

蜘蛛拖丝蛋白基因的构建及在大肠杆菌中的表达

李 敏* 章文贤 黄智华 黄建坤

(福建师范大学生物工程学院,福州 350007)

摘 要 蜘蛛大壶腹线产生的拖丝是非常优良的纤维蛋白,具有独特的强度和弹性。基于拖丝蛋白高度重复序列和部分 cDNA 序列,合成蜘蛛拖丝蛋白基因单体,通过头尾相连的构建策略,得到拖丝蛋白多聚体,与原核高效表达载体 pET30a(+)连接,转化大肠杆菌 BLR(DE3),用 IPTG 诱导表达。表达产物经 His. Bind 树脂金属螯合亲和层析一步纯化,纯度达 90% 以上,表达量为 20mg/L。SDS-PAGE 和蛋白质印迹图谱显示表达产物分子量为 37kD,其值与氨基酸组分分析结果与理论推算值基本符合。

关键词 蜘蛛,拖丝,基因合成,表达

中图分类号 Q78 文献标识码 A 文章编号 1000-3061(2002)03-0331-04

蜘蛛丝光滑、闪亮,耐紫外线,耐高温和低温,具有特殊的溶解性,仅溶于浓硫酸、溴化锂(9.3mmol/L)甲酸(88%),并且对大部分蛋白水解酶具有抗性^[1]。蜘蛛不同的丝腺所产生的丝的功能不尽相同,其中以棒络新妇蛛(*Nephila clavipes*)大壶腹线产生的拖丝所表现出的显著柔韧性、抗疲劳和与钢材相似的张力强度,以及在 200℃ 高温的热稳定性,使蛛丝成为自然界力学性能优良的天然蛋白质纤维,其所具有的强固性和柔韧性是其它纤维材料无法同时比拟的^[2,3]。由于拖丝的特殊性质,作为生物材料应用在生物医学工程中作缝合材料、细胞膜以及组织工程所需的临时搭架等^[4],在军事上可制防弹衣和特殊性能的降落伞。

仿制蛛丝一直是纤维工业孜孜不倦的奋斗目标。多年来,美国、德国等国家的学者对蜘蛛丝的化学组成、分子结构及聚集态结构研究相当深入。蜘蛛拖丝蛋白(以下统称丝蛋白)为高度重复序列的纤维蛋白质,具有小侧链的丙氨酸和甘氨酸的重复分别达到 25% 和 42%,组成丝蛋白的结构单元主要有(Ala)₁₀, Gly-Gly-X(X 为 Ser, Gly, Tyr)和 Gly-Pro-Gly-X-X 等^[5]几种类型。由于蜘蛛人工饲养的不可能性,蛛丝的获取十分困难。直到 20 世纪 90 年代初,通过基因测序方法得到蜘蛛拖丝基因的部分序列以后^[6],研究者利用拖丝蛋白的部分 cDNA 片

段和人工合成拖丝蛋白部分基因,分别在大肠杆菌、毕赤酵母等系统中表达^[7-10],表达产物与天然丝蛋白具有相似性,但表达产率较低。最近 Scheller 等人^[11]将人工合成的拖丝蛋白基因转入土豆和烟草,得到占植物总蛋白量 2% 的丝蛋白。

这里我们报道通过人工设计合成特殊的蜘蛛拖丝蛋白基因序列,克隆至表达载体 pET-30a(+)中,使其在 T7 启动子控制下,经 IPTG 诱导,在大肠杆菌细胞高效表达。

1 材料与方法

1.1 试剂

T4 DNA 连接酶、T4 多核苷酸激酶和 CIP 为 Promega 产品,各种限制酶和卡那霉素为北京华美产品,蛋白质分子量标准、S. Tag AP Western Blot Kit 和 His. Bind 树脂为 Novagen 产品,PMSF 为 Sigma 产品,BCA 蛋白分析试剂为 Pierce 产品,其余化学试剂为国产分析纯。

1.2 菌株和载体

宿主菌 *E. coli* DH 5 α 、BLR(DE3)和表达载体 pET30a(+)购自 Novagen,克隆载体 pCR script(Cam^r)和 pUC19(Amp^r)为本室保存。

