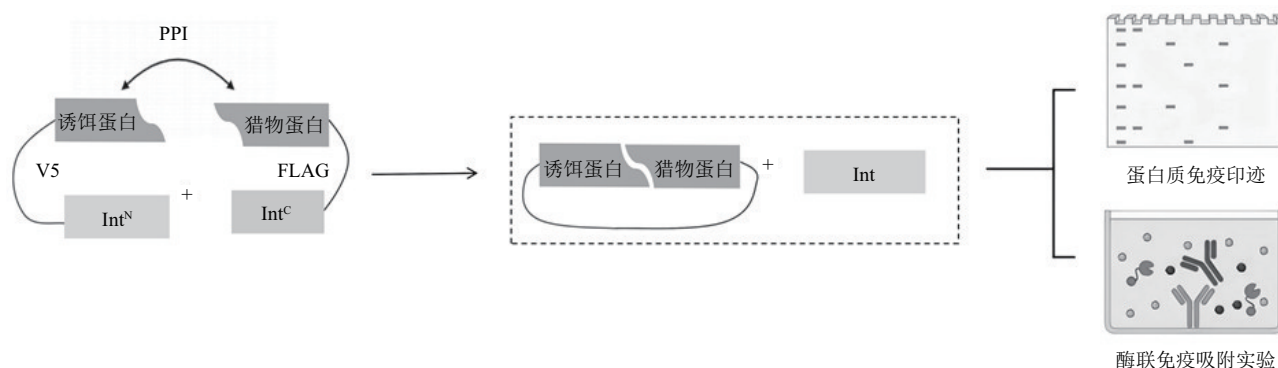


Fig. 7 Principle of intein application in detecting protein-protein interactions

图7 分裂内含肽应用于探测蛋白质-蛋白质相互作用的原理

诱饵蛋白与猎物蛋白相互靠近后诱导内含肽发生蛋白质反式剪接，将V5与FLAG标签相连，随后可通过蛋白质印迹法以及酶联免疫吸附测定检测标签信号，Int^N、Int^C分别为分裂内含肽的N端与C端结构域。

与之类似，输出信号与输入信号之间关系相反，通过特定基因调控元件，如CRISPR结合蛋白9（CRISPR-associated protein 9, Cas9）实现信号输出^[94]。通过整合多种基因元件，还能够构建更为复杂的“IF/AND”“NIMPLY/NOT”等逻辑电路^[95]。

近期，Muir等^[96]将逻辑门控技术拓展至细胞表面，提出靶标识别引发的剪接调节驱动（splicing-modulated actuation upon recognition of targets, SMART）系统，旨在通过调控细胞表面微环境实现精准的细胞行为控制。基于先前分裂内含肽片段与其配对片段中的截短版本结合时（即笼化分裂内含肽）可以抑制片段间相互作用的研究^[97]，以笼化分裂内含肽为核心构建了逻辑门控，将内含肽的N端和C端分别与靶向不同细胞表面抗原的识别模块及断裂功能蛋白质片段连接。只有当识别模块与细胞表面特定抗原共定位时，这种“笼化”状态才会被解除，触发内含肽剪接并重构出活性功能蛋白质，最终实现从抗原组合信号到功能输出的转化（图8）。基于这一原理，Muir等构建了多种逻辑门，如“AND”“NOT”和“OR”，并展示了细胞选择性杀伤和邻近标记等应用。以“AND”门应用为例，针对同时在细胞表面表达人类表皮生长因子受体2（human epidermal growth factor receptor 2, HER2）和EGFR的HER2⁺/EGFR⁺双阳性肿瘤细胞，Muir等将抗HER2的单域抗体片段（HER2-DARPin）与抗EGFR的单域抗体片段（EGFR-DARPin）分别与笼化分裂内含肽片段相融合，仅当这两个抗原共定位时内含肽剪接才能被激

活。重构的活性功能蛋白SpyCatcher003能够特异性结合生物素标记的SpyTag003，并在剪接完成后进一步招募与链霉亲和素相连的生物毒素皂草素，从而导致细胞内吞并最终死亡，实现对肿瘤细胞的选择性杀伤。

3.5 构建精准蛋白以促进生化机制解析

3.5.1 构建修饰蛋白质

应用分裂内含肽技术进行蛋白质修饰近年取得了显著进展，尤其是该技术与化学合成的蛋白质片段融合后，在精准标记和蛋白质功能研究方面展现了巨大潜力。近期，Muir等^[91]应用Cfa分裂内含肽实现了对染色质的无痕精准修饰。将目标蛋白C端与Cfa^N融合，随后通过蛋白质反式剪接引入与Cfa^C相连的铱（iridium, Ir）光催化剂以实现光催化邻近标记。在蓝光激发下，Ir催化剂通过德克斯特激发转移（Dexter excitation transfer）激活重氮环丙烷生成卡宾，随后卡宾能够插入邻近的蛋白质中，最终这些被标记的蛋白质可通过链霉亲和素磁珠被富集（图9），从而在几乎无痕的条件下完成对特定核蛋白微环境的精确映射，实现高精度的蛋白质相互作用捕捉。该技术随后被应用于与染色质相关的蛋白质相互作用分析，如：描绘组蛋白H3.1与其着丝粒特异性变体着丝粒蛋白A（centromere protein A, CENP-A）的相互作用组差异，揭示癌症相关组蛋白突变型H2A E92K对染色质互作组、表观修饰及染色质可及性的扰动作用，实现对JQ-1等小分子表观遗传抑制剂的靶点及脱靶作用分析等。

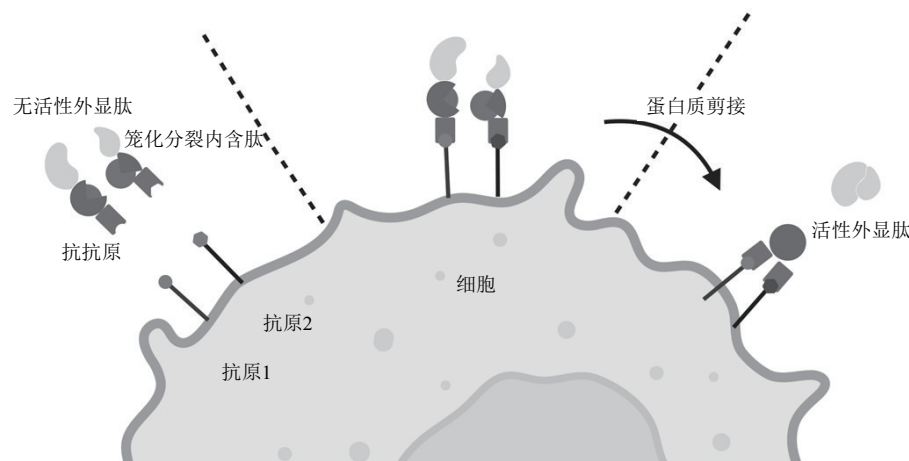


Fig. 8 Schematic of Boolean logic operations on the surface of live cells

图8 在活细胞表面进行布尔逻辑运算原理图

基于AND门的输入-输出系统，其中互补的笼化分裂内含肽片段被连接到靶向细胞表面不同抗原的识别模块上，仅当这些片段在细胞表面共定位时才会触发蛋白质的转剪接反应，重组活性外显肽。

