

研究报告

分枝杆菌 LY-1 内源性表达元件的筛选及其在降低 Cas9 蛋白毒性中的应用

薛苗苗¹, 李会^{*1}, 曹雪岑¹, 张晓梅¹, 王淑丽², 陈立营³, 史劲松¹, 许正宏⁴

1 江南大学生命科学与健康工程学院, 江苏 无锡 214122

2 天津天药业股份有限公司, 天津 300301

3 天津药业研究院股份有限公司, 天津 300301

4 江南大学生物工程学院 工业生物技术教育部重点实验室, 江苏 无锡 214122

薛苗苗, 李会, 曹雪岑, 张晓梅, 王淑丽, 陈立营, 史劲松, 许正宏. 分枝杆菌 LY-1 内源性表达元件的筛选及其在降低 Cas9 蛋白毒性中的应用[J]. 微生物学通报, 2022, 49(8): 3267-3278

Xue Miaomiao, Li Hui, Cao Xuecen, Zhang Xiaomei, Wang Shuli, Chen Liying, Shi Jinsong, Xu Zhenghong. Endogenous expression elements of *Mycobacterium* sp. LY-1: screening and application in reducing Cas9 toxicity[J]. Microbiology China, 2022, 49(8): 3267-3278

摘要:【背景】分枝杆菌 LY-1 因能够将天然植物甾醇代谢转化为重要甾体药物中间体, 目前已成为工业上的优势生产菌株。高效的 CRISPR/Cas9 基因编辑技术是工业菌株代谢工程改造进行产量性状提升的关键。然而由于 Cas9 蛋白的高表达毒性问题且分枝杆菌中已公开报道的可用表达元件较少, 极大地限制了 Cas9 蛋白在该菌株中的适度表达。【目的】筛选内源性表达元件, 利用合适的表达元件启动 Cas9 蛋白的表达, 降低其对菌株的毒性。【方法】依据文献和前期研究获得的分枝杆菌基因转录组水平数据, 并结合启动子在线预测网站 BDGP (https://www.fruitfly.org/seq_tools/promoter.html), 筛选内源性表达元件。以增强型绿色荧光蛋白作为报告基因对表达元件的强度进行评估, 并采用不同强度的表达元件启动 Cas9 蛋白的表达。【结果】获得了 23 个不同表达强度的表达元件, 采用中等强度的表达元件及弱表达元件都降低了 Cas9 蛋白对分枝杆菌 LY-1 的毒性, 实现了 Cas9 蛋白在该菌株中的适度表达。【结论】建立了分枝杆菌 LY-1 内源性表达元件库, 为后续菌株中高效 CRISPR/Cas9 基因编辑技术的构建及关键酶基因调控奠定了良好的基础。

关键词: 分枝杆菌; 内源性表达元件; CRISPR/Cas9; 增强型绿色荧光蛋白; 基因表达与调控

基金项目: 国家重点研发计划(2019YFA0905300); 江苏省青蓝工程; 中央高校基本科研业务费专项资金(JUSRP221025)

Supported by: National Key Research and Development Program of China (2019YFA0905300); Qing Lan Project in Jiangsu Province; Fundamental Research Funds for Central Universities of China (JUSRP221025)

*Corresponding author: E-mail: lihui@jiangnan.edu.cn

Received: 2021-11-25; Accepted: 2022-02-24; Published online: 2022-03-31

Endogenous expression elements of *Mycobacterium* sp. LY-1: screening and application in reducing Cas9 toxicity

XUE Miaomiao¹, LI Hui^{*1}, CAO Xuecen¹, ZHANG Xiaomei¹, WANG Shuli², CHEN Liying³, SHI Jinsong¹, XU Zhenghong⁴

1 School of Life Sciences and Health Engineering, Jiangnan University, Wuxi 214122, Jiangsu, China

2 Tianjin Tianyao Pharmaceutical Limited Company, Tianjin 300301, China

3 Tianjin Pharmaceutical Research Institute Limited Company, Tianjin 300301, China

4 Key Laboratory of Industrial Biotechnology, Ministry of Education, School of Bioengineering, Jiangnan University, Wuxi 214122, Jiangsu, China

Abstract: [Background] *Mycobacterium* sp. LY-1 has become a dominant strain in industrial production because of its ability to metabolize natural phytosterols into important steroid drug intermediates. CRISPR/Cas9 as an efficient gene editing technology is the key to improving the yield and traits of industrial strains through metabolic engineering. However, due to the toxicity resulted from the high expression of Cas9 and the few available expression elements that have been reported in *Mycobacterium*, the moderate expression of Cas9 protein in *Mycobacterium* is greatly limited. [Objective] The endogenous expression elements were selected to activate the expression and reduce the toxicity of Cas9. [Methods] We used the online Berkeley Drosophila Genome Project (BDGP) to predict the endogenous expression elements from the transcriptome data of *Mycobacterium* genes in literature and available studies. The intensity of each expression element was assessed with enhanced green fluorescent protein as the reporter, and the expression of Cas9 protein was initiated with the expression elements of different intensities. [Results] Twenty-three expression elements with different expression intensities were obtained. The medium and weak expression elements reduced the toxicity of Cas9 to *Mycobacterium* sp. LY-1 and realized the moderate expression of Cas9 in the strain. [Conclusion] The endogenous expression element library of *Mycobacterium* sp. LY-1 was established, which laid a good foundation for the subsequent construction of CRISPR/Cas9 tools and the expression regulation of key enzymes in *Mycobacterium*.

Keywords: *Mycobacterium*; endogenous expression elements; CRISPR/Cas9; enhanced green fluorescent protein; gene expression and regulation

分枝杆菌(*Mycobacterium*) LY-1 可以将天然植物甾醇生物转化为重要甾体药物中间体 9 α -羟基雄甾-4-烯-3,17-二酮(9 α -OH-AD), 目前已成为工业上的优势生产菌株^[1-2]。虽然国内外研究者通过菌株诱变选育、发酵工艺优化等手段, 在一定程度上提高了分枝杆菌转化生成甾体药物中间体的合成效率, 但仍存在底物投料浓度低、菌株转化效率不高等瓶颈问题, 限制其工业化

应用^[3-6]。因此, 利用代谢工程改造技术获得高效生产菌株成为首要研究策略。

目前, 分枝杆菌的代谢工程改造主要依靠质粒进行关键酶的增强表达和基因敲除, 存在抗生素对菌株生长有抑制作用和两轮交换耗时长的缺点。CRISPR/Cas9 系统是一种新兴的基因编辑技术, 具有合成简单、使用方便、特异性高的特点, 广泛应用于原核生物和真核生

