

蜘蛛大壶状腺丝蛋白基因的克隆和原核表达 Cloning and Prokaryotic Expression of Major Ampullate Spidroin Gene of Spider

潘红春^{1*} 宋大祥² 周开亚² 朱国萍¹

PAN Hong-Chun^{1*}, SONG Da-Xiang², ZHOU Kai-Ya² and ZHU Guo-Ping¹

1 安徽师范大学生命科学学院分子进化与生物多样性实验室 芜湖 241000

2 南京师范大学遗传资源研究所 南京 210097

1 Laboratory of Molecular Evolution & Biodiversity, College of Life Sciences, Anhui Normal University, Wuhu 241000, China

2 Institute of Genetic Resources, Nanjing Normal University, Nanjing 210097, China

摘 要 以悦目金蛛(*Argiope amoena*)丝腺 SMART RACE cDNA 文库为模板进行 RT-PCR,克隆了 1 条大壶状腺丝蛋白(major ampullate spidroin, MaSp)基因 cDNA 序列。该条 cDNA 序列编码的氨基酸序列可区分为两部分:(1)富含丙氨酸的片段和富含甘氨酸的片段相间排列构成的重复氨基酸序列区,并且富含甘氨酸的片段中有脯氨酸分布(2)约 100 个氨基酸残基组成的 C 末端非重复氨基酸序列区。把 MaSp 基因 cDNA 序列亚克隆到质粒 pET28I(+)中,构建原核表达质粒 pET28I(+)-MaSp,表达质粒转化大肠杆菌 BL2I(DE3),用 IPTG 诱导表达。SDS-PAGE、氨基酸组成测定和 N 末端氨基酸序列测定的结果表明,表达产物为重组 MaSp,表达量约为 40mg/L。还对 C 末端非重复氨基酸序列对重组 MaSp 在水媒介中溶解性的影响进行了探讨。

关键词 悦目金蛛,大壶状腺丝蛋白基因,原核表达,pET28I(+),BL2I(DE3)

中图分类号 Q786 文献标识码 A 文章编号 1000-306I(2007)03-0446-06

Abstract RT-PCR was conducted with one degenerate primer designed according to repetitive regions' amino acid sequence of major ampullate spidroin(MaSp) in spiders and adaptor primer in the SMARTTM cDNA Library Construction Kit. By cloning and sequencing of amplified products, one cDNA clone(GenBank Accession No. AY365017) of *Argiope amoena* MaSp gene was obtained. The deduced amino acid sequence can be distinctly divided into two regions:(1) Repetitive region that consists of an alternating alanine-rich and glycine-rich domain in which many prolines are present; and (2) C-terminal non-repetitive region. The region coding for 272 amino acids of MaSp gene was subcloned into prokaryotic expression vector pET28I(+) and an about 26kD recombinant protein was expressed at high levels in *Escherichia coli* BL2I(DE3) after induction of IPTG. After being purified with metal-affinity chromatography on Ni²⁺-IDA-Sepharose columns as well as gel filtration chromatography, the recombinant protein was confirmed to be predicted MaSp by means of amino acid composition analysis and N-terminal amino acid sequence analysis. The solubility behavior of recombinant MaSp with C-terminal non-repetitive region in the present study is similar to that of recombinant dragline silk proteins without C-terminal non-repetitive region expressed by bacteria and yeast in the other studies. The result shows that absence or presence of C-terminal non-repetitive region is not a crucial factor affecting the solubility of the recombinant MaSp.

Key words *Argiope amoena*, major ampullate spidroin(MaSp) gene, prokaryotic expression, pET28I(+), BL2I(DE3)

Received: October 31, 2006; Accepted: December 28, 2006.

This work was supported by the grants from the Key Laboratory of Biotic Environment & Ecological Safety in Anhui Province, and the Key Laboratory of Conservation and Utilization of Important Biological Resources in Anhui Province.

