

结核分枝杆菌 H37Rv 新基因 Rv2742 克隆表达及纯化

赵加玲^{1,3}, 武舒佳^{2,3}, 王红^{3,4}, 李芊璘⁴, 孙金帅^{3,5}, 常蕾³, 戴二黑⁴, 武军驻²,
张瑶^{3,6}, 徐平^{1,3}

1 武汉大学 药学院, 湖北 武汉 430000

2 武汉大学 基础医学院, 湖北 武汉 430000

3 军事科学院军事医学研究院生命组学研究所 国家蛋白质科学中心 (北京) 北京蛋白质组研究中心 蛋白质组学国家重点实验室, 北京 102206

4 华北理工大学 公共卫生学院和附属石家庄市第五医院, 河北 唐山 063210

5 河北大学 生命科学学院, 河北 保定 071000

6 中山大学 生命科学学院, 广东 广州 510275

赵加玲, 武舒佳, 王红, 等. 结核分枝杆菌 H37Rv 新基因 Rv2742 克隆表达及纯化. 生物工程学报, 2019, 35(9): 1771-1786.
Zhao JL, Wu SJ, Wang H, et al. Cloning, expression and purification of novel gene Rv2742 in *Mycobacterium tuberculosis* H37Rv. Chin J Biotech, 2019, 35(9): 1771-1786.

摘要: Rv2742 是本课题组前期基于蛋白质基因组学策略从结核分枝杆菌 *Mycobacterium tuberculosis* H37Rv 中发现、鉴定的遗漏注释基因。文中旨在建立结核分枝杆菌 H37Rv 漏注释蛋白 Rv2742 的可溶性诱导表达、纯化体系, 为进一步探索 Rv2742 基因参与的生物学功能奠定基础。前期实验发现构建的 pGEX-4T-2-Rv2742、pET-28a-Rv2742、pET-32a-Rv2742 及 pMAL-c2X-Rv2742 原核表达载体均无法实现目的蛋白的诱导表达。但经密码子优化后, 仅有 pMAL-c2X-Rv2742 载体能够实现目的蛋白的可溶性诱导表达。此外, 通过比较不同宿主菌、温度及 IPTG 浓度对目的蛋白表达量的影响, 发现目的蛋白在 Rosetta (DE3) 中, 16 °C 及 0.5 mmol/L IPTG 诱导条件下表达量最高。直链淀粉树脂 (Amylose resin) 亲和层析柱纯化获得较纯的产物, 经 LC-MS/MS 验证确认是 Rv2742 融合蛋白肽段序列。成功获得结核分枝杆菌 H37Rv 新基因 Rv2742 的重组蛋白, 可进一步开展其潜在相互作用及免疫原性研究工作。

关键词: 结核分枝杆菌, 新基因, 原核表达, 亲和纯化

Received: March 12, 2019; **Accepted:** June 14, 2019

Supported by: Chinese National Basic Research Programs (No. 2017YFC0906600), National Megaprojects for Key Infectious Diseases (No. 2018ZX10302302001003), National Natural Science Foundation of China (Nos. 31670834, 31870824, 91839302), Applied and basic research foundation of Guangdong Province (No. 2018A030310257).

Corresponding authors: Ping Xu. Tel: +86-10-61777113; Fax: +86-10-61777050; E-mail: xupingghy@gmail.com

Yao Zhang. Tel: +86-10-61777113; Fax: +86-10-61777050; E-mail: zhangyaows@163.com

国家精准医学重大专项 (No. 2017YFC0906600), 国家传染病重大专项 (No. 2018ZX10302302001003), 国家自然科学基金 (Nos. 31670834, 31870824, 91839302), 广东省基础及应用基础研究博士科研启动项目 (No. 2018A030310257) 资助。

Cloning, expression and purification of novel gene *Rv2742* in *Mycobacterium tuberculosis* H37Rv

Jialing Zhao^{1,3}, Shujia Wu^{2,3}, Hong Wang^{3,4}, Qianlin Li⁴, Jinshuai Sun^{3,5}, Lei Chang³, Erhei Dai⁴, Junzhu Wu², Yao Zhang^{3,6}, and Ping Xu^{1,3}

¹ School of Pharmaceutical Sciences, Wuhan University, Wuhan 430000, Hubei, China

² School of Basic Medical Sciences, Wuhan University, Wuhan 430000, Hubei, China

³ State Key Laboratory of Proteomics, Beijing Proteome Research Center, National Center for Protein Sciences (Beijing), Beijing Institute of Lifeomics, Beijing 102206, China

⁴ School of Public Health and Affiliated Shijiazhuang Fifth Hospital, North China University of Science and Technology, Tangshan 063210, Hebei, China

⁵ School of Life Sciences, Hebei University, Baoding 071000, Hebei, China

⁶ School of Life Sciences, Sun Yat-sen University, Guangzhou 510275, Guangdong, China

Abstract: *Rv2742* is a novel gene identified from *Mycobacterium tuberculosis* H37Rv by the proteogenomics strategy. The aim of this study was to establish a system of soluble expression and purification of the missing protein *Rv2742* in *M. tuberculosis* H37Rv, to provide reference for further research on the biological function of *Rv2742*. The soluble protein was not successfully induced by prokaryotic expression vectors pGEX-4T-2-*Rv2742*, pET-32a-*Rv2742*, pET-28a-*Rv2742* and pMAL-c2X-*Rv2742*. After the codon of novel gene *Rv2742* was optimized according to *E. coli* codon usage frequency, only the recombinant strain containing plasmid pMAL-c2X-*Rv2742* could produce soluble products of *Rv2742* encoding gene. In addition, the expression effects of the desired fusion protein were also analyzed under different conditions including hosts, culture temperatures and IPTG concentrations. The optimum expression conditions were as follows: Rosetta (DE3) host, 16 °C culture temperature and 0.5 mmol/L IPTG. After being purified by affinity chromatography with amylose resin, the fusion protein sequence was confirmed by LC-MS/MS. These results indicated that the novel gene *Rv2742* product could be successfully induced and expressed in a soluble form by the expression system pMAL-c2X with MBP tag. Our findings provide reference for studies on potential interaction and immunogenicity.

Keywords: *Mycobacterium tuberculosis*, novel gene, prokaryotic expression, affinity purification

结核分枝杆菌 (*Mycobacterium tuberculosis*, MTB) 是引起人类结核病的病原菌^[1], 可入侵多种组织、器官如胸膜、淋巴管、泌尿生殖器和骨关节炎等^[2], 但以肺结核最为常见。据 2018 年 WHO 报道, 2017 年全球约有 130 万人死于该病, 有 1 000 万新病例发生。结核病仍然是传染病中的头号杀手。我国依然是全球结核高发病、高耐药、高 HIV 共感染负担国家之一^[3]。因此, 有效的早期诊断新方法和及时治疗是控制结核疫情蔓延的重要手段^[4]。

目前, 临床常用的诊断方法包括涂片显微镜法、细菌培养法、核酸扩增试验 (如 GeneXpert MTB/RIF)、结核菌素皮肤试验 (TST)、 γ 干扰素

释放试验 (IGRA) 等^[5], 但存在涂片染色检出率低、培养周期长、结核抗原特异性差和灵敏度不高等问题。其中, 结核菌素皮肤试验因与卡介苗 (BCG) 及非结核分枝杆菌之间存在交叉反应而导致特异性不高^[6], 而基于 ESAT-6 与 CFP-10 研发的 T-SPOT 试剂盒虽被广泛地用于临床结核菌感染的筛查, 但仍存在“阳性不能确证, 阴性不能排除”的弊端, 不能有效区分活动性结核患者和潜伏感染者^[7]。因此, 寻找能够有效区分活动性结核患者和潜伏感染者高特异性、灵敏性的有效抗原可解决结核临床诊断的难题。

1998 年, Cole 等^[8-9]基于全基因组鸟枪法和系统测序法, 完成了结核分枝杆菌 H37Rv 的基因

组测序和注释工作。2002 年, Camus 等^[10]通过 EMBL、TrEMBL 和 SWISS-PROT 三个数据库序列同源性比较, 补充了 82 个新编码基因。传统的基因组注释是利用已知的基因特征和同源性比较实现的, 难以发现人类未知的、新基因结构特征的基因, 造成这些生物学功能和对菌株鉴别可能重要的基因遗漏, 逃离人类研究的视线, 形成盲区。国内外研究者基于算法优化^[10]、基因预测软件及注释工具开发、转录组学和蛋白质组学^[11-15]、比较基因组学^[16]等策略, 一直在校正和完善 H37Rv 基因组注释数据库。然而, MTB 属于原核生物, 受基因组测序质量影响以及缺乏合适的校正评估方法, 原核生物基因组注释中尚可能存在注释错误^[17](过度注释; ORF 起始、终止位点注释错误; 可变剪接; 核糖体移位; 漏注释等), 给准确解析相应物种的生物学机制带来了困扰^[18]。