物的基因工程改造中^[7-8]。由于 Cas9 蛋白的毒性及低同源重组效率, 导致仅有个别分枝杆菌属(如结核分枝杆菌、海洋分枝杆菌和耻垢分枝杆菌)建立了成熟的 CRISPR/Cas9 基因编辑技术^[9]。基于分枝杆菌未能广泛建立 CRISPR/Cas9 基因编辑技术的原因之一是 Cas9 蛋白的表达对大多数分枝杆菌有较大的毒性^[10], 因此, 在分枝杆菌 LY-1 中构建 CRISPR/Cas9 系统之前筛选表达元件文库, 为调控 Cas9 蛋白的表达及 CRISPR/Cas9 系统在分枝杆菌中的构建提供可靠的分子操作元件, 具有非常重要的意义。

表达元件是包含启动子^[11]、核糖体结合位点(ribosome binding site, RBS)^[12]等能够使目的基因表达的元件^[13]。在分枝杆菌中, 已报道的牛型分枝杆菌 Hsp60 热休克蛋白基因前的表达元件(hsp60)是应用最广泛的一个强表达元件^[14]; 大肠杆菌中诱导型表达元件四环素调控表达系统 Tet[由脱水四环素(anhydrotetracycline, ATC)诱导]也可以控制和驱动分枝杆菌中基因的表达^[15-16]。然而在分枝杆菌 LY-1 中使用 hsp60 与 Tet 启动 Cas9 表达时, 使用强表达元件 hsp60 高表达 Cas9 蛋白, 直接导致无转化子长出; 使用诱导型表达元件 Tet 时, 诱导剂本身对菌株产生较大的毒性。此外, Sun 等^[17]近年探究了外源表达元件在分枝杆菌中的应用, 将 2 个单启动子联合起来形成双启动子, 与合适的 RBS 结合形成强表达元件, 过表达代谢途径中的关键酶, 提高了 9 α -OH-AD 的产量。然而分枝杆菌与大肠杆菌的 σ 70 共有序列差异显著, 而且表达元件的通用性较差, 导致外源表达元件并不能很好地适应分枝杆菌。因此, 分枝杆菌中可用的表达元件较少, 筛选分枝杆菌内源性表达元件可能会有更好的相容性和通用性^[18]。

分枝杆菌 LY-1 是本研究室自主筛选保藏

的用于 9 α -OH-AD 生物合成的生产菌株^[19]。前期初步研究发现, 分枝杆菌 LY-1 在 Cas9 蛋白高表达时几乎无菌体生长。为了降低 Cas9 蛋白对分枝杆菌 LY-1 的毒性, 本研究筛选分枝杆菌 LY-1 中的内源性表达元件, 以增强型绿色荧光蛋白作为报告基因评价表达元件的表达强度, 选取合适的表达元件启动 Cas9 蛋白的表达, 以期在分枝杆菌 LY-1 中 CRISPR/Cas9 系统的构建奠定基础, 也为异源基因的表达提供调控文库。

1 材料与方法

1.1 材料

1.1.1 菌株和质粒

分枝杆菌 *Mycobacterium* sp. LY-1 (CGMCC 13031)、大肠杆菌 JM109 及大肠杆菌和分枝杆菌的穿梭质粒 pMV261 均由本实验室保藏。

1.1.2 主要试剂和仪器

QuickCut 限制酶、PrimeSTAR[®] HS DNA Polymerase, TaKaRa 公司; Phanta HS Master Mix、Green Taq Mix 和 ClonExpress II One Step Cloning Kit, 诺唯赞生物科技股份有限公司; 卡那霉素和 EZ-10 柱式 DNA 纯化试剂盒, 生工生物工程(上海)股份有限公司; 细菌质粒小提试剂盒、柱式琼脂糖凝胶 DNA 回收试剂盒和细菌基因组 DNA 提取试剂盒(离心柱型), 上海捷瑞生物工程有限公司。电转仪和 C100 PCR 仪, Bio-Rad 公司; 多功能酶标仪, Molecular Devices 公司; 紫外分光光度计, 尤尼柯(上海)仪器有限公司; 激光共聚焦显微镜, Leica 公司; 流式细胞仪, BD 公司。