* Corresponding author. Tel: +86-553-5991116; E-mail: panhongchun@126.com

安徽省高校生物环境与生态安全省级重点实验室资助项目和安徽省重要生物资源保护与利用研究省级重点实验室资助项目

许多蜘蛛如园蛛总科(Araneioidea)蜘蛛的丝腺能产几种不同的蛛丝蛋白,它们的氨基酸组成和力学性能均不相同^[1]。如下几个因素使得蜘蛛丝成为备受关注的天然材料(1)蛛丝是在常温、常压下,并以水作为溶剂等温和条件下形成的^2由于纺器喷丝头(spinneret)及喷丝头开关(spigot)上分布有肌肉和神经,因此蜘蛛能通过喷丝头开关控制蛛丝的直径等指标从而影响蛛丝的力学特性,这代表了动物行为直接影响蛋白质特性的途径^3蛛丝是研究蛋白质结构和功能联系的理想材料^4大多数溶剂不能溶解蛛丝(5)大多数蛋白酶对蛛丝蛋白不起作用^[5]。在几种蛛丝中,最值得称道的是大壶状腺丝,大壶状腺丝可作为蛛网的支架网框丝或蜘蛛的“生命线”拖丝(dragline silk),在蜘蛛生命活动中发挥重要作用。大壶状腺丝除具备与高性能化学合成纤维凯夫拉尔(Kevlar)不相上下的强度外,还同时兼备良好的延伸性^[6]。因此大壶状腺丝在材料科学、纺织工业、生物医学、军事、航空航天等领域具有潜在的应用前景和巨大的商业价值^[7]。

目前对蛛丝的研究热点集中在天然蛛丝形成过程及机理^[2]、编码蛛丝蛋白的基因及通过基因工程方法生产重组蛛丝蛋白,再把重组丝蛋白转变为人工丝纤维等几个方面。根据目前对蛛丝蛋白及其编码基因的研究结果表明^[8],蛛丝蛋白由N末端的非重复氨基酸序列、中间段的重复氨基酸序列和C末端的非重复氨基酸序列三段构成,根据中间段重复序列中的氨基酸组成和排列方式等特征,目前已把蛛丝蛋白分为大壶状腺丝蛋白(major ampullate spidroin, MaSp)、小壶状腺丝蛋白(minor ampullate spidroin, MiSp)、鞭状腺丝蛋白(flagelliform spidroin, Flag)、葡萄状腺丝蛋白(aciniform spidroin, AcSp)和管状腺丝蛋白(tubuliform spidroin, TuSp)等几类。可能因为MaSp基因主体由重复序列组成导致克隆长片段MaSp基因相当困难及不同生物类群之间遗传密码子偏爱性的不同,许多研究小组均采用人工合成MaSp基因类似物的途径在大肠杆菌^[9-14]、酵母^[15]或转基因植物^[16]中表达重组大壶状腺丝蛋白类似物。到目前为止,仅有两篇关于表达天然MaSp基因的报道,一篇是把络新妇(*Nephila clavipes*)的MaSp基因在大肠杆菌表达体系中表达^[17],另一篇是把络新妇及王冠园蛛(*Araneus diadematus*)的MaSp基因在哺乳动物细胞表达体系中表达^[18]。哺乳动物细胞表达体系中表达的重组MaSp即使在纯化过程形成沉淀,沉淀的重组MaSp仍能直接溶解于磷酸缓冲液,而在细菌或酵母等微生物表达体系中表达天

然MaSp基因或编码MaSp的人工基因时得到的重组MaSp一旦沉淀后,就不能直接溶解于磷酸缓冲液^[19]。由于在哺乳动物细胞表达体系中表达的重组MaSp中含有C末端非重复氨基酸序列,而以前在细菌或酵母等微生物表达体系中表达的重组MaSp中不包括C末端非重复氨基酸序列,Lazaris等^[18]推测可能是重组MaSp包含的C末端非重复氨基酸序列提高了重组MaSp在水媒介中的溶解性。因此本研究首次在细菌表达系统中表达带有C末端非重复氨基酸序列的重组MaSp,旨在进一步探讨C末端非重复氨基酸序列对重组MaSp在水媒介中溶解性的影响。

1 材料与方法

1.1 菌种、表达质粒和试剂盒

E. coli JM109菌株为本实验室自备;*E. coli* BL21(DE3)菌株、原核表达质粒pET28b(+)均购自Novagen公司;SMART™ RACE cDNA Amplification试剂盒购自Clontech公司。

1.2 MaSp基因cDNA片段的扩增

根据蜘蛛MaSp重复片段的部分氨基酸序列(CYGPCSG)设计和合成1条简并引物P-MaSp:5'-GGNTANGGNCNCGNTCNGG-3'(N=A, T, C or GT)。用作者制备的悦目金蛛丝腺SMART cDNA RACE文库为模板^[20],通过SMART™ RACE cDNA Amplification试剂盒提供的接头引物CDS III:5'-ATTCTAGAGGCCGAGGCCGCCGACATG-(dT)₁₇-3'和作者设计的简并引物P-MaSp进行PCR。