1.3 穿梭载体 pUC-LINK 和 pCR-LINK 的构建

Fig. 1A 为合成的一段衔接子序列,其两端 *Xba*I

位点插入到 pUC19 和 pCR script 上形成 pUC-LINK 和 pCR-LINK。本文中质粒 DNA 提取、限制性酶解、磷酸化、去磷酸化、连接、转化、电泳等实验方法参见分子克隆手册。

1.4 蛛丝蛋白基因单体 DNA 的合成与构建

蜘蛛拖丝蛋白多肽片段及 DNA 序列设计见图 1B, 单体命名为 DSI, 分别在 5'-端和 3'-端加上 *NheI* 和 *SpeI* 位点。两条多核苷酸链分 4 段合成, 寡聚核苷酸链经 5'-末端磷酸化, 互补退火, 连接形成基因单体。合成的丝蛋白单体分别与 pUC-LINK 和 pCR-LINK 的 *NheI/SpeI* 位点连接, 转化 DH5 α , 抗性筛选重组子。与 pUC-LINK 相连接的重组子称为 pUC-AMP, 与 pCR-LINK 相连接的重组子称为 pCR-CAM。

1.5 蛛丝蛋白基因多聚体的构建

pUC-AMP 经 *SphI* 酶解和 CIP 处理形成含有丝蛋白基因插入的线性片段, 称之为“受体”。含有插入子的 pCR-CAM 经 *NheI* 和 *SpeI* 双酶解, 作为“供体”。二者分别经酚/氯仿纯化后, 混合, 连接, 转化 DH5 α , 抗性筛选(Amp^r)。形成的二聚体在此基础上重复上述过程, 由此得到多聚体重组子。此蛛丝蛋白基因多聚体与 pET30a(+) 分别经 *BamH I* 酶解, 连接, 转化 BLR(DE3), 重组子命名为 pNS2。

1.6 重组蛋白的表达和纯化

挑取 pNS2 单菌落培养过夜, 按 2% 比例接种于 500mL LB 培养基(含卡那霉素 30 μ g/mL), 37 $^{\circ}$ C 培养至菌液的 A₆₀₀ 达到 1.5~2.0 时, 加 IPTG 至终浓度为 1 mmol/L, 诱导表达 4~5h。离心收集菌体, 重悬结合缓冲液(20mmol/L Tris-HCl/ 5 mmol/L 咪唑/ 500 mmol/L NaCl/ 1mmol/L PMSF/ 1% Triton-X100, pH8.0)中, 超声波破菌。破菌液于 4 $^{\circ}$ C, 15 000r/min 离心 15 min, 收集上清液。取适量 His. Bind 树脂, 预先用结合缓冲液平衡, 上清液经 0.45 μ m 滤膜过滤后上样, 用 10 倍柱床体积的洗涤缓冲液(20mmol/L Tris-HCl/20mmol/L 咪唑 /500mmol/L NaCl /1mmol/L PMSF, pH8.0) 洗去杂蛋白。用洗脱缓冲液(20mmol/L Tris-HCl/500mmol/L 咪唑 /500mmol/L NaCl /1mmol/L PMSF, pH8.0) 分部收集, 对水透析, 浓缩, 冻干。BCA 蛋白分析试剂测定蛋白质浓度。

1.7 蛋白质印迹分析

蛋白质印迹实验采用 S. Tag 试剂, 蛋白质样品经 SDS-PAGE 后电转移到 PVDF 膜上, 按 Novagen 提供的方法操作, 抗 S-Tag 多肽抗体与碱性磷酸酶组成偶联物, BCIP/NBT 为底物, 直接呈色反应鉴定。

1.8 氨基酸组成分析

氨基酸组成分析使用 Waters Pico-Tag 系统。蛋白质在 6 mol/L HCl, 150 $^{\circ}$ C 条件下水解 85 min, 冻干, PITC 衍生, HPLC 分离并分析。

2 结 果

2.1 蜘蛛拖丝蛋白基因单体和多聚体的构建

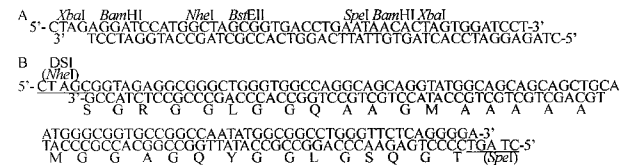


图 1 合成 pUC-LINK 和 pCR-LINK 的衔接子和蜘蛛拖丝蛋白单体 DNA 序列

Fig.1 DNA sequences of synthetic spider silk monomer and pUC-LINK and pCR-LINK adapter