利用蛋白质组学数据, 结合基因组数据、转录组数据来研究基因组注释问题, 被称为蛋白质基因组学 (Proteogenomics)^[19]。近 10 年来, 蛋白质组学直接用于基因组的注释已经越来越受到相关领域的关注。本课题组前期基于深度覆盖的精准蛋白质基因组学技术^[20,24]和比较基因组学策略, 完成了 H37Rv 基因组数据库的重注释, 发现了 22 个结核遗漏注释基因, 矫正了 28 个 N 端注释错误基因 (图 1, 尚未发表)。此发现有利于结核新型特异性免疫原性分子标志物筛选, 为结核病的快速、精准临床检测提供基础理论和原始创新技术支持。

从 TubercuList 库中下载了 H37Rv (<http://genolist.pasteur.fr/TubercuList/>) 基因组注释数据库, 共 4 031 个基因, 包含 4 018 个编码基因和 13 个假基因。六阅读框翻译数据库由中国科学院计算技术研究所 pFind 团队与我们课题组合作开发的蛋白质基因组学软件 pAnno 对 NCBI 中结核分枝杆菌 H37Rv 的基因组文件 (NC000962.3.fna) 按照正链和互补链上连续的 3 个核苷酸翻译一个

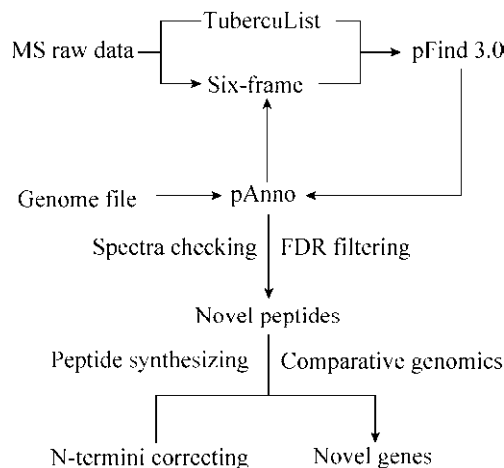


图 1 结核分枝杆菌 H37Rv 蛋白质基因组学分析流程图

Fig. 1 Workflow of *Mycobacterium tuberculosis* H37Rv proteogenomic analysis.

氨基酸的规则翻译, 使用终止子到终止子的翻译模式, 采用 MTB 特殊的 3 种起始密码子 ATG、GTG 和 TTG, 共翻译得到 141 851 个开放阅读框 (Opening reading frame, ORF)。与已注释的蛋白质序列数据库进行合并, 经过严格 FDR 筛选, 结合谱图人工核实、肽段合成验证、转录丰度、基因组多序列比对, 获得了一批质量可信的新肽段, 完成了 H37Rv 新编码基因的验证和原先注释错误基因的矫正。

Rv2742 基因是我们发现的 H37Rv 遗漏注释基因之一 (根据基因在染色体的坐标位置及结核基因命名法进行命名)。前期深度覆盖质谱数据共鉴定到两条肽段即“PNPWQYIR”和“VHNLDPPELVDEHAR”, 化学合成肽段与原先鉴定肽段谱图相似性分别为 0.94 和 0.96, 意味着 Rv2742 基因成功被转录、翻译。T 细胞表位预测 (<https://www.iedb.org/>) 发现, Rv2742 新蛋白中高可信的 T 细胞表位数有 19 个 (>0.5 分), 高可信 T 细胞表位覆盖到该蛋白 66.42% 的序列, 与已知表位库比较, 都是新型的表位, 该蛋白具有很好的免疫原性。此外, 经 NCBI-BLASTP 后, 数据库中没有比对到

相似序列,属于功能未知蛋白。为进一步探索 Rv2742 新基因功能及其潜在临床应用价值,有必要建立其表达、纯化体系。

本研究比较了 4 种不同的原核表达载体 (pGEX-4T-2、pET-32a、pET-28a 和 pMAL-c2X)、密码子利用偏性、5 种宿主 (JM109 (DE3)、BL21、BL21 (DE3)、BL21 (DE3) pLysS 和 Rosetta (DE3))、不同诱导温度 (16 °C 和 28 °C)、不同 IPTG 浓度 (0.5 mmol/L 和 1.0 mmol/L) 等条件下,目的蛋白的表达量。旨在最优实现结核分枝杆菌 H37Rv 新编码基因 Rv2742 克隆、可溶性诱导表达,从而能高效纯化 Rv2742 蛋白,为后续开展该蛋白的潜在相互作用及免疫原性研究奠定基础。后期实验采用的质谱方法是基于“自下而上”(Bottom-up) 鸟枪法的策略^[21],其基本流程是首先将蛋白质酶解为肽段,再用反相液相色谱进行分离,被洗脱的肽段用串联质谱(一级质谱和二级质谱)检测,最后利用搜库软件对样品中的肽段和蛋白质进行鉴定。

1 材料与amp;方法

1.1 材料

1.1.1 菌株及质粒

M. tuberculosis H37Rv 菌株 DNA 来自河北省石家庄市第五医院。大肠杆菌 DH5 α 、BL21 (DE3) 及 BL21 (DE3) pLysS 感受态细胞均购自北京康为世纪生物科技有限公司。大肠杆菌 BL21 及 Rosetta (DE3) 感受态细胞均购自北京全式金生物技术有限公司。JM109 (DE3) 感受态细胞购自北京鼎国昌盛生物技术有限责任公司。pGEX-4T-2 为本实验室保存质粒;pET-28a 质粒由国家蛋白质科学中心田春艳博士馈赠;pET-32a 质粒由德阳生物技术(固安)有限责任公司技术总监赵明治博士馈赠;pMAL-c2X 质粒由华大青兰生物科技(无锡)有限公司馈赠。引物合成和测

序均由北京擎科新业生物技术有限公司完成。

1.1.2 主要试剂

DNA marker (DL 1 000、 λ -Hind III)、pMD18-T 载体试剂盒均购自宝生物工程有限公 司(大连); 2 \times Taq PCR Master Mix、DNA 胶回收试剂盒、质粒 DNA 提取试剂盒均购自天根生化科技有限公 司(北京);限制性内切酶 (*Bam*H I、*Hind* III、*Eco*R I、*Not* I、*Xho* I、*Nde* I 及直链淀粉树脂均购自纽英伦生物技术公司(美国);注射用氨苄西林钠购自石药集团中诺药业有限公司;蛋白 marker (10 kDa) 购自富酶泰斯公司(加拿大);琼脂糖购自 Biowest 公司(法国);耐自消化高活性乙酰化胰蛋白酶来自北京酶知源生物科技有限公司^[22-23]。

1.1.3 主要仪器设备

PCR 仪、凝胶电泳电源、凝胶成像系统、DNA 电泳仪、电泳槽均为 BIO-RAD (美国) 公司产品;冷冻离心机、摇床、LTQ Orbitrap Velos 质谱仪均为 Thermo Fisher Scientific (美国) 公司产品;紫外分光光度计为岛津公司(日本)产品;纯水仪为 Merck Millipore (德国) 公司产品;电热恒温水浴锅为北京长风公司产品;电热恒温培养箱为上海一恒科学仪器有限公司产品;超声波细胞粉碎机为宁波新芝生物科技股份有限公司产品。

1.2 方法

1.2.1 重组表达质粒的构建

以 H37Rv DNA 为模板,引物如表 1 所示,分别扩增获得 Rv2742 新基因片段。PCR 反应体系均为: ddH₂O 9.5 μ L、2 \times Taq PCR Master Mix 12.5 μ L、DNA 1 μ L、上下游引物各 1 μ L。PCR 反应条件: 94 °C 预变性 3 min; 94 °C 变性 1 min, 60 °C 退火 30 s, 72 °C 延伸 1 min, 30 个循环; 72 °C 延伸 5 min。

PCR 产物利用 DNA 回收试剂盒回收,与 pMD18-T 克隆载体于 16 °C 过夜连接。连接体系

为 T4 DNA 连接酶缓冲液 1 μ L、T4 DNA 连接酶 1 μ L、ddH₂O 3 μ L、pMD18-T 克隆载体 0.5 μ L、目的片段回收产物 4.5 μ L。将连接产物转化至 DH5 α 感受态细胞。菌落 PCR 鉴定为阳性的质粒送至北京擎科新业生物技术有限公司进行测序。NCBI 在线比对为阳性的质粒命名为 pMD18T-Rv2742。

上述测序正确的质粒分别与其相应的载体 (pGEX-4T-2、pET-32a、pET-28a 或 pMAL-c2X) 分别利用 *Bam*H I /*Not* I 或 *Eco*R I /*Hind* III 进行双酶切。双酶切体系为 ddH₂O 24 μ L、质粒 20 μ L、NE 缓冲液 3.15 μ L、*Bam*H I (0.3 μ L)/*Not* I (0.5 μ L) 或 *Eco*R I /*Hind* III (各 0.3 μ L) 37 $^{\circ}$ C 双酶切 6 h, DNA 回收试剂盒分别回收目的基因及载体片段, 经 T4 DNA 连接酶 16 $^{\circ}$ C 过夜连接。连接体系为 T4 DNA 连接酶缓冲液 1 μ L, T4 DNA 连接酶 1 μ L, ddH₂O 3 μ L, 表达载体双酶切产物 1 μ L, 目的片段回收产物 4 μ L。将连接产物 10 μ L 转化至 DH5 α 感受态细胞, 菌落 PCR 为阳性的质粒命名为 pGEX-4T-2-Rv2742-1、pET-32a-Rv2742-1、pET-28a-Rv2742-1 及 pMAL-c2X-Rv2742-1。