1.3 PCR产物的分子克隆和测序

使用DNA gel extraction试剂盒对PCR产物进行切胶纯化后,与PMD18-T克隆载体连接,再将连接液转化*E. coli* JM109感受态细胞,挑取阳性克隆。单克隆菌株经扩大培养后,提取并纯化质粒,用通用引物M13和M13 Reserve在LI-COR DNA测序仪上进行双向序列测定。阳性克隆命名为PMD18-T-MaSp。

1.4 表达质粒pET28b(+)-MaSp的构建

以PMD18-T-MaSp为模板,用Pfu DNA聚合酶扩增MaSp基因。扩增所用的上游引物5'-GCACGACC [ATC] GGAGGTTATGGACCC-3'包含1个Nco I识别位点(单下划线标注),1个翻译起始密码子(方框标注)和MaSp基因的前15个核苷酸序列;下游引物(5'-TAATGGATCC [TTA] GTGCGTGGTGGTGGTGA-CCAAGGGCCTGAGCA-3')包含1个6-His编码区段(波浪线标注),1个Bam HI识别位点(单下划线标注)。

注) MaSp 基因编码区最后 15 个核苷酸序列及一个翻译终止密码子(方框标注)。PCR 扩增产物和 pET28k(+) 分别用 *Nco* I 和 *Bam*H I 进行双酶切。酶切反应体系为: 10 × Buffer 5 μ L, 扩增产物或 pET28k(+) 20 μ L, *Nco* I 1 μ L, *Bam*H I 1 μ L, 补水至 50 μ L。酶切产物经 1% 的琼脂糖凝胶电泳后, 用 DNA gel extraction 试剂盒分别回收纯化目的片段和线性化的 pET28k(+), 并用 T4 DNA 连接酶将两者连接。连接反应体系为: 10 × Buffer 1 μ L, 线性化的 pET28k(+) 1 μ L, 目的片段 3 μ L, T4 DNA 连接酶 1 μ L, 补水至 10 μ L, 混匀, 16 $^{\circ}$ C 放置 12h。连接物转化 *E. coli* JM109 感受态细胞, 转化物涂在含 50 μ g/mL 卡那霉素的 LB 平板上生长。随机挑选抗卡那霉素的克隆并扩大培养, 提取的质粒通过 *Nco* I 和 *Bam*H I 双酶切方法筛选和鉴定。阳性克隆命名为 pET28k(+)-MaSp。

1.5 BL21(DE3)pET28k(+)-MaSp 表达菌株的筛选和鉴定

E. coli BL21(DE3) 的感受态细胞用常规 CaCl_2 法制备, 表达质粒 pET28k(+)-MaSp 用热激法转化 *E. coli* BL21(DE3), 转化物均匀地涂于含 50 μ g/mL 卡那霉素的 LB 平板上生长。挑取单克隆接种于 5mL 含 50 μ g/mL 卡那霉素的 LB 液体培养基中, 37 $^{\circ}$ C 摇床培养过夜。再把 5mL 的培养物接种于 50mL 含 50 μ g/mL 卡那霉素的 LB 液体培养基中, 并经 37 $^{\circ}$ C 摇床培养至 OD_{600} 为 0.4 时, 取出 1mL 培养物, 离心后沉淀用 100 μ L SDS 样品缓冲液(100mmol/L Tris-HCl, pH 6.8, 200mmol/L DTT, 4% SDS, 0.2% 溴酚蓝, 20% 甘油)处理后于 -20 $^{\circ}$ C 保存; 然后向余下的 49mL 的培养物中加入 IPTG 至终浓度为 0.05mmol/L, 37 $^{\circ}$ C 摇床培养, 并分别在培养至 1h、2h、3h、4h、5h 时各取出 1mL 培养物, 离心后沉淀用 100 μ L SDS 样品缓冲液处理后 -20 $^{\circ}$ C 保存。培养至 5h 后, 把 44mL 培养物离心(2500r/min, 10min)收集菌体沉淀并冻存于 -20 $^{\circ}$ C 备用。用 5mL 50mmol/L Tris-HCl 缓冲液(pH 8.0, 内含 1mmol/L EDTA)重悬菌体沉淀, 并用超声波破碎仪处理菌体悬浊液(4 $^{\circ}$ C, 30W, 总处理时间 30min, 5s 超声/15s 停顿)。超声处理后的溶液高速离心(4 $^{\circ}$ C, 15000r/min, 20min), 上清液和沉淀分别保存备用。诱导表达过程中不同时间段所取的细菌总蛋白样品和细菌破碎后的上清液、细菌破碎后的沉淀一起通过 SDS-PAGE 进行检测。

