

# SUMO 基因的内在原核启动子活性及其在大肠杆菌蛋白表达系统中的应用

亓燕红<sup>1</sup>, 邹竹荣<sup>2</sup>, 邹华英<sup>3</sup>, 范云六<sup>4</sup>, 张春义<sup>4</sup>

1 山东农业大学农学院, 泰安 271000

2 云南师范大学生命科学学院, 昆明 650092

3 云南省农业科学院 院办档案科, 昆明 650231

4 中国农业科学院生物技术研究所, 北京 100081

**摘要:** SUMO 融合系统已成为目前大肠杆菌重组蛋白生产的重要手段,但在载体构建效率和蛋白可溶性等方面仍有待改进。本研究在 PCR 克隆酿酒酵母 SUMO 基因 *Smt3(Sm)* 时意外发现 *Sm* 具有组成型原核启动子活性;而且经软莓 BPROM 程序预测发现大多数物种 SUMO 基因编码区都具有依赖 $\sigma 70$ 的原核启动子。进一步通过整合 *Sm* 启动子和 *Sm* 3' 末端 *Stu* I 位点特性以及引入 His 标签和超酸增溶标签,构建了基于 *Sm'-LacZ $\alpha$*  融合基因的一系列通用克隆表达载体,并通过蓝白斑筛选和 SDS-PAGE 分析进行了多个靶蛋白基因的克隆和表达,结果表明实现了基因表达载体的快速构建、蛋白高可溶性表达和关联蛋白的共表达等目标。因此,集成诸多特性而改良的 SUMO 融合技术可以成为大肠杆菌蛋白表达系统的有力工具,建立的共表达载体系统还可以用作研究细胞内蛋白间的相互作用。

**关键词:** SUMO 基因, 原核启动子, SUMO 融合系统, 共表达, 大肠杆菌

## Intrinsic prokaryotic promoter activity of SUMO gene and its applications in the protein expression system of *Escherichia coli*

Yanhong Qi<sup>1</sup>, Zhurong Zou<sup>2</sup>, Huaying Zou<sup>3</sup>, Yunliu Fan<sup>4</sup>, and Chunyi Zhang<sup>4</sup>

1 College of Agronomy, Shandong Agricultural University, Tai'an 271000, China

2 School of Life Sciences, Yunnan Normal University, Kunming 650092, China

3 Archives of Administrative Agency, Yunnan Academy of Agricultural Sciences, Kunming 650231, China

4 Biotechnology Research Institute, Chinese Academy of Agricultural Sciences, Beijing 100081, China

**Received:** September 3, 2010; **Accepted:** November 29, 2010

**Supported by:** National Natural Sciences Foundation of China (No. 30400268), National Basic Research Program of China (973 Program) (No. 2007CB108801).

**Corresponding author:** Zhurong Zou. E-mail: zrz09@sina.com

Chunyi Zhang. Tel: +86-10-82106138; E-mail: chunyi.zhang@caas.net.cn

国家自然科学基金 (No. 30400268), 国家重点基础研究发展计划 (973 计划) (No. 2007CB108801) 资助。

**Abstract:** Nowadays, SUMO fusion system is important for recombinant protein production in *Escherichia coli*, yet a few aspects remain to be improved, including the efficacy for vector construction and protein solubility. In this study, we found the SUMO gene *Smt3 (Sm)* of *Saccharomyces cerevisiae* conferred an unexpected activity of constitutive prokaryotic promoter during its PCR cloning, and the gene coding regions of SUMOs in most species had a  $\sigma_{70}$ -dependent prokaryotic promoter embedded, through the prediction via the BPROM program developed by Softberry. By combining the characters of *Sm* promoter activity and the *Stu I* site (added at the 3'-terminal of *Sm*), and introducing a His-tag and a hyper-acidic solubility-enhancing tag, we further constructed a set of versatile vectors for gene cloning and expression on the basis of *Sm'-LacZa* fusion gene. Experimentally started from these vectors, several target genes were subcloned and expressed through blue-white screening and SDS-PAGE analysis. The results manifest a few of expectable advantages such as rapid vector construction, highly soluble protein expression and feasible co-expression of correlated proteins. Conclusively, our optimized SUMO fusion technology herein could confer a large potential in *E. coli* protein expression system, and the simultaneously established co-expression vector systems could also be very useful in studying the protein-protein interactions *in vivo*.

**Keywords:** SUMO gene, prokaryotic promoter, SUMO fusion system, co-expression, *Escherichia coli*

在真核生物泛素家族成员中, SUMO (Small ubiquitin-related modifier, 小泛素相关修饰物) 具有多种重要的生物学功能。SUMO 化修饰介导靶蛋白分子定位和功能调节, 在转录调控、核质转运、信号转导、DNA 修复和细胞周期调控等方面均发挥着重要作用<sup>[1-2]</sup>。

近年来, SUMO 作为 N 端融合标签在大肠杆菌表达系统中得到越来越多的应用<sup>[3-8]</sup>。SUMO 标签不仅能有效提高目标蛋白的表达水平和可溶性<sup>[5]</sup>, 而且整体作为一个空间结构上的识别位点, 被其特异性蛋白酶 Ulp1 从融合表达蛋白上定位 (SUMO 末端二甘氨酸, -GG 基序的后随肽键上) 切除<sup>[3,9]</sup>, 不存在任何非特异性酶切, 目标蛋白上可以不留任何残基, 这对生产具有天然氨基酸序列的重组蛋白极其适用。另外, Ulp1 不仅酶切效率高, 而且在较广范围的 pH 和温度条件下都具有活性<sup>[3]</sup>, 这相比其他切除标签、依赖识别序列的工具蛋白酶 (如 Factor Xa、烟草蚀刻病毒蛋白酶 TEV、肠激酶 EK) 都具有显著的综合优势。类似的泛素 (Ubiquitin) 融合系统虽然具备 SUMO 融合系统的诸多优点<sup>[3,10]</sup>, 但切除泛素标签的去泛素化酶 DUBs 不稳定, 难以生产和廉价制备<sup>[3]</sup>; 而且大肠杆菌中存在一种内源性泛素蛋白酶 ElaD, 特异性切割泛素融合蛋白, 使得泛素标签作用失去意义<sup>[11]</sup>。因此, SUMO 融合技

术已成为目前大肠杆菌重组蛋白表达和纯化的重要手段。

尽管如此, SUMO 融合系统在载体构建效率和蛋白可溶性等方面仍有待改进。在本实验中, 我们发现酿酒酵母 SUMO 基因 *Smt3(Sm)* 具有组成型原核启动子活性。利用此特性, 结合引入 *Sm* 3' 末端 *Stu I* 位点以及 His 标签和超酸增溶标签, 构建了基于 *Sm'-LacZa* 融合基因的系列通用克隆表达载体。通过多个基因克隆和表达实验, 结果表明在改良的 SUMO 融合系统中实现了基因表达载体的快速构建和蛋白高可溶性表达, 一同建立的共表达载体系统还可以用作研究细胞内蛋白间的相互作用。