1.1.3 引物设计及合成

研究所用引物使用 CE Design 软件设计, 引物名称及序列如表 1 所示, 引物均由亦欣生物科技无锡有限公司合成。

表 1 PCR 引物信息

Table 1 The primers for PCR

引物 Primers	引物序列 Primer sequences (5'→3')
261-P04810-EGFP-F	taccagatctttaaactagaGTCAAACGACAGTTTCAGTGGCG
261-P04810-EGFP-R	gcccttgctcaccatggatccTTAGATTGCCTTCCCACGTGTC
261-P10585-EGFP-F	taccagatctttaaactagaGAATTAACCTGGCAACTGGACAAG
261-P10585-EGFP-R	gcccttgctcaccatggatccCTGGAATCCATTTCTTCGGTCA
261-P10665-EGFP-F	taccagatctttaaactagaGCCGCGCTGGGCACCTTC
261-P10665-EGFP-R	gcccttgctcaccatggatccCCCCACTAACTCCTCACAAAACCTAG
261-P10760-EGFP-F	taccagatctttaaactagaCCCGCCTTTGCGCAGGTA
261-P10760-EGFP-R	gcccttgctcaccatggatccGCGCCGACCACCTCCTGC
261-P10825-EGFP-F	taccagatctttaaactagaGAACGCATCGATTTCTTTACTGTAA
261-P00068-EGFP-R	gcccttgctcaccatggatccGACGCCGATGTAGTTCACCTCA
261-P04700-EGFP-F	taccagatctttaaactagaCGCCGGGTTTCGATTACG
261-P04700-EGFP-R	gcccttgctcaccatggatccGGCAGTCAGAGTATCGCGGT
261-P04825-EGFP-F	taccagatctttaaactagaCCAGGCGGGCCGCTGTGT
261-P04825-EGFP-R	gcccttgctcaccatggatccTGCGCTCAACGCACTTCG
261-P04895-EGFP-F	taccagatctttaaactagaGATAGACCTGAGTTCTCACCTGGTT
261-P04895-EGFP-R	gcccttgctcaccatggatccGCAGACCGCGTTACCTTCAC
261-P10590-EGFP-F	taccagatctttaaactagaCCGTGGATTTACTCACCGGTA
261-P10590-EGFP-R	gcccttgctcaccatggatccCAGAATCTCATCTCGCCCCGA
261-P10970-EGFP-F	taccagatctttaaactagaTCACAAAATGTTTGCCTGACCG
261-P10970-EGFP-R	gcccttgctcaccatggatccGTGCCGGTCTTCTCTGGCTA
261-P11010-EGFP-F	taccagatctttaaactagaGGGCGCGTCTTGGCGTTT
261-P11010-EGFP-R	gcccttgctcaccatggatccGTTCTGAATAACATTTCTGAAATTCCC
261-P25915-EGFP-F	taccagatctttaaactagaCAACGGGCATGGGCGGTA
261-P25915-EGFP-R	gcccttgctcaccatggatccGGGATCTCGTTTCTCAGGCC
261-P25925-EGFP-F	taccagatctttaaactagaCGTGACCTGGAACAAAAGCAGT
261-P25925-EGFP-R	gcccttgctcaccatggatccTTGCTTGTGGCACGGGAC
261-P04690-EGFP-F	taccagatctttaaactagaGTTTCGGCGAATGCGGACC
261-P04690-EGFP-R	gcccttgctcaccatggatccGACGGGAATCCTACCTTGTGTG
261-P26180-EGFP-F	taccagatctttaaactagaAGCCAGCAATGCTTGAAACAA
261-P26180-EGFP-R	gcccttgctcaccatggatccGAAGCCAGGCTAGAACACGTTT
261-P09240-EGFP-F	taccagatctttaaactagaCGGCGGCTGGGGCCCCAC
261-P09240-EGFP-R	gcccttgctcaccatggatccGTCCGGCTCCTTTGATCGGG
261-P26490-EGFP-F	taccagatctttaaactagaCCTAGACAGAGATTAGAACACGTTACG
261-P26490-EGFP-R	gcccttgctcaccatggatccGGCTGGACCCTTTCTTCACTC
261-P07820-EGFP-F	taccagatctttaaactagaGGGGCGGTCCCTGCCAGT
261-P07820-EGFP-R	gcccttgctcaccatggatccAGGACCACCTCCGAAGATGTC
261-P06810-EGFP-F	taccagatctttaaactagaCCGACCAGGCAAAAACAGCT
261-P06810-EGFP-R	gcccttgctcaccatggatccTTGTCCCGCGGCGTTGCG
261-hsp60-EGFP-F	ggccaagacaattggcgatccATGGTGAGCAAGGGCGGAGG
261-hsp60-EGFP-R	gacatcgataagctgaattcTTACTTGTACAGCTCGTCCATGCC
261-P10760-Casf-F	ccgtggcgcggcccggtaccCCCGCCTTTGCGCAGGTA
261-P10760-Casf-R	gtacttctgtccatgaattcGCGCCGACCACCTCCTGC
261-P06810-Casf-F	ccgtggcgcggcccggtaccCCGACCAGGCAAAAACAGCT
261-P06810-Casf-R	gtacttctgtccatgaattcTTGTCCCGCGGCGTTGCG
261-P25915-Casf-F	ccgtggcgcggcccggtaccCAACGGGCATGGGCGGTA
261-P25915-Casf-R	gtacttctgtccatgaattcGGGATCTCGTTTCTCAGGCC
261-P10970-Casf-F	ccgtggcgcggcccggtaccTCACAAAATGTTTGCCTGACCG
261-P10970-Casf-R	gtacttctgtccatgaattcGTGCCGGTCTTCTCTGGCTA
261-P00068-Casf-F	ccgtggcgcggcccggtaccATCCGCACCGGCGGTGCG
261-P00068-Casf-R	gtacttctgtccatgaattcGACGCCGATGTAGTTCACCTCA
261-P26490-Casf-F	ccgtggcgcggcccggtaccCCTAGACAGAGATTAGAACACGTTACG
261-P26490-Casf-R	gtacttctgtccatgaattcGGCTGGACCCTTTCTTCACTC
261-P09240-Casf-F	ccgtggcgcggcccggtaccCGGCGGCTGGGGCCCCAC
261-P09240-Casf-R	gtacttctgtccatgaattcGTCCGGTCTTCTTGTGATCGGG
261-P26180-Casf-F	ccgtggcgcggcccggtaccAGCCAGCAATGCTTGAAACAA
261-P26180-Casf-R	gtacttctgtccatgaattcGAAGCCAGGCTAGAACACGTTT
261-P07820-Casf-F	ccgtggcgcggcccggtaccGGGGCGGTCCCTGCCAGT
261-P07820-Casf-R	gtacttctgtccatgaattcAGGACCACCTCCGAAGATGTC

1.1.4 培养基与培养条件

分枝杆菌 LY-1 种子培养基(g/L): 酵母粉 15.0, 硝酸钠 5.4, 磷酸氢二铵 0.6, 甘油 2.0, 用于分枝杆菌种子液的培养; 分枝杆菌 LY-1 感受态培养基(g/L): 酵母粉 15.0, 硝酸钠 5.4, 磷酸氢二铵 0.6, 吐温-80 1.0, 甘油 2.0, 甘氨酸 20.0, 玻璃珠 10.0, 用于分枝杆菌感受态的制备; LB 液体培养基(g/L): 酵母粉 5.0, 氯化钠 10.0, 蛋白胨 10.0, 用于大肠杆菌的培养; LB 固体培养基(g/L): 酵母粉 5.0, 氯化钠 10.0, 蛋白胨 10.0, 琼脂粉 20.0, 用于大肠杆菌及分枝杆菌的平板培养。

大肠杆菌使用 LB 培养基在 37 °C、220 r/min 回旋式摇床中振荡培养, 分枝杆菌使用分枝杆菌种子培养基在 30 °C、120 r/min 往复式摇床中振荡培养。

1.2 方法

1.2.1 构建基于 EGFP 的表达元件报告系统

以 pMV261 质粒为模板, 采用 *Bam*H I 和 *Eco*R I 限制性内切酶对模板上 hsp60 表达元件后的多克隆位点进行双酶切, 获得线性化的 pMV261 质粒。以实验室保藏的 pTK-EGFP 质粒为模板, 利用引物 261-hsp60-EGFP-F 和 261-hsp60-EGFP-R 扩增增强型绿色荧光蛋白 (enhanced green fluorescent protein, EGFP) 基因。由于引物中携带用于重组克隆的同源臂, 因此采用同源重组单片段一步连接酶将线性化 pMV261 质粒与绿色荧光蛋白基因连接, 获得重组质粒 pMV261-hsp60-EGFP, 送至苏州金唯智生物技术有限公司进行测序验证。接着将重组质粒电转入分枝杆菌 LY-1 的感受态细胞中, 涂布于终浓度为 50 μg/mL 的 Kan^r 抗性平板上, 30 °C 静置培养 4-5 d, 待菌体长出。挑取单菌落于分枝杆菌种子培养基中培养, 使用流式细胞仪和激光共聚焦显微镜对表达了增强型绿色荧光蛋白