1.6 重组 MaSp 的大量表达与初步纯化

把含有已经证明能表达重组 MaSp 的 BL21(DE3)pET28k(+)-MaSp 单克隆菌株的培养物接种

于 500mL 含 50 μ g/mL 卡那霉素的 LB 液体培养基中, 并经 37 $^{\circ}$ C 摇床培养至 OD_{600} 为 0.4 时, 向培养物中加入 IPTG 至终浓度为 0.05mmol/L, 37 $^{\circ}$ C 摇床培养, 培养至 5h 后, 离心(2500r/min, 10min)收集菌体沉淀并冻存于 -20 $^{\circ}$ C 备用。用 Ni-NTA His-Bind 树脂层析柱分离纯化重组蛋白, 收集洗脱液并进行 SDS-PAGE 分析, 洗脱液用去离子水透析 24h。

1.7 保持重组 MaSp 溶解性稳定的缓冲液体系的筛选

洗脱液用去离子水透析 24h 后, 重组丝蛋白全部沉淀, 沉淀的重组丝蛋白不能直接溶解于去离子水或磷酸缓冲液。用适量的 6mol/L 盐酸胍溶液溶解重组丝蛋白沉淀, 参照 Fahnestock 等^[9]的方法用不同 pH 值(1~14 梯度)的磷酸缓冲液透析过夜, 观察重组丝蛋白在不同 pH 值条件下的溶解性。

1.8 用分子筛技术对重组 MaSp 的二次纯化

分子筛纯化在中低压层析系统(Biologic LP liquid chromatograph, Bio-Rad)上进行。把凝胶填料 Sephadex G-100(Pharmacia)装入 2.0 cm × 80.0 cm 规格的层析柱后, 用 pH 值为 4.0 的磷酸缓冲液(1.7 步骤中筛选结果表明重组 MaSp 在 pH 值为 3~5 范围内的磷酸缓冲液中溶解性最为稳定, pH 4.0 磷酸缓冲液: 1% (W/V) NaCl, 0.075% (W/V) KCl, 0.14% (W/V) Na_2HPO_4 , 0.0125% (W/V) KH_2PO_4)平衡层析柱。每次纯化时上样量大约为含 50mg 重组蛋白的透析液 1~2mL, 洗脱液也为 pH 4.0 的磷酸缓冲液。

1.9 重组 MaSp 的鉴定

1.9.1 氨基酸组成测定 将约 1mg 经分子筛纯化的重组 MaSp 样品放入 6mol/L HCl 溶液中, 在 110 $^{\circ}$ C 下水解 24h, 在日立 835-50 型氨基酸分析仪上测定其氨基酸组成。

1.9.2 N 末端氨基酸序列测定 首先将经分子筛纯化的重组 MaSp 进行 SDS-PAGE 电泳, 然后电转移至 PVDF 膜上。靶蛋白所在位置的斑点被小心剪下, 委托中国协和医科大学基础医学院中心实验室在蛋白质测序系统 Procise™ 491(PE Applied Biosystems, USA)上测定 N 末端 10 个氨基酸残基的序列。

2 结果

2.1 MaSp 基因的克隆

用接头引物和 P-MaSp 进行的 RT-PCR 获得了几个条带, 这些条带分别经切胶纯化并进行克隆、DNA 序列测定后证明其中只有 1 个条带为目的条带(图 1)。从这个目的条带的克隆库中随机挑取 30

个单克隆并分别测序,筛选出了1条MaSp基因cDNA序列,该序列提交GenBank,获得的登录号为AY365017。这条cDNA序列编码的氨基酸序列可区分为两部分(图2)(1)N端富含丙氨酸的片段和富含甘氨酸的片段相间排列而构成重复序列区,并且富含甘氨酸的片段中有脯氨酸分布(2)约100个氨基酸残基组成的C末端非重复氨基酸序列。根据其编码的氨基酸序列特征可以初步判断本研究获得的这条cDNA序列与已发表的其他蜘蛛的大壶状腺丝蛋白-2(major ampullatespidroin-2, MaSp2)基因同源,特别是C末端非重复氨基酸序列与其他蜘蛛的同源片段氨基酸序列比对后显示一定的保守性(图3)。另外,悦目金蛛MaSp基因cDNA序列在密码子的使用上偏爱性非常明显,有78.8%的密码子第三个碱基为A或T。