- A. DNA sequence of synthetic adapter inserted into pUC19 and pCR-Script to create pUC-LINK and pCR-LINK
B. DNA sequence of synthetic spider silk monomer

Fig. 1A 为合成的一段寡核苷酸, 有多个酶切位点, 作为衔接子插入到载体 pUC19 和 pCR script 的 *XbaI* 位点上, 形成 pUC-LINK 和 pCR-LINK 穿梭载体, 由此增加两个特殊的酶切位点: *NheI* 和 *SpeI* 为多聚体构建所需。Fig. 1B 为设计的拖丝蛋白单体 DNA 序列, 由 99bp 组成, 编码 33 个氨基酸, 在其 5'-端和 3'-端分别加上 *NheI* 和 *SpeI* 限制酶位点。利用 *NheI* 和 *SpeI* 切割 DNA 后形成可匹配的“CTAG”粘性末端的特点, 以头尾相连的方式连接, 所形成的“杂合位点”不再被这两种酶切割(图 2)。

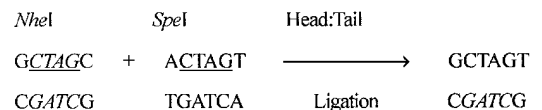


图 2 *NheI* 和 *SpeI* 头尾相连形成杂合位点

Fig.2 Hybrid formation via the head to tail ligation of *NheI* and *SpeI*

对每一次头尾连接的结果, 分别通过 *BamH I* 酶解和 *NheI/SpeI* 双酶解, 电泳鉴定片段长度和插入方向性。在本实验中, 最终所得到的多聚体片段为 800bp 左右。克隆在 pUC19 载体上的多聚体片段用 *BamH I* 酶切, 低熔点琼脂糖回收与经相同酶切的载体 pET30a(+) 连接, 其 5' 端与表达载体上编码 6 \times His 和 S. Tag 的基因相接, 保持阅读框架不变, 构建表达质粒 pNS2。

2.2 蜘蛛拖丝蛋白基因在 *E. coli* 中的表达、纯化和鉴定

从 SDS-PAGE 结果(图 3) 可看到, 破菌上清液

通过 His. Bind 金属螯合柱层析后,含 $6 \times \text{His}$ 编码基因的融合蛋白结合到柱上。从标准曲线可见表达产物的质量在 37kD 左右。为进一步证实目的蛋白得到正确的表达,我们对重组蛋白进行 Western 印迹分析,结果显示表达产物有特异性条带(图 4),重组丝蛋白表达水平为 20mg/L。

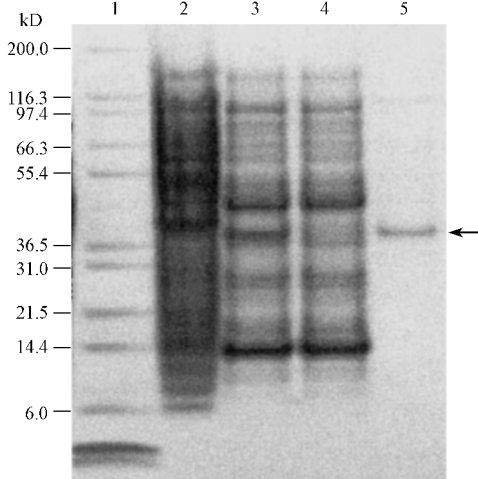


图 3 SDS-聚丙烯酰胺凝胶电泳检测重组蜘蛛拖丝蛋白的表达

Fig.3 SDS-PAGE analysis of the expression of recombinant dragline silk of spider

1. Protein molecular weight markers ; 2. The whole expressed protein by heating ; 3. Supernatant before His. Bind resin ; 4. Flow through after loading to His. Bind resin ; 5. Purified protein