## 1 材料与方法

### 1.1 材料

大肠杆菌菌株 DH5 $\alpha$ 、BL21(DE3) 以及 pUC19、pET32a(+)、pACYC184 和 pEGFP-N1 等质粒为本实验室保存, 本工作中所建载体以外的其他载体均为本实验室先前构建和保存。限制性内切酶、*pfu* DNA 聚合酶、T4 DNA 连接酶和 DNA marker 购自 MBI Fermentas 公司; DNA 快速纯化回收试剂盒和蛋白分子量 marker 购自北京鼎国生物技术有限公司; 氨苄青霉素 (Amp)、氯霉素 (Cm)、异丙基- $\beta$ -D-硫代半乳糖苷 (IPTG) 为 Genview 公司

产品；其他生化试剂均为国产分析纯。本工作中所有引物 (表 1) 的合成以及 DNA 序列分析均由上海生工生物工程技术有限公司完成。

表 1 本研究中所用引物

Table 1 Primers used in this study

| Primer name | Primer sequence (5'-3')                                       |
|-------------|---|
| SmNdFw      | ATAgCATATgTCggACTCgAAgTCAA                                    |
| SmStRv      | ATggTCAggCCTCCAATCTgTTCTCTgTgAg                               |
| HSNdFw      | gAACCATATgAgAggATCgCATCACCATCAC<br>CATCACACTATgTCggACTCgAAgTC |
| LZXhRv      | gCATTATCTCgAgTTAgCgCCATTCgCCATTC                              |
| GfpFw       | TATggTgAgCAAgggCgAggAg  |
| GfpRv       | gAACTTAACTTgTACAgCTCgTCCAT                                    |
| SmNsRv      | CTAgAgTCgACACCATggTCaggCCTCCAATC                              |
| LZNsFw      | CATggTgTCgACTCTAgAggATC                                       |
| pAYFw       | gCTggATAICTgTCAAACATgAgAATTACAAC                              |
| pAYRv       | gCTggATATCACACggTCACACTgCTTC                                  |
| Et1Fw       | TATggAATCACTCgTCCTCgTg  |
| Et1Rv       | gTCTTAAACCgAATTCACgCTgCTTgAg                                  |
| TevFw       | ggTgAATCACTgTTCAAaggAC  |
| TevRv       | ggTTAgTTCATCgTTgggTCgCTTC                                     |
| EkFw        | CATgTTggAggAAgTgATTCAAgAg                                     |
| EkRv        | ggCTTAAATgTAggAAACTTTgAATCCATTC                               |
| Pt7Fw       | CCgCgAAATTAATACgACTCACTA                                      |
| Tt7Rv       | ggATATAgTTCCTCCTTTCAgCA                                       |
| Ulp1Fw      | CCTTgTTCCTgAATTAATgAAAAgAgC                                   |
| Ulp1Rv      | gCCTAACTTCAAAGCgTCggTAAAAATC                                  |

Forward primer is named with the suffix 'Fw', and reverse primer with 'Rv'. Thin-solid underline indicates the introduced restriction enzyme site. Double-solid underline indicates the DNA sequence of His-tag. Bold-solid underline indicates the first gene codon of target protein. Bold-dot underline indicates the tri-bases complementary to the stop gene codon of target protein.

## 1.2 方法

### 1.2.1 载体的构建

所用基因操作方法均参照分子克隆实验手册<sup>[12]</sup>。载体构建如结果中详述，并通过 DNA 测序验证。

### 1.2.2 原核启动子预测分析

在 NCBI 网站上查询并提取各物种 SUMO 基因的 mRNA 序列，分别对其基因编码区序列 (CDS) 通过软莓网站 (www.softberry.com) 提供的原核启动子预测程序 BPROM 和设定的参数 (LDF  $\geq 0.2$

指含有启动子) 进行分析，得到预测原核启动子的强度 (LDF 值)、-10 和-35 元件等信息，并通过 EXCEL 软件整理分析。

### 1.2.3 蛋白表达及 SDS-PAGE 分析

将表达载体转化至大肠杆菌 BL21(DE3) 中，挑取单菌落 37 °C 培养过夜。然后按照 4% 量接种到含相应抗生素浓度的 LB 液体培养基中 (对 pET 载体单独表达使用 100 mg/L Amp，对 pAY 载体单独表达使用 34 mg/L Cm；对 pET 和 pAY 载体共表达使用 100 mg/L Amp 和 34 mg/L Cm)，37 °C 培养至 OD<sub>600</sub> 达 0.6，加入 IPTG 至终浓度 1 mmol/L，37 °C 诱导表达 4 h。收集 10 mL 的诱导表达细菌培养液，离心后加入 3 mL PBS 缓冲液悬浮，并用超声波裂解。取细胞裂解物 20  $\mu$ L，15 000 r/min、4 °C 离心 5 min，分为上清 (S) 和沉淀 (P)，用 12% SDS-PAGE 进行电泳检测，并用凝胶成像系统生成图形文件。

### 1.2.4 蛋白体外酶切反应

将 SUMO 融合蛋白的 pET 载体表达细菌与酵母 Ulp1 蛋白酶的载体 pET(Ulp1) 表达细菌分别按上述方法制备得到裂解物上清。两者上清一般按体积比 20 : 1 混匀，放置 25 °C 反应 3 h，然后用 12% SDS-PAGE 分析反应情况。

## 2 结果

### 2.1 酵母 SMUO 基因 *Smt3* 具有组成型原核启动子活性

利用引物 SmNdFw/SmStRv 扩增酿酒酵母 SUMO 基因 *Smt3*(*Sm*) 的 DNA 片段平末端克隆到 pUC19 的 *Hinc* II (GTC↓GAC) 位点，转化大肠杆菌 DH5 $\alpha$ ，发现在 X-gal+IPTG 的 LB 平板上长出来的菌落几乎全都为蓝色。经 PCR 鉴定 *Sm* 的重组克隆，命名为 pUC(Sm-LacZ $\alpha$ )。

通过测序分析，发现 *Sm* 插入 (图 1，两竖箭头之间) 打断了 pUC19 上 *LacZ $\alpha$*  的翻译阅读框，即由核糖体结合位点 RBS(lac) [AGGA] 介导的翻译 (图 1)。但是，*Sm* 的反向引物 SmStRv 在其 3' 末端 (即