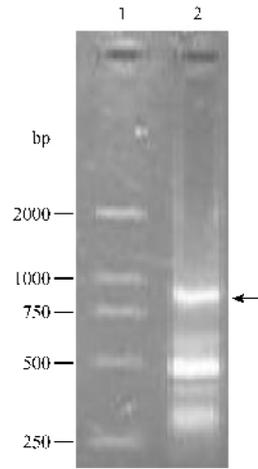


图1 RT-PCR产物琼脂糖凝胶电泳

Fig.1 Agarose gel electrophoresis of RT-PCR products
1: DL2000 marker; 2: PCR products (Arrow points to target band).

```

CGA GGT TAT GGA CCC GGA TCT GGA CAA CAA GGA GGA GGC GGT CAA CAA GGA CCT GGA GGT CTA GGA CCA TAT GGA CCT GGT GCA 90
G G Y G P G S G Q Q G P G G A G Q Q G P G G I, G P Y G P G A 30
GCA GCA CCA GCC GCA GGT AGC TAT GGA CCA AGA GCA GGC CAA CAG AGA CCA GGA GCT GGT CAA CAA GGA CCT GCA GCT GGA CAA CAA GGT 180
A A A A A G S Y G P R A G Q Q R P G A G Q Q G P G A G Q Q G 60
CCA GGT GGA GCA GGC CAA CGA GGA CCG TAC GGA CCC GGG GCA GCA GGC GCA GGC GCT GGC GGC TAT GGA CAA CAA GGA CCT GGA 270
P G G A G Q R G P Y G P G A A A A A A A A G G Y G Q Q G P G 90
GCT GGA CAA CAA GGA CCA GGA GTA GGA CAA CAA GGT CCA GGT GGA GCA GGC CAA CGA GGA CCA TAC GCA CCC GGG GCA GCA GGC GCA GCT 360
A G Q Q G P G V G Q Q G P G G A G Q R G P Y G P G A A A A A 120
GCA GCC GCT GGT GGC TAT GGA CCA AGA TAT GGA CAA CAA GGA CCT GGA CCA GGA CCA TAT GGC CCA GGA CCA GGA GCC ACA GCT GCT GCC 450
A A A G G Y G P R Y G Q Q G P G A G P Y G P G A G A T A A A 150
GCT GGA GGA TAT GGA CCA GGA GCT GGA CAG CAA GGA CCA AGA AGT CAG GCT CCA GTC GCA TCC GCT GCA GCC GCT GCA CTT TCT TCT CCT 540
A G G Y G P G A G Q Q G P R S Q A P V A S A A A A R L S S P 180
CAA GCC GCT TCT AGA CTT TCA TCT GCT GTT TCA ACT TTG GTA TCG AGT GGA CCT ACT AAT CCT GCA TCA CTT TCT AAT GCT ATC GGT AGC 630
Q A G S R V S S A V S T L V S S G P T N P A S L S N A T G S 210
GTT GTT TCG CAG GTT AGT GCA AGC AAT CCC GGC CTT CTT AGT TGC GAT GTT CTT GTG CAA GCA TTA CTG GAA ATC GTA TCC GCC CTC GTA 720
Y V S Q V S A S N P G L P S C D V L V Q A L L E I V S A L V 240
CAT ATC CTT GGA TCT TCC AGC ATT GGC CAA ATT AAC TAT TCC GCT TCT TCT CAG TAT GCC CGA TTA GTT GGC CAA TCT ATT GCT GAG GCC 810
H I L G S S S I G Q I N Y S A S S Q Y A R L V G Q S I A Q A 270
CTT GGT TAA TTT CAT AAA TGA TTT CAA TTT TTC TAT TAA TGT AAT GAC AGA TTT GTA ATC TTT CTG AAT AAA TTT TGA AGC ATA TTT AAA 900
L G 300
AAA (polyA)
    
```

图2 悦目金蛛 MaSp 基因 cDNA 序列及推导的氨基酸序列

Fig.2 cDNA sequence of *Argiope amoena* MaSp gene and deduced amino acid sequence
The Ala-rich sequences in the repetitive region of MaSp were underlined; poly(A) signal site and poly(A) were indicated by drawing double lines under them.