的菌株进行荧光测定。

1.2.2 分枝杆菌的基因操作

分枝杆菌的质粒转化使用电击转化法。分枝杆菌基因组的提取使用上海捷瑞生物工程有限公司的细菌基因组 DNA 提取试剂盒。

1.2.3 分枝杆菌表达元件的筛选及插入

经文献调研发现, 与分枝杆菌 LY-1 16S rRNA 基因相似性呈 100% 的分枝杆菌 VKM Ac-1817D, 其全基因组以植物甾醇为底物做了转录组水平分析^[20], 结果显示有 3 个区域的基因表达上调最明显, 分别在分枝杆菌 VKM Ac-1817D 全基因组的 1.1×10^6 、 2.3×10^6 和 5.5×10^6 bp 处。原核生物基因的高转录表明该基因上游可能存在高强度的表达元件, 所以本研究对分枝杆菌 VKM Ac-1817D 基因组这 3 个区域 $\pm 0.5 \times 10^6$ bp 内的基因上游进行了表达元件的搜索。研究室前期还对分枝杆菌 LY-1 甾醇代谢关键酶基因做了转录水平分析^[4], 发现在以植物甾醇为底物时, 有 5 个表达明显上调的关键酶基因, 分别为基因 *fadA5*、*kstD3*、*kshB*、*kshA2* 和 *hsdB*, 因此将表达元件区域的筛选定位于这 5 个关键酶基因上游处。

筛选得到的表达元件区域利用启动子预测评分网站 BDGP (https://www.fruitfly.org/seq_tools/promoter.html) 选择原核生物, 评分不低于 0.8 的条件进行预测, 最终从中筛选出 23 个评分在 0.8 及以上的表达元件核心基因序列。以预测得到的表达核心基因序列上游延伸 10-60 bp, 下游延伸至起始密码子 ATG 前, 以其整体作为分枝杆菌内源性表达元件, 这样既保证预测的启动子元件能够启动转录, 又可以利用基因本身搭配启动子的 RBS 序列启动翻译。提取分枝杆菌 LY-1 的基因组作为 PCR 模板, 设计引物扩增筛选得到的 23 个分枝杆菌内源性表达元件基因片段。报告系统质粒 pMV261-hsp60-EGFP 采用 *Xba*I 和 *Bam*H I 双酶切进行质粒线性化, 切胶回

收目的片段以去除 hsp60 表达元件基因序列, 然后将扩增好的分枝杆菌内源性表达元件与线性化的质粒同源重组连接, 构建不同启动子调控增强型绿色荧光蛋白表达的重组质粒(图 1)。

1.2.4 酶标仪检测重组菌的荧光值

将构建好的重组质粒电转入分枝杆菌 LY-1 中, 挑取阳性菌落于分枝杆菌 LY-1 的种子培养基中培养, 获得 23 株重组菌, 然后重组菌按 1% 的接种量接种于分枝杆菌种子培养基中, 培养至对数生长期(OD_{600} 在 1.0–2.0), 每组实验设置 3 个平行。以 pMV261 空质粒为空白对照, hsp60 作为标准高强度表达元件, 待菌体生长至 OD_{600} 为 1.5 时取样, 经 PBS 洗涤重悬、细玻璃珠振荡 5 min, 再静置 2 min 后取 100 μ L 加样至 96 孔板中, 酶标仪测波长 600 nm 处的吸光值及在 485 nm 激发波长、535 nm 发射波长处的荧光值, 以单位吸光度下的荧光强度大小作为判定表达元件强弱的标准, 实现分枝杆菌内源性表达元件的强度测定。

1.2.5 构建含 Cas9 蛋白的重组质粒

采用穿梭质粒 pMV261 作为载体, 密码子偏好性分析及密码子优化 Cas9 基因序列由生工生物工程(上海)股份有限公司化学合成至质粒 pMV261 的 *EcoR* I 和 *Hind* III 酶切位点, 密码子优化的 Cas9 基因以 *Casf* 表示, 以此来区分密码子未优化的 *Cas9*^[21]。选择质粒上的 *Kpn* I 和 *EcoR* I 的酶切位点进行质粒线性化, PCR 扩增 P10760、P06810、P25915 强表达元件序列, P10970、P00068、P26490 中等表达元件和 P09240、P26180、P07820 弱表达元件序列, 重组质粒构建方法同 1.2.1。

2 结果与分析

2.1 基于 EGFP 表达元件报告系统的构建

以 *EGFP* 为报告基因, 按照 1.2.1 中的方法在大肠杆菌 JM109 中构建重组质粒 pMV261-hsp60-*EGFP*。将重组质粒 pMV261-hsp60-*EGFP*

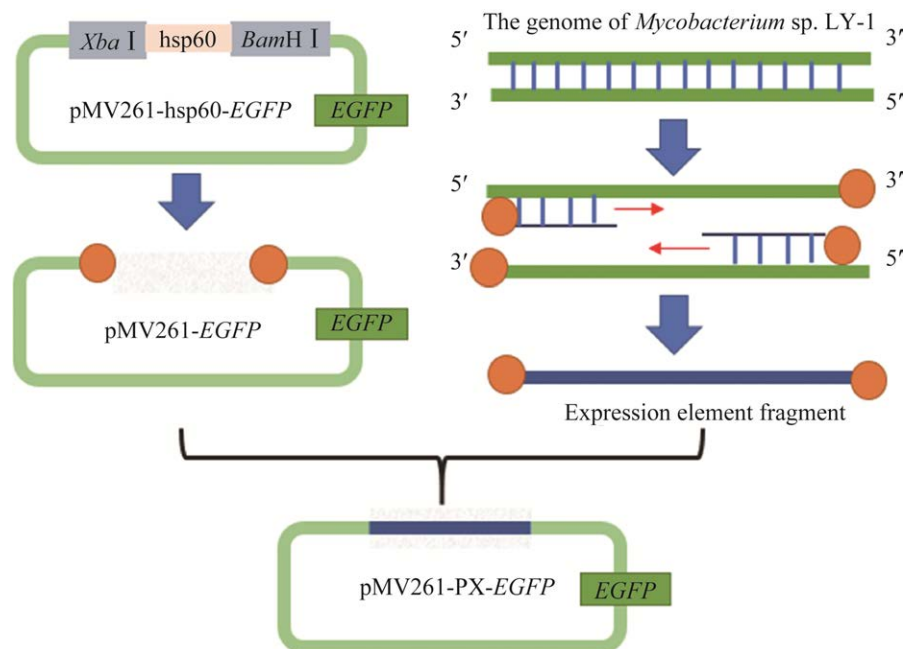


图 1 内源性表达元件重组质粒构建图

Figure 1 Construction process of recombinant plasmid with endogenous expression element. “PX” is the name of the expression element.