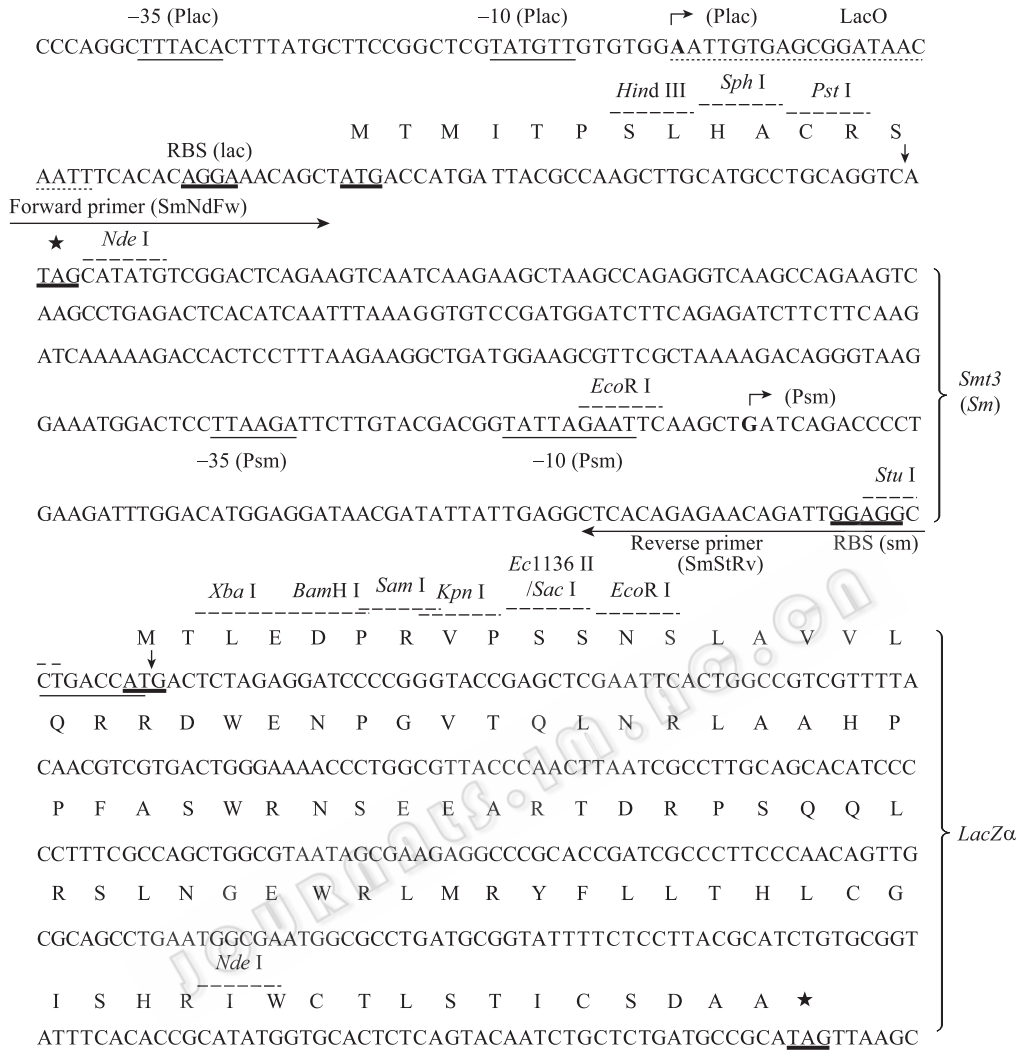


图 1 质粒 pUC19(Sm-LacZα) 中 SUMO 基因 *Smt3* 与 *LacZα* 融合的顺序分析

Fig. 1 Sequence analysis of the fusion gene of *Smt3* and *LacZα* in plasmid pUC19 (Sm-LacZα).

-GG 基序密码子处) 引入 *Stu I* 位点 (AGG↓CCT) 产生了一个经典的 RBS, 即 RBS(sm) [GGAGG], 同时该引物 5'端附加保护碱基对应的互补碱基 (-AT 3') 与克隆位点 *Hinc II* 的平末端 (5' GAC-) 融合产生一个 ATG 密码子, 与 RBS(Sm) 的间隙长为 7 bp, 这些刚好为其下游 *LacZα* 表达提供了有效的翻译元件 (图 1), 从而导致蓝色菌落形成。不过, 这种翻译模板除了由 *Lac* 启动子 (Plac) (图 1) 在 IPTG 诱导下产生的 *Sm-LacZα* 转录产物, 会不会在 *Sm* 上存在其他启动子产生新的转录产物并由 RBS(sm) 介

导翻译 *LacZα* 呢?

通过对 *Sm*(306 bp) 进行启动子预测分析, 发现 *Sm* 存在依赖  $\sigma_{70}$  的原核启动子 Psm, 包含 -35(Psm)、-10(Psm) 元件, 分别为 TTAAGA、TATTAGAAT (其后面 6 bp 为核心序列) (图 1), 可能引导下游 *LacZα* 表达。由于受固有 *Lac* 启动子的转录干扰, 这种预测需要进一步证实。

随后以 pUC(Sm-LacZα) 为模板, 利用引物 SmNdFw 和 LZXhRv, 扩增出 *Sm-LacZα* 片段, 经 *Nde I* 和 *Xho I* 酶切, 与同样酶切的 pET32a(+) 连

接,转化大肠杆菌 DH5 $\alpha$ ,发现在 X-gal+IPTG 的 LB 平板上也能长出蓝色菌落,而且经鉴定全都为阳性克隆,取名 pET(Sm-LacZ $\alpha$ )。由于 pET32a(+) 使用依赖噬菌体 T7 RNA 聚合酶的 T7 启动子(后随 LacO 序列),DH5 $\alpha$  和载体自身不表达 T7 RNA 聚合酶,因而即便存在 IPTG 也不会由 T7 启动子转录 Sm-LacZ $\alpha$ ,那么这个导致蓝色菌落形成的 LacZ $\alpha$  表达唯可能由 Sm 启动子 Psm 的作用所致。

另外,以 pEGFP-N1 为模板,利用引物 GFPfw 和 GFPrv 扩增 GFP (绿色荧光蛋白基因) 编码区片段,分别平端插入到 pET(Sm-LacZ $\alpha$ ) 的 Stu I 和 Ecl136 II 位点(图 1),转化大肠杆菌 DH5 $\alpha$ ,通过蓝白斑筛选和菌落 PCR 鉴定正确阳性克隆 pET(Sm-GFP) stu 和 pET(Sm-GFP) ecl。GFP 片段在两个位点的插入都导致其后面的 LacZ $\alpha$  不表达,形成白色菌落。分别挑取 2 种白菌落划线在普通 LB 平板上培养,发现 pET(Sm-GFP) ecl 的菌落逐渐变成绿色,而 pET(Sm-GFP) stu 的菌落颜色则没什么变化。通过序列分析,GFP 在两个位点的插入都与其 5'端 Sm 形成一个正确阅读框的融合基因,而且都有可能由 Sm 启动子引导 GFP 转录物产生。但是,在 Stu I 位点插入导致 RBS(sm)(图 1)与 GFP 起始密码子 ATG 紧密相邻,GFP 不能在此处翻译;而在 Ecl136 II 位点插入则可以通过 RBS(sm)和其后的 ATG 翻译出 GFP 蛋白并累积,使 pET(Sm-GFP) ecl 菌落逐渐变成绿色。这表明在诱导物 IPTG 不存在的情况下,Sm 启动子能够组成型发挥转录活性,叠加在 Stu I 位点的 RBS(sm)是其下游基因翻译的重要功能元件。总之,上述结果表明酵母 SUMO 基因 Smt3 具有明显的组成型原核启动子活性。