<i>Argiope amoena</i>	RLSSPQAGSRVSSAVSTLVSSGPTNPASLSNALGSVVSQVSASNPGLPS 49
<i>Latrodectus hesperus</i>	ALSSPTTHARTSSHASTLISSGPTNAAALSNVTSNAVVSQVSASNPSSS
<i>Nephila clavipes</i>	RLASPDGSRVASAVSNLVSSGPTSSAALSSVTSNAVVSQVSASNPGLSG
<i>Araneus diadematus</i>	RLSSPAASSRVSSAVSSLVSSGPTKHAALSNTSSVVSQVSASNPGLSG
<i>Uloborus diversus</i>	RLNSPASTSRVASAVSSLASAGAPSVGSLSSVLSLSSSVSASNPGLSG
<i>Argiope amoena</i>	CDVLVQALLEIVSALVHLLGSSSIGQINYSASSQYARLVGQ-SIAQALG 98
<i>Latrodectus hesperus</i>	CDVLVQALLEITTALTSTLDSSVGVVNYGSSGQYAVGQ-SMQQAMG
<i>Nephila clavipes</i>	CDVLVQALLEIVSACVTLLSSSIGQVNYGAASQYAVVGVQ-SVLSAF-
<i>Araneus diadematus</i>	CDVLVQALLEVVVALVSTLLGSSSTGQTNYGASQYAVVGVQ-SVAQALA
<i>Uloborus diversus</i>	CELLVQVLLLEVVALVALLGSAVGVVVDVSSSQYAGLVSS-ALVQAL-

图3 悦目金蛛与其他4种蜘蛛 MaSp C末端非重复氨基酸的同源比较

Fig.3 Homology analysis of C-terminal non-repetitive region of MaSp between *Argiope amoena* and other four species of spiders
Amino acid sequences of MaSp were from GenBank (*Latrodectus Hesperus*: DQ409058; *Nephila clavipes*: AY654293; *Araneus diadematus*: U47855; *Uloborus diversus*: DQ399334).

2.2 MaSp 基因原核表达载体的构建

通过 PCR 方法在 MaSp 基因编码区的侧翼加上两个内切酶位点 *Nco* I 和 *Bam* H I、6-His 编码区段和翻译终止密码子后,再亚克隆进原核表达质粒 pET28I(+)(图 4)。

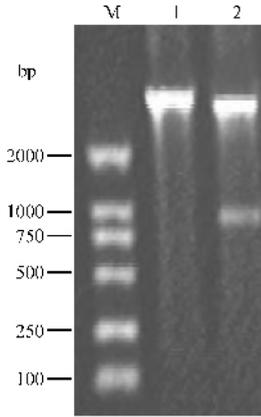


图 4 重组表达质粒的酶切分析

Fig.4 Restriction analysis of recombinant vector pET28I(+)-MaSp

M:DL2000 marker; 1:pET28I(+)-MaSp; 2:pET28I(+)-MaSp/*Nco* I + *Bam* H I.

2.3 重组 MaSp 的表达及纯化

IPTG 诱导表达的结果表明(图 5)重组 MaSp 在细菌细胞中主要以可溶性形式存在,而少部分形成

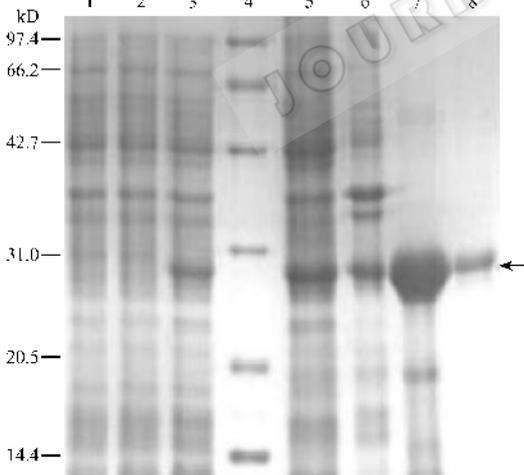


图 5 悦目金蛛 MaSp 诱导表达及纯化的 SDS-PAGE 分析

Fig.5 SDS-PAGE analysis of recombinant MaSp expression and purification

Arrow: pointing to recombinant MaSp; lane 1: the whole cell protein of BL21(DE3) pET28I(+), for 5 h following induction of IPTG; lane 2: the whole cell protein of BL21(DE3) pET28b(+)-MaSp for 0 h following induction of IPTG; lane 3: the whole cell protein of BL21(DE3) pET28I(+)-MaSp for 5 h following induction of IPTG; lane 4: molecule weight marker of protein; lane 5: the soluble fraction of BL21(DE3) pET28I(+)-MaSp for 5 h following induction of IPTG after sonic treatment; lane 6: the insoluble fraction of BL21(DE3) pET28b(+)-MaSp for 5 h following induction of IPTG after sonic treatment; lane 7: recombinant protein purified with Ni-NTA His-bind resin column; lane 8: recombinant protein purified with gel filtration.