电转入分枝杆菌 LY-1 中获得重组菌株。以带有空白质粒 pMV261 的分枝杆菌 LY-1 为对照, 利用流式细胞仪和激光共聚焦观察重组菌中的荧光显示情况。如图 2 所示, 与对照相比, 在激光共聚焦显微镜下观察, 重组菌株显示出绿色荧光; 流式细胞仪检测结果也显示, 重组菌有明显的绿色荧光菌落(图 3)。这些均表明绿色荧光蛋白在分枝杆菌 LY-1 中成功表达, 可以将其作为报告基因验证表达元件的强度。

2.2 内源性表达元件的特征序列分析及重组质粒的构建

BPROM 网站是细菌 $\sigma 70$ 启动子识别程序, 准确率和特异性能够达到 80%, 最适合用于 ORF 开始的直接上游区域, 用于改进细菌中的基因和操纵子预测。采用该网站对筛选得到的表达元件核心序列的-10 区和-35 区进行预测, 结果显示筛选得到的 23 个内源性表达元件均存在细菌保守的-10 区和-35 区(表 2)。

以重组质粒 pMV261-hsp60-EGFP 为模板, 分别用不同的表达元件替换原有重组质粒中的

hsp60 表达元件, 构建不同内源性表达元件启动绿色荧光蛋白表达的重组质粒。如图 4 所示, 23 个内源性表达元件重组质粒均构建成功。

2.3 分枝杆菌内源性表达元件强度的表征

按照方法 1.2.3 对分枝杆菌内源性表达元件的强度进行表征, 结果如图 5 所示。筛选出的 23 个表达元件中存在 4 个高强度表达元件(P10760、P06810、P25915 和 P04690), 15 个中

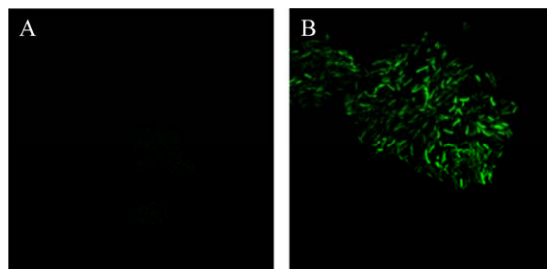


图 2 分枝杆菌 LY-1 (pMV261-hsp60-EGFP) 菌株的激光共聚焦结果 A: 空白对照(pMV261); B: 重组菌(pMV261-hsp60-EGFP)

Figure 2 Laser confocal results of *Mycobacterium* sp. LY-1 (pMV261-hsp60-EGFP) strain. A: Blank control (pMV261); B: Recombinant (pMV261-hsp60-EGFP).

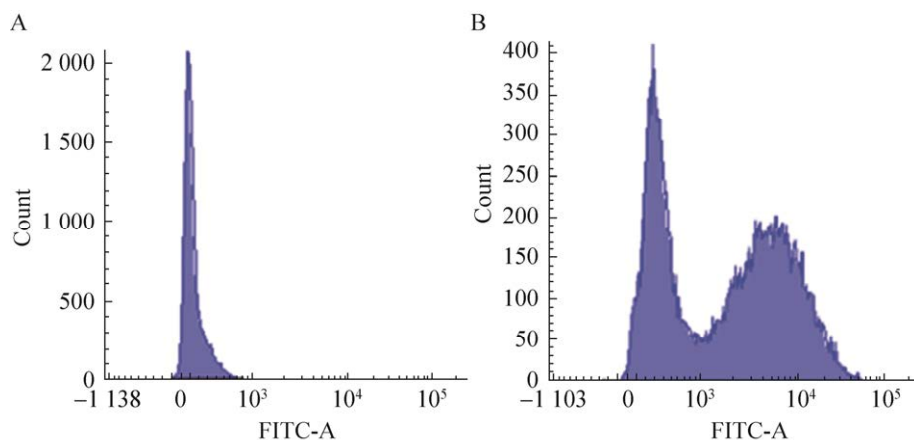


图 3 分枝杆菌 LY-1 (pMV261-hsp60-EGFP) 菌株的流式细胞仪结果 A: 空白对照(pMV261); B: 重组菌(pMV261-hsp60-EGFP)

Figure 3 Flow cytometry results of *Mycobacterium* sp. LY-1 (pMV261-hsp60-EGFP) strain. A: Blank control (pMV261); B: Recombinant (pMV261-hsp60-EGFP).