## 2.2 不同物种 SUMO 基因的原核启动子预测分析

SUMO 是真核生物中广泛存在的一类蛋白,酿酒酵母 SUMO 基因 Smt3 具有原核启动子活性,我们推测其他物种的 SUMO 基因可能具有同样特性。

通过软莓程序 BPROM 对 NCBI 核酸数据库中几乎所有的 SUMO 基因编码区(CDS)进行启动子

预测分析,发现 121 个被测对象中约 90%都具有原核启动子活性(LDF $\geq$ 0.2),在不同范围 LDF 值的 SUMO CDS 的比例依次为 3% (>5)、4% (4~5)、9% (3~4)、19% (2~3)、38% (1~2) 和 17% (0.2~1)(图 2)。进一步将不同 SUMO CDS 按物种划分成 4 类:真菌(Fungi)、植物(Plant)(绝大多数为被子植物)、无脊椎动物(Invertebrate)和脊椎动物(Vertebrate),然后分类统计它们在不同 LDF 范围所占比例。结果显示,真菌和植物中 85%的 SUMO CDS 具有原核启动子活性,而在脊椎和无脊椎动物中这个比例高达 95%。LDF>3 的 SUMO CDS 在各类物种所占比例分别为 30%(真菌)、18%(植物)、27%(无脊椎动物)和 2%(脊椎动物),说明除脊椎动物外其他物种的 SUMO CDS 有较大比例具有较高的原核启动子活性,尤其在真菌和无脊椎动物中表现更为突出,LDF>4 的比例分别为 19%和 16%(图 2)。

## 2.3 基于 Sm'-LacZ $\alpha$ 融合基因的通用克隆表达载体构建及其特性分析

以 pET(Sm-LacZ $\alpha$ ) 为模板,使用引物对 HSNdFw/SmNsRv 和 LZNsFw/LZXhRv,通过引物重叠延伸法<sup>[13]</sup>合成新的融合基因片段 Sm'-LacZ $\alpha$ 。通过 Nde I 和 Xho I 酶切,与同样酶切的 pET32a(+) 连接,构建成 pET(Sm'-LacZ $\alpha$ )(图 3A)。Sm'是由引物 HSNdFw 使 Sm 5'端加上了一个 His 标签序列(H6),有利于表达蛋白纯化;同时在 Sm'-LacZ $\alpha$  融合处又引入了 Nco I 和 Sal I/Hinc II 位点,使得该处多克隆位点(MCS)数目更多,便于目的基因(如 TEV)多种形式克隆和表达。在 Sm 3'末端 -GG 基序密码子处,引入了一个 Stu I 位点(-ggAGG↓C▲CT-,下划线代表-GG 基序密码子,▲代表 Ulp1 切点;大写序列 AGG↓CCT 代表 Stu I 位点,↓代表平末端切点)(图 1、图 3A),该酶切平末端 5'-Sm' ggAGG-3'能与任何靶蛋白编码序列 target 的 PCR 产物 5'-N target-3' (在 target 5'端仅需要通过 PCR 引物添加任何一种碱基 N,表 1) 连接,即可得到正确阅读框的

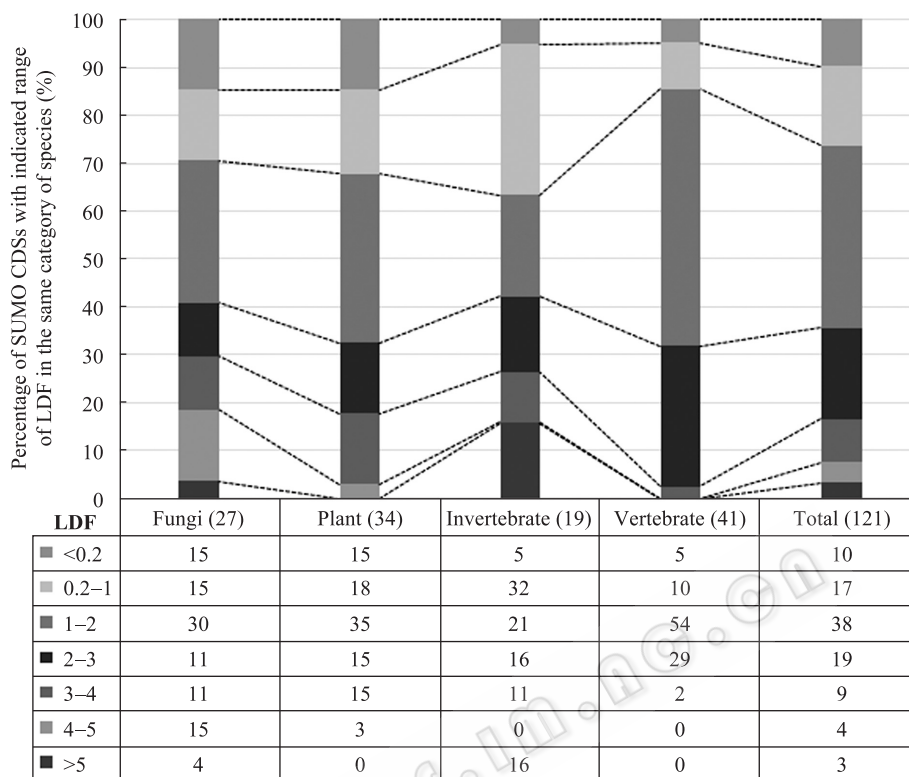


图 2 含某指定范围预测启动子 LDF 值的 SUMO 基因 CDS 在同类物种中所占比例

Fig. 2 Percentages of SUMO CDSs with indicated range of LDFs of the predicted promoters in the same category of species.

*Sm'*-target 融合基因 (即 5'-*Sm'* ggAGGN $\blacktriangle$ target-3'), 融合表达蛋白刚好在 -GG 基序 (由 ggA GGN 编码) 后被 Ulp1 酶切开, 不会残留任何氨基酸在靶蛋白的 N 端。由于 *Sm* 标签还只能相对有限地提高靶蛋白的可溶性<sup>[7-8]</sup>, 我们分别将超酸增溶标签 *Msb* (*E. coli* msyB)、*Yd* (*E. coli* yjgD)、*Od* (*E. coli* rpoD 的 N 端结构域)<sup>[7-8]</sup> 的 PCR 产物用 *Nde* I 酶切, 然后插入到 pET(*Sm'*-LacZ $\alpha$ ) 的 *Nde* I 位点 (图 3A), 构建成通用增溶表达载体 pET(*MsbSm'*-LacZ $\alpha$ )、pET(*YdSm'*-LacZ $\alpha$ ) 和 pET(*OdSm'*-LacZ $\alpha$ ), 超酸标签 *Msb*、*Yd*、*Od* 与 *Sm'* 分别形成复合超酸增溶标签 *MsbSm'*、*YdSm'* 和 *OdSm'*。由于 *Sm* 原核启动子活性, 所有这些载体的大肠杆菌 DH5 $\alpha$  转化菌落在 X-gal+IPTG 平板上都呈现蓝色。总之, 这些载体组成的改良 SUMO 融合系统理论上不但通过蓝白斑筛