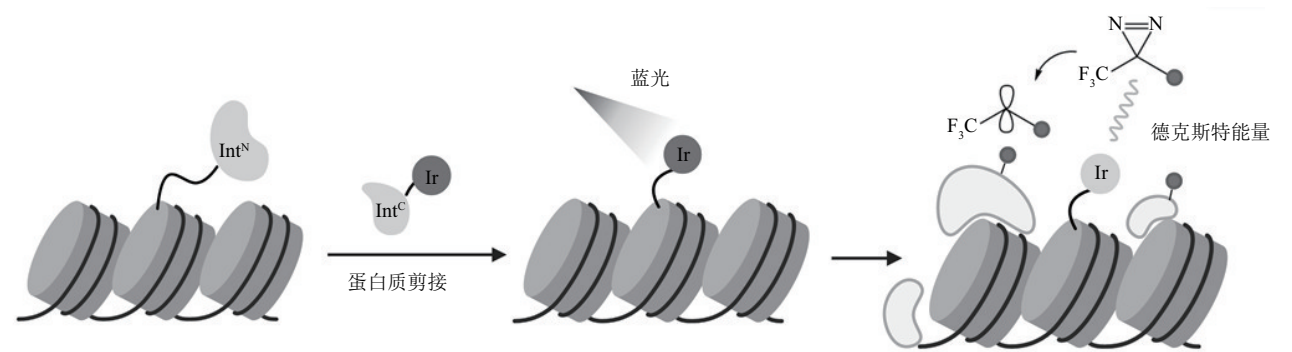


Fig. 9 Split intein-based photo-proximity labeling strategy

图9 基于分裂内含肽的光邻近标记策略

基于分裂内含肽的光邻近标记的技术原理。其中分裂内含肽的N端片段（Int^N）与目标蛋白相连，而Int^C与铱（Ir）光催化剂相连，通过分裂内含肽介导的蛋白质反式剪接可将Ir光催化剂引入核蛋白。在蓝光照射下，这些光催化剂会通过德克斯特激发转移（Dexter excitation transfer）机制激活重氮环丙烷，使其失去分子氮并形成卡宾，随后标记Ir附近的蛋白质。

除了精准标记，分裂内含肽技术更多被应用于构建结构明确的蛋白质样品。精准结构的蛋白质样品被广泛应用于时空分布、相互作用网络、作用机制等基础生物学问题的解析^[98-101]，以及诊断工具与治疗方法的研发与评估^[102-103]。例如，本课题组近期开发了系列精准结构的泛素或类泛素修饰的蛋白质样品或探针，推动了泛素或类泛素相关的结构机制、靶标筛选等研究^[102-103]。然而，诸如基因重组表达或细胞提取等传统生物方法通常难以提供特定定位点携带翻译后修饰（post-translation modification, PTM）的蛋白质样品^[104]。借助分裂

内含肽的反式剪接技术，可高效实现蛋白质或蛋白质结构域之间的选择性连接，能够在体外制备特定定位点携带PTM的蛋白质，进而为解析生命体内分子事件的机制提供支撑。

作为组蛋白H2A的一种变体，macro H2A在DNA损伤应答和转录抑制过程中发挥关键作用^[105]，但其结构中含有不溶的组蛋白结构域与可溶的macro结构域，两者的差异导致化学合成困难。为此，Muir等^[106]利用分裂内含肽VidaL建立了macro H2A-核小体的合成方法。首先，构建了macro H2A1.1的N端难溶组蛋白结构域与短VidaL-

N 片段 (macro H2A1.1₁₋₁₁₉-Vid^N) 以及 C 端可溶 macro 结构域与长 VidaL-C (Vid^C-macro H2A1.1₁₂₀₋₃₇₂) 片段连接的两个融合片段。前者与野生型 H2B 折叠形成 macro H2A1.1₁₋₁₁₉-Vid^N-H2B 二聚体后, 通过反式剪接反应与后者连接, 生成完整的 macro H2A1.1-H2B 二聚体。随后, 将 macro H2A1.1-H2B 二聚体与 H3-H4 四聚体、Widom 601 DNA 组装, 获得天然结构的 macro H2A1.1 核小体 (图 10)。体

外实验表明, macro H2A1.1 能被聚腺苷二磷酸核糖聚合酶 1 (poly ADP-ribose polymerase 1, PARP1) 高效修饰, 且无需组蛋白乙酰化因子 1 (histone parylation factor 1, HPF1) 参与, 揭示了 macro H2A1.1 在 DNA 损伤应答相关的聚腺苷二磷酸核糖 (poly-ADP ribose, PAR) 修饰中可能具有特殊作用模式。

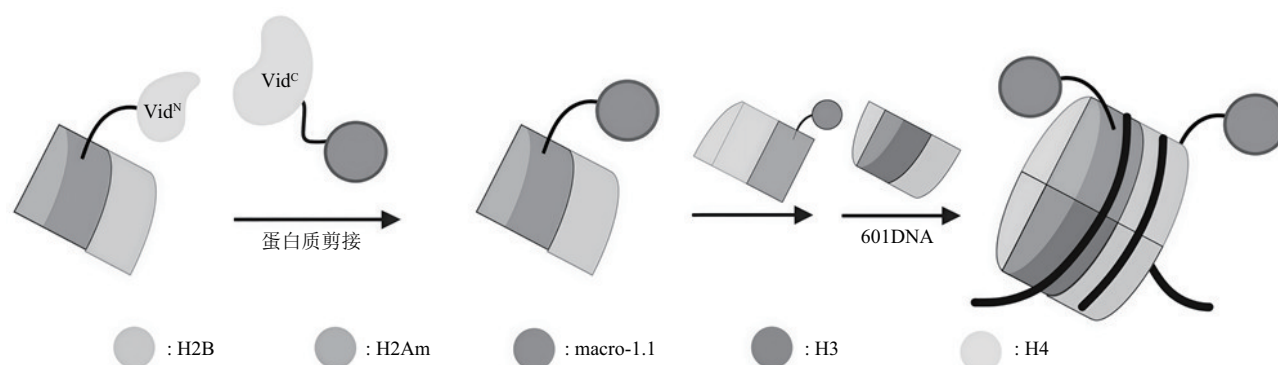


Fig. 10 Strategy for *in vitro* recombination of macro H2A1.1 nucleosome mediated by VidaL

图10 VidaL介导体外重组macro H2A1.1核小体策略

通过分裂内含肽介导的蛋白质反式剪接反应生成macroH2A1.1-H2B二聚体, 随后与H3、H4以及601 DNA组装成macro H2A1.1核小体, 其中Vid^N、Vid^C分别为VidaL分裂内含肽的N端与C端结构域。

微管蛋白作为细胞骨架的重要组成部分, 其携带的PTM对细胞功能维持以及发育至关重要。为克服马来酰亚胺等传统化学方法在制备微管蛋白PTM中的局限^[107], Fierz与Aumeier等^[108]整合分选酶和内含肽介导的串联转肽策略, 获取了特定位点携带多聚谷氨酸修饰的微管蛋白。他们首先构建了C端尾部 (C-terminal tail, CTT) 被串联转酰胺肽标签 (tandem transamidation peptide tag, TTP-tag) 取代的 α 微管蛋白N端球状部分 (残基1~439), 通过工程化细菌分选酶 (engineered sortase A, eSrt2A) 将Cfa的N端 (Cfa^N) 与N端球状部分进行拼接, 随后借助Cfa的反式剪接反应将化学合成的携带多聚谷氨酸修饰的微管蛋白C端尾部 (残基440~451) 与N端部分进行连接, 进而得到C端携带精准PTM的完整微管蛋白 (图 11), 并随后用于解析聚谷氨酰化对去酪氨酸化的调控机制, 发现聚谷氨酰化能够增强去酪氨酸化酶的活性, 且修饰链长影响去酪氨酸化的效率。

3.5.2 构建精准蛋白复合物

除获取翻译后修饰蛋白外, 近期通过基于分裂

内含肽的蛋白质反式剪接技术还被拓展至化学合成人工设计的蛋白质复合物, 用于模拟关键生化过程中的反应中间体^[109]。通过直接孵育重组表达^[110-112]或化学合成^[113]的蛋白质能获得结构明确的复合物样品, 例如通过该策略解析了泛素特异性肽酶 16 (ubiquitin-specific peptidase 16, USP16) 识别 H2AK119Ub 核小体并实现去泛素化的机制^[112], 但是对于组分间仅存在弱/瞬时相互作用的复合物, 直接孵育难以获得高质量的结构分析样品。为此, 有团队近期开发了共价捕获策略, 以分裂内含肽介导的蛋白质反式剪接实现对复合物的共价稳定或捕获, 为分子机制解析提供了可行策略。组蛋白 H2B 的泛素化在调控基因转录激活、高阶染色质结构以及 DNA 损伤应答等细胞过程方面发挥关键作用^[114-115], 在酵母体内负责该泛素化过程的泛素修饰酶为泛素连接酶 (E3) Bre1 与泛素结合酶 (E2) Rad6, 人体内为 E3 酶无名指蛋白 20 (ring finger protein 20, RNF20) /无名指蛋白 40 (ring finger protein 40, RNF40) 与 E2 酶 hRAD6A。然而, 该过程涉及的酶促复合物具有瞬时且高度动

