

一类新的编码 PRPs 基因的分离 及其在棉花纤维等组织细胞中的表达 *

许文亮 ** 黄耿青 ** 王秀兰 汪 虹 李学宝 ***

(华中师范大学生命科学学院, 武汉 430079)

摘要 富含脯氨酸的细胞壁蛋白(proline-rich proteins, PRPs)在植物中广泛分布, 它们在建造围绕特定细胞类型的细胞壁结构上起着很重要的作用. 从棉花的 cDNA 文库中分离了 5 个编码富含脯氨酸的细胞壁蛋白质基因, 这 5 个基因推断的氨基酸序列最普遍的特点就是脯氨酸含量非常高. 根据氨基酸组成、富含脯氨酸的重复单元和结构域组织的特点, 将这 5 个蛋白质分成 2 个亚类: 一类(包括 GhPRP3-6)与典型的 PRPs 结构相似, 由 N 端疏水区(或信号肽)与不同富含脯氨酸的重复序列组成; 另一类(GhPRPL)与典型的 PRPs 结构不同, 这个蛋白质的 N 端为亲水序列, GhPRPL 在靠近 C 端有 8 个 5 肽(类似 PPKKE)的重复基序, 与典型 PRPs 所含有的重复序列 PPVYK 非常相似. 实时 RT-PCR(Real-time RT-PCR)分析表明, GhPRP3 和 GhPRP5 在 10 dpa 纤维中特异表达, 而 GhPRPL 在子叶中优势表达. GhPRP4 和 GhPRP6 在所分析的组织中都有表达, GhPRP4 mRNA 在下胚轴中最丰富, 在花药中次之, 而 GhPRP6 在 10 dpa 纤维中表达最强, 在 10 dpa 胚珠中次之. 此外, GhPRP3, GhPRP5 基因表达受纤维发育调节, 表明它们可能在棉纤维发育中起重要作用.

关键词 富含脯氨酸蛋白, 基因表达, 棉花

学科分类号 Q 75

植物细胞壁是一个高度复杂和动态的结构, 它们决定细胞的大小和形状, 在植物的生长和发育中起着非常重要的作用. 植物细胞壁主要由纤维素、半纤维素、果胶和蛋白质组成. 蛋白质占细胞壁干重的 10% 左右, 初生细胞壁比次生细胞壁蛋白质含量高, 这些蛋白质在植物的不同器官、组织和细胞中有独特的分布. 细胞壁蛋白质主要由结构蛋白和酶蛋白组成, 结构蛋白被认为形成一个独立的决定壁结构的网络进而来帮助合适的壁装配, 这些蛋白质在结构上高度重复, 富含某一两个氨基酸^[1-6].

富含脯氨酸的蛋白质(proline-rich proteins, PRPs)代表一类富含 proline (Pro-) 和 hydroxyproline (Hyp-) 的结构蛋白, 这类蛋白质在双子叶植物如大豆^[7]、胡萝卜^[8]和紫花苜蓿^[9]及单子叶植物如玉米^[10]和小麦^[11]等植物中广泛存在, 而且被认为在建造围绕特定细胞型的初生壁结构方面起作用. Jose 和 Puigdomenech^[12]将 PRPs 进一步分成 3 个亚类: a. 整个蛋白质富含 Pro-rich 的重复序列缺少半胱氨酸; b. 不含 Pro-rich 的重复序列, 只含小的 Pro 基序; c. HyPRP-Hybrid PRPs. 这类蛋白质最普遍的特点就是 Pro 残基含量高, 至少有 2 个连续的 Pro 排

列在氨基酸链中. 它们最先在胡萝卜贮藏根中被认为是外界伤害诱导的基因产物, 后来的研究显示, 它们在很多植物中表达, 参与植物发育的不同方面. 例如, 种子萌发, 豆荚形成, 早期结瘤等. PRPs 基因的表达还受伤害、真菌、乙烯、细胞培养、干旱和光照影响, 它们一般位于木质部、纤维、石细胞、表皮、糊粉层和结节实质中^[2,8,13,14]. 但到目前为止, 对它们的功能研究尚未见报道.

棉花是世界上首要的纤维作物和重要的油料作物, 各主要产棉国都十分重视棉花发育和遗传的研究. 棉纤维是由棉花胚珠外珠被表皮层的单细胞发育而成, 其分化和发育过程可分为纤维起始、初生细胞壁形成(或纤维细胞的伸长)、次生细胞壁的加厚和脱水成熟等 4 个时期. 棉花是研究植物细胞伸长和细胞壁生物发生的理想模型^[15-17]. 我们在棉花的 cDNA 库中分离了 5 个编码 PRPs 的基因, 蛋白

*国家自然科学基金(30400022), 国家重点基础研究发展计划(973)(2004CB117304)和武汉市青年科技晨光计划(20065004116-26)资助项目.

** 并列第一作者.

*** 通讯联系人. Tel: 027-67862443, E-mail: xbli@mail.ccnu.edu.cn

收稿日期: 2006-11-03, 接受日期: 2007-01-16

质结构分析显示, 其中的 4 个与细胞壁 PRP 类蛋白质结构特点相似, 但另外 1 个 *GhPRPL* 其 N 端是亲水的氨基酸多肽, 除了 Pro 含量非常高外, 与常规的细胞壁 PRPs 结构不一样, 可能是一个新的 PRP 类蛋白质. 这些基因的表达研究结果显示, 其中的 3 个基因为组织特异性或优势表达, 另外 2 个基因在研究的 7 个组织中均有表达, 显示这类基因可能在棉花纤维细胞或其他组织细胞发育中起重要作用.

