

研究报告

嗜热 II 型内含子 Tel3c/4c-RT 结构域活性位点分析

花登雄^{1,2}, 王鑫鑫^{2,3}, 陈峥宏^{2,3}, 洪伟^{2,3}, 汪依航², 俞青凤¹, 谷俊莹^{*1,4}, 崔古贞^{*1,2,3}

1 贵州医科大学医学检验学院, 贵州 贵阳 550004

2 贵州医科大学基础医学院病原生物学重点实验室, 贵州 贵阳 550025

3 地方病与少数民族疾病教育部重点实验室 贵州省医学分子生物学重点实验室, 贵州 贵阳 550004

4 贵州医科大学附属医院, 贵州 贵阳 550004

花登雄, 王鑫鑫, 陈峥宏, 洪伟, 汪依航, 俞青凤, 谷俊莹, 崔古贞. 嗜热 II 型内含子 Tel3c/4c-RT 结构域活性位点分析[J]. 微生物学通报, 2022, 49(8): 3049-3061

Hua Dengxiong, Wang Xinxin, Chen Zhenghong, Hong Wei, Wang Yihang, Yu Qingfeng, Gu Junying, Cui Guzhen. Analysis of active sites from the structural domain of thermophilic group II intron Tel3c/4c-RT[J]. Microbiology China, 2022, 49(8): 3049-3061

摘要:【背景】嗜热 II 型内含子是一类由内含子 RNA 和内含子编码蛋白(intron encoded protein, IEP)组成且在高温条件下能够在染色体上高频移动的反转录转座子, 目前已被开发为高效基因打靶工具 Thermotargetron, 阐明其活性催化位点, 对深入研究其“归巢”机制及开发新型遗传工具具有重要意义。【目的】筛选嗜热 II 型内含子 Tel3c/4c-RT 结构域关键活性位点, 并获得失活反转录功能的内含子编码蛋白突变体。【方法】先利用生物信息学技术分析并筛选可能影响 Tel3c/4c-RT 反转录功能的关键氨基酸位点; 然后对筛选到的关键氨基酸位点进行定点突变, 并以 Thermotargetron 质粒为基础构建失活反转录功能的突变型嗜热 II 型内含子打靶系统; 最后以大肠杆菌 *lacZ* 基因为例, 通过蓝白斑计数分析突变型 Thermotargetron 系统的打靶效率, 体内验证 Tel3c/4c-RT 结构域关键活性位点突变对嗜热 II 型内含子打靶效率的影响。【结果】共筛选到 15 个可能影响反转录活性的氨基酸位点, 包括 D194、I195、S196、G197、C198、F199、Q241、G242、R274、Y275、A276、D277、D278、L324 和 G325。其中, 当 G242 和 R274 这 2 个位点突变后, 嗜热 II 型内含子 Tel3c/4c 的“归巢”效率几乎完全丧失, 表明 G242 和 R274 是影响嗜热 II 型内含子 Tel3c/4c-RT 结构域反转录功能的最核心催化位点。【结论】筛选并获得影响嗜热 II 型内含子

基金项目: 国家自然科学基金(31760318, 32160015); 贵州省科技计划项目([2019]1441, [2020]1Z067); 贵州医科大学培育项目(20NSP0004); 大学生创新训练项目(S202010660052)

Supported by: National Natural Science Foundation of China (31760318, 32160015); Science and Technology Project of Guizhou Province ([2019]1441, [2020]1Z067); Guizhou University of Medical Sciences Development Project (20NSP0004); National Training Program of Innovation and Entrepreneurship for Undergraduates (S202010660052)

*Corresponding authors: E-mail: GU Junying: gujunying@gmc.edu.cn; CUI Guzhen: cuiguzhen@gmc.edu.cn

Received: 2021-11-26; Accepted: 2022-02-24; Published online: 2022-04-21

Tel3c/4c-RT 结构域反转录功能的关键催化位点, 为深入研究嗜热 II 型内含子的“归巢”机制及其功能应用开发奠定了良好的基础。

关键词: 大肠杆菌; 嗜热 II 型内含子; Thermotargetron; Tel3c/4c; 内含子编码蛋白; 反转录结构域

Analysis of active sites from the structural domain of thermophilic group II intron Tel3c/4c-RT

HUA Dengxiong^{1,2}, WANG Xinxin^{2,3}, CHEN Zhenghong^{2,3}, HONG Wei^{2,3}, WANG Yihang², YU Qingfeng¹, GU Junying^{*1,4}, CUI Guzhen^{*1,2,3}

1 School of Clinical Laboratory Science, Guizhou Medical University, Guiyang 550004, Guizhou, China

2 Key Laboratory of Microbiology and Parasitology of Education Department of Guizhou, School of Basic Medicine, Guizhou Medical University, Guiyang 550025, Guizhou, China

3 Key Laboratory of Endemic and Ethnic Diseases, Ministry of Education, Key Laboratory of Medical Molecular Biology of Guizhou Province, Guizhou Medical University, Guiyang 550004, Guizhou, China

4 The Affiliated Hospital of Guizhou Medical University, Guiyang 550004, Guizhou, China

Abstract: [Background] Thermophilic group II introns are a class of retrotransposons that are composed of intron RNA and intron-encoded protein (IEP) and can move on chromosomes at high frequency under high temperature, and nowadays, they have been used to develop Thermotargetron, an efficient gene-targeting tool. Therefore, it is highly important to elucidate their active catalytic sites for studying the “retrohoming” mechanism and developing new genetic tools. [Objective] To screen the key active sites from the domain of thermophilic group II intron Tel3c/4c-RT and obtain IEP mutants with inactivated reverse transcription function. [Methods] Firstly, bioinformatics method was used to analyze and screen the key amino acid sites that might affect the reverse transcription function of Tel3c/4c-RT. Then, site-directed mutations were performed on the selected key amino acid sites, and Thermotargetron plasmids were used to construct a targeting system of thermophilic group II intron with inactivated reverse transcription function. Finally, taking the *lacZ* gene of *Escherichia coli* as example, the targeting efficiency of the Thermotargetron mutant system was analyzed by blue-white screening, and the effect of the key active site of the Tel3c/4c-RT on the targeting efficiency of the thermophilic group II intron was verified *in vivo*. [Results] A total of 15 amino acid sites that may affect reverse transcriptional activity were screened out: D194, I195, S196, G197, C198, F199, Q241, G242, R274, Y275, A276, D277, D278, L324 and G325. The “retrohoming” of Tel3c/4c was almost completely lost when the G242 and R274 were mutated, suggesting that the G242 and R274 were the core catalytic sites affecting the reverse transcription function of Tel3c/4c-RT. [Conclusion] The core catalytic sites affecting the reverse transcription function of Tel3c/4c-RT were screened out, which laid a good foundation for in-depth study of the “retrohoming” mechanism of the thermophilic group II introns and the further development.