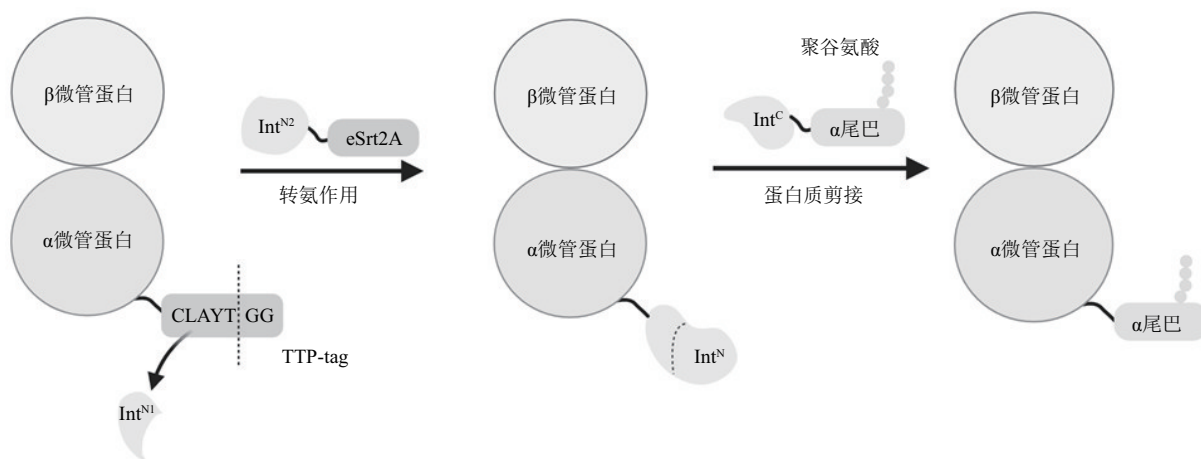


Fig. 11 Tandem transamidation strategy to produce semi-synthetic tubulin

图11 串联转酰胺策略用于制备半合成微管蛋白

图中被虚线分割的黑色方框指串联转酰胺标签TTP-tag, 含有工程化细菌分选酶eSrt2A的识别序列, 虚线左侧CLAYT为内含肽N端结构域部分序列 (Int^{N1}), 随后由eSrt2A介导其与其余部分 (Int^{N2}) 连接, 并与C端结构域 (Int^C) 通过蛋白质反式剪接引入携带聚谷氨酸的C末端尾巴 (α尾巴)。

态的特性, 导致其催化构象难以捕获, 在很大程度上阻碍了催化机制的阐释。为解决该难题, Liu等^[116]发展了基于分裂内含肽GP41-1的化学合成技术, 共价捕获了泛素从泛素修饰酶转移到核小体(H2B)的关键中间体。在该策略中, 他们构建了含Int^N的Bre1与含Int^C的Rad6-ub修饰核小体核心颗粒(nucleosome core particle, NCP), 利用GP41-1介导的剪接反应形成稳定的催化中间体复合物(Bre1-Rad6-Ub-NCP)。随后, 该策略也成功拓展至人源RNF20/RNF40-Rad6-Ub-NCP复合物的构建。利用上述共价复合物, Liu等揭示了酵母与人H2B上位点特异性泛素化的机制, 鉴定出Bre1与Rad6之间非经典的非疏水相互作用, 将Rad6定位在H2B的赖氨酸(lysine, Lys)残基上方, 同时在E3酶与核小体之间的界面上发现Bre1与RNF20/RNF40不仅识别H2A-H2B的酸性板块, 还以二聚体中非常有趣的新基因(really interesting new gene, RING)结构域中的Arg作为锚定残基直接结合核小体超螺旋位置6.6(superhelix location 6.6, SHL6.6)处的DNA。该研究揭示了核小体DNA在介导E3酶识别组蛋白H2B及Ub的关键作用细节, 为理解人源RNF20/RNF40的癌症驱动突变提供了结构基础。

本课题组近期开发出三功能分子2-((2-氯乙

基)氨基)乙烷-1-硫醇(2-((2-chloroethyl)amino)ethane-1-thiol, CAET)介导组蛋白泛素化的新策略^[117-118], 高效合成了单泛素化与双单泛素化组蛋白, 并对DNA甲基转移酶1(DNA methyltransferase 1, DNMT1)与不同泛素修饰NCP的相互作用进行了系统分析。在此基础上, 利用CAET结合分裂内含肽GP41-1建立了核小体泛素化相关催化复合物的共价捕获策略, 重构了单亚基的RING型E3连接酶无名指蛋白168(ring finger protein 168, RNF168)与其协同E2酶泛素结合酶H5c(ubiquitin-conjugating enzyme H5c, UbcH5c)泛素化核小体的催化中间体^[119]。由于RNF168-UbcH5c-核小体复合物之间瞬时且较弱的相互作用, 无法通过直接孵育3种蛋白质以获得高分辨率的冷冻电镜结构。为此, 通过该策略将融合了GP41-1内含肽C端片段的E2酶(Int^C-Ubc5c)、Ub与组蛋白H2AK15C变体通过CAET化学偶联, 形成共价连接的三元复合物, 并组装成核小体, 随后与融合了GP41-1的N端结构域的E3酶(RNF168-Int^N)孵育, 通过剪接反应获得共价连接的RNF168-UbcH5c~Ub-核小体复合物(图12), 并由此获得了高分辨率的复合物冷冻电镜结构。结构解析发现, RNF168的RING结构域通过碱性螺旋结合核小体的酸性口袋, 使UbcH5c的活性中心直

接位于 H2A 的 K13/15 上方，揭示了 RNF168 结合核小体的独特“螺旋锚定”模式，促进了对染色质

泛素化的催化机制认知。

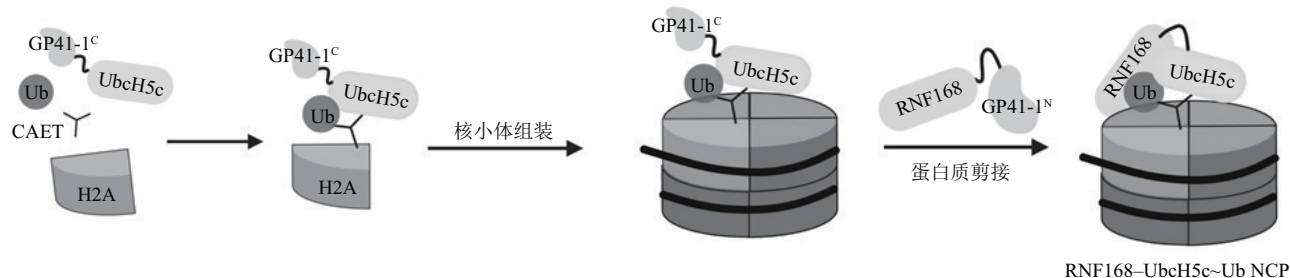


Fig. 12 Strategy for reconstructing the reaction intermediate of RNF168-UbcH5c-catalyzed nucleosome mono-ubiquitination

图12 重构RNF168-UbcH5c介导的核小体单泛素化反应中间体策略

通过三功能分子CAET连接各组分，随后组装成核小体，通过分裂内含肽介导的蛋白质反式剪接构建RNF168-UbcH5c介导的核小体泛素化的催化中间体。Ub：泛素（ubiquitin）；NCP：为核小体核心颗粒（nucleosome core particle）；GP41-1^N、GP41-1^C分别为分裂内含肽GP41-1的N端与C端结构域。