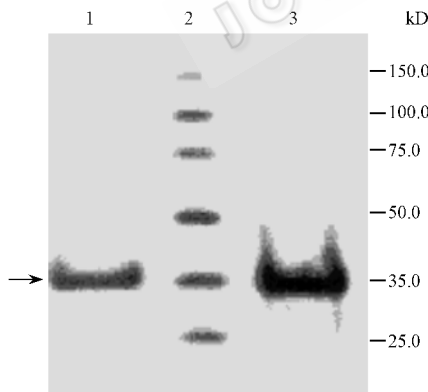


图 4 Western blot 检测表达产物的特异性

Fig.4 Western Blot analysis of the specificity of the expression product

1. Purified product ; 2. Protein molecular weight markers ; 3. Supernatant before His. Bind resin

pET30a(+) N-端含 49 个氨基酸, C-端有 24 个氨基酸, 多聚体 DNA 片段长度约 800bp, 编码的氨基酸为 267 个, 加上载体的氨基酸, 约 340 个氨基酸。从理论上计算, 重组表达产物的分子量约

37kD, 实验结果与理论推算值基本吻合。亲和层析纯化表达产物的纯度达 90% 以上, 除了主要带出现外, 在 SDS-PAGE 和 S. Tag Western 印迹实验中也观察到不甚明显的呈梯度排列的较小分子量带, 这种情况与有高度重复序列的天然蛋白质很相似。

2.3 重组丝蛋白氨基酸组分分析

重组丝蛋白表达产物的氨基酸组分分析的结果基本上与理论值相吻合(表 1)。甘氨酸和丙氨酸占总氨基酸 50% 以上。

表 1 丝蛋白氨基酸组成(仅列主要氨基酸)

Table 1 Amino acid composition of silk protein (only the major amino acids are listed)

Amino acid	Expected/%	Observed/%
Glycine	32.3	32.1
Alanine	23.5	22.7
Serine	6.2	7.2
Methionine	5.9	6.5
Arginine	4.2	3.6
Gln	7.4	7.2

3 讨论

蛋白质纤维有不同的二级结构, 包括 α -螺旋, β -转角, β -片层等, 蜘蛛拖丝以具有反向平行的 β -片层结构为主。多聚丙氨酸能形成 β -片层, 占蜘蛛丝总晶体结构的 30%, 甘氨酸富集区形成 α -螺旋。丝蛋白主要由有序的结晶区(小分子的侧链为主, 多为 β -片层)和无序的非结晶区(大分子侧链为主)交替排列, 这种结构方式可使丝承受的张力并不直接落在多肽链的共价键上, 从而赋予蛛丝很好的抗张强度、较高的强力和韧性^[12]。

Xu 等从 *N. clavipes* 拖丝丝蛋白部分 cDNA 序列中, 分别得到两种不同的 cDNA, 命名其编码的蛋白质为 Spidroin I 和 Spidroin II^[6, 43]。尽管所得到的两者的 cDNA 序列都是不完全的, 但其编码的蛋白质序列在天然的蛛丝蛋白中都大量重复存在。Spidroin I 和 II 的重复序列都富含甘氨酸和谷氨酰胺, 但 Spidroin II 在富含脯氨酸区还具有 β -转角连接, 并与由多聚丙氨酸片段组成的 β -片层相交替。Spidroin I 和 II 的结构为拖丝的结构和功能的研究提供了初始模型。

我们设计并合成的 33 肽蜘蛛拖丝蛋白基因单体, 比较了 Spidroin I 和 *N. clavipes* 天然拖丝蛋白部分 cDNA 序列^[8], 在中间段加入 5 个多聚丙氨酸, 并在多聚丙氨酸两端加上蛋氨酸, 期望通过恰当的

空间构象,增加蛋白质的溶解性,以利于表达产物的分离纯化。从文献中了解到,拖丝蛋白丙氨酸是比较特殊的,其编码规律常常以 GCA 开始,而以 GCT 结束,偶而也出现 GCC 或 GCG。此外,天然拖丝蛋白基因编码有高度的选择性,绝大多数氨基酸的密码子的第 3 位很少甚至不用 C 或 G。因此,在密码子的选择上既要考虑拖丝蛋白的编码特点,又要注意大肠杆菌对密码子的偏爱性。由于拖丝蛋白基因高 G + C 含量,往往导致在原核表达系统中表达量偏低,我们在设计上考虑了除甘氨酸和丙氨酸以外,对其它几种较常出现于拖丝蛋白重复序列的氨基酸,尽量增加 A/T 的比例。实验中所得到的拖丝蛋白多聚体为八聚体,我们也曾得到高于八聚体的其它多聚体形式,但表达产物量偏低(2 ~ 5mg/L),表达产物呈梯度降解也较为明显,分析认为这是大分子质粒和高分子多聚物在大肠杆菌中不稳定所致。由于高分子量蛋白质的获得是空间结构研究的需要,我们还要进一步深入研究高分子量表达产物的稳定性。