Keywords: *Escherichia coli*; thermophilic group II introns; Thermotargetron; Tel3c/4c; intron-encoded protein; reverse transcription domain

II 型内含子(group II intron)属于一类反转录转座子,由具有核酶活性的内含子 RNA 和具有多酶活性的内含子编码蛋白(intron encoded protein, IEP)组成^[1],广泛存在于细菌、古菌及真核生物的线粒体和叶绿体中,其中以细菌 II 型内含子最常见^[1-3]。在已知的细菌 II 型内含子中,研究最为详尽的主要有两类:一类是来源于乳酸乳球菌(*Lactococcus lactis*)的嗜中温 II 型内含子 L1.LtrB^[4-5];另一类是由来源于嗜热蓝藻菌(*Thermosynechococcus elongatus*) Tel3c 的内含子 RNA 和来源于嗜热脂肪地芽孢杆菌(*Geobacillus stearothermophilus*)的反转录酶(Tel4c-RT)共同组装而成的杂合型嗜热 II 型内含子 Tel3c/4c^[2,6-7]。上述两类内含子因其具有在基因组 DNA 高效迁移的特性,已被研究人员开发为靶向基因失活工具 Targetron (嗜中温)和 Thermotargetron (嗜高温),并在多种微生物的遗传改造中获得广泛运用^[8-14]。

II 型内含子通过“归巢”(retrohoming)机制实现在基因组中的定向迁移,“归巢”过程如图 1 所示:首先,内含子 RNA 与 IEP 蛋白相互作用,组成有功能活性的核糖核蛋白复合物(ribonucleoprotein, RNP);然后,RNP 中内含子 RNA 通过碱基互补配对原则识别双链 DNA 靶位点,接着 IEP 蛋白在 RNA 的引导下反向切割双链 DNA 并产生切口;随后,IEP 蛋白自身携带的反转录酶活性以切口处产生的 3'-OH 末端作为延伸引物,以内含子 RNA 为模板反转录合成 cDNA;最后,依靠宿主自身的 DNA 修复机制将 cDNA 特异性地插入到靶位点,实现 II 型内含子在基因组靶位点上的“归巢”^[11,15]。

为了深入研究 II 型内含子的功能,拓展 II 型内含子的应用范围,前期研究分析了 L1.ltrB 内含子编码蛋白反转录结构域(reverse transcription

domain, RT)的核心催化位点,构建了一系列失活反转录功能的突变体,并分析了 IEP 蛋白突变后对 II 型内含子“归巢”效率的影响及其在基因组编辑中的潜在作用,结果预示 L1.ltrB 内含子编码蛋白反转录功能失活后完全丧失了内含子的“归巢”功能,具有在靶位点引入 DNA 损伤的潜能,而且有被开发为新型基因编辑工具的潜力^[16-18]。Tel3c/4c 嗜热 II 型内含子与中温的 L1.ltrB 内含子具有相似的结构与功能,其 IEP 蛋白的反转录结构域(Tel3c/4c-RT)也由 9 个相对保守的基序构成^[4,19],因此,解析影响 Tel3c/4c-RT 反转录功能的关键活性位点对拓展嗜热 II 型内含子的功能及开发新型基因组编辑工具等具有重要意义。

1 材料与方法

1.1 材料

1.1.1 培养基及培养方法

大肠杆菌(*Escherichia coli*) DH5 α 、BL21(DE3)均使用 LB 培养基进行培养,质粒构建及诱导表达时培养基中加入终浓度为 30 $\mu\text{g/mL}$ 的氯霉素;大肠杆菌 37 $^{\circ}\text{C}$ 有氧培养,液体培养时摇床转速 200 r/min。蓝白斑筛选时,在固体 LB 培养基中加入氯霉素(30 $\mu\text{g/mL}$)、X-gal (40 $\mu\text{g/mL}$)和 IPTG (0.1 mmol/L)。

1.1.2 主要试剂和仪器

高保真 DNA 聚合酶、T4 DNA 连接酶和 2 \times Taq Master Mix,南京诺唯赞生物科技股份有限公司;DNA Marker,北京全式金生物科技股份有限公司;Dpn I 限制性核酸酶,赛默飞技术(北京)有限公司;质粒提取试剂盒、普通 DNA 纯化试剂盒和凝胶回收试剂盒,天根生化科技(北京)有限公司;Tryptone、Yeast Extract、IPTG、X-gal 和氯霉素,北京索莱宝科技有限公司。

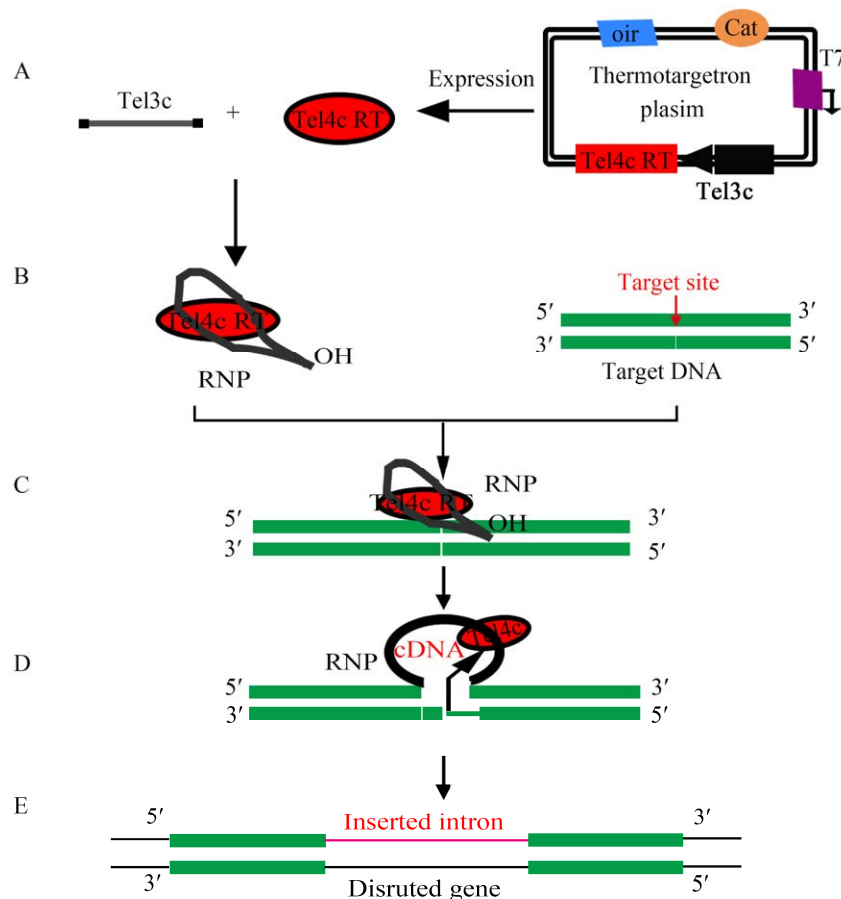


图 1 II 型内含子“归巢”原理 A: II 型内含子 RNA 和 IEP 蛋白组装, 形成具有功能活性的 RNP 复合物; B: 在 IEP 蛋白辅助下, RNP 识别 DNA 靶位点; C: 内含子 RNA 反向插入到 dsDNA 中; D: 反义链切割, 在 IEP 的 DNA 核酸内切酶作用下切开双链 DNA 的反义链, 并在 IEP 反转录酶的作用下以内含子 RNA 为模板, 3'外显子为引物合成 cDNA; E: DNA 修复, 包括模板链的切除和互补链的合成