4 总结与展望

近年来，分裂内含肽介导的蛋白质反式剪接技术已经从一种有趣的翻译后现象发展成为化学生物学与合成生物学领域中高效的蛋白质连接工具。尽管已取得显著进展并呈现广阔的应用前景，该技术仍存在很多挑战有待克服：首先，高性能分裂内含肽的多样性不足，对剪接机制和构效关系理解不足使得理性设计受限，定向进化则多以整体内含肽为初始底物，针对天然分裂内含肽的高效改造体系尚未完全建立；其次，分裂片段的聚集倾向（尤其是IntN）虽然能够通过外加增溶标签、外显肽伴侣^[120]等策略得到缓解，但面向大规模体外重组与工业化生产场景，仍需开发更普适的低聚集性分裂内含肽系统或聚集抑制策略，以进一步提升可溶性表达效率与剪接效率；第三，现有分裂内含肽普遍存在对外显肽的序列依赖性，尽管已有理性设计以增强其对序列变化的耐受能力^[120]，但现阶段依然非常有限，从而导致在实际应用中对目标蛋白质产生不可避免的氨基酸突变，存在影响目标蛋白质功能的潜在风险；第四，其介导的蛋白质反式剪接通常局限于常规线性氨基酸序列的编辑或合成，对普遍存在的天然的非线性序列的蛋白质（如泛素或类泛素修饰的蛋白质）的精准构建仍存在困难。

展望未来，分裂内含肽相关领域研究可能存在以下发展方向。其一，借助人工智能技术革新分裂内含肽的传统设计范式，实现定制化性能（如：低

聚集倾向及低外显肽依赖性）的分裂内含肽的从头设计。结合机器学习对多源数据（如：序列、结构、动力学以及剪接活性）的整合分析，可构建高效的大规模筛选与验证平台，加速高性能正交对的挖掘。随着RFDiffusion^[121]、AlphaFold系列^[122-123]、trRosetta^[124]等工具的发展，蛋白质设计已经突破传统理性设计与定向进化的瓶颈，打破天然折叠范式的约束，或许实现真正从头设计具有定制化剪接活性的“人工剪接元件”已不再遥远。其二，突破分裂内含肽的天然剪接机制，产生新型反应路径，拓展其可合成蛋白质的范围。如精准结构的泛素或类泛素修饰蛋白是翻译后修饰解码中的核心底物，其特殊的非线性结构及对应活性探针的精确设计使其获取高度依赖于化学合成，但普遍存在技术要求高、合成效率低等固有局限^[120]。分裂内含肽介导的蛋白质反式剪接具有化学合成难以比拟的优势，若能通过创新反应路径实现非天然肽键、非线性序列的剪接，可能为泛素化等特殊结构蛋白的精准构建提供潜在的高效策略。其三，化学调控分裂内含肽的选择性剪接，并与其他蛋白质片段连接技术进行正交融合，实现超大蛋白质的化学合成。在蛋白质合成领域，尽管结合基于酰肼的蛋白质片段连接^[125-126]、连接后脱硫^[127-128]以及临时支架^[129-130]等策略以及一些新兴的其他技术^[131]，已经能够实现500个氨基酸以下的单链蛋白质分子的获取^[132]，然而高效合成多结构域蛋白和超大蛋白质仍然存在挑战。分裂内含肽基于片段识别的模板效应与高效

剪接能力 (可在 $\mu\text{mol/L}$ 浓度下实现高效连接) 的技术优势, 结合多剪接反应间的有序调控, 有可能突破传统化学合成 (如自然化学连接方法) 对超大蛋白质制备的限制, 为生物制药与功能生物材料领域提供全新技术路径。

参 考 文 献

- [1] Aranko A S, Oeemig J S, Iwai H. Structural basis for protein trans-splicing by a bacterial intein-like domain—protein ligation without nucleophilic side chains. *FEBS J*, 2013, **280**(14): 3256-3269
- [2] Hall T M, Porter J A, Young K E, *et al.* Crystal structure of a Hedgehog autoprocessing domain: homology between Hedgehog and self-splicing proteins. *Cell*, 1997, **91**(1): 85-97
- [3] Kane P M, Yamashiro C T, Wolczyk D F, *et al.* Protein splicing converts the yeast TFP1 gene product to the 69-kD subunit of the vacuolar H(+)-adenosine triphosphatase. *Science*, 1990, **250**(4981): 651-657
- [4] Hirata R, Ohsumi Y, Nakano A, *et al.* Molecular structure of a gene, *VMA1* encoding the catalytic subunit of H(+)-translocating adenosine triphosphatase from vacuolar membranes of *Saccharomyces cerevisiae*. *J Biol Chem*, 1990, **265**(12): 6726-6733
- [5] Pihl R, Zheng Q, David Y. Nature-inspired protein ligation and its applications. *Nat Rev Chem*, 2023, **7**(4): 234-255
- [6] Perler F B. InBase: the intein database. *Nucleic Acids Res*, 2002, **30**(1): 383-384
- [7] Hodges R A, Perler F B, Noren C J, *et al.* Protein splicing removes intervening sequences in an Archaea DNA polymerase. *Nucleic Acids Res*, 1992, **20**(23): 6153-6157
- [8] Wang H, Wang L, Zhong B, *et al.* Protein splicing of inteins: a powerful tool in synthetic biology. *Front Bioeng Biotechnol*, 2022, **10**: 810180
- [9] Li Y. Split-inteins and their bioapplications. *Biotechnol Lett*, 2015, **37**(11): 2121-2137
- [10] Lei B, Wang S, Zhang X, *et al.* Novel protein ligase based on dual split intein. *Biochem Biophys Res Commun*, 2024, **720**: 150097
- [11] Mootz H D. Split inteins as versatile tools for protein semisynthesis. *ChemBioChem*, 2009, **10**(16): 2579-2589
- [12] Thiel I V, Volkmann G, Pietrovski S, *et al.* An atypical naturally split intein engineered for highly efficient protein labeling. *Angew Chem Int Ed*, 2014, **53**(5): 1306-1310
- [13] Sun W, Yang J, Liu X Q. Synthetic two-piece and three-piece split inteins for protein trans-splicing. *J Biol Chem*, 2004, **279**(34): 35281-35286
- [14] Appleby J H, Zhou K, Volkmann G, *et al.* Novel split intein for trans-splicing synthetic peptide onto C terminus of protein. *J Biol Chem*, 2009, **284**(10): 6194-6199
- [15] Wu H, Hu Z, Liu X Q. Protein trans-splicing by a split intein encoded in a split DnaE gene of *Synechocystis* sp. PCC6803. *Proc Natl Acad Sci USA*, 1998, **95**(16): 9226-9231
- [16] Shah N H, Muir T W. Split inteins: nature's protein ligases. *Isr J Chem*, 2011, **51**(8/9): 854-861
- [17] Zettler J, Schütz V, Mootz H D. The naturally split Npu DnaE intein exhibits an extraordinarily high rate in the protein trans-splicing reaction. *FEBS Lett*, 2009, **583**(5): 909-914
- [18] Carvajal-Vallejos P, Pallissé R, Mootz H D, *et al.* Unprecedented rates and efficiencies revealed for new natural split inteins from metagenomic sources. *J Biol Chem*, 2012, **287**(34): 28686-28696
- [19] Lauro F M, DeMaere M Z, Yau S, *et al.* An integrative study of a meromictic lake ecosystem in Antarctica. *ISME J*, 2011, **5**(5): 879-895
- [20] Giriat I, Muir T W. Protein semi-synthesis in living cells. *J Am Chem Soc*, 2003, **125**(24): 7180-7181
- [21] McDermott A, Windeln L M, Valentine J S D, *et al.* Next generation SICLOPPS screening for the identification of inhibitors of the HIF-1 α /HIF-1 β protein-protein interaction. *ACS Chem Biol*, 2024, **19**(10): 2232-2239
- [22] Wu Y, Ikeno S. Cyclization of short peptides designed from late embryogenesis abundant protein to improve stability and functionality. *ChemBioChem*, 2025, **26**(8): e202401013
- [23] Shepherd C, Lawson-Williams M, Holland A, *et al.* Conditional split inteins: adaptable tools for programming protein functions. *Int J Mol Sci*, 2025, **26**(2): 586
- [24] Baloban M, Manoilov K Y, Karasev M M, *et al.* Photoswitchable intein for light control of covalent protein binding and cleavage. *Nat Commun*, 2025, **16**(1): 8263
- [25] Wang Y, Shi Y, Hellinga H W, *et al.* Thermally controlled intein splicing of engineered DNA polymerases provides a robust and generalizable solution for accurate and sensitive molecular diagnostics. *Nucleic Acids Res*, 2023, **51**(11): 5883-5894
- [26] Aranko A S, Wlodawer A, Iwai H. Nature's recipe for splitting inteins. *Protein Eng Des Sel*, 2014, **27**(8): 263-271
- [27] Klabunde T, Sharma S, Telenti A, *et al.* Crystal structure of GyrA intein from *Mycobacterium xenopi* reveals structural basis of protein splicing. *Nat Struct Biol*, 1998, **5**(1): 31-36
- [28] Iwai H, Mikula K M, Oeemig J S, *et al.* Structural basis for the persistence of homing endonucleases in transcription factor IIB inteins. *J Mol Biol*, 2017, **429**(24): 3942-3956
- [29] Beyer H M, Mikula K M, Kudling T V, *et al.* Crystal structures of CDC21-1 inteins from hyperthermophilic archaea reveal the selection mechanism for the highly conserved homing endonuclease insertion site. *Extremophiles*, 2019, **23**(6): 669-679
- [30] Eryilmaz E, Shah N H, Muir T W, *et al.* Structural and dynamical features of inteins and implications on protein splicing. *J Biol Chem*, 2014, **289**(21): 14506-14511
- [31] Wu Q, Gao Z, Wei Y, *et al.* Conserved residues that modulate protein trans-splicing of Npu DnaE split intein. *Biochem J*, 2014, **461**(2): 247-255
- [32] Beyer H M, Iwai H. Structural basis for the propagation of homing endonuclease-associated inteins. *Front Mol Biosci*, 2022, **9**: 855511
- [33] Boral S, Maiti S, Basak A J, *et al.* Structural, dynamic, and