表 2 启动子特征分析

Table 2 Promoter characteristics analysis

基因 ID Gene ID	表达元件 Expression element	功能 Function	预测的-35 区和-10 区 Predicted -35 zone and -10 zone (5'→3')	评分 Score
G155_04810	P04810	F420-dependent oxidoreductase coenzyme	ggcggcttggctggctgtcacacatcgagtagaacaggttctaa caatcg	0.93
G155_10585	P10585	Putative protein	tgccaactggacaagatttctcagctcgtggaattccactacga ggcggcgcaatact	0.83
G155_10665	P10665	Putative protein	ctgtttagatcagcagcgatcaatttctagaagcaattctagttt gt	0.87
G155_10760	P10760	Enoyl-CoA hydratase	ggacttgggaatgcactaccgataggtatatacagtatgtatgtctg agcg	0.84
G155_10825	P10825	Putative cytochrome P450 hydroxylase	cgattcttactgtaatgtttacggaatctactcagagcgatcgg tc	0.86
G155_10835	P10835	Aldehyde dehydrogenase	ggatgactgtaaacattcaggctggtcggacaatcgcttcgac ctgtg	0.92
G155_10900	P10900	Rifampicin monooxygenase	ggcttggcgcaaggccccgggtggtgcagaatcgctccc gcagttc	0.81
G155_26060	P26060	Putative oxidoreductase/short-chain dehydrogenase	accgttgacaaccggggcgcagcttccgtatagttaccaacc ggagc	0.91
G155_00068	P00068	Putative protein	ttgaaaattcgctactgattgcaatcgcaaacatgaacttgaat cgcggccgaa	0.94
G155_04700	P04700	Long-chain fatty acid AMP ligase, <i>Mycobacterium</i> subgroup FadD28	cgtagatgtgattcataaggctactcaaatatagttgctggctg	0.89
G155_04825	P04825	Putative protein	acaggcttcgcataggcagctcaattagatagctcgaacatcg gggat	0.80
G155_04895	P04895	Ferredoxin	gtctgactggattcgtcaggtgtgctgattctgtggctgcgcg cagc	0.91
G155_10590	P10590	Leucovorin deformylase	gatttactaccggtacatgcaatttggcaattaggagttcatc att	0.92
G155_10970	P10970	Putative protein	aatgttttgcgtgaccgaccgggctcgtctgatcgcgaactg ccaaggcgactttagtagcaaccgcagcgggtctgatcag accgacc	0.95
G155_11010	P11010	The glutathione reductase-like protein NrdH required for the reduction of ribonucleotide reductase class Ib	ggcgttttggcgaagcggcgttccggcaaaattctgccc ggctg	0.97
G155_25915	P25915	Protein EspJ, a component of the type VII secretion system ESX-1	ttgtgacaggtgggtcaccggggtccgtaaacccgcagcag tg	0.87
G155_25925	P25925	Proline- and alanine-rich protein EspK, a component of the type VII secretion system ESX-1	gccctcattaaggcgatcgcaatcaatcacacaataaacctat catga	0.86
G155_04690	P04690	Putative protein	ttcgttccgaaacggcgggaagaacaacagaataacccga tagcagcgggagccg	0.85
G155_26180	P26180	Putative acetyl-CoA acetyltransferase FadA5	acggcttgaccgcagatcaaggcagtgatacatccgcgttgt cagagt	0.82
G155_09240	P09240	3-sterone- Δ 1-dehydrogenase	cggcgggtgaccaccatcacccttggccaagattgcggtgca gctcac	0.81
G155_26490	P26490	3-ketone steroid-9- α -hydroxylase reductase subunit	cttgtaatgacggctggaccagcgaagtacgttttgaacatg ttct	0.86
G155_07820	P07820	Rieske (2Fe-2S) domain protein	gtttggggctcaacgtcagtcgcccggacatcttcggaggtggt cct	0.83
G155_06810	P06810	Hydroxysteroid dehydrogenase	ggggttgggttggctcgggggaggtgtacctcttaaggccc cgca	0.80

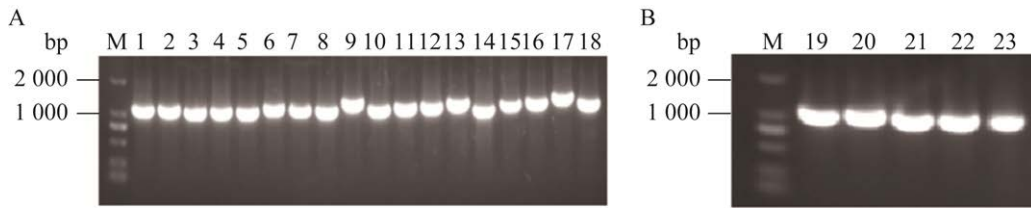


图4 内源性表达元件重组质粒构建 A: 文献中3个高转录水平上调区域所预测得到的18个表达元件; B: 5个关键基因前所预测得到的表达元件

Figure 4 Construction of recombinant plasmid of endogenous expression element. A: 18 expression elements predicted from 3 up-regulated regions of high transcription level in the literature; B: The predicted expression elements of five key genes.

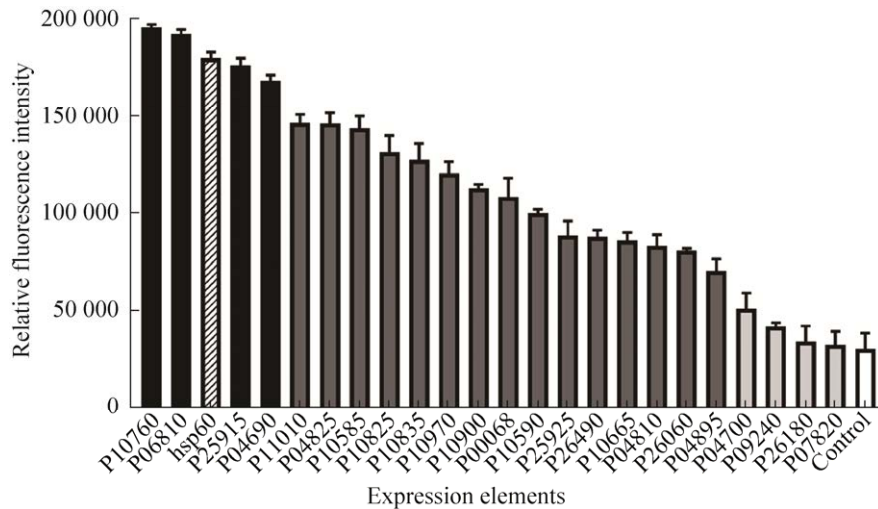


图5 分枝杆菌 LY-1 中各内源性表达元件表达 EGFP 的荧光强度

Figure 5 Fluorescence intensity of EGFP expressed by endogenous expression elements in *Mycobacterium* sp. LY-1.

强度表达元件(P11010、P04825、P10585、P10825、P10835、P10970、P10900、P00068、P10590、P25925、P26490、P10665、P04810、P26060和P04895)和4个弱表达元件(P04700、P09240、P26180和P07820)。表达元件的表达效果范围为1–16倍,其中有2个表达元件P10760、P06810的效果高于已报道的hsp60强表达元件。

2.4 内源性表达元件在降低 Cas9 蛋白毒性中的应用

按照方法 1.2.5 构建不同强度的表达元件用于表达 Cas9 蛋白的重组质粒,并将构建好的重

组质粒分别电转入分枝杆菌 LY-1 中,平板上培养 4–5 d,观察其菌落数;并分别挑取单菌落接入分枝杆菌 LY-1 的种子培养基中培养 4 d,检测各重组菌的生长情况。如图 6 显示,在强表达元件(P10760、P06810和P25915)的调控下,平板上几乎无菌落长出,而且菌体生长受到严重的抑制,证实了 Cas9 蛋白的高表达对菌株有较大的毒性。在中等表达元件(P10970、P00068和P26490)的调控下,菌落数与生长状况有明显的提升;而使用弱表达元件(P09240、P26180和P07820)表达 Cas9 蛋白时,菌落数和菌株生长曲