1 材料和方法

1.1 植物材料

棉花 (*Gossypium hirsutum* Coker 312) 种子经 70% 乙醇表面灭菌 1 min 后, 再用 10% H_2O_2 溶液浸泡 1~2 h, 随后用无菌水洗 3~4 次, 于 28℃ 培养箱中浸泡 24 h 后, 将种子置于 1/2 MS 培养基上萌发 (16 h 光照 / 8 h 黑暗, 28℃). 幼苗生长 5~7 天后, 取根、下胚轴、子叶用于提取 RNA. 花瓣、花药, 2 dpa (day post anthesis) 纤维、5 dpa 纤维、10 dpa 纤维、15 dpa 纤维、20 dpa 纤维和 10 dpa 胚珠材料取自大田种植的棉花植株.

1.2 *GhPRP 3~6*、*GhPRPL* cDNA 分离与鉴定

从棉花 cDNA 文库^[8]中随机挑选 2 000 多个 cDNA 克隆, 通过核苷酸序列分析以及编码蛋白质序列比较分析, 分离获得 5 个编码 PRP 的 *GhPRP* 基因(cDNA), 命名为 *GhPRP3~6*、*GhPRPL*.

1.3 序列分析

用 DNASTar (DNA star, Madison, WI) 分析核苷酸和氨基酸序列, 用 Clustal W (<http://www.ebi.ac.uk>) 进行多个核苷酸和蛋白质序列比对分析, 用 Motifscan (http://myhits.isb-sib.ch/cgi-bin/motif_scan) 分析基因的结构域, 用 protparam (<http://www.expasy.org/tools/protparam.html>) 分析蛋白质的氨基酸组成, 用 Kyte-Doolittle Hydropathy Plots (http://fasta.bioch.virginia.edu/fasta_www/grease.htm) 分析蛋白质的亲水疏水性, 用 SigalP (<http://www.cbs.dtu.dk/services/SignalP/>) 分析可能的信号肽序列, 用 NetNGlyc (<http://www.cbs.dtu.dk/services/NetNGlyc/>) 分析可能的 N-glycosylation 位点.

1.4 RNA 分离和 Real-time RT-PCR 分析

按改良的 CTAB 法提取棉花组织总 RNA, DNase(RQ1 RNase-Free DNase, Promega 公司)消化后用 RNeasy Plant Mini Kit(Qiagen 公司)纯化. 应用实时荧光定量 RT-PCR 技术(荧光定量 PCR 仪型号

为 MJ Research, Opticon 2), 参照文献[19]进行 *GhPRPs* 的表达分析. 首先, 将棉花不同组织(根、下胚轴、子叶、花瓣、花药、2 dpa 纤维、5 dpa 纤维、10 dpa 纤维和 10 dpa 胚珠、15 dpa 纤维、20 dpa 纤维等)的总 RNA(2μg/ 样)用 SuperScript™ II RNase H⁻ Reverse Transcriptase (Invitrogen Life Technologies) 反转录成 cDNA. 然后, 以 cDNA 为模板, 用基因特异的引物 (表 1) 和 Real-time PCR Master Mix (TOYOBO, Japan) 进行 Real-time PCR 反应. 棉花 polyubiquitin 基因(*GhUBI*)作为 RT-PCR 反应的内标^[19], 目标基因每一个循环的扩增都被 SYBR-Green 荧光检测. 每个基因的表达水平相对值按公式 $Y=10^{\Delta C_t/3.57} \times 100\%$ 计算 (其中 $\Delta C_t = Gh_{UBI} - Gh_{PRP}$, 3.57 是利用 *GhUBI* 制备的标准曲线 $y = -0.28x + 9.87$ 中斜率的倒数, 表示基因表达相差 10 倍的 PCR 循环数). 重复 3 次, 统计分析实验结果.

Table 1 Primers used in gene-specific RT-PCR of *GhPRP* genes

Genes	Primers
<i>GhPRP3</i>	5' GTA CCA TTG CCA CCT CCG ATG 3' 5' TGTGATGGCAGAGTTTCATTCTC 3'
<i>GhPRP4</i>	5' GCATATCAGCAGGCTTACTACC 3' 5' GGCTCAGATCAGATAAACAGTAC 3'
<i>GhPRP5</i>	5' GAG AAT TCT GAG AAG GAA CCT 3' 5' TAACTGAATGACAATAATTCAAG 3'
<i>GhPRP6</i>	5' CACCAGTATATGAACCTCCAAAG 3' 5' GGTGCATTAATGGTGGCTTAGGT 3'
<i>GhPRPL</i>	5' CCACCAGTGTATGAGCCTAAG 3' 5' GAGTACTGCCATAAGCACTCT 3'
<i>GhUBI</i>	5' CTGAATCTTCGCTTTTCACGTTATC 3' 5' GGGATGCAAATCTTCGTGAAAAC 3'

2 结 果

2.1 *GhPRP* cDNA 分离鉴定

采用随机挑选克隆的方法, 对 2 000 多个 cDNA 克隆测序, 通过核苷酸序列分析以及编码蛋白质序列比较等生物信息学分析, 筛选到 5 个编码 PRP 的 cDNA 阳性克隆, 命名为 *GhPRP3*、*GhPRP4*、*GhPRP5*、*GhPRP6*、*GhPRPL* (Accession numbers in GenBank: EF095704~EF095708). 为了解这类细胞壁蛋白质基因在棉花生长和发育, 尤其是在纤维发育中的可能作用, 我们对这 5 个基因进行了分析. *GhPRP3* cDNA 包含 333 bp 的开放阅读框(open reading frame, ORF), 编码一个 111 个氨基酸的蛋白质. 该蛋白质分子质量为 11.74 ku, pI 为 4.78, 富含 Pro (32.4%), Leu (17.1%), Ser