- functional characterization of a DnaX mini-intein derived from *Spirulina platensis* provides important insights into intein-mediated catalysis of protein splicing. *Biochemistry*, 2020, **59**(50): 4711-4724
- [34] Beyer H M, Mikula K M, Li M, *et al.* The crystal structure of the naturally split gp41-1 intein guides the engineering of orthogonal split inteins from *cis*-splicing inteins. *FEBS J*, 2020, **287**(9): 1886-1898
- [35] Perler F B, Olsen G J, Adam E. Compilation and analysis of intein sequences. *Nucleic Acids Res*, 1997, **25**(6): 1087-1093
- [36] Tori K, Perler F B. Expanding the definition of class 3 inteins and their proposed phage origin. *J Bacteriol*, 2011, **193**(8): 2035-2041
- [37] Shah N H, Eryilmaz E, Cowburn D, *et al.* Naturally split inteins assemble through a “capture and collapse” mechanism. *J Am Chem Soc*, 2013, **135**(49): 18673-18681
- [38] Stevens A J, Sekar G, Gramespacher J A, *et al.* An atypical mechanism of split intein molecular recognition and folding. *J Am Chem Soc*, 2018, **140**(37): 11791-11799
- [39] Southworth M W, Benner J, Perler F B. An alternative protein splicing mechanism for inteins lacking an N-terminal nucleophile. *EMBO J*, 2000, **19**(18): 5019-5026
- [40] McNeal T A, Weinberger J, Liman G L S, *et al.* Controllable intein splicing and N-terminal cleavage at mesophilic temperatures. *Front Bioeng Biotechnol*, 2025, **13**: 1543573
- [41] Prabhala S V, Mayone S A, Moody N M, *et al.* A convenient self-removing affinity tag method for the simple purification of tagless recombinant proteins. *Curr Protoc*, 2023, **3**(10): e901
- [42] Amitai G, Callahan B P, Stanger M J, *et al.* Modulation of intein activity by its neighboring extein substrates. *Proc Natl Acad Sci USA*, 2009, **106**(27): 11005-11010
- [43] Mathys S, Evans T C, Chute I C, *et al.* Characterization of a self-splicing mini-intein and its conversion into autocatalytic N- and C-terminal cleavage elements: facile production of protein building blocks for protein ligation. *Gene*, 1999, **231**(1/2): 1-13
- [44] Chong S, Shao Y, Paulus H, *et al.* Protein splicing involving the *Saccharomyces cerevisiae* VMA intein. *J Biol Chem*, 1996, **271**(36): 22159-22168
- [45] Qi X, Meng Q, Liu X Q. Spontaneous C-cleavage of a mini-intein without its conserved N-terminal motif A. *FEBS Lett*, 2011, **585**(15): 2513-2518
- [46] Wu H, Xu M Q, Liu X Q. Protein trans-splicing and functional mini-inteins of a cyanobacterial dnaB intein. *Biochim Biophys Acta*, 1998, **1387**(1/2): 422-432
- [47] Aranko A S, Oeemig J S, Zhou D, *et al.* Structure-based engineering and comparison of novel split inteins for protein ligation. *Mol Biosyst*, 2014, **10**(5): 1023-1034
- [48] Appleby-Tagoe J H, Thiel I V, Wang Y, *et al.* Highly efficient and more general *cis*- and trans-splicing inteins through sequential directed evolution. *J Biol Chem*, 2011, **286**(39): 34440-34447
- [49] Mills K V, Lew B M, Jiang S, *et al.* Protein splicing in trans by purified N- and C-terminal fragments of the *Mycobacterium tuberculosis* RecA intein. *Proc Natl Acad Sci USA*, 1998, **95**(7): 3543-3548
- [50] Southworth M W, Adam E, Panne D, *et al.* Control of protein splicing by intein fragment reassembly. *EMBO J*, 1998, **17**(4): 918-926
- [51] Humberg C, Yilmaz Z, Fitzian K, *et al.* A cysteine-less and ultra-fast split intein rationally engineered from being aggregation-prone to highly efficient in protein trans-splicing. *Nat Commun*, 2025, **16**(1): 2723
- [52] Zhan Q, Shi C, Jiang Y, *et al.* Efficient splicing of the CPE intein derived from directed evolution of the *Cryptococcus neoformans* PRP8 intein. *Acta Biochim Biophys Sin*, 2023, **55**(8): 1310-1318
- [53] Lockless S W, Muir T W. Traceless protein splicing utilizing evolved split inteins. *Proc Natl Acad Sci USA*, 2009, **106**(27): 10999-11004
- [54] Bhagawati M, Terhorst T M E, Füsser F, *et al.* A mesophilic cysteine-less split intein for protein *trans*-splicing applications under oxidizing conditions. *Proc Natl Acad Sci USA*, 2019, **116**(44): 22164-22172
- [55] Tsois A C, Papandreou N C, Iconomidou V A, *et al.* A consensus method for the prediction of ‘aggregation-prone’ peptides in globular proteins. *PLoS One*, 2013, **8**(1): e54175
- [56] Stevens A J, Brown Z Z, Shah N H, *et al.* Design of a split intein with exceptional protein splicing activity. *J Am Chem Soc*, 2016, **138**(7): 2162-2165
- [57] Zhang B, Zheng Y, Chu G, *et al.* Backbone-installed split intein-assisted ligation for the chemical synthesis of mirror-image proteins. *Angew Chem Int Ed*, 2023, **62**(33): e202306270
- [58] Gu R, Lin Y, Di R, *et al.* A unique THN motif is critical for enabling efficient C-terminal traceless cleavage. *Adv Sci (Weinh)*, 2025, **12**(26): e2501991
- [59] Cheriyan M, Pedamallu C S, Tori K, *et al.* Faster protein splicing with the *Nostoc punctiforme* DnaE intein using non-native extein residues. *J Biol Chem*, 2013, **288**(9): 6202-6211
- [60] Shi J, Muir T W. Development of a tandem protein trans-splicing system based on native and engineered split inteins. *J Am Chem Soc*, 2005, **127**(17): 6198-6206
- [61] Shah N H, Eryilmaz E, Cowburn D, *et al.* Extein residues play an intimate role in the rate-limiting step of protein trans-splicing. *J Am Chem Soc*, 2013, **135**(15): 5839-5847
- [62] Stevens A J, Sekar G, Shah N H, *et al.* A promiscuous split intein with expanded protein engineering applications. *Proc Natl Acad Sci USA*, 2017, **114**(32): 8538-8543
- [63] Sekar G, Stevens A J, Mostafavi A Z, *et al.* A conserved histidine residue drives extein dependence in an enhanced atypically split intein. *J Am Chem Soc*, 2022, **144**(41): 19196-19203
- [64] Dassa B, Amitai G, Caspi J, *et al.* Trans protein splicing of cyanobacterial split inteins in endogenous and exogenous combinations. *Biochemistry*, 2007, **46**(1): 322-330
- [65] Shah N H, Vila-Perelló M, Muir T W. Kinetic control of one-pot trans-splicing reactions by using a wild-type and designed split intein. *Angew Chem Int Ed*, 2011, **50**(29): 6511-6515
- [66] Pinto F, Thornton E L, Wang B. An expanded library of orthogonal