选非常便于目的基因 PCR 产物的快速克隆和蛋白可溶性表达, 而且尤其适合含天然氨基酸序列靶蛋白的重组生产。

此外, 用引物 PayFw 和 PayRv 扩增低拷贝型质粒 pACYC184<sup>[14]</sup>, 得到去除四环素抗性基因的大片段, 然后用 *Eco*R V 酶切, 与上面 *Sm'*-LacZ $\alpha$  融合基因 PCR 片段连接, 转化大肠杆菌 DH5 $\alpha$ , 通过蓝白斑筛选和菌落 PCR 鉴定构建了重组质粒 pAY(*Sm'*-LacZ $\alpha$ ) (图 3B)。与载体 pET(*Sm'*-LacZ $\alpha$ ) 一样, 它也可以通过蓝白斑筛选用作目的基因片段 (如 *Ulp1*) 的快速克隆和表达。由于 pAY(*Sm'*-LacZ $\alpha$ ) 派生载体含 pACYC184 的复制起始区 p15A, 与其他类型如 pET(*Sm'*-LacZ $\alpha$ ) 派生载体 (含 *ColE1* 复制起始区) 相容, 两者可在同一细菌中共表达, 可以用于两个 (或几个) 表达靶蛋白的相互作用分析。

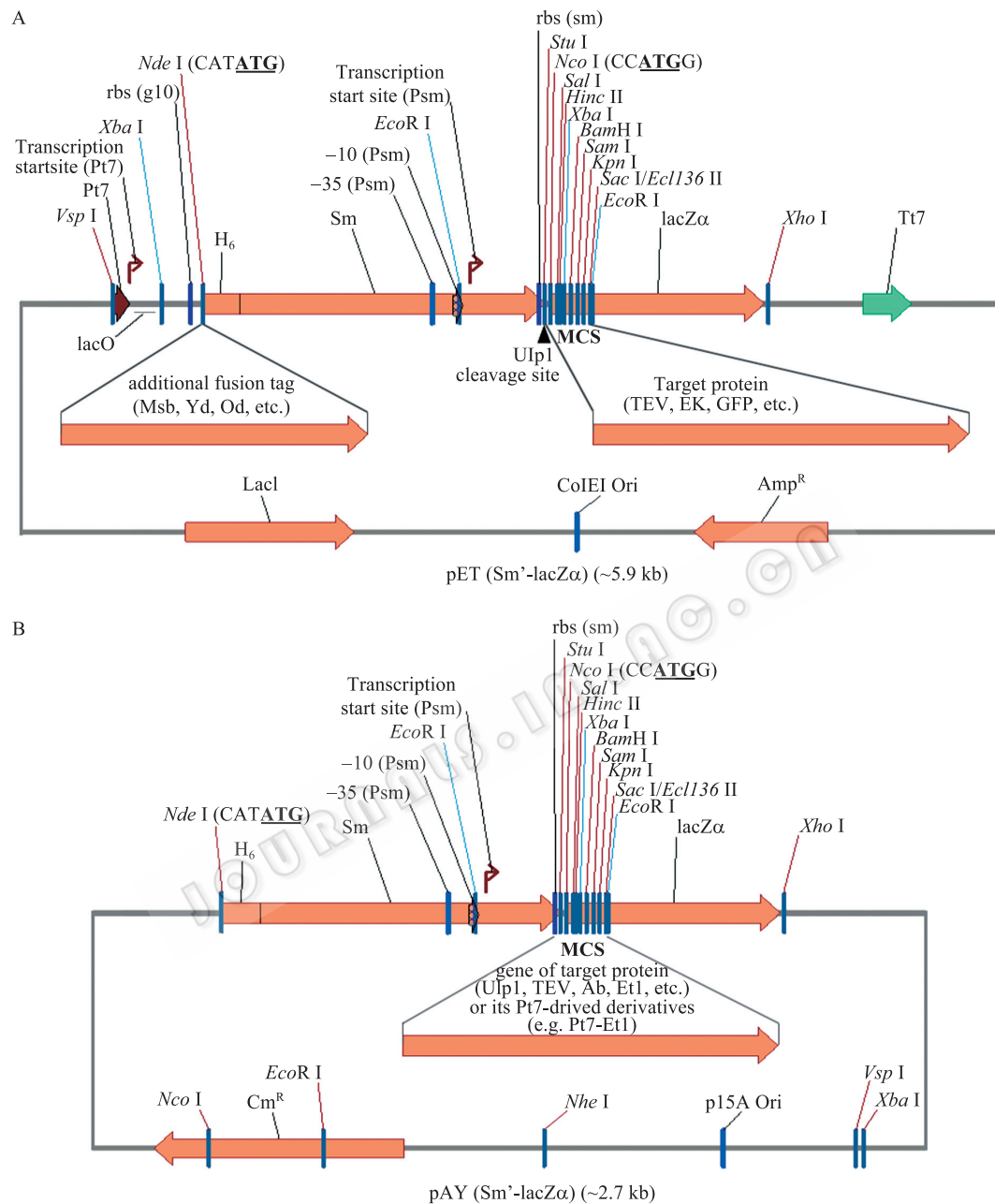


图3 载体 pET(Sm'-LacZα) (A) 和 pAY(Sm'-LacZα) (B) 示意图

Fig. 3 Diagrams of vectors pET(Sm'-LacZα) (A) and pAY(Sm'-LacZα) (B).

#### 2.4 基于 pET(Sm'-LacZα) 系列的基因表达载体构建及其蛋白表达分析

根据上述 pET(Sm'-LacZα) 系列载体的特性, 分别利用引物对 GfpFw/GfpRv、Et1Fw/Et1Rv、TevFw/TevRv、EkFw/EkRv 从 pET(GFP)、pET(Et1)、

pET(EK)、pET(TEV) 扩增出靶蛋白 GFP、Et1 (TEL-SAM, 转录抑制因子 TEL 的 N 端 SAM 结构域)、EK、TEV 的基因片段, 插入到载体的 *Stu* I 位点, 通过蓝白斑筛选和菌落 PCR 鉴定, 依次得到重组质粒 pET(Sm'-GFP)、pET(MsbSm'-GFP)、

pET(Sm'-Et1)、pET(YdSm'-Et1)、pET(OdSm'-Et1)、pET(Sm'-EK)、pET(MsbSm'-EK)、pET(YdSm'-EK)、pET(OdSm'-EK)、pET(Sm'-TEV)、pET(MsbSm'-TEV)、pET(YdSm'-TEV) 和 pET(OdSm'-TEV)。