<i>GhPRPL</i>	ATGCTCTACAACACACTTGCTAGTATTCCTACTTGGAGTGGTGACTCTCACCCTCCACA	60
<i>GhPRP6</i>		
<i>GhPRP5</i>	-----ATGCCCAAATCGCCGTTTCGCTT	24
<i>GhPRP4</i>		
<i>GhPRP3</i>		
<i>GhPRPL</i>	TTTGGTACCTACGAGTCACCAAAATTATGGGAAACCCCTACTCCTGAATTCAAGCCTCCC	120
<i>GhPRP6</i>	-----ATGGGGGAG	9
<i>GhPRP5</i>	GGCCTCGGCGATCAGATGCCTGACATGATTGAAGAATTGATCAGCAAGGGACAGCAGCTT	84
<i>GhPRP4</i>		
<i>GhPRP3</i>		
<i>GhPRPL</i>	-----ATGAACATAAGCCCCCTGTCTATGAACCACCA--AAGAAGGA	40
<i>GhPRP6</i>	AAGGTGAAACCAACCACCATATGAACCTAAGCCACCAAGTGTATGAACCACCA--AAAAAGGA	179
<i>GhPRP5</i>	AAGGTGACGATAATGGTGC--TGAAGGTCGACCTTCAAGTGTAGACGCTGTACAGAAGGT	68
<i>GhPRP4</i>	GATGCAGTTCAATTTACTTTATGAA--GTTGGCCTT--GTGGACAAGTCCCCCTGTGCC	141
<i>GhPRP3</i>		
<i>GhPRPL</i>	AAAGCCTGAGCCTAAGCCACCAAGTTTATACACCT--CCAAAGAAA----GAGAAGCCTGA	94
<i>GhPRP6</i>	GAAGCCTGAACCTAAGCCACCGGTTTATGACCT--CCAAAGAAA----GAGAAGCCTGG	233
<i>GhPRP5</i>	GAAGCAAGTACTTTGCAAATTCCTCAAATACGAGACCAATATACGACGAGAAGGCCAA	128
<i>GhPRP4</i>	CTGCTGAAATCTTTCTTGAGGATGCAAAGAAGGCTGCATCTTCTATTTTAGATGATCCC	201
<i>GhPRP3</i>		
<i>GhPRPL</i>	--ACCCAAACCACAGTATA--TGAACC----TCCAAAGAAGGAAAAGCCTGAGCCT--AAGC	147
<i>GhPRP6</i>	--ACCCAAACCACAGTGTAT--TGAACC----TCCAAAGAAGGAGAAGCCTGAGCCA--AAGC	286
<i>GhPRP5</i>	CACGGTGACAATCAAGGTAGTTTGTCTGCGATCCAGAGAAGATGAGGGGCAAGA--T--ACGT	186
<i>GhPRP4</i>	AATAATACTGGCCGGCTGCGCAACT--TGCTGCACGCAAGAGCAATCTGCACTTCGGGC	260
<i>GhPRP3</i>	-----ATGGCAGCTTCCAACGT	17
<i>GhPRPL</i>	CACCAAGTTTATACACCAC--CAA--AGAAAGAGAAGCCTGAGCCCAACCC--ACCAGTATAT	203
<i>GhPRP6</i>	CACCAAGTTTATACACCTC--CAA--AGAAAGAGGAGCCTAAACCTAAACCC--ACCAGTATAT	342
<i>GhPRP5</i>	TGCAAGGGTGGCGATTCCATCA--AGAGCATTGAGATCAAACCACTCC--AA--AGCCCAA	242
<i>GhPRP4</i>	TGTCATCAAGTGCAATGAAGAGTACAACCTGAGGCTGAGTTTCCGCTGAAAACCTCAA	320
<i>GhPRP3</i>	TTTCATTTTATGACCTGCTTC--ATTTCTTGGGTTTTCAAGCATGGAGGTTAGCTAG	76
<i>GhPRPL</i>	* * * * *	
<i>GhPRP6</i>	GAACC--CCAAAG----AAGGAAAAGCCTGAACCTAAGCCACCAAGTTTATACACCACCAA	258
<i>GhPRP5</i>	GAACC--TCCAAAG----AAGGAGAAACCTGAGCCAAACCAACCGATTATACACCTCCAA	397
<i>GhPRP4</i>	GGACC--CGAGAAATCTGAGAAGGAACCTGAAAAGAAAACAGAGAAGCCTACAGAGTCCG	301
<i>GhPRP3</i>	GAACCGCCTTGAGCAGCTAGAGAAGACCAAGACTGAAAAGAGAAAACCAAGTTGTAGTCCC	380
<i>GhPRPL</i>	CAACC--CGAAATCTTACGCAAGTGCCACCTCCATTTCACCAATACCATTGCCACCTC	133
<i>GhPRP6</i>	* * * * *	
<i>GhPRP5</i>	AGA--AAGAGAAGCCTGAGCCCAAAACCAAGTGTATGGACCACCAAA--GAAGGAGAAGC	316
<i>GhPRP4</i>	AGA--AAGAGAAGCCTGAACCCAAACCAACCCGAT--ATGAACCTCCAAA--GAAGGAGAAGC	454
<i>GhPRP3</i>	AGA--AGAAGTCTGAAAAGCCCAAGAAACACCA--AAGTCACCGGC--GAACAGGCTT	358
<i>GhPRPL</i>	TGCTAACAAGCGAAGCGGTGTAATAATGGTGGTCCGATGCCTCTGCTAAAGCTGGCCG	440
<i>GhPRP6</i>	CAT----TGTCGCCAATACCAATGTACACAGTACCA--TTGCTACCTCC--ATTGTGCG	184
<i>GhPRP5</i>	* * * * *	
<i>GhPRPL</i>	C-----ACCGGCTTATA-----	328
<i>GhPRP6</i>	CCGAGCCAAAGCCACCGGTTTATACACCTCCAAAGAAAAGAGAAAACCTGAACCCAAACCA	514
<i>GhPRP5</i>	CTCCTCCG--GAAGGCTTTTGTGTCACGGA-----T	387
<i>GhPRP4</i>	TTTGACGAATGCATATGTCTCATCTTCCGGCACCT-----CCTC	481
<i>GhPRP3</i>	CAATACCAATGCCACCTCCATTTT-----C	211
<i>GhPRPL</i>	* * * * *	
<i>GhPRP6</i>	-----AGCCTCCAAAGAGCCACCAAGTGTATGAGCCTAAGCCACCAAAAGCCACCGG	379
<i>GhPRP5</i>	CAGTGTATGAACCTCCAAAAAACCTCCAATGTACGAACCTAAGCCACCAAAACCAACCGG	574
<i>GhPRP4</i>	TGTTATCATGGTATCGTGGTGGCCCTTGCTACTTTGG--TGGACCACTCCGCCCCCAT	445
<i>GhPRP3</i>	CATTGTAAAGGTCTCCCTCCCA--CACCCAGTACCCTGCTCCAGTCCAGGATACC--CATC	539
<i>GhPRPL</i>	CAGTACCAATGCCATCTGCATATTTACCGGTACCATTGCCACCTCCGATGCACCACTTC	271
<i>GhPRP6</i>	* * * * *	
<i>GhPRP5</i>	TTTACACGCCACCAAGAAAGAGAAGCCGAACCTAAACCAACCAAGTACATGAGTCTCCCA	439
<i>GhPRP4</i>	TTTATACACCTCCAAAGAAAGAGAAACCTGAACCCATACCAACCAATGTATGAACCTCCAA	634
<i>GhPRP3</i>	ACTATTGGCCTT-----ATGGTAGGCC--AGTTTATGTTAACTGGG--GCGGCTGCGG	494
<i>GhPRPL</i>	CCCACCTCCATGTATGGAAGCAGGAGCCACCCACCAACCCCTATGCTTACTCACCAGA	599
<i>GhPRP6</i>	CTTTAATCCCA-----GAACCACCAACCTTCTTCAATCCCAAGAA-----CTTAGA	318
<i>GhPRP5</i>	** * * *	
<i>GhPRPL</i>	AGAAA--CCACC--GTATGGTCACTATCCAGGACACCTCCATTGGGGAAGCCTCAATA	494
<i>GhPRP6</i>	AAAAG--CCTCCAATGTGTGAACCTAAGCCACCAAGCCACCGGTTTATACACCTCCAAA	692
<i>GhPRP5</i>	CGGAA--GCACCTA--CTGTTATTGTT--TTGAAGAAACCCACAATGCTCAGTCATGTAA--	549
<i>GhPRP4</i>	AGCAGTCCCTCTCCCTTGTCTGGTTCATACCTGGAGCTCCCATGAATTTATCCGCATA	659
<i>GhPRP3</i>	ACTA--CCACCACTTCGTAA-----	336
<i>GhPRPL</i>	* * * * *	
<i>GhPRP6</i>	-G-----	495
<i>GhPRP5</i>	-GAAAGAAAAATCAGACCAAGCCAACCATGTATCAACCTCCCAACCAACCATATATG	751
<i>GhPRP4</i>	CGGGGGCTACGGTAATGGTTTGGCACCAGCATATCAGCAGGCTTACTACCGATAG-----	714
<i>GhPRP3</i>		
<i>GhPRPL</i>	AGCCAAACCACCTAAGCCACCATTA	777
<i>GhPRP6</i>		
<i>GhPRP5</i>		
<i>GhPRP4</i>		
<i>GhPRP3</i>		