Figure 1 The “retrohoming” principle of group II intron. A: Group II intron RNA and IEP protein were assembled to form a functional active RNP complex; B: RNP recognizes DNA target sites with the assistance of IEP protein; C: Intron RNA is reversely inserted into dsDNA; D: Antisense strand cutting, the antisense strand of dsDNA is cut by the endonuclease activity of IEP, and the cDNA is synthesized using the intron RNA as template and 3' exon as primer using the reverse transcriptase activity of IEP; E: DNA repair, including excision of template strands and synthesis of complementary strands.

PCR 仪, Thermo Fisher Scientific 公司; 电泳仪和凝胶成像仪, 北京六一生物科技有限公司; 双控温干式恒温器, 杭州米欧公司; 恒温振荡器, 上海一恒科学仪器有限公司。

1.1.3 引物

研究所用引物均由生工生物工程(上海)股

份有限公司合成, 引物序列见表 1。

1.2 方法

1.2.1 生物信息学分析

来源于嗜热 II 型内含子编码蛋白 Tel4c-RT 的反转录结构域序列(登录号为 NP_681951)为模板, 在保守结构域数据库(conserved domain

表 1 本研究所用引物

Table 1 Primers used in this study

Primers name	Sequences (5'→3')	Notes
D194A-U	AACACGTCCTTGACGCT gcg ATATCCGGAT	To mutate the D194A
D194A-D	cgc AGCGTCAAGGACGTGTTTCAGCCGACTT	
I195A-I	AACACGTCCTTGACGCTGAC gcg TCCGGATGCT	To mutate the I195A
I195A-D	cgc GTCAGCGTCAAGGACGTGTTTCAGCCGACTT	
S196A-U	CTTGACGCTGACATA gcg GGATGCTTTG	To mutate the S196A
S196A-D	cgc TATGTCAGCGTCAAGGACGTGTTCA	
G197A-U	CTTGACGCTGACATATCC gcg TGCTTTGATA	To mutate the G197A
G197A-D	cgc GGATATGTCAGCGTCAAGGACGTGTTCA	
C198A-U	TGACGCTGACATATCCGG gcg TTTGATAACA	To mutate the C198A
C198A-D	TGTTATCAA cg cTCCGGATATGTCAGCGTCA	
Q241A-U	ACCCATGCTGGGACACCT gcg GGAGGGGTAAT	To mutate the Q241A
Q241A-D	cg cAGGTGTCCCAGCATGGGTGGGGAAGAGTT	
G242A-U	ACACCTCAG g cgGGGGTAATCTCCCCAGTTCTT	To mutate the G242A
G242A-D	AAGAACTGGGGAGATTACCC cg cCTGAGGTGT	
R274A-U	g cgTATGCTGACGATTTTGTCTG	To mutate the R274A
R274A-D	ACAAAATCGTCAGCATA cg cGATGAGGTTGA	
Y275A-U	AACCTCATCCG g cgGCTGACGATTTTG	To mutate the Y275A
Y275A-D	CAAAATCGTCAGC cg cTCGGATGAGGTT	
A276G-U	GTCAACCTCATCCGATAT gga GACGATTTTGT	To mutate the A276G
A276G-D	cct CATATCGGATGAGGTTGACTTTTTGACCTC	
L324A-U	GAGGAAGGCTTCGACTTT g cgGGATGGAACATTC	To mutate the L324A
L324A-D	cg cAAAGTCGAAGCCTTCCTCAATATGGACGAT	
G325A-U	AAGGCTTCGACTTTCTC g cgTGGAACATTC	To mutate the G325A
G325A-D	GAATGTTCCA cg cGAGAAAGTCGAAGCCTT	
lacZ-R	GCCATTTTTTGGATGGACCATTTT	To detect the <i>lacZ-60a/369a</i> site of the <i>lacZ</i> gene
lacZ-F	ATGACCATGATTACGGATT	
Tel4c-U	AGTGAAAGCTTTGCAATGGC	To detect the Tel4c-RT fragment
Tel4c-D	CTCGGCTGTGCACCTGTTTG	

注: 加粗字母表示突变位点碱基

Note: The bold letter indicates the base of the mutation site.

databases, CDD)中进行多序列比对与结构预测, 筛选出可能影响 RT 活性的关键氨基酸位点, 并对筛选出的关键氨基酸位点进行同源建模。首先, 利用 Swiss-Model 在线网站 (<https://swissmodel.expasy.org/>)检索并获取嗜热 II 型内含子 Tel3c/4c-RT 蛋白晶体结构; 然后, 以野生型 Tel3c/4c-RT 晶体结构为模型进行同源建模, 模拟构建关键氨基酸突变后的 Tel4c-RT

蛋白三维结构; 最后, 利用 Swiss-PDB Viewer 软件比较关键氨基酸位点突变前后三维空间结构的变化。

1.2.2 定点突变及突变型 Thermotargetron 载体构建

根据 1.2.1 中生物信息学预测出可能影响 RT 活性的关键氨基酸位点信息, 以载体 pACD2X-TT1A-lacZ60a/369a 的 Tel3c/4c-RT 为

模板,使用反向定点突变引物(表 1)对载体 Tel3c-RT 区域的氨基酸位点进行定点突变(除 A276 位点突变为甘氨酸外,其余位点均突变为丙氨酸)。PCR 条件:95 °C 30 s;95 °C 30 s,58 °C 30 s,70 °C 6 min,30 个循环;72 °C 4 min。PCR 结束后加入 1 μ L 的 *Dpn* I 限制性内切酶,37 °C 条件下反应 1 h 去除模板质粒,随后转化大肠杆菌 DH5 α ,挑取单克隆培养后提取质粒并测序验证突变位点变化情况,构建突变型 Thermotargetron 载体 pACD-M-lacZ-60a 和 pACD-M-lacZ-369a (M 表示氨基酸突变位点)。