- split inteins enables modular multi-peptide assemblies. *Nat Commun*, 2020, **11**(1): 1529
- [67] Dassa B, London N, Stoddard B L, *et al.* Fractured genes: a novel genomic arrangement involving new split inteins and a new homing endonuclease family. *Nucleic Acids Res*, 2009, **37**(8): 2560-2573
- [68] Bowen C H, Dai B, Sargent C J, *et al.* Recombinant spidroins fully replicate primary mechanical properties of natural spider silk. *Biomacromolecules*, 2018, **19**(9): 3853-3860
- [69] Humberg C, Kröger J, Pietrokovski S, *et al.* Three-segment protein labeling using a highly efficient and cysteine-less split intein identified with computational prediction of aggregation properties. *Angew Chem Int Ed*, 2025, **64**(51): e202515821
- [70] Raguram A, Banskota S, Liu D R. Therapeutic *in vivo* delivery of gene editing agents. *Cell*, 2022, **185**(15): 2806-2827
- [71] Tasfaout H, Halbert C L, McMillen T S, *et al.* Split intein-mediated protein trans-splicing to express large dystrophins. *Nature*, 2024, **632**(8023): 192-200
- [72] Kolesnik V V, Nurtudinov R F, Oloruntimehin E S, *et al.* Optimization strategies and advances in the research and development of AAV-based gene therapy to deliver large transgenes. *Clin Transl Med*, 2024, **14**(3): e1607
- [73] Mich J K, Ryu J, Wei A D, *et al.* Interneuron-specific dual-AAV *SCN1A* gene replacement corrects epileptic phenotypes in mouse models of Dravet syndrome. *Sci Transl Med*, 2025, **17**(790): eadn5603
- [74] Tornabene P, Trapani I, Centrulo M, *et al.* Inclusion of a degron reduces levelsof undesired inteins after AAV-mediated proteintrans-splicing in the retina. *Mol Ther Methods Clin Dev*, 2021, **23**: 448-459
- [75] Ferla R, Pugni E, Lupo M, *et al.* Retinal gene therapy for Stargardt disease with dual AAV intein vectors is both safe and effective in large animal models. *Sci Adv*, 2025, **11**(13): eadt9354
- [76] Tornabene P, Trapani I, Minopoli R, *et al.* Intein-mediated protein trans-splicing expands adeno-associated virus transfer capacity in the retina. *Sci Transl Med*, 2019, **11**(492): eaav4523
- [77] Shubina-Oleinik O, Nist-Lund C, French C, *et al.* Dual-vector gene therapy restores cochlear amplification and auditory sensitivity in a mouse model of DFNB16 hearing loss. *Sci Adv*, 2021, **7**(51): eabi7629
- [78] Zhou Y, Zhang C, Xiao W, *et al.* Systemic delivery of full-length dystrophin in Duchenne muscular dystrophy mice. *Nat Commun*, 2024, **15**(1): 6141
- [79] Łoboda A, Chamberlain J S, Dulak J. Genetic strategies for therapy of Duchenne muscular dystrophy. *Mol Ther Nucleic Acids*, 2025, **36**(4): 102759
- [80] Ramos J N, Hollinger K, Bengtsson N E, *et al.* Development of novel micro-dystrophins with enhanced functionality. *Mol Ther*, 2019, **27**(3): 623-635
- [81] Weinmann J, Weis S, Sippel J, *et al.* Identification of a myotropic AAV by massively parallel *in vivo* evaluation of barcoded capsid variants. *Nat Commun*, 2020, **11**(1): 5432
- [82] Chen P J, Liu D R. Prime editing for precise and highly versatile genome manipulation. *Nat Rev Genet*, 2023, **24**(3): 161-177
- [83] Chemello F, Chai A C, Li H, *et al.* Precise correction of Duchenne muscular dystrophy exon deletion mutations by base and prime editing. *Sci Adv*, 2021, **7**(18): eabg4910
- [84] Tachida Y, Manian K V, Butcher R, *et al.* Systematic empirical evaluation of individual base editing targets: validating therapeutic targets in USH2A and comparison of methods. *Mol Ther*, 2025, **33**(4): 1466-1484
- [85] Levy J M, Yeh W H, Pendse N, *et al.* Cytosine and adenine base editing of the brain, liver, retina, heart and skeletal muscle of mice *via* adeno-associated viruses. *Nat Biomed Eng*, 2020, **4**(1): 97-110
- [86] Yeh W H, Shubina-Oleinik O, Levy J M, *et al.* *In vivo* base editing restores sensory transduction and transiently improves auditory function in a mouse model of recessive deafness. *Sci Transl Med*, 2020, **12**(546): eaay9101
- [87] Ling H, Gray C B, Zambon A C, *et al.* Ca²⁺/Calmodulin-dependent protein kinase II δ mediates myocardial ischemia/reperfusion injury through nuclear factor- κ B. *Circ Res*, 2013, **112**(6): 935-944
- [88] Lebek S, Chemello F, Caravia X M, *et al.* Ablation of CaMKII δ oxidation by CRISPR-Cas9 base editing as a therapy for cardiac disease. *Science*, 2023, **379**(6628): 179-185
- [89] Hua Y, Tay N E S, Ye X, *et al.* Protein editing using a coordinated transposition reaction. *Science*, 2025, **388**(6742): 68-74
- [90] Beyer J N, Serebrenik Y V, Toy K, *et al.* Intracellular protein editing enables incorporation of noncanonical residues in endogenous proteins. *Science*, 2025, **388**(6746): eadr5499
- [91] Seath C P, Burton A J, Sun X, *et al.* Tracking chromatin state changes using nanoscale photo-proximity labelling. *Nature*, 2023, **616**(7957): 574-580
- [92] Yao Z, Aboulizadeh F, Kroll J, *et al.* Split intein-mediated protein ligation for detecting protein-protein interactions and their inhibition. *Nat Commun*, 2020, **11**(1): 2440
- [93] Schaerli Y, Gili M, Isalan M. A split intein T7 RNA polymerase for transcriptional AND-logic. *Nucleic Acids Res*, 2014, **42**(19): 12322-12328
- [94] Kawasaki S, Ono H, Hirokawa M, *et al.* Programmable mammalian translational modulators by CRISPR-associated proteins. *Nat Commun*, 2023, **14**(1): 2243
- [95] Muldoon J J, Kandula V, Hong M, *et al.* Model-guided design of mammalian genetic programs. *Sci Adv*, 2021, **7**(8): eabe9375
- [96] Kofoed C, Erkalo G, Tay N E S, *et al.* Programmable protein ligation on cell surfaces. *Nature*, 2025, **645**(8081): 793-800
- [97] Gramespacher J A, Stevens A J, Nguyen D P, *et al.* Intein zymogens: conditional assembly and splicing of split inteins *via* targeted proteolysis. *J Am Chem Soc*, 2017, **139**(24): 8074-8077
- [98] Wang Y, Fan J, Meng X, *et al.* Development of nucleus-targeted histone-tail-based photoaffinity probes to profile the epigenetic interactome in native cells. *Nat Commun*, 2025, **16**(1): 415
- [99] Tao S, He W, Li C, *et al.* Synthesis of an E2-Ub-nucleosome