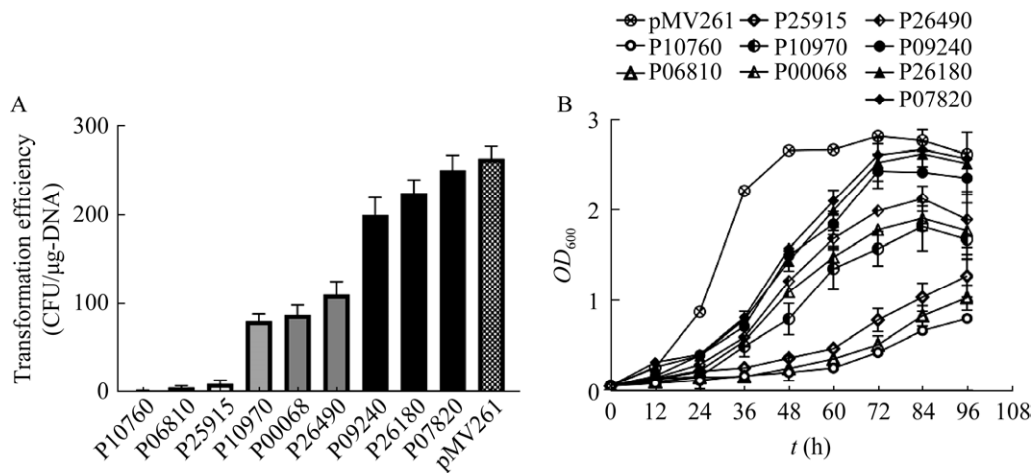


图6 分枝杆菌 LY-1 中不同内源性表达元件表达 Cas9 蛋白对菌株的影响 A: 不同内源性表达元件表达 Cas9 蛋白的重组菌菌落数; B: 不同内源性表达元件表达 Cas9 蛋白的重组菌生长曲线。pMV261: 含空载体的重组菌; P10760: 内源性表达元件 P10760 表达 Cas9 蛋白的重组菌; 其余同理

Figure 6 Effects of different endogenous expression elements expressing Cas9 protein on the strain in *Mycobacterium* sp. LY-1. A: The colony count of recombinant bacteria expressing Cas9 protein with different endogenous expression elements; B: The growth curve of recombinant bacteria expressing Cas9 protein with different endogenous expression elements; pMV261: Recombinant bacteria with empty vector; P10760: Recombinant bacteria expressing Cas9 protein with endogenous expression element P10760, the rest are the same.

线的最高点几乎接近对照,表明弱表达元件表达 Cas9 蛋白显著降低了 Cas9 蛋白对菌株的毒害作用。综上所述,分枝杆菌 LY-1 中筛选得到的中等或弱强度内源性表达元件均能够有效降低 Cas9 蛋白对菌株的毒性,为分枝杆菌 LY-1 中构建 CRISPR/Cas9 基因编辑方法提供了良好的开端。

3 讨论与结论

分枝杆菌作为生产甾体药物中间体的优势菌株,亟待开发便捷、高效的基因编辑工具。CRISPR-Cas9 系统在分枝杆菌中的构建主要面临的挑战有:(1) Cas9 蛋白对分枝杆菌有较大的毒性;(2) 分枝杆菌内源性同源重组能力较低。解决 Cas9 蛋白的毒性问题是分枝杆菌能否成功构建该系统的前提。研究者通常采用诱导型表达元件表达 Cas9 蛋白^[22-23],通过调整诱导剂的浓

度来控制蛋白的表达量,降低其对菌株的毒性。由于诱导剂对一些菌株有毒性且存在渗漏表达的情况^[24],所以诱导型表达元件不具有广泛的通用性。因此,在分枝杆菌中筛选不同强度的内源性表达元件,不仅为调控 Cas9 蛋白的表达提供一系列表达元件,也为分枝杆菌代谢途径的改造工程提供了工具,具有非常重要的意义。

启动子与 RNA 转录起始有关,-35 区是 RNA 聚合酶因子的识别位点,-10 区是 RNA 聚合酶与 DNA 的结合位点,基因的高转录水平主要归因于较强的启动子^[25-26]。RBS 与蛋白翻译有关,存在于目标基因上游 5-10 bp 处。蛋白的表达不仅需要启动子对蛋白基因进行转录,还需要包含 RBS 序列以便于核糖体 RNA 的识别和结合。因此,本研究对基因组高转录水平区域进行筛选并将表达元件序列延伸至目标基因 ATG

前, 使之包含天然的 RBS。对得到的表达元件进行启动子特征序列分析, 结果显示皆包含-35 区和-10 区 2 个保守区, 这表明此种筛选方法可以准确地得到潜在的细菌内源性表达元件区域。

蛋白表达元件强度的调控是控制蛋白表达的一种重要手段^[27], 为了调控分枝杆菌中 Cas9 蛋白的表达, 需筛选一系列不同强度的表达元件。本研究筛选出 23 个不同强度的内源性表达元件, 并采用强、中、弱 3 种不同的表达元件启动 Cas9 蛋白的表达, 结果表明降低表达元件的强度, Cas9 蛋白对菌株的毒害作用明显减弱, 为 CRISPR/Cas9 系统在分枝杆菌中的构建提供可能。同时, 分枝杆菌代谢途径中关键酶的表达元件也可以利用筛选得到的内源性表达元件进行替换, 提高甾体药物中间体的产量或降低副产物的含量。因此, 本研究为分枝杆菌蛋白表达提供了丰富的表达元件选择, 为之后基因编辑方法的构建及代谢通路的改造奠定了基础。