1.2.2 Rv2742 基因密码子优化及优化产物重组表达质粒的构建

依据大肠杆菌密码子偏好性, 将 Rv2742 基因

序列进行密码子优化。对优化后的基因片段进行引物设计 (表 1), 操作步骤同前。将连接产物转化至 DH5 α 感受态细胞, 菌落 PCR 验证为阳性的质粒分别命名为 pGEX-4T-2-Rv2742-2、pET-32a-Rv2742-2、pET-28a-Rv2742-2、pMAL-c2X-Rv2742-2。

1.2.3 重组质粒原核表达及鉴定

重组质粒分别转化 BL21 (DE3) 宿主菌, 挑取单菌落于 5 mL LB (含 100 μ g/mL Amp/Kana) 液体培养基过夜培养。以起始 $OD_{600}=0.1$ 接种于 50 mL LB (含 100 μ g/mL Amp/Kana) 液体培养基中。37 $^{\circ}$ C 振荡培养至 OD_{600} 为 0.6–0.8 时, 加入终浓度为 1 mmol/L 的 IPTG, 28 $^{\circ}$ C、200 r/min 诱导 6–8 h。分别收取诱导前后 14 OD 菌体, 离心后用预冷的 PBS 洗涤两次, 加入等体积裂解液 (PBS+1 mmol/L PMSF+0.5% Triton X-100) 超声破碎, 分别取上清和沉淀进行 SDS-PAGE, 比较 IPTG 诱导前后目的蛋白的表达情况。

1.2.4 Rv2742 新蛋白诱导表达条件的优化

将重组质粒 pMAL-c2X-Rv2742-2 取 0.5 μ L 分别转化至表达宿主菌 JM109 (DE3)、BL21、BL21 (DE3)、BL21 (DE3) pLysS 及 Rosetta (DE3) 中, 28 $^{\circ}$ C、终浓度为 1 mmol/L 的 IPTG 诱导表达 6–8 h, 比较不同宿主菌对目的蛋白表达量的影响。基于

表 1 引物序列

Table 1 Primer sequence

Expression vector	Primer name	Primer sequence (5'-3')	Restriction site	Size (bp)
pGEX-4T-2-1	F-1	<u>GGATCC</u> ATGAGCGACAACGCAATCC	<i>Bam</i> H I	25
pET-32a-1	R-1	GCGGCCGCTTAGTCGTTCCGGTCCG	<i>Not</i> I	25
pET-28a-1	F-2	CGGAATTCATGAGCGACAACGCAAT	<i>Eco</i> R I	25
pMAL-c2X-1	R-2	AAGCTTTTAGTCGTTCCGGTCCGGC	<i>Hind</i> III	25
pGEX-4T-2-2	F-3	CCGGAATTCATGAGCGATAACGCGA	<i>Eco</i> R I	25
pET-32a-2	R-3	TTGCGGCCGCTCAATCGTTAGGATC	<i>Not</i> I	25
pET-28a-2	F-4	CATATGATGAGCGATAACGCGATTC	<i>Nde</i> I	25
	R-4	CTCGAGTCAATCGTTAGGATCAGGG	<i>Xho</i> I	25
pMAL-c2X-2	F-5	<u>GAATTC</u> ATGAGCGATAACGCGATTC	<i>Eco</i> R I	25
	R-5	AAGCTTTCATCGTTAGGATCAGGG	<i>Hind</i> III	25

.-1 represented the primer before codon optimization while -2 represented the primer after codon optimization. The single underlined sequences indicate restriction enzyme site.

表达量较高的宿主进行不同温度(16 °C 和 28 °C)、不同 IPTG 浓度(终浓度为 0.5 mmol/L 和 1.0 mmol/L)过夜诱导,比较诱导条件对目的蛋白表达量的影响。10% SDS-PAGE 检测目的蛋白表达情况。

1.2.5 Rv2742 融合蛋白的纯化及 LC-MS/MS 鉴定

将诱导后上清经直链淀粉树脂(Amylose resin)柱(100 μL)4 °C 孵育 1 h,用清洗液(20 mmol/L Tris-HCl (pH 7.4)、200 mmol/L NaCl、1 mmol/L EDTA、1 mmol/L 二硫苏糖醇、1 mmol/L 苯甲基磺酰氟、0.5% Triton X-100)500 μL 洗 3 次,用洗脱液(清洗液+10 mmol/L 麦芽糖)100 μL 洗脱 2 次,纯化得到目的融合蛋白,10% SDS-PAGE 检测。经考马斯亮蓝 R-250 染色并脱色后检测目的融合蛋白的纯化情况。

将纯化的目的条带切成 1 mm³ 胶粒,脱色、脱水、蒸干、经乙酰化胰蛋白酶(10 ng/μL)37 °C 过夜消化、抽提肽段、蒸干后置-80 °C 冻存。

蒸干后的肽段使用上样缓冲液(1% ACN+1% FA+98% ddH₂O)进行充分溶解。通过超高压液相色谱进行分离,其中分离柱采用内径 75 μm 长 15 cm 的 C₁₈ 反相色谱分析柱,内部 C₁₈ 填料内径 3 μm。洗脱组分经纳升级电喷雾离子源接口喷出进入 LTQ Orbitrap Velos 分析。毛细管离子传输温度为 250 °C,电喷雾电压为 1.8 kV。质谱采用一级质谱数据依赖的二级质谱扫描模式(Data dependent MS/MS scan)碰撞诱导裂解(Collision-induced dissociation, CID)模式碎裂一级离子。一级质谱扫描质核比范围为 300–1 600 (*m/z*),分辨率设置为 30 000;自动增益控制(Automatic gain control, AGC)为 10⁶。依次选取一级信号强度最高的 20 个离子进行二级碎裂分析,碰撞归一化能量(Normalized collision energy, NCE)为 35%;AGC 为 10⁴;最大离子注入时间为 30 ms;动态排除(Dynamic exclusion)为

40 s(40s 内不重复扫描已检测母离子)^[24]。

1.2.6 数据库搜索

pFind3 (<https://github.com/pFindStudio/pFind3/issues>)对质谱产生的数据文件(.raw)进行蛋白质数据库搜索。数据库由从 UniProt (Version: 201601) 下载目标 *E. coli* 完整蛋白质组序列、结核分枝杆菌 H37Rv 遗漏注释蛋白 Rv2742 序列、pMAL-c2X 载体标签 MBP 蛋白序列和常见污染构成。搜库参数设置如下: 1) 胰酶特异性酶切 (Trypsin KR_C); 2) 固定修饰为半胱氨酸的烷基化修饰 Carbamidomethyl[C] (+57.021 46 Da); 3) 可变修饰为甲硫氨酸氧化修饰 Oxidation[M] (+15.994 92 Da); 4) 母离子质量误差 20 ppm; 5) 子离子质量误差 0.5 Da; 6) 允许最大漏切位点数目为 2 个; 7) 7 AA ≤ 肽段长度 ≤ 100 AA; 8) 肽段最大修饰为 2 种。搜库结果采用目标-诱饵库策略进行过滤,并设定肽段和蛋白质鉴定假阳性率(False discovery rate, FDR)均小于 1%^[25]。

1.2.7 IPTG 诱导表达 Rv2742 蛋白生长曲线的测定

将重组质粒 pMAL-c2X-Rv2742-2 取 0.5 μL 转化至 BL21 (DE3),挑取单菌落扩大培养于 5 mL LB (含 100 μg/mL Amp) 液体培养基过夜培养。采用比浊法,分光光度计波长为 600 nm 检测。用 LB 液体培养基调零,以起始 OD₆₀₀ 为 0.1 接种于 5 mL LB (含 100 μg/mL Amp) 液体培养基中,37 °C、200 r/min 振荡培养。待菌体生长至对数期,实验组加入终浓度为 1 mmol/L 的 IPTG 诱导,设置 3 组生物学重复,每隔 1.5 h 检测 OD 值。

2 结果与分析

2.1 重组表达质粒的构建

重组表达质粒构建方法类似,以 pMAL-c2X-Rv2742-2 为例。以密码子优化后的 Rv2742 序列为模板扩增,获得 Rv2742 基因片段

约 420 bp, 与预期的目的片段大小一致(图 2A), 表明成功获得目的基因 *Rv2742*。纯化的 *Rv2742* 目的片段与 pMD 18-T 连接转化后进行菌落 PCR 验证, 在目的条带处成功获得特异性条带 (图 2B), 且测序序列正确。*EcoR* I 和 *Hind* III 双酶切 pMD 18T-*Rv2742* 质粒 6 h, 酶切片段约为 420 bp, 与预期大小相符 (图 2C)。*EcoR* I 和 *Hind* III 双酶切质粒 pMAL-c2X, 酶切片段约为 6 478 bp (图 2D), 与预期相符。将双酶切后回收的目的片段与载体片段连接转化进行菌落 PCR 验证, 在目的条带处获得特异性条带 (图 2E), 说明成功构建重组表达质粒 pMAL-c2X-*Rv2742*-2。