包涵体(图 5:lane 5 和 lane 6)。SDS-PAGE 分析结果表明重组 MaSp 的分子量与通过 MaSp 基因 cDNA 序列推导的氨基酸序列预期分子量 25.53kD 基本一致(图 5:lane 3 和 lane 8)。重组 MaSp 经镍离子亲和层析初步纯化后发现还有少量杂蛋白条带(图 5:lane 7)所以再用分子筛进行二次纯化(图 5:lane 8),二次纯化后的产率大约为 40mg/L。

2.4 重组 MaSp 的鉴定

经二次纯化后的重组 MaSp 氨基酸组分和含量的测定结果与通过 MaSp 基因 cDNA 序列推导的氨基酸的组分和含量较为吻合(表 1)。重组蛋白的 N 端 10 个氨基酸残基序列的测定结果为 GGYGPGSGQQ 与根据 cDNA 序列推导的 MaSp 的 N 端前 10 个氨基酸残基序列相同(图 2),这一结果直接证实了本研究得到的重组蛋白为重组 MaSp。

表 1 重组 MaSp 氨基酸组成分析

Table 1 Amino acid composition of recombinant MaSp

Amino acid	Observed/%	Predicted/%
G	15.2	16.4
A	14.8	15.4
S	7.7	9.2
Y	7.0	7.3
N+D	4.1	2.3
E+Q	14.1	15.1
H	3.9	3.9
R	6.0	5.3
T	2.0	1.2
P	8.2	10.6
V	5.4	5.0
M	0.0	0.5
I	3.1	2.6
L	6.5	5.2
F	0.0	0.0
K	1.0	0.0
C	1.1	0.4

3 讨论

从悦目金蛛 MaSp 基因及推导的氨基酸序列来看(图 2),至少有三个原因使得 MaSp 基因在细菌表达系统中表达存在一定的困难和风险。其一, MaSp 氨基酸序列中甘氨酸和丙氨酸比例过大(蜘蛛基因组中存在大量成串编码 tRNA^{Ala} 的基因^[21]),而大肠杆菌中如没有足够的 tRNA^{Ala} 和 tRNA^{Gly},则蛋白质的翻译可能提前终止;其次, MaSp 基因主要由重复序列构成,因此其转录的 mRNA 可能形成复杂的二级结构从而阻碍蛋白质翻译的正常进行;第三,蜘蛛细胞与大肠杆菌之间密码子偏爱性的差异也可能使得蛋白质表达难以进行。本研究在做预试验时首先采

用原核表达载体 pQE30(具有 T5 启动子)在大肠杆菌 M15 菌株中表达 MaSp 基因,结果表明这个表达体系不能表达重组 MaSp,转而采用具有强启动子的原核表达载体 pET28b(具有 T7 启动子)在大肠杆菌 BL21(DE3)菌株中成功表达了重组 MaSp。看来,对于像 MaSp 基因这种特殊的基因,选用具有强转录启动子的原核表达载体是成功表达的关键。

目前普遍把 6 × His 设计在重组蛋白序列的 N 或 C 端,然再用镍离子亲和层析柱来纯化重组蛋白,这种方法显著地提高了纯化效率,许多文献称这种方法能“一步法纯化”重组蛋白^[22]。本研究把 6 × His 设计在重组 MaSp 蛋白的 C 末端,在纯化过程中探索和优化了亲和层析的条件,但经镍离子亲和层析柱的纯化产物仍有少量杂蛋白条带(图 5, lane 7)。看来,要想得到高纯度的靶蛋白用于如蛋白质结构分析等要求较高的试验,最好采用其他方法如分子筛方法对镍离子亲和层析柱纯化的重组蛋白进行二次纯化。

为探讨 C 末端非重复氨基酸序列对重组大壶状腺丝蛋白溶解性的影响,本研究首次在细菌中表达了带有 C 末端非重复氨基酸序列的重组 MaSp。在用洗脱液从镍离子亲和层析柱上收集重组 MaSp 时,从层析柱流出的带有靶蛋白的洗脱液通常为牛奶状液体,洗脱液用去离子水透析 24h 后,重组蛋白全部沉淀,沉淀的重组蛋白不能直接溶解于去离子水或磷酸缓冲液等水媒介。用盐酸胍溶液溶解重组 MaSp 沉淀,再用不同 pH 值的磷酸缓冲液透析的结果表明,重组丝蛋白在 pH 值 3 ~ 5 范围内的磷酸缓冲液中溶解性最为稳定。本研究表达的带有 C 末端非重复氨基酸序列的重组 MaSp 与其他作者在细菌中表达的、不带有 C 末端非重复氨基酸序列的重组 MaSp 相比^[19],其在水媒介中的溶解性并未改善,至少说明 C 末端非重复氨基酸序列的有或无不是重组 MaSp 在水媒介中溶解性的决定因素。

REFERENCES(参考文献)

[1] Denny MW. Silks-Their Properties and Functions. The Mechanical Properties of Biological Materials. Cambridge :Cambridge University Press ,1980.