Fig. 1 Comparison of the nucleic acid sequences (coding region) of cotton GhPRPs

The nucleotide sequences were aligned with the Clustal W program. Gaps were introduced to maximize the identity. The conserved residues are marked by asterisks.

(10.8%); *GhPRP4* cDNA 包含 711 bp 的开放阅读框, 编码一个 237 个氨基酸, 分子质量为 25.82 ku、*pI* 为 9.46, 富含 Pro (15.2%)、Leu (8.0%)、Lys (7.2%) 的蛋白质; *GhPRP5* cDNA 包含 546 bp 的开放阅读框, 编码一个 182 个氨基酸、分子质量为 20.57 ku、*pI* 为 8.84, 富含 Pro (13.2%)、Lys (13.7%)、Glu (8.8%) 的蛋白质; *GhPRP6* cDNA 包含 774 bp 的开放阅读框, 编码一个 258 个氨基酸、分子质量为 29.35 ku、*pI* 为 9.55、富含 Pro (35.16%)、Lys (19.8%)、Glu (11.6%) 的蛋白质; *GhPRPL* cDNA 包含 492 bp 的开放阅读框, 编码一个 164 个氨基酸、分子质量为 19.35 ku、*pI* 为 10.93、富含 Pro (19.8%)、Lys (12.8%)、His (9.8%)、Ser (9.8%) 的蛋白质. 这些基因编码的蛋白质最显著的特点就是脯氨酸含量高.