1.2.3 Thermotargetron 打靶效率分析

将野生型 Thermotargetron 载体 pACD-lacZ-60a/369a 和突变型 Thermotargetron 载体 pACD-M-lacZ-60a/369a (M 表示突变氨基酸位点)分别转化大肠杆菌 BL21(DE3)感受态细胞,挑取单克隆菌落接种于 5 mL 的 LB 培养基中(含 30 μ g/mL 氯霉素),37 °C、200 r/min 培养过夜,取 40 μ L 过夜培养菌液加入 4 mL 含有 30 μ g/mL 氯霉素的 LB 培养基中,37 °C、200 r/min 培养 2 h,随后向试管中加入 IPTG 终浓度为 1.0 mmol/L 的 IPTG,48 °C 诱导 5 h,稀释不同的浓度梯度后涂布 LB 培养基(氯霉素:30 μ g/mL;X-gal:40 μ g/mL;IPTG:0.1 mmol/L),37 °C 培养过夜,通过菌落蓝白斑计算野生型/突变型载体的“归巢”效率[“归巢”效率=白斑/(白斑+蓝斑) \times 100%],实验 3 次平行重复,根据蓝白斑计数结果取平均值为最终打靶效率。

1.2.4 菌落 PCR 验证

分别挑取野生型载体和突变型载体平板上的蓝色菌落和白色菌落,以 lacZ-R/lacZ-F (lacZ-60a/369a 检测引物)为引物进行菌落 PCR 检测。PCR 条件:95 °C 3 min;95 °C 30 s,56 °C 30 s,72 °C 30 s,30 个循环;72 °C 3 min。通过菌落 PCR 进一步验证 Tel3c/4c-RT 结构域突

变后对“归巢”的影响。

2 结果与分析

2.1 Tel3c/4c-RT 结构域同源序列比对及关键活性位点筛选

Tel3c/4c 内含子编码蛋白由反转录结构域(RT)、核酸酶结构域(En)、DNA 结构域(D)及一个未知功能结构域(X)组成,在三维空间结构和功能活性上相互独立^[7],该蛋白质同时具有反转录酶、DNA 聚合酶、核酸内切酶等多种酶活性(图 2A)。为了筛选无反转录活性的嗜热 II 型内含子,我们以 Tel3c/4c 嗜热 II 型内含子为基础,以 RT 结构域的氨基酸序列(登录号为 NP_681951)为模板,利用 NCBI 蛋白序列比对及 CDD 保守结构域数据库进行分析,结果表明,Tel3c/4c-RT 在多种细菌及古菌中高度保守,其中,G242 预测为核酸结合位点,D194、I195、S196、G197、C198、F199、Q241 和 D277 预测为 NTP 结合位点,D277、D278 预测为 Mg²⁺结合位点,Y275、L324 和 G325 预测为与反转录功能直接相关的活性位点(图 2B)。此外,根据空间结构分析,我们推测 R274 和 A276 这 2 个位点也可能与 Tel3c/4c-RT 的反转录活性有关,因此,我们共筛选到 15 个可能的关键氨基酸活性位点(图 2B),为后续定点突变及功能验证提供靶点与依据。

2.2 Tel3c/4c-RT 定点突变及突变型 Thermotargetron 载体构建

以上述筛选到的 15 个可能影响 Tel3c/4c-RT 反转录功能的关键氨基酸位点为依据,以 Thermotargetron 载体为基础,利用定点突变技术构建了 12 个突变型 Thermotargetron 系统(15 个关键氨基酸位点预测结果中 F199、D277 和 D278 这 3 个位点突变体尚未构建成功),除 A276 位点突变为甘氨酸(G)外,其余位点均突