通过 IPTG 诱导, 这些载体都获得了很好地表达, 融合表达蛋白的可溶性与其对应的非融合表达靶蛋白相比都不同程度地获得了提高。其中, Sm' 能轻微增强蛋白可溶性, 但其各种复合超酸增溶标签 (MsbSm'、YdSm'、OdSm') 的增溶效果则极其显著

(图 4), 这与我们先前工作报道的结论一致<sup>[7-8]</sup>。

将每个高可溶性表达融合蛋白的细菌裂解物上清分别与表达 Ulp1 酶的细菌裂解物上清进行体外反应, 经 SDS-PAGE 分析, 发现这些含 Sm 的融合蛋白都能与 Ulp1 很好地反应, Sm' 及其各种复合标签 (MsbSm'、YdSm'、OdSm') 都能有效地从融合蛋白上切离, 并得到相应的靶蛋白产物 (GFP、Et1、EK、TEV) (图 5), 这说明 Sm 的 N 端附加蛋白序列并不影响 Ulp1 对它的特异性酶切, 而且也证实

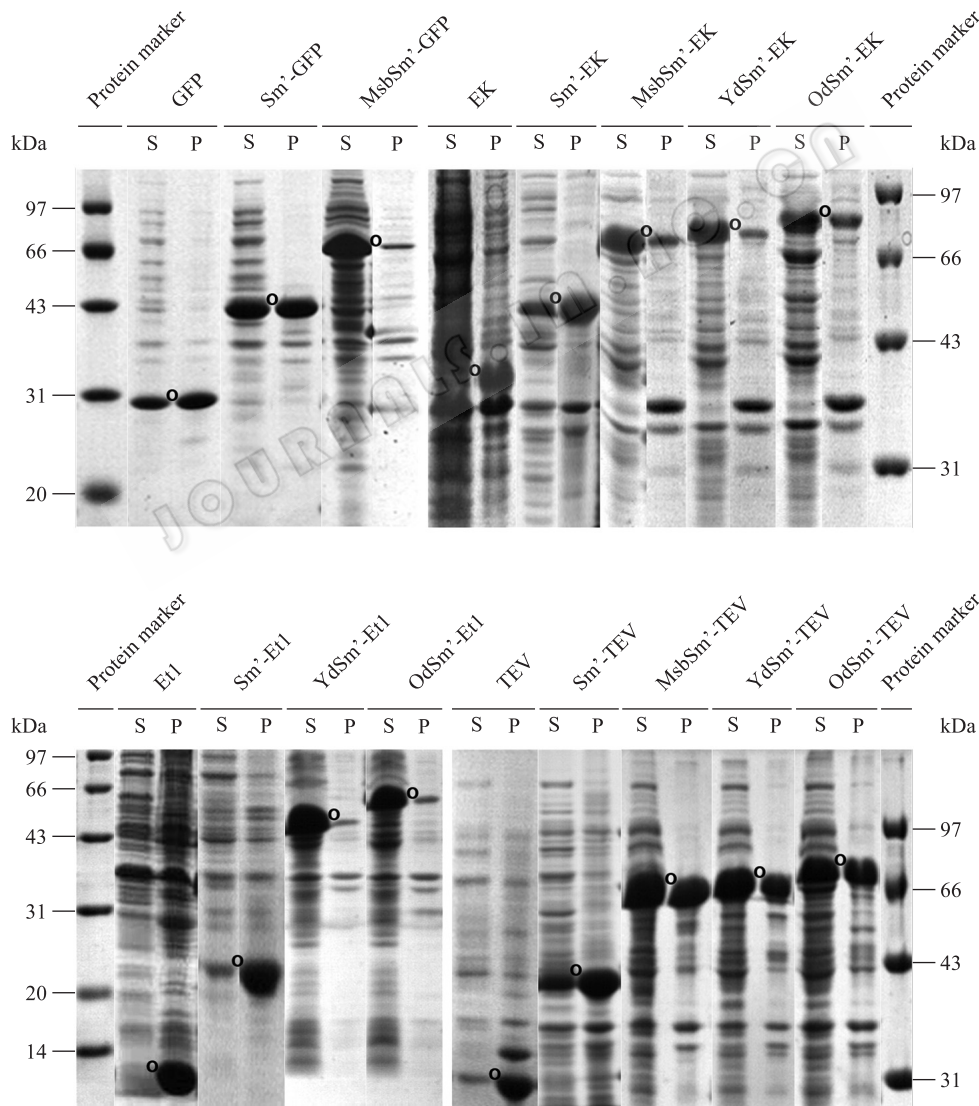


图 4 SDS-PAGE 分析各种表达载体在大肠杆菌 BL21(DE3) 中的表达

Fig. 4 Expression of various constructs in *E. coli* BL21(DE3) analyzed by SDS-PAGE. <sup>o</sup> indicates the expressed target protein in non-fusion or fusion form; S: supernatant of the cell lysate; P: pellet of the cell lysate.



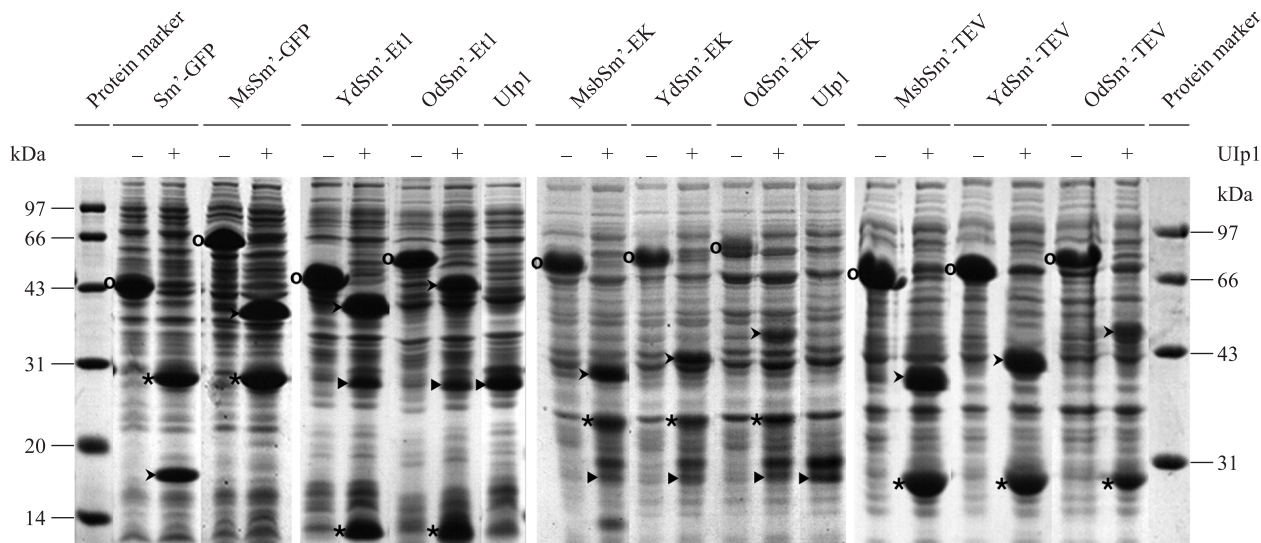


图5 SDS-PAGE 分析 Sm 融合蛋白与 Ulp1 表达细菌的裂解物上清体外酶切反应

Fig. 5 *In vitro* enzymatic reactions of the supernatants of cell lysates of Sm fusion proteins and Ulp1, analyzed by SDS-PAGE.  $\circ$  indicates Sm fusion protein; \* indicates target protein;  $\triangleright$  indicates Sm' or its combinatorial tag (e.g. YdSm');  $\blacktriangleright$  indicates Ulp1 overused in some reactions.