这 5 个基因在核苷酸水平上除了 *GhPRP6* 和 *GhPRPL* 有 71% 的同源性外, 其他序列之间只有 3%~15% 的同源性(图 1), 而且, Blastn 分析也未发现与它们同源性较高的基因序列.

2.2 GhPRP 3~6、GhPRPL 结构分析

典型的植物 PRPs 的结构一般如图 2 所示. 图

2A 显示的是两结构域 PRPs: N 端信号肽和脯氨酸重复区; 图 2B 显示的是三结构域 PRPs (Hybrid PRPs): N 端信号肽, 脯氨酸重复区和 C 端富含半胱氨酸区. 分析表明, *GhPRP3*~*6* 为典型 PRPs 的两结构域特征(图 2A), 而 *GhPRPL* N 端不具有信号肽, 但 C 端含有 8 个类似 PPVYK 的重复序列(图 2C).

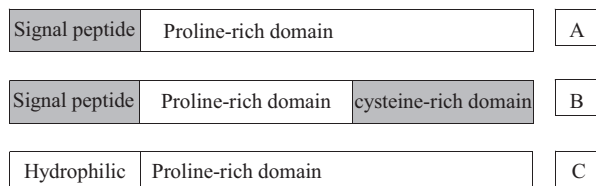


Fig. 2 Domain organization of proline-rich proteins

Typical PRPs composed of two main domains is shown in A such as *GhPRP3* and *GhPRP6* in this study; typical hybrid PRPs composed of three main domains is shown in B such as *GhHyPRP1* (Accession number in GenBank: DQ825801); while a PRP like protein in this study is shown in C such as *GhPRPL*.

2.2.1 典型 PRPs 的结构特点.

如图 3 所示, *GhPRP3*~*6* 蛋白质序列的 N 端有一段由 20 多个氨基酸残基组成的疏水序列, 经 SignalP (<http://www.cbs.dtu.dk/services/SignalP/>) 分析

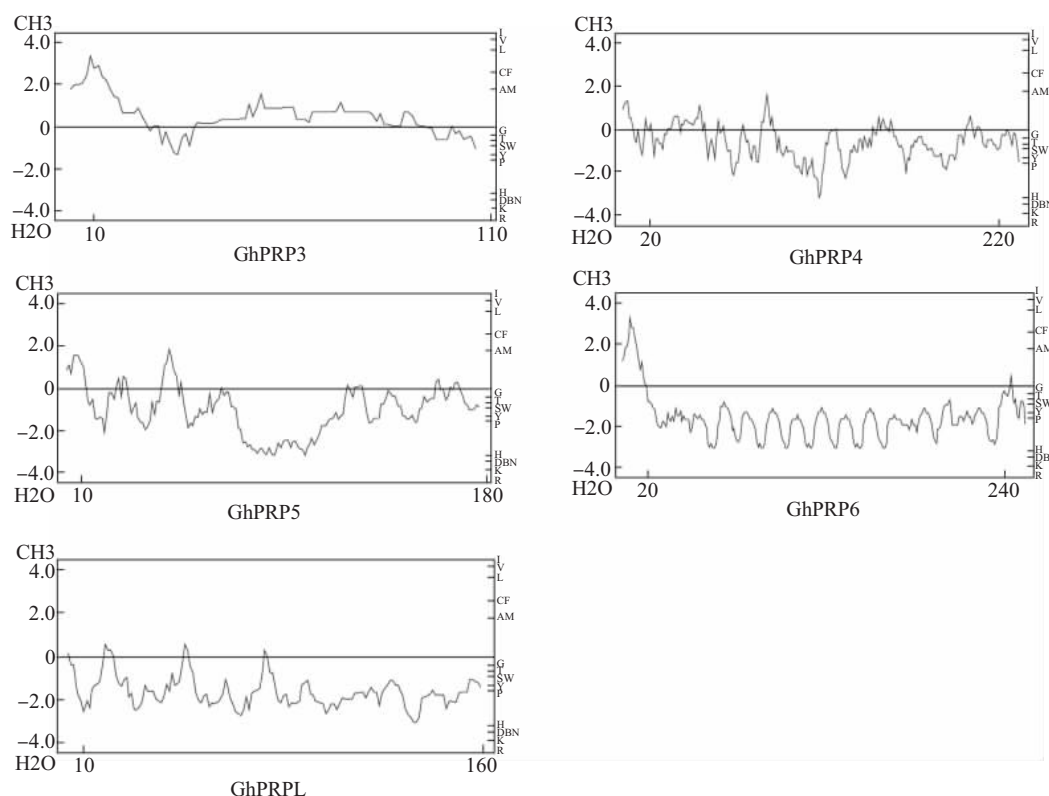


Fig. 3 Comparison of hydropathy plots of GhPRPs

The hydropathy value of polypeptide was calculated by the algorithm of Kyte and Doolittle over a window of nine aa residues and was plotted as a function of aa position (http://fasta.bioch.virginia.edu/o_fasta/grease.htm). Values above the horizontal line indicate the hydrophobic regions, and those below the axis show hydrophilic regions. The hydrophobic regions tend to be embedded inside the membranes.