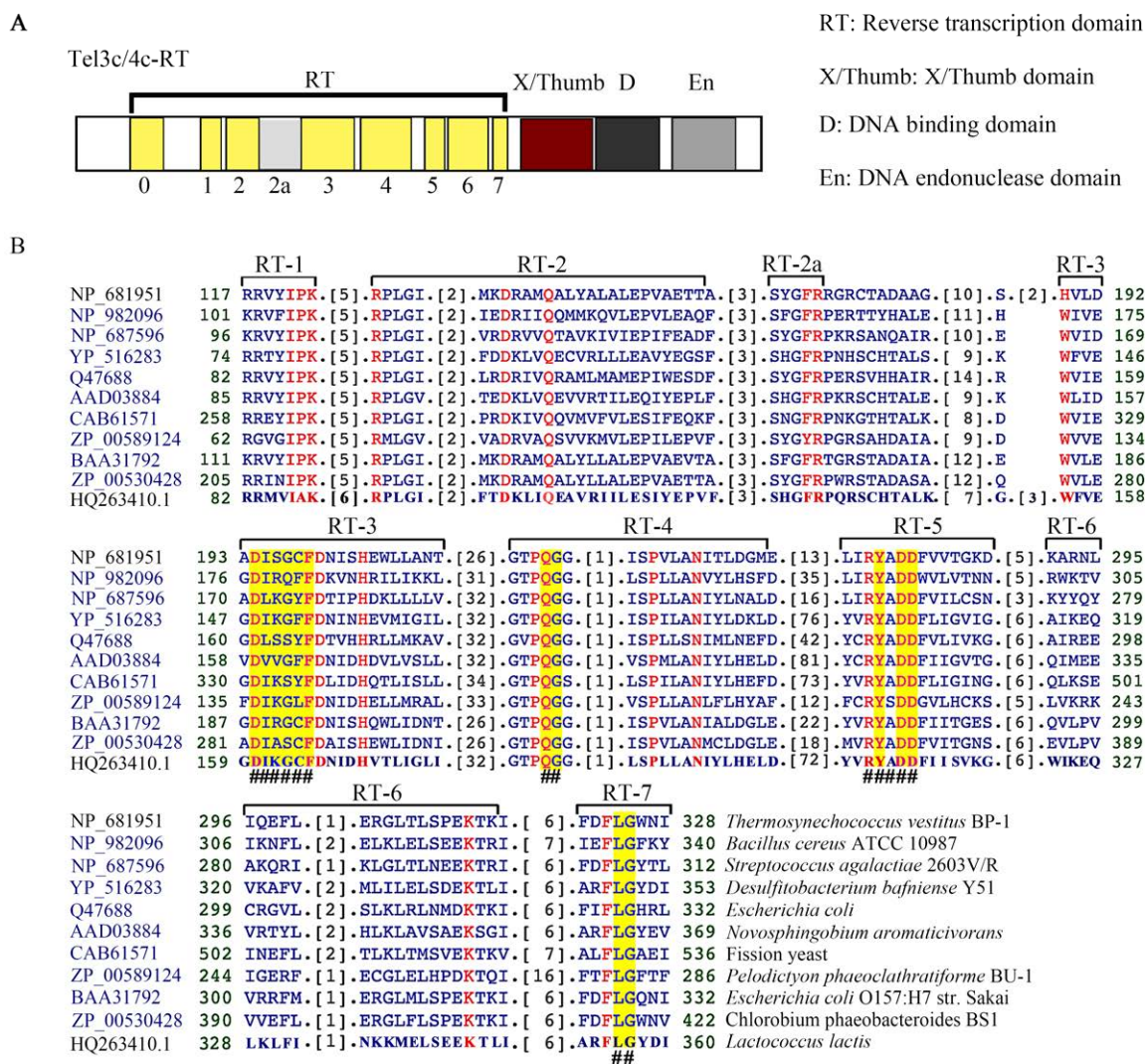


图 2 Tel3c/4c-RT 结构示意图及结构域特征序列分析 A: 嗜热 II 型内含子 Tel3c/4c-RT 结构示意图, 阿拉伯数字 0-7 表示 RT 结构域的保守基序; B: Tel3c/4c-RT 结构域特征序列分析, 黄色背景表示 Tel3c/4c-RT 活性位点, 红色字母表示高度保守位点, 蓝色字母表示相对保守位点, 非保守氨基酸已省略, 方括弧中数字表示省略的氨基酸个数, 绿色数字表示氨基酸所在位点, #表示该位点可能是结合位点或 RT 活性直接相关的氨基酸位点, NP_681951 表示 Tel3c/4c-RT 序列, HQ263410.1 表示嗜中温乳酸乳球菌 L1.ltrB 的 RT 序列

Figure 2 Schematic diagram of Tel3c/4c-RT structure and analysis of domain characteristic sequence. A: Schematic of the Tel3c/4c-RT structure of the thermophilic group II intron, the Arabic numerals 0-7 represent the conservative motif of the RT domain; B: The yellow background indicates the active site of Tel3c/4c-RT, the red letter indicates the highly conserved site, the blue letter indicates the relatively conserved site, the non-conserved amino acid has been omitted, the number in square bracket indicates the number of omitted amino acids, the green number indicates the amino acid site, # indicate the amino acid sites directly associated with RT activity, NP_681951 represents the Tel3c/4c-RT sequence, HQ263410.1 represents the RT sequence of L1.ltrB of *L. lactis*.

变为丙氨酸(A), 包括 D194A、I195A、S196A、G197A、C198A、Q241A、G242A、R274A、Y275A、A276G、L324A 和 G325A (表 2)。

为了验证上述位点突变后对 Tel3c/4c 内含子“归巢”效率的影响, 我们以大肠杆菌为例, 选择 *lacZ-60a* 和 *lacZ-369a* 这 2 个位点作为靶位点, 构建了 24 个突变型打靶载体 pACD-M-*lacZ-60a* 和 pACD-M-*lacZ-369a* (M 表示 12 个不同的氨基酸突变位点, 表 2), 所有突变位点均经测序证实。

2.3 Tel3c/4c-RT 关键活性位点突变对嗜热 II 型内含子“归巢”效率的影响

为了体内证实 Tel3c/4c-RT 关键氨基酸位点突变对其“归巢”效率的影响, 将野生型 Thermotargetron 载体及各种 Tel3c/4c-RT 突变型 Thermotargetron 载体分别转化 *E. coli* BL21(DE3), 通过蓝白斑筛选及菌落 PCR (图 3)验证 Tel3c/4c-RT

突变位点的氨基酸功能。结果表明: 野生型 Thermotargetron 载体在 *lacZ-60a* 和 *lacZ-369a* 位点的“归巢”效率分别为 $91.76\% \pm 7.08\%$ 和 $98.05\% \pm 1.67\%$, 而 Tel3c/4c-RT 突变后其“归巢”效率均有不同程度降低(表 2)。其中, G242 和 R274 这 2 个位点突变后, 在 *lacZ-60a* 和 *lacZ-369a* 位点的“归巢”功能几乎完全丧失(表 2, 图 3), 表明 G242 和 R274 突变后内含子的“归巢”过程被终止, G242 和 R274 可能是影响 Tel3c/4c-RT 活性的最核心氨基酸催化位点。

2.4 G242 和 R274 突变对 Tel3c/4c-RT 空间构象的影响

上述体内实验证实 G242 和 R274 突变完全失活了 Tel3c/4c-RT 的功能, 说明 G242 和 R274 这 2 个位点是影响 Tel3c/4c-RT 反转录功能的最核心催化位点, 为进一步分析 G242 和 R274 这 2 个位点突变对 Tel3c/4c-RT 空间结构的影响, 我们以 Tel3c/4c-RT 的晶体结构(SMTL ID: 6me0.1, 图 4A 所示)信息为基础, 用同源建模方法构建了 R274A、G242A 突变后的 Tel3c/4c-RT 三维空间结构, 并分析两者突变前后三维空间结构的变化, 结果显示: 野生型 G242 与 Q142 存在 3.4 Å 的氢键链接、与 P240 分别存在 3.1 Å 和 3.4 Å 的氢键相互作用, 当 G242 突变后, G242 与 Q142、P240 之间的氢键消失; 可能受到 G242 氨基酸突变的影响, 在 G242 突变的三维结构中发现 T239 与 E241 之间的氢键由原来 3.5 Å 缩短到 3.4 Å, 而且 T239 与 G242 新产生 3.2 Å 的氢键连接; S246 与 G243 及 E241 与 N80、S81 之间的氢键作用也因 G242 氨基酸的改变而消失(图 4B)。野生型 R274 与 D255、L254、D158、S161、A276 通过氢键产生相互作用, 当 R274 突变后, R274 与 D255、L254、D158、S161、A276 之间丧失氢键相互作用; 同样可能受到 R274 氨基酸突变的影响,

表 2 Tel3c/4c-RT 突变体“归巢”效率

Table 2 “Retrohoming” efficiency of Tel3c/4c-RT mutants (%)

Mutant site	pACD-M- <i>lacZ-369a</i>	pACD-M- <i>lacZ-60a</i>
WT	98.05±1.67	91.76±7.08
D194A	14.72±12.27	90.33±1.07
I195A	90.60±4.92	71.51±5.06
S196A	94.89±5.69	88.92±2.39
G197A	74.03±27.86	75.37±2.97
C198A	76.95±0.53	30.86±19.98
Q241A	95.98±1.45	83.32±2.70
G242A	1.75±2.83	0.00
R274	0.00	0.00
Y275A	57.20±22.62	59.22±3.77
A276G	98.64±0.49	85.65±4.38
L324A	82.11±16.17	74.78±9.75
G325A	5.86±8.61	93.89±2.17

注: 表中数据为平行重复 3 次的结果

Note: The data in this table are the results of three parallel repeats.