REFERENCES (参考文献)

- [1] Tillinghast E K, Chase S F, Townley M A. Water extraction by the major ampullate duct during silk formation in the spider, *Argiope aurantia* Lucas. *J Insect Physiol*, 1984, **30** :591 ~ 596
- [2] Vollrath F. Biology of spider silk. *Int J Biol Macromol*, 1999, **24** :

81 ~ 88

- [3] Cunniff P M, Fossey S A, Auerbach M A *et al.* Mechanical and thermal properties of dragline silk from the spider *Nephila clavipes*. *Polym Adv Technol*, 1995, **5** :401 ~ 410
- [4] Baldwin S P, Saltzman W M. Materials for protein delivery in tissue engineering. *Adv Drug Deliv Reviews*, 1998, **33** :71 ~ 86
- [5] Hayashi Y, Lewis R V. Evidence from flagelliform silk cDNA for the structure basis of elasticity and modular nature of spider silks. *J Mol Biol*, 1998, **275** :773 ~ 784
- [6] Xu M, Lewis R V. Structure of a protein superfiber: spider dragline silk. *Proc Natl Acad Sci*, 1990, **87** :7120 ~ 7124
- [7] Prince J T, McGrath K P, Kaplan D L. Construction, cloning and expression of genes encoding spider dragline silk. *Biochem*, 1995, **34** :10879 ~ 10885
- [8] Arcidacono S, Mello C, Kaplan *et al.* Purification and characterization of recombinant spider silk expressed in *Escherichia coli*. *Appl Microbiol Biotechnol*, 1998, **49** :31 ~ 38
- [9] Fahnestock S R, Bedzyk L A. Production of synthetic spider dragline silk protein in *Pichia pastoris*. *Appl Microbiol Biotechnol*, 1997, **47** :33 ~ 39
- [10] Hinman M B, Jones J A, Lewis R V. Synthetic spider silk: a modular fiber. *Trends Biotechnol*, 2000, **18** :374 ~ 379
- [11] Scheller J, Guhrs K H, Grosse F *et al.* Production of spider silk proteins in tobacco and potato. *Nat Biotechnol*, 2001, **19** :573 ~ 577
- [12] Gosline J M. The mechanical design of spider silks: from fibroin sequence to mechanical function. *J Exp Biol*, 1999, **23** :3295 ~ 3303
- [13] Hinman M B, Lewis R V. Isolation of a clone encoding a second dragline silk fibroin. *J Biol Chem*, 1992, **267** :19320 ~ 19324

Study on Construct and Expression of Synthetic Genes Encoding Spider Dragline Silk in *Escherichia coli*

LI Min* ZHANG Wen-Xian HUANG Zhi-Hua HUANG Jian-Kun
(College of Biological Engineering, Fujian Normal University, Fuzhou 350007, China)

Abstract Dragline spider silk produced from *Nephila clavipes* major ampullate is a natural fibrous protein with specific mechanical properties such as high tensile strength and elasticity. Synthetic gene monomer encoding recombinant spider silk protein, based on the known repetitive protein sequence and partial cDNA sequence of dragline silk, was constructed and expressed. DNA monomer sequences were multimerized to encode high molecular weight synthetic spider silks using a "head-to-tail" construction strategy. Multimer was cloned into pET30a(+), a prokaryotic high potency expression vector, and induced with IPTG. The protein from 8-unit repeat was produced in *Escherichia coli* at levels up to 20 mg/L. The protein was easily purified with high recovery by using an metal ion affinity chromatography and purity was over 90%. The results of SDS-PAGE and Western blot suggested that the mass of the expression product was about 37kD. This value and amino acid analysis were consistent with those of theoretic calculation.

Key words spider, dragline silk, gene synthesis, expression

Received 09-11-2001

This work was supported by Grant from the Key Project of the Fujian Provincial Natural Science Foundation(No. 2001F006).

* Corresponding author. Tel/Fax 86-591-3425127; E-mail: ML12@sina.com. © 中国科学院微生物研究所期刊联合编辑部 <http://journals.im.ac.cn>