为信号肽区，可能起着锚定蛋白质或分泌蛋白质到胞外细胞壁上的作用，接着的脯氨酸重复区有不同的特点：GhPRP3 含有 7 个 9 肽(类似 PPPFPPIPL)的重复序列；GhPRP4 含有 2 个 5 肽(类似 PPENL)和 3 个 22 肽(类似 PPPFVRSPSHTQYPAPVPGYPS)的重复基序；GhPRP5 没有规律的富含脯氨酸的重

复序列，但含有一个富含脯氨酸的 19 肽(PCYFGGPPPPYYWPYGRP)；GhPRP6 含有 12 个 15 肽(类似 PPVYEPPKKEKPEPK)的重复序列；这些富含脯氨酸的重复序列与已有的植物 PRPs 中存在的 proline-rich repeats 绝大多数是不同的(表 2)，暗示新的功能单位蕴藏在这些蛋白质中。

Table 2 Classification of the known PRPs in higher plants based on proline-rich repeats

Class	Proline-rich repeats	Distribution of proline-rich repeats	Examples and references
I	PPXZ(K/T)	Tandem or in cluster	SbPRP1/2 (Hong <i>et al</i> , 1989), GhPRPL this study
II	PPTY and PPTPRPS	Dispersed or in cluster	MsPRP5(Gyorgyey, <i>et al</i> , 1997), ZmHyprp(Jose-Estanyol, <i>et al</i> , 1992)
III	PPV, P(V/T)YK, KKPCPP	Dispersed or in cluster	AtPRP2/4 (Fowler, <i>et al</i> , 1999)
IV	PEPK	Tandem or dispersed	WPRP1(Raines, <i>et al</i> , 1991),
V	PKPE, P(V/E)PPK	Dispersed	OsPRP1~4 (Wang, <i>et al</i> , 2006)
VI	P(I/V)PL	Dispersed	GhPRP3, this study
VII	PPP(F/M/L)	Dispersed	GhPRP3, GhPRP4, this study
VIII	PPVY(E/A/T)PPKKE KP(E/G/K)PK	Tandem or dispersed	GhPRP6, this study

2.2.2 PRPL(proline-rich protein like)的结构特点.

GhPRPL 蛋白质序列的 N 端是亲水的序列(图 3)，这与一般的富含脯氨酸的细胞壁结构蛋白质不

同，但在 C 端区 Pro 非常丰富，40 个氨基酸残基里含有 26 个 Pro(65%)，有 8 个 5 肽(类似 PPKKE)的重复基序，其中的 3 个 5 肽重复(PPAYK，



Fig. 4 Comparison of the predicted amino acid sequences of cotton GhPRP genes

Multiple alignment of amino acid sequences which were deduced from the five GhPRP genes. Gaps were introduced to maximize the identity. The conserved residues are marked by asterisks.

PPVYE, PPVYT)与典型的 PRPs 所含有的 PPVYK 非常类似. N 端不具有信号肽, 不大可能定位于细胞壁上, 这种类似植物 PRPs 但又不同于植物 PRPs 的结构特点尚未见有报道.

这 5 个蛋白质序列之间有 13%~31% 的同源性, 除了 GhPRPL 与 GhPRP6 之间有 31% 的同源性外, 其他序列之间一般只有百分之十几的同源性(图 4), 而且与已有报道的植物 PRPs 的同源性也较低, 显示是新的富含脯氨酸的蛋白质.

2.3 5 个 *GhPRP* 基因在棉花组织中的表达谱

2.3.1 *GhPRP* 基因在不同棉花组织中的差异表达. 利用定量 RT-PCR 技术对这些基因在棉花不同组织中的表达情况进行了研究. 结果显示, *GhPRP3*, *GhPRP5* 在纤维细胞中特异表达, 而在其他组织中表达非常微弱或几乎检测不到. *GhPRP4* 在所检测的组织中都有表达, 但在下胚轴中表达最高, 花药中其次, 在胚珠、根中也有表达. *GhPRP6* 在所检测的 7 个组织中都有表达, 在纤维中最强, 在胚珠中其次, 在根中也有较强表达. *GhPRPL* 在子叶中表达量最高, 在根中其次, 在花瓣和花药中表达很低, 在 10 dpa 纤维、胚珠中检测不到表达(图 5). 这些结果显示这类基因的表达是组织特异性或细胞特异性的.

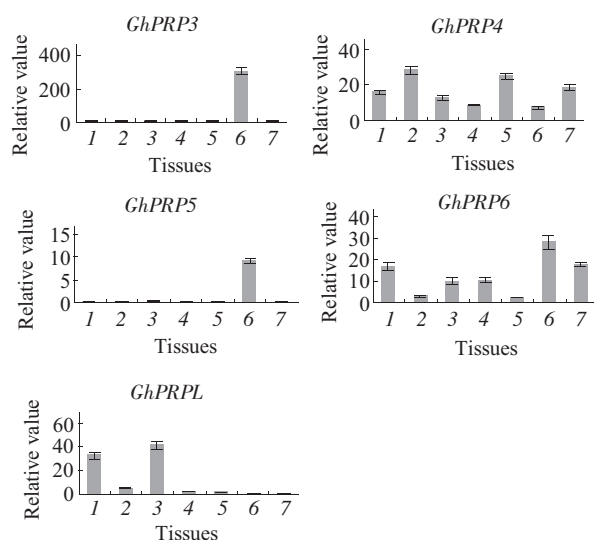


Fig. 5 Quantitative realtime RT-PCR analysis of *GhPRPs* transcripts in tissues

1: Cotton roots; 2: Hypocotyls; 3: Cotyledon; 4: Petal; 5: Anther; 6: 10 dpa fiber; 7: 10 dpa ovule.