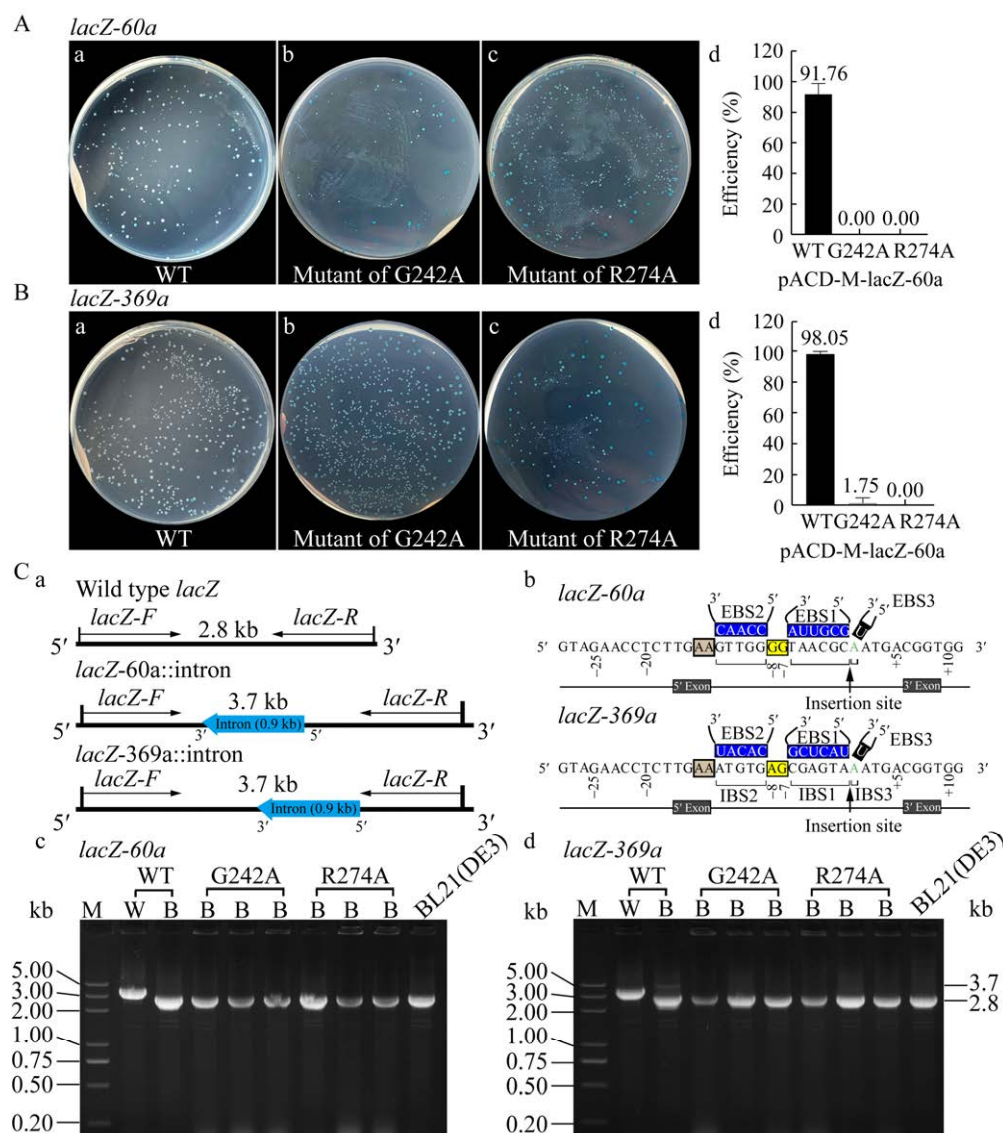


图 3 突变型 Tel3c/4c 嗜热 II 型内含子功能验证 A: *lacZ-60a* 位点蓝白斑筛选, a: 野生型; b: G242A 突变型; c: R274A 突变型; d: “归巢”效率。B: *lacZ-369a* 位点蓝白斑筛选, a: 野生型; b: G242A 突变型; c: R274A 突变型; d: “归巢”效率。C: 菌落 PCR 验证, a: 菌落 PCR 检测示意图; b: *lacZ* 基因插入位点识别序列示意图; c: *lacZ-60a* 位点菌落 PCR; d: *lacZ-369a* 位点菌落 PCR; WT: 野生型; G242A: G242 突变型; R274A: R274 突变型; M: 2k Plus Marker; W: 白色菌落; B: 蓝色菌落; BL21(DE3): BL21(DE3)空白对照; IBS: 内含子结合序列; EBS: 外显子结合序列; Exon: 外显子

Figure 3 Functional verification of the mutated Tel3c/4c thermophilic group II intron. A: Blue-white spot screening for *lacZ-60a* site. a: Wild type; b: G242A mutant type; c: R274A mutant type; d: “Retrohoming” efficiency. B: Blue-white spot screening for *lacZ-369a* site. a: Wild type; b: G242A mutant type; c: R274A mutant type; d: “Retrohoming” efficiency. C: Colony PCR detection. a: Schematic diagram of colony PCR detection; b: Recognition sequence diagram of *lacZ* gene insertion site; c: Colony PCR for *lacZ-60a* site; d: Colony PCR for *lacZ-369a* site; WT: Wild type; G242A: G242 mutant type; R274A: R274 mutant type; M: 2k Plus Marker; W: White colony; B: Blue colony; BL21(DE3): BL21(DE3) blank control; IBS: Intron binding sequence; EBS: Exon binding sequence.

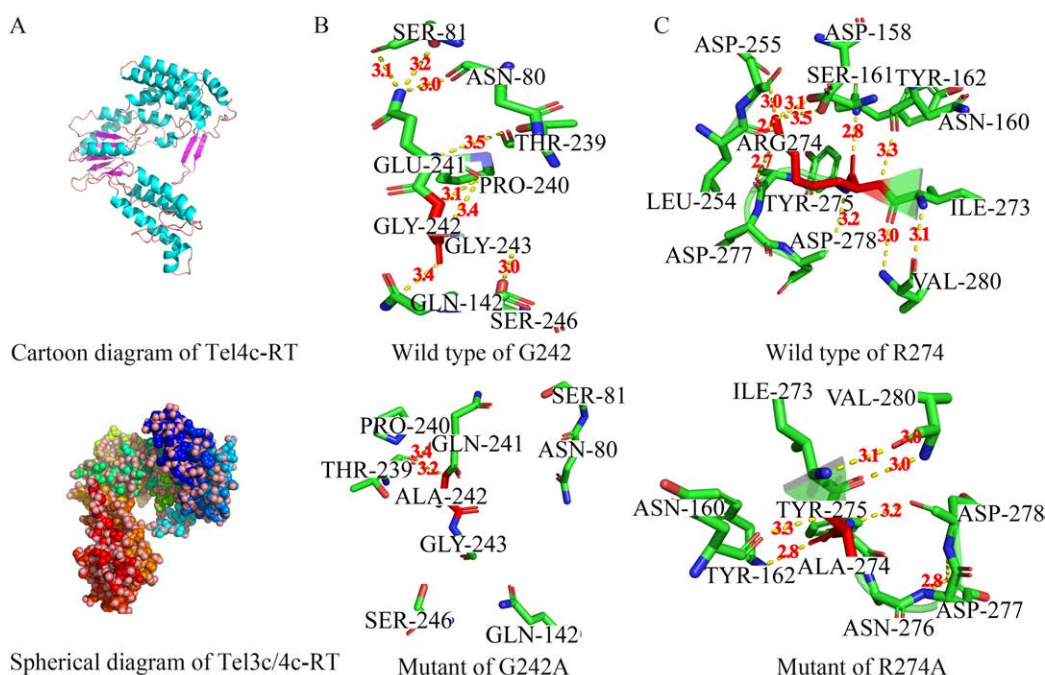


图 4 G242 和 R274 野生型与突变型 Tel3c/4c-RT 三维空间结构分析 A: Tel4c-RT 卡通结构图和球形结构图; B: G242 突变前后三维结构比对图; C: R274 突变前后三维结构比对图

Figure 4 3D structure analysis of G242 and R274 wild type and mutant Tel3c/4c-RT. A: Cartoon and spherical diagram of Tel4c-RT; B: Comparison of 3D structure of G242 before and after mutation; C: Comparison of 3D structure of R274 before and after mutation.