- conjugate to capture the E3 ligase PHF7-catalyzed H3K14 ubiquitination intermediate. *Sci China Chem*, 2025, **68**(11): 5871-5880
- [100] Ai H, Sun M, Liu A, *et al.* H2B Lys34 ubiquitination induces nucleosome distortion to stimulate Dot1L activity. *Nat Chem Biol*, 2022, **18**(9): 972-980
- [101] Shi Q, Deng Z, Zhang L, *et al.* Promotion of RNF168-mediated nucleosomal H2A ubiquitylation by structurally defined K63-polyubiquitylated linker histone H1. *Angew Chem Int Ed*, 2025, **64**(1): e202413651
- [102] Yang Z, Xiao Y, Shi Y, *et al.* Advances in the chemical synthesis of human proteoforms. *Sci China Life Sci*, 2025, **68**(9): 2515-2549
- [103] Zuo C, Shi W W, Chen X X, *et al.* Chimeric protein probes for C5a receptors through fusion of the anaphylatoxin C5a core region with a small-molecule antagonist. *Sci China Chem*, 2019, **62**(10): 1371-1378
- [104] Ai H, Pan M, Liu L. Chemical synthesis of human proteoforms and application in biomedicine. *ACS Cent Sci*, 2024, **10**(8): 1442-1459
- [105] Sun Z, Bernstein E. Histone variant macroH2A: from chromatin deposition to molecular function. *Essays Biochem*, 2019, **63**(1): 59-74
- [106] Burton A J, Haugbro M, Parisi E, *et al.* Live-cell protein engineering with an ultra-short split intein. *Proc Natl Acad Sci USA*, 2020, **117**(22): 12041-12049
- [107] Sirajuddin M, Rice L M, Vale R D. Regulation of microtubule motors by tubulin isotypes and post-translational modifications. *Nat Cell Biol*, 2014, **16**(4): 335-344
- [108] Ebberink E, Fernandes S, Hatzopoulos G, *et al.* Tubulin engineering by semi-synthesis reveals that polyglutamylation directs detyrosination. *Nat Chem*, 2023, **15**(8): 1179-1187
- [109] Zhang L, Deng Z, Du Y, *et al.* RAD18-catalysed formation of ubiquitination intermediate mimic of proliferating cell nuclear antigen PCNA. *Bioorg Med Chem*, 2025, **117**: 118016
- [110] Tong Z, Ai H, Xu Z, *et al.* Synovial sarcoma X breakpoint 1 protein uses a cryptic groove to selectively recognize H2AK119Ub nucleosomes. *Nat Struct Mol Biol*, 2024, **31**(2): 300-310
- [111] Zhang T, Zheng J, Tong Z, *et al.* Unique gluing effect of ASXL1 K351 monoubiquitination stimulates the PR-DUB activity. *Nat Chem Biol*, 2026. DOI: 10.1038/s41589-025-02126-5
- [112] Ai H, He Z, Deng Z, *et al.* Structural and mechanistic basis for nucleosomal H2AK119 deubiquitination by single-subunit deubiquitinase USP16. *Nat Struct Mol Biol*, 2024, **31**(11): 1745-1755
- [113] Deng Z, Tao S, Du Y, *et al.* Allosteric activation of RNF20/RNF40-RAD6A-mediated H₂BK₁₂₀ monoubiquitylation by H₂BS₁₁₂ GlcNAcylation. *Nat Chem Biol*, 2026. DOI: 10.1038/s41589-025-02109-6
- [114] Fuchs G, Oren M. Writing and reading H2B monoubiquitylation. *Biochim Biophys Acta*, 2014, **1839**(8): 694-701
- [115] Uckelmann M, Sixma T K. Histone ubiquitination in the DNA damage response. *DNA Repair*, 2017, **56**: 92-101
- [116] Deng Z, Ai H, Sun M, *et al.* Mechanistic insights into nucleosomal H2B monoubiquitylation mediated by yeast Bre1-Rad6 and its human homolog RNF20/RNF40-hRAD6A. *Mol Cell*, 2023, **83**(17): 3080-3094.e14
- [117] Zheng Q, Wang T, Mao J, *et al.* A bifunctional molecule-assisted synthesis of mimics for use in probing the ubiquitination system. *Nat Protoc*, 2023, **18**(2): 530-554
- [118] Li Z, Tong Z, Gong Q, *et al.* The expedient, CAET-assisted synthesis of dual-monoubiquitinated histone H3 enables evaluation of its interaction with DNMT1. *Chem Sci*, 2023, **14**(21): 5681-5688
- [119] Ai H, Tong Z, Deng Z, *et al.* Mechanism of nucleosomal H2AK13/15 monoubiquitination and adjacent dual monoubiquitination by RNF168. *Nat Chem Biol*, 2025, **21**(5): 668-680
- [120] Gramespacher J A, Stevens A J, Thompson R E, *et al.* Improved protein splicing using embedded split inteins. *Protein Sci*, 2018, **27**(3): 614-619
- [121] Watson J L, Juergens D, Bennett N R, *et al.* *De novo* design of protein structure and function with RFdiffusion. *Nature*, 2023, **620**(7976): 1089-1100
- [122] Jumper J, Evans R, Pritzel A, *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature*, 2021, **596**(7873): 583-589
- [123] Abramson J, Adler J, Dunger J, *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, 2024, **630**(8016): 493-500
- [124] Du Z, Su H, Wang W, *et al.* The trRosetta server for fast and accurate protein structure prediction. *Nat Protoc*, 2021, **16**(12): 5634-5651
- [125] Fang G M, Li Y M, Shen F, *et al.* Protein chemical synthesis by ligation of peptide hydrazides. *Angew Chem Int Ed*, 2011, **50**(33): 7645-7649
- [126] Huang Y C, Fang G M, Liu L. Chemical synthesis of proteins using hydrazide intermediates. *Natl Sci Rev*, 2016, **3**(1): 107-116
- [127] Han D, Deng X, Cui Y, *et al.* Superfast protein desulfurization triggered by low-energy visible light. *Angew Chem Int Ed*, 2025, **64**(24): e202502884
- [128] Han D, Cui Y, Deng X, *et al.* Mechanically triggered protein desulfurization. *J Am Chem Soc*, 2025, **147**(5): 4135-4146
- [129] Li J B, Tang S, Zheng J S, *et al.* Removable backbone modification method for the chemical synthesis of membrane proteins. *Acc Chem Res*, 2017, **50**(5): 1143-1153
- [130] Liu Y, Han D, Liu L. Temporary structural supports for chemical protein synthesis. *Angew Chem Int Ed*, 2025, **64**(25): e202504405
- [131] Chu G C, Liang L J, Zhao R, *et al.* Ferricyanide-promoted oxidative activation and ligation of protein thioacids in neutral aqueous media. *CCS Chem*, 2024, **6**(8): 2031-2043
- [132] Dong S, Zheng J S, Li Y, *et al.* Recent advances in chemical protein synthesis: method developments and biological applications. *Sci China Chem*, 2024, **67**(4): 1060-1096