2.3.2 *GhPRP3* 和 *GhPRP5* 基因表达受棉纤维细胞发育调节. 为进一步研究 *GhPRP3*, *GhPRP5* 在纤维

发育阶段的表达情况, 我们研究了这 2 个基因在 2 dpa, 5 dpa, 10 dpa, 15 dpa, 20 dpa 纤维中的表达情况, 结果如图 6 所示, 在纤维发育的早期 (2 dpa) 几乎检测不到 *GhPRP3* mRNA, 在 5 dpa 纤维中有少量表达, 在 10 dpa 纤维中表达量达到最大, 随着纤维的不断延伸, 表达逐渐降低. *GhPRP5* 在纤维发育的早期 (2 dpa) 有很微弱的表达, 在 5 dpa 纤维中表达最强, 10 dpa 纤维中表达有所下降, 在纤维延伸的后期表达维持在一个较低的水平. 上述结果表明, 这 2 个基因的表达受纤维发育阶段调节.

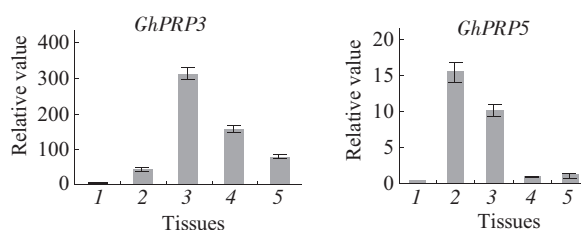


Fig. 6 Quantitative realtime RT-PCR analysis of *GhPRP3* and *GhPRP5* gene expression during fiber elongation period

1: Fiber at 2 dpa; 2: Fiber at 5 dpa; 3: Fiber at 10 dpa; 4: Fiber at 15 dpa; 5: Fiber at 20 dpa.

3 讨 论

除了富含甘氨酸的蛋白质 (glycine-rich proteins, GRPs) 外, 在植物的细胞壁中至少可以将结构蛋白分成两大类: 一类是不溶的细胞壁蛋白如伸展蛋白 (extensins), 一类是可溶的细胞壁蛋白如阿拉伯半乳聚糖蛋白 (arabinogalactan proteins, AGPs), 连同其他一些或多或少糖基化的细胞壁蛋白如 PRPs, 杂合 PRPs (hybrid PRPs, HyPRPs) 在高等植物中普遍存在^[14]. Extensins 是一类在双子叶植物中得到广泛研究的细胞壁结构蛋白, 一般含有 Ser~(Hyp)₄ 的重复基序, 富含 Hyp, Ser, Val, Tyr, Lys 和 His^[2,4,14]. 绝大多数 AGP 糖蛋白中的蛋白质含量不到 10% 而含有多于 90% 的糖类, 一般含有几个 Ala~Hyp 或类似的重复单元, 富含 Hyp, Ala, Thr, Gly 和 Ser, 但不同的物种或组织其氨基酸含量不一样, 一个最显著的特点就是所有 AGPs 都可以和 β -glucosyl Yariv 试剂结合^[2,4,14]. 我们在棉花的 cDNA 文库中分离了 5 个编码富含脯氨酸的蛋白质基因, 分析组成蛋白质的结构域, 发现它们没有 extensins, AGPs 的主要结构特点, 基于其编码蛋白的高脯氨酸含量以及脯氨酸重复序列的特点, 我们认为它们是一类编码 PRPs 的基因. 这 5

个蛋白质中的 4 个(*GhPRP3*~*6*)显示典型细胞壁 PRPs 的结构特性(图 2A), 其中 *GhPRP3*、*GhPRP4*、*GhPRP6* 属于 PRPs 的第一亚类, *GhPRP5* 属于 PRPs 的第二亚类^[12], 但所含有的脯氨酸重复序列各有特点(表 2), 与目前报道的 5 大类 PRPs 的脯氨酸重复基序有较大差异^[20], 表明是新的 PRPs. 对于 *GhPRPL*, 它虽然含有类似典型 PRPs 的重复序列(PPVYK), 但它们不含有 N 端信号肽(图 2C), 这种结构特点提出了一个很有意思的问题, 它们可能是细胞壁蛋白质吗? 如果是, 它们是如何转运到细胞壁上的, 如果不是, 又是一类什么样的蛋白质? 有待进一步研究.

植物中的细胞壁 PRPs 一般都由一个基因家族编码, 如烟草 *CELP1*~*CELP5*、*NtTTS1*~*NtTTS2*, 花烟草 *NaPRP1*~*NaPRP4*, 葡萄 *VvPRP1*~*VvPRP2*^[21~25]. 在大豆中至少分离了 3 个编码 PRPs 的基因: *SbPRP1*、*SbPRP2* 和 *SbPRP3*, 尽管在核苷酸和氨基酸序列上显示很高的同源性, 但它们在不同的植物器官或同一器官的不同区域都显示很大的表达差异. *SbPRP1* 在成熟的下胚轴, 根和未成熟的种皮中优势表达; *SbPRP2* 在胚轴顶部和幼嫩的悬浮培养细胞中大量表达; *SbPRP3* 在气生部位如叶子中高量表达, 而在根中没有表达. 这些结果显示 PRPs 是发育调节和器官特异性表达^[26~28]. 我们的研究结果也表明, 所分离的棉花 PRP 基因是组织或细胞特异性表达的. 不同的基因在不同的组织或细胞中呈现优势表达, 说明这类基因可能分别在棉花不同组织发育中具有一定的生物学功能.