在 R274 突变的三维结构中发现 D277 与 A276、D278 之间均产生了 2.8 Å 的氢键连接(图 4C), 这些氢键及结构上的变化可能影响了 IEP 蛋白与模板、核苷酸及 Mg^{2+} 之间的结合。从目前唯一获得晶体结构的嗜热 II 型内含子反转录酶 GsI-IIC-RT 的三维空间结构可知, PQG 基序和 YADD 基序分别是 II 型内含子与底物核酸结合和 Mg^{2+} 结合的关键催化位点, 而且在所有 II 型内含子反转录酶中高度保守^[20-22]。R274 紧邻 YADD 基序, 构成 RYADD 催化基序, G242 突变影响了 Tel3c/4c-RT 与底物的结合, R274 突变影响了 RYADD 基序与 Mg^{2+} 的结合, 而 Mg^{2+} 是 II 型内含子发挥功能所必需的, 因此, G242 和 R274 突变影响了 Tel3c/4c-RT 的反转录功能, 造成归巢功能丧失。为了解析 Tel3c/4c-RT 催化

位点突变后的真实空间结构, 仍需制备突变型晶体或利用冷冻电镜进一步解析相关结构。

3 讨论

我们在验证 Tel3c/4c-RT 突变型 Thermotargetron “归巢”效率时发现, 与野生型相比, 突变型 Thermotargetron “归巢”效率均存在不同程度的降低(表 2), 说明预测的氨基酸位点对 Tel3c/4c-RT 的反转录活性具有一定的影响, 其中 G242 和 R274 这 2 个位点的突变几乎完全失活了 Tel3c/4c-RT 的功能(表 2), 因此, G242 和 R274 这 2 个位点是 Tel3c/4c-RT 的最核心催化位点。同时, 我们发现同一突变体在不同靶位点的“归巢”效率存在较大差异, 如: D194A 和 C198A 这 2 个突变位点在 *lacZ-369a* 位点和 *lacZ-60a*

位点的“归巢”效率分别为 $14.72\% \pm 12.27\%$ 和 $90.33\% \pm 1.07\%$ 及 $76.95\% \pm 0.53\%$ 和 $30.86\% \pm 19.98\%$, 尤其是 G325A 突变位点, 在 *lacZ-369a* 位点和 *lacZ-60a* 位点的“归巢”效率分别为 $5.86\% \pm 8.61\%$ 和 $93.89\% \pm 2.17\%$, 两者相差甚远(表 2)。我们推测, 引起这种差异的原因是由细菌 II 型内含子碱基识别特点决定的, 细菌 II 型内含子识别靶位点是通过碱基互补配对原则实现的, 而且对靶位点的碱基具有一定的偏好性和依赖性^[2,23], 导致内含子对不同靶点的切割效率不同, 即使是野生型细菌 II 型内含子, 在不同靶位点的“归巢”效率也存在一定的差异(表 2)。当我们对 Tel3c/4c-RT 关键位点进行突变后, Tel3c/4c-RT 的反转录活性可能完全或部分丧失, 由于受到内含子反转录功能活性下降或丧失的影响, 最终导致 Thermotagetrone 在不同靶点间“归巢”时, 内含子对靶点的切割效率差异被进一步放大。此外, 在另一研究中, 我们发现 Thermotagetrone 打靶位点的-7 位和-8 位的碱基组成对打靶效率也存在一定的影响(图 3C-b), 该位点虽然不与内含子 RNA 序列互补, 但不同的碱基可能影响 IBS/EBS 之间的相互识别, 这可能也是造成 Thermotagetrone 在不同打靶位点效率显著差异的原因之一。

在嗜热 II 型内含子“归巢”的过程中 Tel3c/4c-RT 发挥着重要功能。Tel3c/4c-RT 由 4 个功能和空间结构相对独立的结构域组成^[23-24], 因此, 对某个结构域突变后在理论上可能并不影响其他结构域的功能活性。我们团队前期成功筛选了嗜中温 II 型内含子编码蛋白(登录号为 HQ263410.1)的反转录结构域核心催化位点(C164、G214、D308、D309)^[16-17], 当这些 RT 核心催化位点突变后, 嗜中温 II 型内含子的“归巢”活性完全丧失。与此同时, 我们还发现 IEP 反转录突变后对 II 型内含子其他结构域功能影

响较小, 且突变体在打靶位点具有能引入双链 DNA 损伤的可能性^[16-17]。基因编辑的关键在于能够在基因组特定位点引起 DNA 损伤^[25-26], 因此, 在理论上只要能在 DNA 特定位点引入 DNA 损伤的元件就具有被开发成基因编辑工具的潜能。由嗜热 II 型内含子的“归巢”过程可知, IEP 蛋白可特异地切割基因组上的特定 DNA 位点, 满足基因编辑元件的要求, 具有被开发为基因编辑元件的可能性。但不足之处在于, II 型内含子切割靶位点后将内含子 RNA 插入到靶位点, 使得该系统不能如 CRISPR 等其他基因编辑工具那样实现对靶基因的任意编辑(包括基因敲除、基因敲入、点突变等)^[27], 因此, 若能在不影响 Tel3c/4c-RT 其他功能的前提下失活其反转录酶活性, 阻断内含子 RNA 在 DNA 靶点的反转录插入, 仅在靶点引入 DNA 双链损伤, 并利用其耐高温特性, 便有可能将其改造为嗜热型基因组编辑工具, 因此, 本研究为深入了解嗜热 II 型内含子的“归巢”机制及新型工具开发奠定了良好的基础。

4 结论

本研究利用生物信息学方法, 结合定点突变技术及体内功能验证, 筛选到 15 个影响嗜热 II 型内含子 Tel3c/4c-RT 反转录功能的关键氨基酸位点, 其中 G242 和 R274 这 2 个位点突变完全丧失了嗜热 II 型内含子的功能, 是影响 Tel3c/4c-RT 反转录活性的最核心催化位点, 本研究为嗜热 II 型内含子的结构、功能研究及新型工具开发奠定了基础。

REFERENCES

- [1] Toro N, Molina-Sánchez MD, Nisa-Martínez R, Martínez-Abarca F, García-Rodríguez FM. Bacterial group II introns: identification and mobility assay[J]. Transposons and Retrotransposons, 2016, 1400: 21-32.