2.2 四种表达载体均未实现 *Rv2742* 目的蛋白的诱导表达

经 Glycine-SDS-PAGE 或 Tricine-SDS-PAGE^[26] 电泳结果分析, 如图 3A 所示, 发现在 28 °C 1 mmol/L IPTG 诱导 6 h 后, pGEX-4T-2 系统表达的 *Rv2742* 目的融合蛋白 (分子量为 40.85 kDa) 主要分布在沉淀中, 但在诱导后上清中并没有明显的新增条带, 说明以 pGEX-4T-2 系统表达的 *Rv2742* 目的融合蛋白主要以包涵体形式存在。如

图 3B、3C、3D 所示, 分别以 pET-32a、pET-28a、pMAL-c2X 系统表达的 *Rv2742* 目的融合蛋白 (分子量分别为 28.16、18.04、56 kDa) 在诱导后上清与沉淀中均未出现明显的新增特异性条带, 说明以 pET-32a、pET-28a 以及 pMAL-c2X 表达的 *Rv2742* 目的融合蛋白均未能成功实现诱导表达。综上所述, 以 H37Rv 基因组 DNA 克隆的 *Rv2742* 基因序列为模板所构建的 pGEX-4T-2-*Rv2742*-1、pET-32a-*Rv2742*-1、pET-28a-*Rv2742*-1、pMAL-c2X-*Rv2742*-1 原核表达系统均未能实现 *Rv2742* 目的蛋白可溶性诱导表达。

2.3 *Rv2742* 基因序列密码子优化前后比较

参考大肠杆菌的密码子偏好性, 对 *Rv2742* 序列进行密码子优化, 经 DNAMAN 序列比对密码子优化前后序列, 结果如 4A 所示。参照 GenScript (<http://www.genscript.com/>) 中 *E. coli* 同义密码子的相对频率将密码子分为 3 类: 高频密码子、常用密码子及稀有密码子^[27], 并对优化前后的密码子进行统计分类, 结果如 4B 所示, 表明优化前的 *Rv2742* 含 8.15% 的稀有密码子且高频密码子所占比例小 (28.15%)。对序列的密码子适

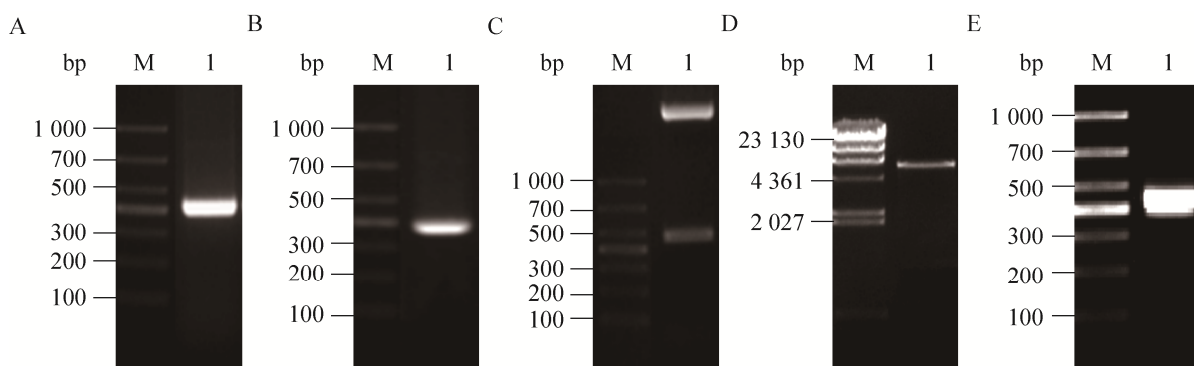


图 2 重组表达质粒 pMAL-c2X-*Rv2742*-2 构建及鉴定

Fig. 2 Construction and identification of recombinant plasmid pMAL-c2X-*Rv2742*-2. (A) PCR amplification of *Rv2742* gene. Lane M: DL 1 000 DNA marker; lane 1: PCR product of *Rv2742* gene. (B) Detection of positive bacteria pMD18T-*Rv2742* by PCR. Lane M: DL 1 000 DNA marker; lane 1: PCR product of pMD18T-*Rv2742*. (C) Plasmid pMD18T-*Rv2742* digested by *EcoR* I /*Hind* III. Lane M: DL 1 000 DNA marker; lane 1: plasmid pMD18T-*Rv2742* digested by *EcoR* I /*Hind* III. (D) Plasmid pMAL-c2X digested by *EcoR* I /*Hind* III. Lane M: λ -*Hind* III; lane 1: plasmid pMAL-c2X digested by *EcoR* I /*Hind* III. (E) Construction of recombinant plasmid pMAL-c2X-*Rv2742*-2. lane M: DL 1 000 DNA marker; lane 1: Detection of positive bacteria recombinant plasmid pMAL-c2X-*Rv2742*-2 by PCR.

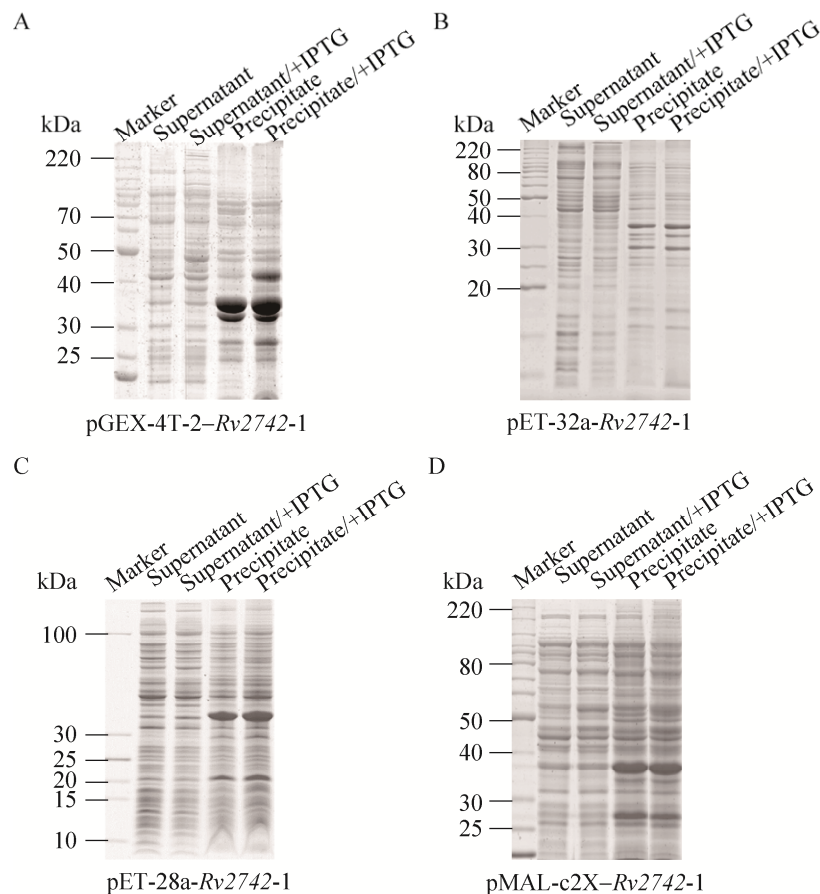


图 3 分别利用 pGEX-4T-2、pET-32a、pET-28a、pMAL-c2X 四种表达载体表达 Rv2742 重组蛋白的 Glycine-SDS-PAGE/Tricine-SDS-PAGE 图谱 (Rv2742 密码子优化前)

Fig. 3 Glycine-SDS-PAGE/Tricine-SDS-PAGE analysis of expression of Rv2742 recombinant protein by pGEX-4T-2, pET-32a, pET-28a and pMAL-c2X expression vectors before Rv2742 codon optimization. (A) Glycine-SDS-PAGE analysis of expression of Rv2742 recombinant protein by pGEX-4T-2 expression vectors before Rv2742 codon optimization. (B) Glycine-SDS-PAGE analysis of expression of Rv2742 recombinant protein by pET-32a expression vectors before Rv2742 codon optimization. (C) Tricine-SDS-PAGE analysis of expression of Rv2742 recombinant protein by pET-28a expression vectors before Rv2742 codon optimization. (D) Glycine-SDS-PAGE analysis of expression of Rv2742 recombinant protein by pMAL-c2X expression vectors before Rv2742 codon optimization.