Ulp1 酶的反应条件非常温和<sup>[3,8]</sup>, 在细菌 PBS 裂解上清中都有很好的活性。这种 SUMO/Ulp1 体外反应实验可以用来分析不同标签对靶蛋白的增溶作用机理<sup>[8]</sup>。另外, 这类融合蛋白表达产物如果通过 His 标签亲和层析纯化, 然后与纯化的 Ulp1 酶体外反应, 将能够有效地制备各种所需的靶蛋白 (尤其是含天然氨基酸序列的功能蛋白), 与我们以前报道的脂联素结果一致<sup>[15]</sup>。

## 2.5 基于 pAY(Sm'-LacZ $\alpha$ ) 的基因表达载体构建及其蛋白表达分析

基于上述 pAY(Sm'-LacZ $\alpha$ ) 的特性, 根据需要进行快速简便地将不同目的基因片段克隆到该质粒的 MCS 上 (图 3B), 通过蓝白斑筛选和菌落 PCR 鉴定得到重组质粒。

当需要诱导型表达时, 可以将目的基因表达盒 (诱导型启动子+目的基因+转录终止子) 插入到 MCS 上的任何单一酶切位点。本实验利用引物对 Pt7Fw/Tt7Rv, 分别以 pET(Et1)、pET(Et1-GFP)、pET(YA)、pET(MST) 为模板, 扩增其对应基因表达盒片段 Pt7\_Et1\_Tt7、Pt7\_Et1-GFP\_Tt7、Pt7\_YA\_Tt7、Pt7\_MST\_Tt7 (Pt7 为 t7 启动子, Tt7 为 t7 转录终止

子, 位于 pET 载体上, 图 3A)。将这些 PCR 片段分别平末端克隆在 pAY(Sm'-LacZ $\alpha$ ) 的 MCS 内 *Hinc* II 位点 (可以选择其他平末端酶位点如 *Stu* I、*Sma* I 和 *Ecl*136 II), 构建成载体 pAY(Et1)、pAY(Et1-GFP)、pAY(YA) 和 pAY(MST)。如同 pET 系列载体一样, 它们都是受 IPTG 诱导表达。在不加 IPTG 的情况下, Pt7 不会发生转录, 同时由插入片段上游 Sm 启动子 Psm 起始的转录遇到 Pt7 后 *LacO* 序列时也会被阻断, 因而没有目的基因表达产物; 当 IPTG 诱导时, 这些 pAY 载体都能很好地表达 (图 6)。由于 pAY 载体与 pET 载体使用不同的复制起始区 (p15A、ColEI), 因而能够相容于同一细菌中共表达。分别将 pET(EY)/pAY(Et1)、pET(EY)/pAY(Et1-GFP)、pET(Trx-Ab)/pAY(YA)、pET(Ab-GFP)/pAY(YA)、pET(TEV)/pAY(MST) 进行共表达, 发现 IPTG 诱导后所有目的基因都不同程度地获得高水平表达 (图 6), 这种共表达实验可以用来研究两 (几) 个关联蛋白在细胞内的相互作用。此外, 所有 pAY 载体在任何情况下都有一条明显的 Cm 抗性基因 (Cm<sup>R</sup>) 表达条带 (图 6), 这可能与 Cm<sup>R</sup> 上游组成型启动子活性密切相关。

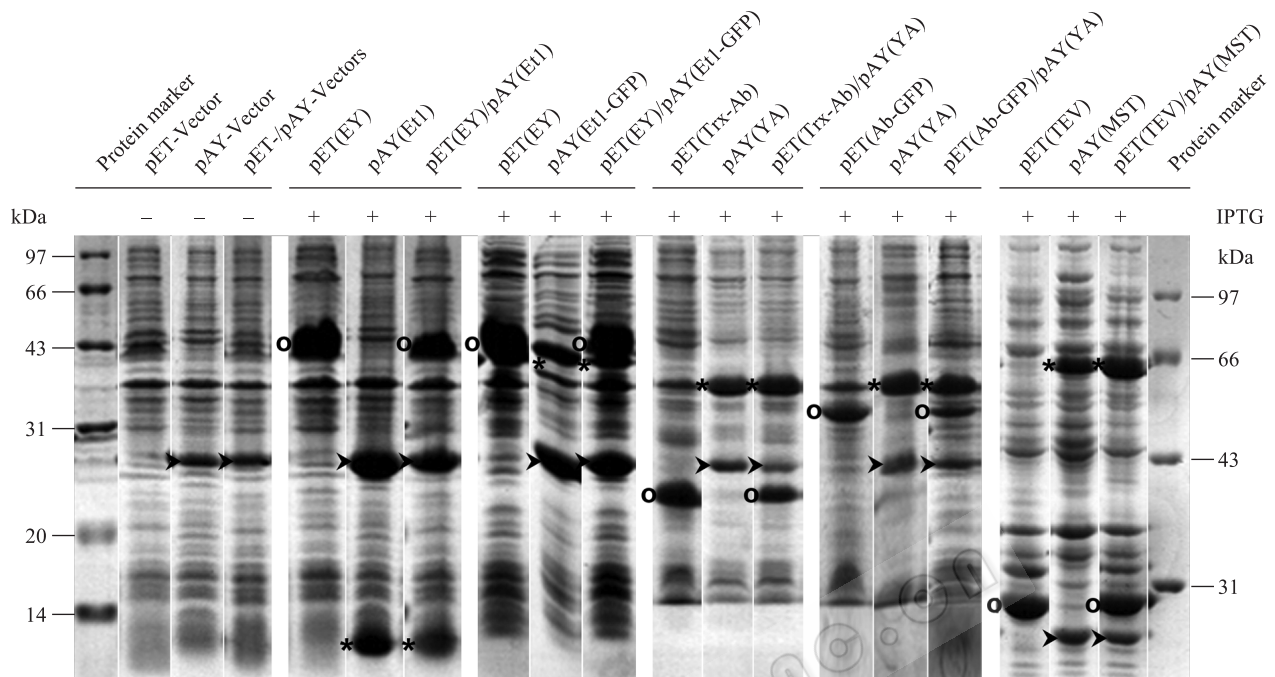


图 6 SDS-PAGE 分析派生于 pAY(Sm'-LacZa) 的 pAY 系列载体与其相关 pET 载体的共表达

Fig. 6 Co-expressions of pAY (Sm'-LacZa)-derived pAY vectors and their correlated pET vectors, analyzed by SDS-PAGE. <sup>o</sup> indicates the recombinant protein expressed from pET vector; \* indicates the recombinant protein expressed from pAY vector; > indicates the protein product of Cm<sup>R</sup> (chloramphenicol acetyltransferase).