REFERENCES

- [1] Fernández-Cabezón L, Galán B, García JL. New insights on steroid biotechnology[J]. *Frontiers in Microbiology*, 2018, 9: 958
- [2] Galán B, Uhlía I, García-Fernández E, Martínez I, Bahillo E, De La Fuente JL, Barredo JL, Fernández-Cabezón L, García JL. *Mycobacterium smegmatis* is a suitable cell factory for the production of steroidal synthons[J]. *Microbial Biotechnology*, 2017, 10(1): 138-150
- [3] 马洋, 王向栋, 王萌慧, 李会, 史劲松, 许正宏. 高效转化植物甾醇为 9 α -OH-AD 的分枝杆菌诱变选育及工艺优化[J]. *生物工程学报*, 2017, 33(7): 1198-1206
Ma Y, Wang XD, Wang MH, Li H, Shi JS, Xu ZH. Mutation breeding of high 9 α -hydroxy-androst-4-ene-3,17-dione transforming strains from phytosterols and their conversion process optimization[J]. *Chinese Journal of Biotechnology*, 2017, 33(7): 1198-1206 (in Chinese)
- [4] 王向栋, 李会, 陈志蔚, 蔡兆培, 史劲松, 许正宏. 分枝杆菌 LY-1 转化植物甾醇产 9 α 羟基雄烯二酮的发酵工艺优化[J]. *生物加工过程*, 2019, 17(2): 177-183
Wang XD, Li H, Chen ZW, Cai ZP, Shi JS, Xu ZH. Fermentation process of *Mycobacterium* sp. LY-1 for the transformation of phytosterol to 9 α -hydroxy-4-androstenedione[J]. *Chinese Journal of Bioprocess Engineering*, 2019, 17(2): 177-183 (in Chinese)
- [5] Li H, Wang XD, Zhou LF, Ma Y, Yuan WJ, Zhang XM, Shi JS, Xu ZH. Enhancing expression of 3-ketosteroid-9 α -hydroxylase oxygenase, an enzyme with broad substrate range and high hydroxylation ability, in *Mycobacterium* sp. LY-1[J]. *Applied Biochemistry and Biotechnology*, 2019, 187(4): 1238-1254
- [6] Zhang Y, Zhou XL, Wang XM, Wang L, Xia ML, Luo JM, Shen YB, Wang M. Improving phytosterol biotransformation at low nitrogen levels by enhancing the methylcitrate cycle with transcriptional regulators PrpR and GlnR of *Mycobacterium neoaurum*[J]. *Microbial Cell Factories*, 2020, 19(1): 13
- [7] Corts AD, Thomason LC, Gill RT, Gralnick JA. Efficient and precise genome editing in *shewanella* with recombineering and CRISPR/Cas9-mediated counterselection[J]. *ACS Synthetic Biology*, 2019, 8(8): 1877-1889
- [8] Liang YX, Jiao S, Wang MM, Yu HM, Shen ZY. A CRISPR/Cas9-based genome editing system for *Rhodococcus ruber* TH[J]. *Metabolic Engineering*, 2020, 57: 13-22
- [9] Chimukuche NM, Williams MJ. Genetic manipulation of non-tuberculosis *Mycobacteria*[J]. *Frontiers in Microbiology*, 2021, 12: 633510
- [10] Sun BB, Yang JJ, Yang S, Ye RD, Chen DJ, Jiang Y. A CRISPR-Cpf1-assisted non-homologous end joining genome editing system of *Mycobacterium smegmatis*[J]. *Biotechnology Journal*, 2018, 13(9): 1700588
- [11] Brázda V, Bartas M, Bowater RP. Evolution of diverse strategies for promoter regulation[J]. *Trends in Genetics*, 2021, 37(8): 730-744
- [12] 李法君. 启动子研究进展[J]. *生物学教学*, 2017, 42(7): 2-3
Li FJ. Advances in promoter research[J]. *Biology Teaching*, 2017, 42(7): 2-3 (in Chinese)
- [13] 胡廷章, 罗凯, 甘丽萍, 石汝杰. 植物基因启动子的类型及其应用[J]. *湖北农业科学*, 2007, 46(1): 149-151
Hu TZ, Luo K, Gan LP, Shi RJ. Classification and application of the plant gene promoters[J]. *Hubei Agricultural Sciences*, 2007, 46(1): 149-151 (in Chinese)
- [14] Al-Zarouni M, Dale JW. Expression of foreign genes in *Mycobacterium bovis* BCG strains using different promoters reveals instability of the hsp60 promoter for expression of foreign genes in *Mycobacterium bovis* BCG strains[J]. *Tuberculosis*, 2002, 82(6): 283-291

- [15] Ehrst S, Guo XV, Hickey CM, Ryou M, Monteleone M, Riley LW, Schnappinger D. Controlling gene expression in mycobacteria with anhydrotetracycline and Tet repressor[J]. *Nucleic Acids Research*, 2005, 33(2): e21
- [16] Yan MY, Yan HQ, Ren GX, Zhao JP, Guo XP, Sun YC. CRISPR-Cas12a-assisted recombineering in bacteria[J]. *Applied and Environmental Microbiology*, 2017, 83(17): e00947-e00917
- [17] Sun H, Yang JL, Song H. Engineering mycobacteria artificial promoters and ribosomal binding sites for enhanced sterol production[J]. *Biochemical Engineering Journal*, 2020, 162: 107739
- [18] Barker LP, Porcella SF, Wyatt RG, Small PLC. The *Mycobacterium marinum* G13 promoter is a strong sigma 70-like promoter that is expressed in *Escherichia coli* and mycobacteria species[J]. *FEMS Microbiology Letters*, 1999, 175(1): 79-85
- [19] 马洋. 利用植物甾醇生成 9 α -OH-AD 的菌种筛选及其特性研究[D]. 无锡: 江南大学硕士学位论文, 2017
Ma Y. Screening of a strain transforming phytosterols into 9 α -OH-AD and its characteristics study[D]. Wuxi: Master's Thesis of Jiangnan University, 2017 (in Chinese)
- [20] Bragin EY, Shtratnikova VY, Schelkunov MI, Dovbnya DV, Donova MV. Genome-wide response on phytosterol in 9-hydroxyandrostenedione-producing strain of *Mycobacterium* sp. VKM Ac-1817D[J]. *BMC Biotechnology*, 2019, 19(1): 39
- [21] Quax TEF, Claassens NJ, Söll D, Van Der Oost J. Codon bias as a means to fine-tune gene expression[J]. *Molecular Cell*, 2015, 59(2): 149-161
- [22] Cho JS, Choi KR, Prabowo CPS, Shin JH, Yang D, Jang J, Lee SY. CRISPR/Cas9-coupled recombineering for metabolic engineering of *Corynebacterium glutamicum*[J]. *Metabolic Engineering*, 2017, 42: 157-167
- [23] Hartz P, Gehl M, König L, Bernhardt R, Hannemann F. Development and application of a highly efficient CRISPR-Cas9 system for genome engineering in *Bacillus megaterium*[J]. *Journal of Biotechnology*, 2021, 329: 170-179
- [24] Liang YX, Wei YW, Jiao S, Yu HM. A CRISPR/Cas9-based single-stranded DNA recombineering system for genome editing of *Rhodococcus opacus* PD630[J]. *Synthetic and Systems Biotechnology*, 2021, 6(3): 200-208
- [25] Feklistov A, Darst SA. Structural basis for promoter -10 element recognition by the bacterial RNA polymerase σ subunit[J]. *Cell*, 2011, 147(6): 1257-1269
- [26] Hinton DM. Prokaryotic Transcription[M]. *Encyclopedia of Cell Biology*. Amsterdam: Elsevier, 2016: 468-480
- [27] Li L, Meng HY, Zhang JM, Liu Y, Zou QJ, Gao Y, Yang HQ, Lai LX. A tunable, rapid, and precise drug control of protein expression by combining transcriptional and post-translational regulation systems[J]. *Journal of Genetics and Genomics*, 2020, 47(11): 705-712