应指数 (Codon adaptation index, CAI) 计算, 发现新编码基因 Rv2742 的 CAI 从 0.73 升高到 0.92, 如图 4C 所示, 结果显示 Rv2742 经密码子优化后更适于大肠杆菌的原核表达, 这为 Rv2742 的原核表达提供有利的理论依据。统计结果发现优化前 Rv2742 的密码子适应度普遍偏低, 且最适于 *E. coli* 表达的密码子占 51% (91–100), 而不适于 *E. coli* 表达的密码子占 9% (0–30), 结果如图 4D 所示, 优化后减少了 Rv2742 的稀有密码子和适应度低的密码子, 不适于 *E. coli* 表达的密码子仅占 2%

(0–30), 同时增加了高频密码子的使用, 使得最适密码子增加到 86% (91–100), 有效地提高了其转录和翻译的效率。

2.4 Rv2742 新基因密码子优化后实现了目的蛋白的诱导表达

以 Rv2742 密码子优化后的序列为模板, 将构建成功的重组质粒 pGEX-4T-2-Rv2742-2、pET-32a-Rv2742-2、pET-28a-Rv2742-2 分别转化至大肠杆菌 BL21 (DE3) 中, 28 °C、终浓度为 1 mmol/L IPTG 诱导 6 h 后, 经 Glycine-SDS-PAGE 或 Tricine-

SDS-PAGE 分析, 如图 5A、5B、5C 所示, IPTG 诱导后的上清和沉淀中均未出现明显新增条带, 表明密码子优化后, pGEX-4T-2、pET-32a、pET-28a

原核表达系统均未能实现 Rv2742 重组蛋白的诱导表达 (分子量分别为 40.85 kDa、28.16 kDa、18.04 kDa)。而将含重组质粒 pMAL-c2X-Rv2742-2

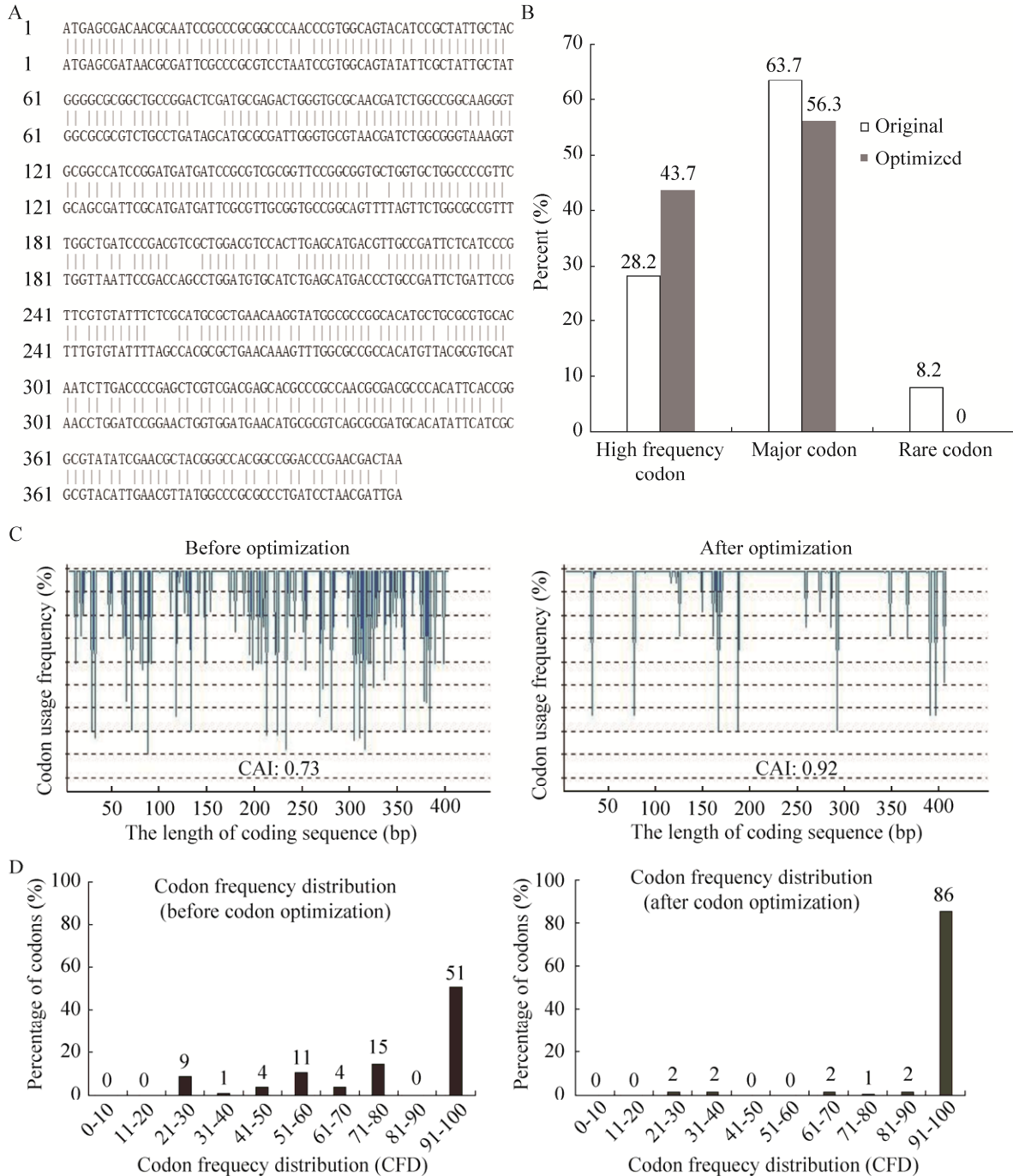


图 4 Rv2742 基因序列密码子优化前后差异的比较

Fig. 4 Comparison of differences from Rv2742 gene sequence before and after codon optimization. (A) Blast analysis of the differences from Rv2742 gene sequence before and after codon optimization. (B) Comparison of the proportion of high frequency codon, major codon and rare codon before and after codon optimization. (C) Comparison of the CAI value before and after codon optimization. (D) Comparison of codon frequency distribution before and after codon optimization.

的大肠杆菌 BL21 (DE3) 在 28 °C、终浓度为 1 mmol/L IPTG 诱导 6 h。10% Glycine-SDS-PAGE 显示, 与未加 IPTG 诱导前相比, 诱导后上清和沉淀均在 56 kDa 出现特异性新增条带, 其大小与目的蛋白预期相符。初步说明目的片段在大肠杆菌 pMAL-c2X 表达系统中得到了表达, 且表达产物能以可溶性蛋白形式出现, 但表达量不高(图 5D)。

2.5 宿主菌 Rosetta (DE3) 实现了目的蛋白的高表达及诱导条件的优化

将重组质粒 pMAL-c2X-Rv2742-2 分别转化

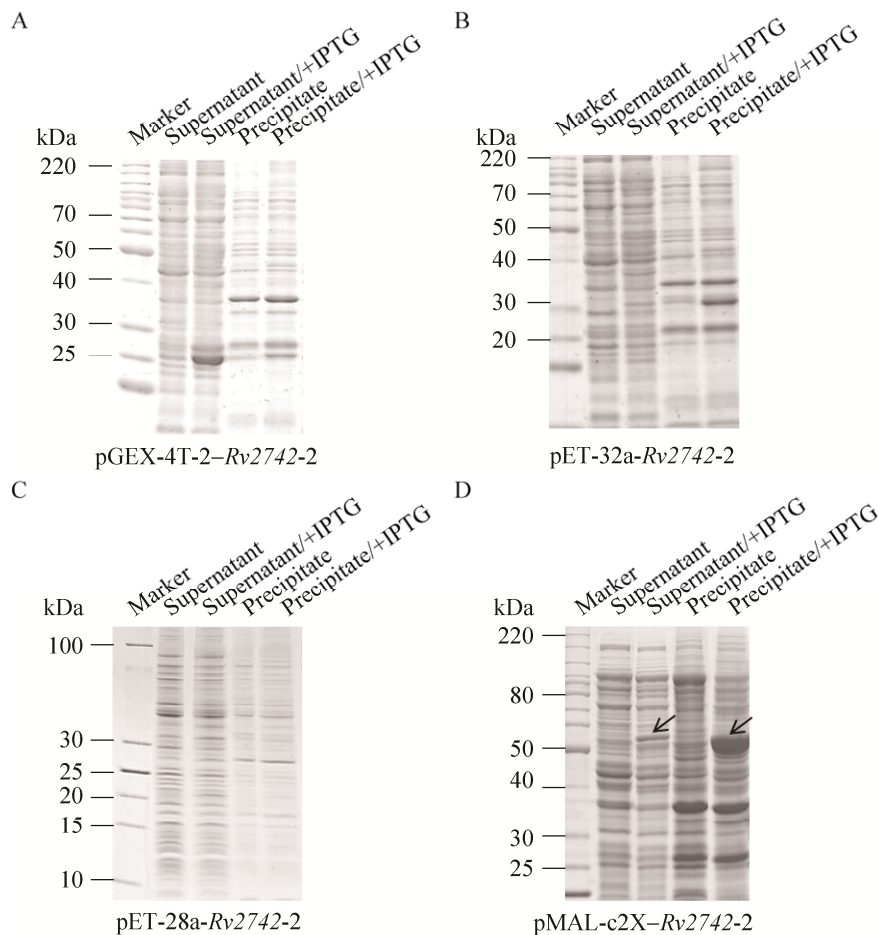


图 5 分别利用 pGEX-4T-2、pET-32a、pET-28a、pMAL-c2X 四种表达载体表达 Rv2742 重组蛋白的 SDS-PAGE/Tricine-SDS-PAGE 图谱 (*Rv2742* 密码子优化后)

Fig. 5 SDS-PAGE/Tricine-SDS-PAGE analysis of expression of Rv2742 recombinant protein by pGEX-4T-2, pET-32a, pET-28a, pMAL-c2X expression vectors after *Rv2742* codon optimization. (A) SDS-PAGE analysis of expression of Rv2742 recombinant protein by pGEX-4T-2 expression vectors after *Rv2742* codon optimization. (B) SDS-PAGE analysis of expression of Rv2742 recombinant protein by pET-32a expression vectors after *Rv2742* codon optimization. (C) Tricine-SDS-PAGE analysis of expression of Rv2742 recombinant protein by pET-28a expression vectors before *Rv2742* codon optimization. (D) SDS-PAGE analysis of expression of Rv2742 recombinant protein by pMAL-c2X expression vectors after *Rv2742* codon optimization.