当需要组成型表达时, 可以将目的基因编码区插入在 pAY(Sm'-LacZa) MCS 的 *Nco* I 后面任何单一酶切位点 (图 3B), 并使基因翻译阅读框从 *Nco* I 上的 ATG 密码子开始, 这样目的基因就可以由 *Sm* 启动子 *Psm* 组成型转录, 并由 RBS(sm) 和其后 *Nco* I 上 ATG 密码子介导翻译 (图 1)。在一些细菌体内共表达分析系统中, 某种蛋白 (如工具酶) 并不需要高水平的表达, 这样有必要选择普通启动子而不是强启动子如 *Pt7*。基于此, 我们利用引物对 *Ulp1Fw*/*Ulp1Rv* 从 pET(*Ulp1*) 模板中扩增出 *Ulp1* 片段, 插入到 pAY(Sm'-LacZa) 的 *Hinc* II 位点, 构建了 *Ulp1* 组成型表达载体 pAY(c-*Ulp1*) (c-代表由 *Sm* 启动子介导的组成型表达), 并与上述含 *Sm* 标签的 pET 系列载体如 pET(*YdSm'*-*Et1*) 共表达, 体内分析标签对靶蛋白的作用, 得到的结果 (未显示) 与我们先前报道的一致<sup>[8]</sup>。

### 3 讨论

近年来, SUMO 融合技术在大肠杆菌表达系统中得到越来越多的应用, 但仍处在不断的改进中<sup>[3-8]</sup>。在本实验中, 我们发现酵母 SUMO 基因 *Smt3* (*Sm*) 具有组成型原核启动子活性, 包含明显的 -35、-10 特征序列 (图 1)。而且, 通过软件 BPROM 程序预测发现, 大多数物种 SUMO 基因编码区普遍都具有依赖  $\sigma_{70}$  的原核启动子, 但活性存在较大差异 (图 2), 这种现象可能与 SUMO 基因作为直向同源基因 (Ortholog) 在不同类物种的进化程度相关。

利用 SUMO 标签本身特点及 *Sm* 内含原核启动子特性, 通过在 *Sm* 3'末端 (-GG 基序密码子处) 引入 *Stu* I 位点和进一步在 *Sm* 的 5'端附加 His 标签和超酸增溶标签 (Msb、Yd、Od), 我们构建了基于 *Sm'-LacZa* 融合基因和 pET32a(+) 背景的系列通用克隆表达载体 (图 3A), 并由多个基因克隆和表达实

验得到检验(图4、5)。由此组成的改良SUMO融合系统集成诸多特性,如通过蓝白斑筛选快速构建基因表达载体、SUMO和超酸增溶标签促进蛋白可溶性表达、SUMO/Ulp1反应系统制备任何含天然氨基酸序列的靶蛋白、利用His标签纯化表达蛋白、以及pET系统固有的蛋白高水平表达等等,因此可以成为大肠杆菌重组蛋白表达的有力工具;与另一改良的SUMO融合系统<sup>[6]</sup>相比,在某些方面可能更具优势。

另外,我们还构建了基于 $Sm'$ - $LacZ\alpha$ 融合基因、派生于质粒pACYC184(低拷贝、含p15A复制起始区)的通用克隆表达载体pAY( $Sm'$ - $LacZ\alpha$ )(图3B)。通过蓝白斑筛选鉴定,由此按需构建的pAY系列目的基因表达载体可以进行单独表达或与其他含不同复制起始区(如ColEI)的质粒表达载体如pET系列共表达(图6)、组成型或诱导型表达,以用于蛋白研究的不同用途如分析细胞内蛋白间的相互作用关系<sup>[8]</sup>等。

## REFERENCES

- [1] Hay RT. SUMO: a history of modification. *Mol Cell*, 2005, 18(1): 1-12.
- [2] Johnson ES. Protein modification by SUMO. *Annu Rev Biochem*, 2004, 73: 355-382.
- [3] Malakhov MP, Mattern MR, Malakhova OA, et al. SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins. *J Struct Funct Genomics*, 2004, 5(1/2): 75-86.
- [4] Butt TR, Edavettal SC, Hall JP, et al. SUMO fusion technology for difficult-to-express proteins. *Protein Expr Purif*, 2005, 43(1): 1-9.
- [5] Marblestone JG, Edavettal SC, Lim Y, et al. Comparison of SUMO fusion technology with traditional gene fusion systems: Enhanced expression and solubility with SUMO. *Protein Sci*, 2006, 15(1): 182-189.
- [6] Lee CD, Sun HC, Hu SM, et al. An improved SUMO fusion protein system for effective production of native proteins. *Protein Sci*, 2008, 17(7): 1241-1248.
- [7] Su Y, Zou ZR, Feng SY, et al. The acidity of protein fusion partners predominantly determines the efficacy to improve the solubility of the target proteins expressed in *Escherichia coli*. *J Biotech*, 2007, 129(3): 373-382.
- [8] Zou ZR, Cao LJ, Zhou P, et al. Hyper-acidic protein fusion partners improve solubility and assist correct folding of recombinant proteins expressed in *Escherichia coli*. *J Biotech*, 2008, 135(4): 333-339.
- [9] Mossessova E, Lima CD. Ulp1-SUMO crystal structure and genetic analysis reveal conserved interactions and a regulatory element essential for cell growth in yeast. *Mol Cell*, 2000, 5(5): 865-876.
- [10] Catanzariti AM, Soboleva TA, Jans DA, et al. An efficient system for high-level expression and easy purification of authentic recombinant proteins. *Protein Sci*, 2004, 13(5): 1331-1339.
- [11] Catic A, Misaghi S, Korbelt GA, et al. ElaD, a deubiquitinating protease expressed by *E. coli*. *PLoS ONE*, 2007, 2(4): e381.
- [12] Sambrook J, Russell DW. *Molecular Cloning: A Laboratory Manual*. 3rd Ed. New York: Cold Spring Harbor Laboratory Press, 2001.
- [13] Horton RM, Hunt HD, Ho SN, et al. Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. *Gene*, 1989, 77(1): 61-68.
- [14] Rose RE. The nucleotide sequence of pACYC184. *Nucleic Acids Res*, 1988, 16(1): 355.
- [15] Zhou P, Feng SY, Zou ZR, et al. Soluble expression of human adiponectin in *Escherichia coli*. *J Sichuan Univ: Nat Sci Ed*, 2010, 47(1): 161-166.  
周培, 封淑颖, 邹竹荣, 等. 人脂联素在大肠杆菌中的可溶性表达. *四川大学学报:自然科学版*, 2010, 47(1): 161-166.