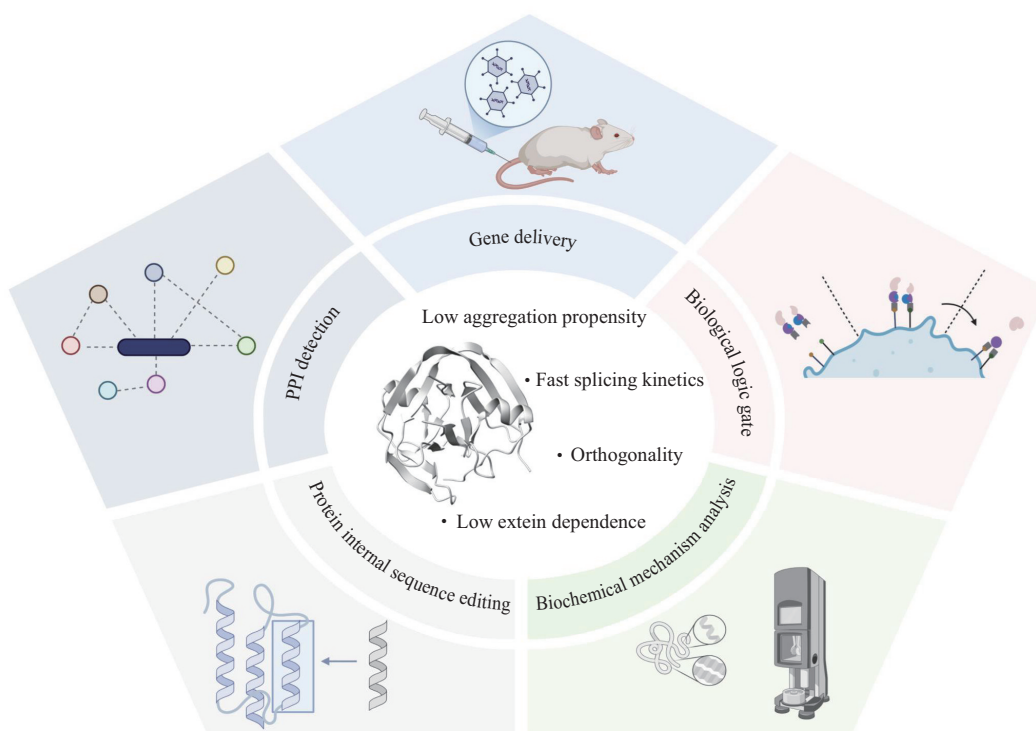
Innovative Development and Cutting-edge Applications of Split Intein Technology*

GAN Jin-Qiu^{1,2)}, DENG Xiang-Yu²⁾, WANG Xin-Yan²⁾, LI Jia-Bin^{1)**}

⁽¹⁾College of Pharmaceutical Sciences, Soochow University, Suzhou 215123, China;

⁽²⁾Department of Chemistry, Tsinghua University, Beijing 100084, China)

Graphical abstract



Abstract Inteins are unique protein insertion sequences capable of self-excision, enabling the covalent ligation of flanking extein peptides *via* amide bond formation. This process proceeds spontaneously without requiring external enzymes, cofactors, or chemical reagents, granting inteins exceptional biocompatibility and traceless performance in protein engineering applications. Split inteins represent a specialized and versatile subclass whose splicing domains are encoded by two separate gene fragments rather than a single continuous open reading frame. These fragments, known as the N-terminal (IntN) and C-terminal (IntC) split inteins, associate through non-covalent interactions including hydrophobic forces, hydrogen bonds, and van der Waals forces to assemble into an active three-dimensional structure, which then drives efficient extein ligation and enables protein trans-splicing. Protein trans-splicing mediated by split inteins has become a cornerstone for traceless protein ligation owing to its

* This work was supported by grants from The National Key R&D Program of China (2022YFA0913200), The National Natural Science Foundation of China (22477092), and Basic Research Program of Jiangsu (BK20250150).

** Corresponding author.

Tel: 86-512-65882089, E-mail: lijiaabin@suda.edu.cn

Received: January 26, 2026 Accepted: April 22, 2026

high specificity and irreversibility, fundamentally reshaping strategies for protein modification, assembly, and functional regulation. Compared with traditional chemical ligation methods, split intein systems require no complex chemical derivatization of peptide fragments and can operate efficiently at micromolar concentrations under physiological conditions, thus avoiding structural and functional damage caused by organic reagents. In contrast to enzymatic ligation tools such as sortase, split inteins eliminate the need for additional enzymes or cofactors, simplifying reaction systems, reducing costs, and minimizing non-specific side products. These distinctive advantages render split inteins highly promising for applications in chemical biology, synthetic biology, and biopharmaceutical development. In recent years, deepened mechanistic understanding has established structure-guided rational design as the primary approach to overcoming key limitations of split inteins, including intrinsic aggregation propensity, strict extein sequence dependence, and limited splicing efficiency. Bioinformatic tools have been used to identify aggregation-prone regions in the IntN fragment, and site-directed mutagenesis of hydrophobic residues, relocation of split sites, or removal of misfolding-prone sequences has substantially reduced *in vitro* aggregation and improved soluble expression and assembly activity. Rational engineering of catalytic residues and adjacent flexible loops has relaxed strict amino acid preferences at extein junctions, enhancing sequence tolerance and reducing the risk of functional impairment in target proteins. Consensus design based on multiple sequence alignments has yielded ultra-fast splicing variants such as Cfa DnaE and Cat-TerL, which exhibit significantly accelerated kinetics and improved tolerance to denaturing conditions. Meanwhile, advances in structural biology have further clarified the conformational dynamics and catalytic mechanisms of splicing, supporting the precise design of high-performance intein modules. On this basis, electrostatic interaction tuning and metagenomic screening have yielded multiple mutually orthogonal split intein pairs, enabling selective multi-fragment protein ligation and providing new routes for the efficient synthesis of large multi-domain functional proteins. With these engineered split inteins offering continuously improved performance and expanded applicability, protein trans-splicing has been widely applied in numerous cutting-edge areas of protein research and biomedicine. In gene delivery, split intein-based systems overcome the packaging limit of adeno-associated viral vectors, enabling the accurate reconstitution of large therapeutic proteins and base editors in target cells, thereby enhancing the efficacy and scope of gene therapy for genetic diseases. In internal protein sequence editing, split inteins mediate precise sequence replacement and modification in flexible regions or loops of target proteins, without the need for complex multi-step ligation and protein refolding involved in traditional protein semisynthesis. In protein-protein interaction studies, intein-mediated splicing covalently captures transient and weak intracellular complexes, enabling sensitive, high-throughput interaction detection and drug screening. In synthetic biology, conditionally controllable splicing systems support the construction of diverse intracellular and cell-surface biological logic gates for the precise regulation of cellular behavior. In mechanistic biochemical research, split inteins enable photocatalytic proximity labeling and site-specific tagging, allowing the preparation of homogeneous protein samples carrying precise post-translational modifications such as ubiquitination and polyglutamylation for chromatin interactome analysis and epigenetic studies. Moreover, covalent trapping strategies using split inteins stabilize transient enzymatic intermediates, providing unprecedented insights into molecular mechanisms such as nucleosome ubiquitination that are difficult to elucidate using conventional methods. This review systematically summarizes key technological advances in split inteins over the past decade, highlighting engineering strategies, mechanistic insights, and the development of orthogonal components. It comprehensively surveys emerging applications at the frontiers of protein research, analyzes current core challenges, and proposes future directions, particularly emphasizing artificial intelligence-driven *de novo* design and novel splicing pathways to break existing technical bottlenecks. By enabling traceless, efficient, and versatile protein manipulation, split inteins continue to serve as indispensable tools that drive innovation in protein engineering and fundamental life science research.

Key words split intein, protein trans-splicing, protein ligation tool, chemical biology

DOI: 10.3724/j.pibb.2026.0043

CSTR: 32369.14.pibb.20260043