白在 16 °C 低温诱导时表达量更高 (图 6B)。在 Rosetta (DE3) 宿主菌及 16 °C 过夜诱导条件下, 我们比较了终浓度为 1 mmol/L 和 0.5 mmol/L IPTG 对目的蛋白表达量的影响, 结果表明目的蛋白在终浓度为 0.5 mmol/L IPTG 诱导条件下表达量最高 (图 6B)。综上所述, 在宿主菌 Rosetta (DE3)、16 °C 及终浓度为 0.5 mmol/L IPTG 过夜诱导条件下能更有效地实现 Rv2742 蛋白的融合表达。

2.6 Rv2742 目的融合蛋白的纯化及 LC-MS/MS 鉴定

在 Rosetta (DE3) 宿主菌中, 16 °C、终浓度为 0.5 mmol/L IPTG 过夜诱导表达的上清超声裂解后亲和纯化 (图 7A), 在 56 kDa 处获得纯度较高的 Rv2742 目的融合蛋白。经 LC-MS/MS 鉴定, pFind 3 搜库鉴定到 4 条 Rv2742 目的蛋白高可信肽段, 分别为 MSDNAIRP、PNPWQYIR、YGPRDPND、VHNLDPPELVDEHAR (图 7C)。计

算得到 Rv2742 目的蛋白鉴定肽段序列覆盖度达 29.63% (图 7B), 而 Rv2742 融合蛋白鉴定肽段序列覆盖度高达 61.46%。说明经过宿主菌、低温、低浓度 IPTG 诱导条件优化后, 纯化得到含量更高的 Rv2742 融合蛋白, 为后续进一步的功能研究提供必要的条件。

2.7 IPTG 诱导表达 Rv2742 蛋白生长曲线的测定

将重组质粒 pMAL-c2X-Rv2742-2 取 0.5 μL 转化至大肠杆菌 Rosetta (DE3) 中。如图 8 所示, 未加 IPTG 诱导组 (-IPTG) 在整个时间段内保持正常生长, 经过对数期后达到平台期。而与对照组 (-IPTG) 相比, 加入 IPTG 诱导 (+IPTG) 组的生长情况明显受到抑制, 诱导后 0–1.5 h 范围内 OD 值只有少量增加且诱导 3 h 后菌体 OD 值基本维持稳定。由于 IPTG 是一种极强的诱导剂, 不被细菌代谢因而十分稳定, 在胞内诱导外源蛋白

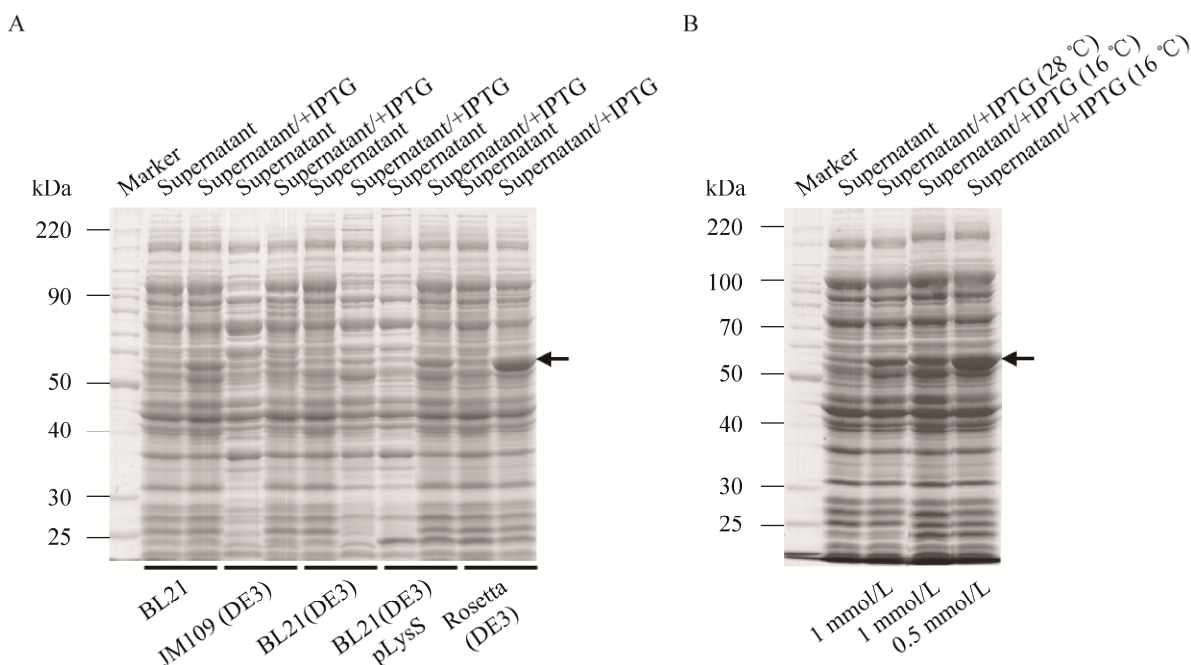


图 6 Rv2742 融合蛋白在五中不同宿主菌中诱导表达及在 Rosetta (DE3) 宿主菌中诱导条件的优化

Fig. 6 Comparison of Rv2742 expression based on five different bacterial hosts and the optimization of expression conditions in the host Rosetta (DE3). (A) SDS-PAGE analysis of Rv2742 expression by different hosts including JM109 (DE3), BL21, BL21 (DE3), BL21(DE3)pLysS and Rosetta (DE3). (B) The optimization of Rv2742 expression conditions for Rosetta (DE3) including culture temperatures and IPTG concentrations.