[2] Vollrath F, Knight DP. Liquid crystalline spinning of spider silk. *Nature*, 2001, **410** :541 - 548.

[3] Craig CL. Evolution of arthropod silks. *Annu Rev Entomol*, 1997, **42** :231 - 267.

[4] Gosline JM, Guerette PA, Ortlepp CS, et al. The mechanical design of spider silks : from fibroin sequences to mechanical

function. *The Journal of Experimental Biology*, 1999, **202** :3295 - 3303.

[5] Work RW, Young CT. The amino acid compositions of major and minor ampullate silks of certain orb-web-building spiders (Araneae, Araneidae). *J Arachnol*, 1987, **15** :65 - 80.

[6] Tirrell DA. Putting a new spin on spider silk. *Science*, 1996, **271** :39 - 40.

[7] Winkler S, Kaplan DL. Molecular biology of spider silk. *Molecular Biotechnology*, 2000, **74** :85 - 93.

[8] Pan HC, Song DX, Zhou KY. Progress of spider silk protein. *Acta Arachnologica Sinica*, 2006, **15** (1) :52 - 59.

[9] Prince JT, McGrath KP, DiGirolamo CM, et al. Construction, cloning, and expression of synthetic genes encoding spider dragline silk. *Biochemistry*, 1995, **34** :10879 - 10885.

[10] Lewis RV, Hinman M, Kothakota S, et al. Expression and purification of a spider silk protein : a new strategy for producing repetitive proteins. *Protein Exp Purific*, 1996, **7** :400 - 406.

[11] Fahnestock SR, Irwin SL. Synthetic spider dragline silk proteins and their production in *Escherichia coli*. *Appl Microbiol Biotechnol*, 1997, **47** :23 - 32.

[12] Fukushima Y. Genetically engineered syntheses of tandem repetitive polypeptides consisting of glycine-rich sequence of spider dragline silk. *Biopolymers*, 1998, **43** (4) :269 - 279.

[13] Winkler S, Szeza S, Avtges P, et al. Designing recombinant spider silk proteins to control assembly. *Int J Biol Macromol*, 1999, **24** (2/3) :265 - 270.

[14] Li M(李敏), Zhang WX(章文贤), Huang ZH(黄智华), et al. Study on construct and expression of synthetic genes encoding spider dragline silk in *Escherichia coli*. *Chinese Journal of Biotechnology* (生物工程学报), 2002, **18** (3) :331 - 334.

[15] Fahnestock SR, Bedzyk LA. Production of synthetic spider dragline silk protein in *Pichia pastoris*. *Appl Microbiol Biotechnol*, 1997, **47** :33 - 39.

[16] Scheller J, Gührs KH, Grosse F, Conrad U. Production of spider silk proteins in tobacco and potato. *Nature Biotechnology*, 2001, **19** :573 - 577.

[17] Arcidiacono S, Mello C, Kaplan D, et al. Purification and characterization of recombinant spider silk expressed in *Escherichia coli*. *Appl Microbiol Biotechnol*, 1998, **49** :31 - 38.

[18] Lazaris A, Arcidiacono S, Huang Y, et al. Spider silk fibers spun from soluble recombinant silk produced in mammalian cells. *Science*, 2002, **295** :472 - 476.

[19] Fahnestock SR, Yao ZJ, Bedzyk LA. Microbial production of spider silk proteins. *Reviews in Molecular Biotechnology*, 2000, **74** :105 - 119.

[20] Pan HC, Song DX, Zhou KY. Construction and characterization of SMART RACE cDNA library of silk gland in *Argiope amoena*. *Acta Arachnologica Sinica*, 2005, **14** (2) :65 - 71.

[21] Luciano E, Candelas GC. An alanine tRNA gene cluster from *Nephila clavipes*. *Gene*, 1996, **171** :301 - 302.

[22] Chu XS, Li D. Cloning, expressing, and purification of His-tagged rat mevalonate kinase. *Protein Expression and Purification*, 2003, **27** :165 - 170.