- [2] Mohr G, Kang SY-S, Park SK, Qin YD, Grohman J, Yao J, Stamos JL, Lambowitz AM. A highly proliferative group IIC intron from *Geobacillus stearothermophilus* reveals new features of group II intron mobility and splicing[J]. *Journal of Molecular Biology*, 2018, 430(17): 2760-2783
- [3] Lambowitz AM, Belfort M. Mobile bacterial group II introns at the crux of eukaryotic evolution[J]. *Microbiology Spectrum*, 2015, 3(1): MDNA3-MDNA0050-2014
- [4] McNeil BA, Semper C, Zimmerly S. Group II introns: versatile ribozymes and retroelements[J]. *Wiley Interdisciplinary Reviews: RNA*, 2016, 7(3): 341-355
- [5] Truong DM, Hewitt FC, Hanson JH, Cui XX, Lambowitz AM. Retrohoming of a mobile group II intron in human cells suggests how eukaryotes limit group II intron proliferation[J]. *PLoS Genetics*, 2015, 11(8): e1005422
- [6] Zhao C, Pyle AM. The group II intron maturase: a reverse transcriptase and splicing factor go hand in hand[J]. *Current Opinion in Structural Biology*, 2017, 47: 30-39
- [7] Mohr S, Ghanem E, Smith W, Sheeter D, Qin YD, King O, Polioudakis D, Iyer VR, Hunnicke-Smith S, Swamy S, et al. Thermostable group II intron reverse transcriptase fusion proteins and their use in cDNA synthesis and next-generation RNA sequencing[J]. *RNA: New York, N Y*, 2013, 19(7): 958-970
- [8] Cui GZ, Zhang J, Hong W, Xu CG, Feng YG, Cui Q, Liu YJ. Improvement of clostron for successive gene disruption in *Clostridium cellulolyticum* using a *pyrF*-based screening system[J]. *Applied Microbiology and Biotechnology*, 2014, 98(1): 313-323
- [9] Cui GZ, Hong W, Zhang J, Li WL, Feng YG, Liu YJ, Cui Q. Targeted gene engineering in *Clostridium cellulolyticum* H10 without methylation[J]. *Journal of Microbiological Methods*, 2012, 89(3): 201-208
- [10] Kuehne SA, Heap JT, Cooksley CM, Cartman ST, Minton NP. Clostron-mediated engineering of *Clostridium*[J]. *Methods in Molecular Biology*: Clifton, N J, 2011, 765(4): 389-407
- [11] Mohr G, Hong W, Zhang J, Cui GZ, Yang YF, Cui Q, Liu YJ, Lambowitz AM. A targetron system for gene targeting in thermophiles and its application in *Clostridium thermocellum*[J]. *PLoS One*, 2013, 8(7): e69032
- [12] 崔古贞, 花登雄, 管玉祝, 鲍江舰, 张馨月, 谷俊莹, 陈峥宏, 洪伟. 解纤维梭菌脱水四环素诱导的 Clostron 基因打靶系统构建[J]. *微生物学通报*, 2021, 48(9): 2982-2990
- [13] 崔古贞, 洪伟, 洪伟. 温度诱导 Targetron 系统用于大肠杆菌高效基因失活[J]. *生物工程学报*, 2020, 36(8): 1659-1671
- [14] 赵行行, 程玉梅, 吴昌学, 任玮, 饶凤琴, 周倩, 崔古贞, 齐晓岚, 洪伟. 温度诱导 Targetron 系统用于大肠杆菌高效基因失活[J]. *生物工程学报*, 2020, 36(8): 1659-1671
- [15] 崔古贞, 陈相好, 洪伟, 张峥嵘, 蔡廷娜, 陈峥宏. 内含子编码蛋白 Mg^{2+} 结合位点功能分析及验证[J]. *生物技术通报*, 2020, 36(10): 165-172
- [16] 崔古贞, 陈相好, 洪伟, 张峥嵘, 蔡廷娜, 陈峥宏. 内含子编码蛋白 Mg^{2+} 结合位点功能分析及验证[J]. *生物技术通报*, 2020, 36(10): 165-172
- [17] 陈相好, 张峥嵘, 刘芳, 陈峥宏, 洪伟, 蔡廷娜, 谷俊莹, 崔古贞. L1.LtrB 内含子编码蛋白反转录结构域关键催化位点分析及功能验证[J]. *微生物学报*, 2019, 59(12): 2357-2366
- [18] 崔古贞, 陈峥宏, 陈相好, 洪伟, 张峥嵘, 蔡廷娜, 刘芳. 一种基于 II 型内含子的基因编辑方法: CN201811110692.X[P]. 2021-04-20
- [19] Belfort M, Lambowitz AM. Group II intron RNPs and reverse transcriptases: from retroelements to research

- tools[J]. Cold Spring Harbor Perspectives in Biology, 2019, 11(4): a032375
- [20] Stamos JL, Lentzsch AM, Lambowitz AM. Structure of a thermostable group II intron reverse transcriptase with template-primer and its functional and evolutionary implications[J]. Molecular Cell, 2017, 68(5): 926-939.e4
- [21] Lentzsch AM, Stamos JL, Yao J, Russell R, Lambowitz AM. Structural basis for template switching by a group II intron-encoded non-LTR-retroelement reverse transcriptase[J]. The Journal of Biological Chemistry, 2021, 297(2): 100971
- [22] Lentzsch AM, Yao J, Russell R, Lambowitz AM. Template-switching mechanism of a group II intron-encoded reverse transcriptase and its implications for biological function and RNA-Seq[J]. Journal of Biological Chemistry, 2019, 294(51): 19764-19784
- [23] Lambowitz AM, Zimmerly S. Group II introns: mobile ribozymes that invade DNA[J]. Cold Spring Harbor Perspectives in Biology, 2011, 3(8): a003616
- [24] Yao J, Truong DM, Lambowitz AM. Genetic and biochemical assays reveal a key role for replication restart proteins in group II intron retrohoming[J]. PLoS Genetics, 2013, 9(4): e1003469
- [25] Zhang HX, Zhang Y, Yin H. Genome editing with mRNA encoding ZFN, TALEN, and Cas9[J]. Molecular Therapy, 2019, 27(4): 735-746
- [26] Khalil AM. The genome editing revolution: review[J]. Journal, Genetic Engineering & Biotechnology, 2020, 18(1): 68
- [27] Shivram H, Cress BF, Knott GJ, Doudna JA. Controlling and enhancing CRISPR systems[J]. Nature Chemical Biology, 2021, 17(1): 10-19