的表达,会对细胞产生一定的毒性。其次,加入 IPTG 后,大量外源蛋白的表达造成菌体原有代谢活动的失衡和重新分配。对于菌体内源蛋白来说,

是一种对游离氨基酸以及核糖体的竞争和掠夺,这必然会影响内源蛋白的合成速度,导致菌体的生长受到明显抑制,生长速度降低。

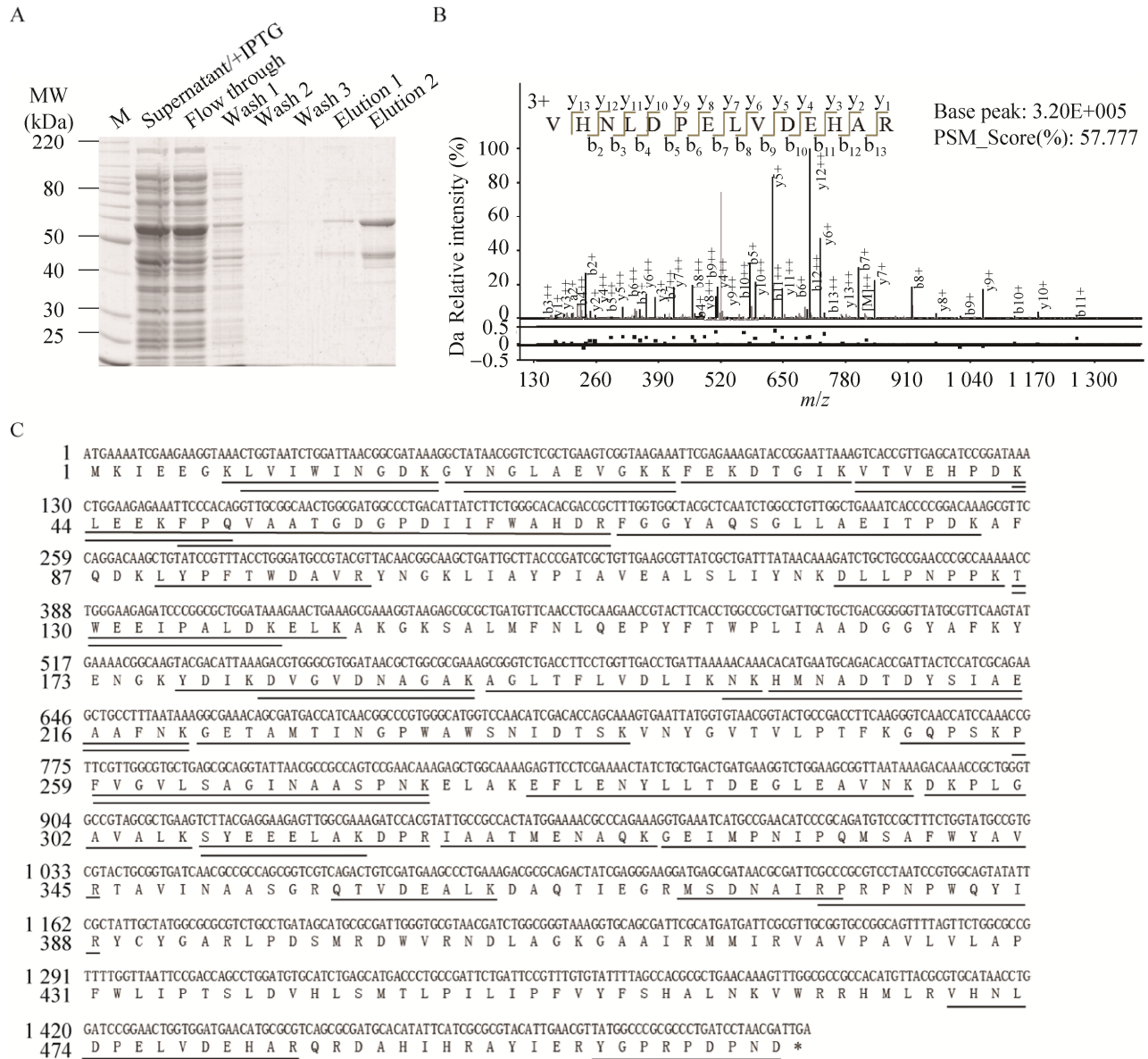


图7 Rv2742 目的融合蛋白纯化及质谱检测

Fig. 7 Purification and LC-MS/MS identification of Rv2742 fusion protein. (A) Purification of Rv2742 fusion protein. Lane M: protein marker; lane supernatant/+IPTG: supernatant of lysate of BL21(DE3)-pMAL-c2X-Rv2742-2 induced by 1 mmol/L IPTG; lane Flow through: supernatant after induction flow through the amylose resin column; lane Wash1, 2, 3: supernatant after induction washed by Wash Buffer; lane Elution1, 2: supernatant after induction eluted by Elution Buffer. (B) Sequence coverage of Rv2742 fusion protein (61.46%) by LC-MS/MS identification. (The single underlined sequences indicated the peptides were identified by LC-MS/MS). (C) The spectrum of Rv2742 target peptide "VHNLDPVLVDEHAR". (D) SDS-PAGE analysis of purification of Rv2742 fusion protein in the host bacteria Rosetta (DE3).

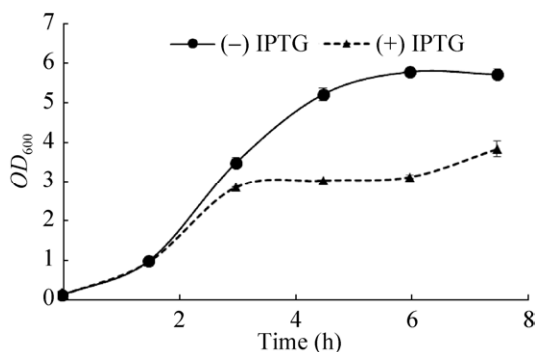


图8 IPTG 诱导表达 Rv2742 蛋白生长曲线的测定
Fig. 8 Determination of growth curve of expression of Rv2742 fusion protein induced by 0.5 mmol/L IPTG.

3 讨论

大肠杆菌是一种基因组 G+C 含量低 (<50%) 的革兰氏阴性菌, 而结核分枝杆菌 (*M. tuberculosis*) 是一种基因组 G+C 含量高 (>65%) 的革兰氏阳性放线菌^[28], 放线菌基因组的典型特征就是基因组 G+C 含量高。结核分枝杆菌 G₃ 和 C₃ 结尾末端的密码子具有很强的偏倚性, 第 3 个位点的密码子 (G+C) 含量高达 83%^[29]。密码子偏倚性影响翻译过程中氨基酸插入的准确性、多肽折叠和 mRNA 序列的稳定性等几方面^[30]。结核分枝杆菌目的基因在大肠杆菌中不易表达可能主要是由于 G+C 含量高和独特密码子的使用^[31]。生物体中基因所使用的密码子和 tRNA 的丰度有着强的正相关性。依照物种的偏好性对 mRNA 的序列进行优化, 将目的基因的密码子替换成宿主细胞常用的密码子, 能够增加 tRNA 的结合效率从而提高蛋白质表达水平。

本次实验研究的结核新基因 Rv2742 编码蛋白分子量本身不大, pET-32a 和 pET-28a 虽然同属 pET 载体, 标签大小对小蛋白的表达有一定程度的影响, 所以在实验过程中考虑到使用不同种类和大小标签的 pET 载体。但仅在密码子优化后, 利用 pMAL-c2X 原核表达系统才在上清和沉淀中

都实现了表达, 沉淀中表达含量要比可溶性部分高。后续实验需要进一步考虑对以包涵体形式表达的蛋白进行变性和复性。诱导后上清虽然实现了可溶性诱导表达, 但表达含量较低, 可从温度、IPTG 浓度、抗生素浓度、诱导时间、宿主菌等多方面进行优化, 从而提高其表达量。本次实验从宿主菌、温度、IPTG 浓度等方面进行优化, 在 Rosetta (DE3) 宿主菌中的表达量尤为显著。低温及低浓度诱导剂的条件有利于蛋白的可溶性表达。Rosetta (DE3) 是携带氯霉素抗性质粒 BL21 的衍生菌, 补充大肠杆菌缺乏的 6 种稀有密码子 (AUA、AGG、AGA、CUA、CCC、GGA) 对应的 tRNA, 提高外源基因在该系统中的表达水平, 为其他结核遗漏注释基因的表达、纯化提供了较好的宿主选择性。

在进行原核表达时, 目前多以融合蛋白形式表达, 具有融合标签的表达载体可促进蛋白质可溶性表达和正确折叠恢复成天然结构。融合蛋白标签的选择很重要, 源于其可能影响天然蛋白质的相互作用、翻译后修饰、重组蛋白的溶解度和细胞定位等^[32], 其应用取决于对特异性、溶解度、结合和洗脱条件等的要求^[33]。添加非肽融合配体具有作为溶解度增强剂作用的优点^[34]。最常用的融合标签有: 麦芽糖结合蛋白 (MBP)、转录终止/抗终止蛋白 NusA)、硫氧还蛋白 (Trx)、谷胱甘肽巯基转移酶 (GST)、ubiquitin、SUMO 等^[35], 但这些融合伴侣作为溶解度增强剂起作用的原因尚不清楚。在 MBP 标签存在情况下, 其具有内在的分子伴侣活性促进溶解影响其融合伴侣的折叠^[36]。谷胱甘肽巯基转移酶 (GST) 可以结合到谷胱甘肽树脂上进行纯化, 可以保护目的蛋白免受蛋白酶降解, 增加蛋白的稳定性。但 GST 是一个很弱的溶解度增强剂^[37], 易产生包涵体。而 NusA、MBP 和 Trx 显示出较好的溶解度增强特性, 但它们的大分子量可能对蛋白质溶解度造成一定的影

响^[38]。pET-32a 含有 *trxA* 基因并携带有用于蛋白生产和纯化的双融合伴侣。TrxA 是一种存在于大肠杆菌胞质中高度可溶性表达的细胞内热稳定性蛋白,能显著增加重组蛋白的可溶性,减少包涵体形成。但 TrxA 没有内在的亲亲和特性,因此在蛋白质纯化时需要附加额外的融合标签,如 His6 标签^[39]。His 标签在天然和变性的纯化条件下被用来帮助溶解和折叠^[40]。

启动子在控制相关基因的转录起始中起重要作用。本实验研究中运用了 4 种原核表达系统均能被 IPTG 诱导,表达载体 pGEX-4T-2 和 pMAL-c2X 均带有强大的 *tac* 启动子。*tac* 启动子由 *trp* 启动子-35 区和 *lac UV5* 启动子-10 区融合而成的杂合启动子,受 *lac* 阻抑物调控^[41]。在 pET 系列表达载体中,外源基因在表达时受 T7 噬菌体 RNA 聚合酶调控,编码序列在多克隆位点插入,置于天然 T7 RNA 聚合酶启动子 ($\phi 10$ 启动子) 或所谓的“T7 *lac* 启动子”的控制之下,后者是带有 *lac* 操纵子 (*lacO*) 序列的天然 T7 RNA 聚合酶启动子的衍生体,*lac* 阻抑物的结合能阻断转录起始^[42]。

对于特异性抗原的筛选,一直是结核分枝杆菌研究中的重点。基于串联质谱的蛋白质组学被应用于注释基因的鉴定、新基因的发现、个人基因组学和疾病相关研究^[43]。本实验通过对结核分枝杆菌 H37Rv 新基因 *Rv2742* 进行克隆、诱导表达和纯化,纯化后的目的蛋白有可能直接作为抗原或制备抗体建立免疫学检测方法,为特异性诊断试剂盒和新型疫苗的研制奠定基础。H37Rv 新基因 *Rv2742* 在结核三大数据库 (UniProt、NCBI、TubercuList) 中均没有记录,其功能目前尚未可知,但其表达丰度较高,只是基于成熟的预测模型没有预测出来。这种高表达漏注释基因产物可能在结核分枝杆菌中发挥了某些特殊的生物学功能,值得进一步探索。

目的基因密码子优化、可溶性融合标签筛选

及宿主菌、温度、IPTG 浓度等表达条件的优化是实现蛋白在大肠杆菌中可溶性诱导表达的有效策略^[44]。

REFERENCES

- [1] Rijo P, Simões MF, Francisco AP, et al. Antimycobacterial metabolites from *Plectranthus royleanone* derivatives against *Mycobacterium tuberculosis* strains. *Chem Biodivers*, 2010, 7(4): 922–932.
- [2] Eddabra P, Benhassou HA. Rapid molecular assays for detection of tuberculosis. *Pneumonia*, 2018, 10(1): 4.
- [3] World Health Organization. Global tuberculosis report 2018. Geneva: World Health Organization, 2018.
- [4] Gupta S, Kakkar V. Recent technological advancements in tuberculosis diagnostics - a review. *Biosens Bioelectron*, 2018, 115: 14–29.
- [5] Singhania A, Wilkinson WJ, Rodrigue M, et al. The value of transcriptomics in advancing knowledge of the immune response and diagnosis in tuberculosis. *Nat Immunol*, 2018, 19(11): 1159–1168.
- [6] Cohn DL, O'Brien RJ, Geiter LJ, et al. Targeted tuberculin testing and treatment of latent *tuberculosis* infection. *MMWR Morb Mortal Wkly Rep*, 2000, 49(6): 1–54.
- [7] Bosshard V, Roux-Lombard P, Perneger T, et al. Do results of the T-SPOT.TB interferon- γ release assay change after treatment of tuberculosis? *Respir Med*, 2009, 103(1): 30–34.
- [8] Cole ST, Barrell BG. Analysis of the genome of *Mycobacterium tuberculosis* H37Rv. *Novartis Found Symp*, 1998, 217: 160–172.
- [9] Cole ST, Brosch R, Parkhill J, et al. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature*, 1998, 393(6685): 537–544.
- [10] Camus JC, Pryor MJ, Médigue C, et al. Re-annotation of the genome sequence of *Mycobacterium tuberculosis* H37Rv. *Microbiology*,

- 2002, 148(10): 2967–2973.
- [11] Brent MR. Steady progress and recent breakthroughs in the accuracy of automated genome annotation. *Nat Rev Genet*, 2008, 9(1): 62–73.
- [12] Goymer P. Genomics: annotating with proteomes. *Nat Rev Genet*, 2008, 9(6): 418.
- [13] de Souza GA, Målen H, Søfteland T, et al. High accuracy mass spectrometry analysis as a tool to verify and improve gene annotation using *Mycobacterium tuberculosis* as an example. *BMC Genomics*, 2008, 9: 316.
- [14] Rison SCG, Mattow J, Jungblut PR, et al. Experimental determination of translational starts using peptide mass mapping and tandem mass spectrometry within the proteome of *Mycobacterium tuberculosis*. *Microbiology*, 2007, 153(2): 521–528.
- [15] Schubert OT, Mouritsen J, Ludwig C, et al. The Mtb proteome library: a resource of assays to quantify the complete proteome of *Mycobacterium tuberculosis*. *Cell Host Microbe*, 2013, 13(5): 602–612.
- [16] Castellana N, Bafna V. Proteogenomics to discover the full coding content of genomes: a computational perspective. *J Proteomics*, 2010, 73(11): 2124–2135.
- [17] Zhang CP, Xu P, Zhu YP. Progress in proteogenomics of prokaryotes. *Chin J Biotech*, 2014, 30(7): 1026–1035 (in Chinese).
张成普, 徐平, 朱云平. 原核生物蛋白质基因组学研究进展. *生物工程学报*, 2014, 30(7): 1026–1035.
- [18] Aivaliotis M, Gevaert K, Falb M, et al. Large-scale identification of N-terminal peptides in the halophilic archaea *Halobacterium salinarum* and *Natronomonas pharaonis*. *J Proteome Res*, 2007, 6(6): 2195–2204.
- [19] Renuse S, Chaerkady R, Pandey A. Proteogenomics. *Proteomics*, 2011, 11(4): 620–630.
- [20] Kelkar DS, Kumar D, Kumar P, et al. Proteogenomic analysis of *Mycobacterium tuberculosis* by high resolution mass spectrometry. *Mol Cell Proteomics*, 2011, 10(12): M111.011627.
- [21] Domon B, Aebersold R. Mass spectrometry and protein analysis. *Science*, 2006, 312(5771): 212–217.
- [22] Wu FL, Zhao MZ, Zhang Y, et al. Recombinant acetylated trypsin demonstrates superior stability and higher activity than commercial products in quantitative proteomics studies. *Rapid Commun Mass Spectrom*, 2016, 30(8): 1059–1066.
- [23] Zhao MZ, Wu FL, Xu P. Development of a rapid high-efficiency scalable process for acetylated *Sus scrofa* cationic trypsin production from *Escherichia coli* inclusion bodies. *Protein Expr Purif*, 2015, 116: 120–126.
- [24] He CT, Jia CX, Zhang Y, et al. Enrichment-based proteogenomics identifies microproteins, missing proteins, and novel smORFs in *Saccharomyces cerevisiae*. *J Proteome Res*, 2018, 17(7): 2335–2344.
- [25] Chi H, Liu C, Yang H, et al. Comprehensive identification of peptides in tandem mass spectra using an efficient open search engine. *Nat Biotechnol*, 2018, 36(11): 1059–1061.
- [26] Schägger S. Tricine-SDS-PAGE. *Nat Protoc*, 2006, 1(1): 16–22.
- [27] Luo WJ, Deng C, Li YC, et al. Soluble expression of Tandem Hybrid Ubiquitin-binding Domains (ThUBD) in prokaryotic cytoplasm of *Escherichia coli* BL21(DE3). *Mil Med Sci*, 2016, 40(10): 795–800 (in Chinese).
罗维佳, 邓晨, 李衍常, 等. 串联杂种泛素结合结构域蛋白(ThUBD)在大肠杆菌中的可溶性高效表达. *军事医学*, 2016, 40(10): 795–800.
- [28] Piubelli L, Campa M, Temporini C, et al. Optimizing *Escherichia coli* as a protein expression platform to produce *Mycobacterium tuberculosis* immunogenic proteins. *Microb Cell Fact*, 2013, 12: 115.
- [29] Andersson SGE, Sharp PM. Codon usage in the *Mycobacterium tuberculosis* complex. *Microbiology*, 1996, 142(4): 915–925.
- [30] Ayyar BV, Arora S, Ravi SS. Optimizing antibody expression: the nuts and bolts. *Methods*, 2017, 116: 51–62.
- [31] Kaur J, Kumar A, Kaur J. Strategies for optimization of heterologous protein expression in *E. coli*: roadblocks and reinforcements. *Int J Biol Macromol*, 2018, 106: 803–822.
- [32] Papanephytou CP, Kontopidis G. Statistical approaches to maximize recombinant protein

- expression in *Escherichia coli*: a general review. *Protein Expr Purif*, 2014, 94: 22–32.
- [33] Zhao XY, Li GS, Liang SF. Several affinity tags commonly used in chromatographic purification. *J Anal Methods Chem*, 2013, 2013: 581093.
- [34] Hammarström M, Hellgren N, van den Berg S, et al. Rapid screening for improved solubility of small human proteins produced as fusion proteins in *Escherichia coli*. *Protein Sci*, 2002, 11(2): 313–321.
- [35] Rosano GL, Ceccarelli EA. Recombinant protein expression in *Escherichia coli*: advances and challenges. *Front Microbiol*, 2014, 5: 172.
- [36] Kapust RB, Waugh DS. *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Sci*, 1999, 8(8): 1668–1674.
- [37] Hammarström M, Woestenenk EA, Hellgren N, et al. Effect of N-terminal solubility enhancing fusion proteins on yield of purified target protein. *J Struct Funct Genomics*, 2006, 7(1): 1–14.
- [38] Costa SJ, Almeida A, Castro A, et al. The novel Fh8 and H fusion partners for soluble protein expression in *Escherichia coli*: a comparison with the traditional gene fusion technology. *Applied microbiology and biotechnology*, 2013, 97(15): 6779–6791.
- [39] Kosobokova EN, Skrypnik KA, Kosorukov VS. Overview of fusion tags for recombinant proteins. *Biochemistry (Moscow)*, 2016, 81(3): 187–200.
- [40] Arnau J, Lauritzen C, Petersen GE, et al. Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins. *Protein Expr Purif*, 2006, 48(1): 1–13.
- [41] Amann E, Brosius J, Ptashne M. Vectors bearing a hybrid *trp-lac* promoter useful for regulated expression of cloned genes in *Escherichia coli*. *Gene*, 1983, 25(2/3): 167–178.
- [42] Sambrook J, Russell DW. *Molecular Cloning: A Laboratory Manual*. 3rd ed. New York: Cold Spring Harbor Laboratory Press, 2001: 1218–1219.
- [43] Zhang K, Fu Y, Zeng WF, et al. A note on the false discovery rate of novel peptides in proteogenomics. *Bioinformatics*, 2015, 31(20): 3249–3253.
- [44] Itkonen JM, Urtti A, Bird LE, et al. Codon optimization and factorial screening for enhanced soluble expression of human ciliary neurotrophic factor in *Escherichia coli*. *BMC Biotechnol*, 2014, 14: 92.

(本文责编 郝丽芳)