棉花 H6 蛋白是一个较早从棉花中分离的富含脯氨酸的蛋白质, 含有 17 个由 Ala(or Ser)-Thr(or Ser)-Pro-Pro-Pro 组成的重复基序, 主要在纤维细胞中表达, 在初生细胞壁形成早期也存在, 然而, H6 蛋白在次生细胞壁形成时积累很多, 显示在转译水平的调节和在次生细胞壁装配上起作用, 可能参与棉花纤维次生细胞壁的建成与发育^[29]. *FSPRP* 也是一个特异在纤维细胞中表达的基因^[30]. Tan 等^[31]从棉花中克隆了 2 个编码 PRP 的基因 *GhPRP1* 和 *GhPRP2*, 在氨基酸水平上显示 80% 的同源性, RT-PCR 结果显示 *GhPRP1* 在棉纤维和根中表达, 而 *GhPRP2* 只在根中表达. 此外, Feng 等^[32]利用 suppression subtractive hybridization 的方法分离了 6 个 PRP 蛋白质基因, 其中的 2 个在 10 dpa 纤维中大量表达. 目前已有的研究显示, 相当数量的基因参与棉纤维的发育过程^[18,19,32~40]. 在我们的研究中发

现, 有 3 个基因(*GhPRP3*, *GhPRP5*, *GhPRP6*)在 10 dpa 纤维中特异性或优势表达, 而且, *GhPRP3* 和 *GhPRP5* 基因表达是纤维发育阶段调节的. 利用 RNAi 等技术, 我们将进一步研究这类基因在棉花生长和发育尤其是纤维发育中的作用.

参 考 文 献

- 1 Minorsky P V. The wall becomes surmountable. *Plant Physiol*, 2002, **128** (2): 345~353
- 2 Cassab G I. Plant cell wall proteins. *Annu Rev Plant Physiol Plant Mol Biol*, 1998, **49** (1): 281~309
- 3 Keller B. Structural cell wall proteins. *Plant Physiol*, 1993, **101** (4): 1127~1130
- 4 Showalter A M. Structure and function of plant cell wall proteins. *The Plant Cell*, 1993, **5** (1): 9~23
- 5 Lamport D T A. Cell wall proteins. *Adv Bot Res*, 1965, **2** (1): 215~218
- 6 Varner J E, Lin L S. Plant cell wall architecture. *Cell*, 1989, **56** (2): 231~239
- 7 Hong J C, Nagao R T, Key J L. Characterization of a proline-rich cell wall protein gene family of soybean. A comparative analysis. *J Biol Chem*, 1990, **265** (5): 2470~2475
- 8 Chen J, Varner J E. Isolation and characterization of cDNA clones for carrot extensin and proline-rich 33-kDa protein. *Proc Natl Acad Sci USA*, 1985, **82** (13): 4399~4403
- 9 Györgyey J, Németh K, Magyar Z, *et al.* Expression of a novel-type small proline-rich protein gene of alfalfa is induced by 2, 4-dichlorophenoxyacetic acid in dedifferentiated callus cells. *Plant Mol Biol*, 1997, **34** (4): 593~602
- 10 Vignols F, Josè-Estanyol M, Caparrós-Ruiz D, *et al.* Involvement of a maize proline rich protein in secondary cell wall formation as deduced from its specific mRNA localization. *Plant Mol Biol*, 1999, **39** (5): 945~952
- 11 Raines C A, Lloyd J C, Chao S, *et al.* A novel proline-rich protein from wheat. *Plant Mol Biol*, 1991, **16** (6): 663~670
- 12 Jose M, Puigdomenech P. Transley review No.55: structure and expression of genes coding for structural proteins of the plant cell wall. *New Phytol*, 1993, **125** (2): 259~282
- 13 Menke U, Renault N, Mueller-Roeber B. StGCPRP, a potato gene strongly expressed in stomatal guard cells, defines a novel type of repetitive proline-rich proteins. *Plant Physiol*, 2000, **122** (3): 677~686
- 14 Jose-Estanyol M, Puigdomenech P. Plant cell wall glycoproteins and their genes. *Plant Physiol Biochem*, 2000, **38** (1~2): 97~108
- 15 Basra A S, Malik C P. Development of the cotton fiber. *Int Rev Cytol*, 1984, **89** (1): 65~113
- 16 Kim H J, Triplet B A. Cotton fiber growth in planta and *in vitro*. Models for plant cell elongation and cell wall biogenesis. *Plant Physiol*, 2001, **127** (4): 1361~1366
- 17 Wilkins T A, Bulak A A. The cotton fiber transcriptome. *Physiologia Plantarum*, 2005, **124** (3): 295~300

- 18 Li X B, Cai L, Cheng N H, *et al.* Molecular characterization of the cotton GhTUB1 gene that is preferentially expressed in fiber. *Plant Physiol*, 2002, **130** (2): 666~674
- 19 Li X B, Fan X P, Wang X L, *et al.* The cotton ACTIN1 gene is functionally expressed in fibers and participates in fiber elongation. *The Plant Cell*, 2005, **17** (3): 859~875
- 20 Wang R, Chong K, Wang T. Divergence in spatial expression patterns and in response to stimuli of tandem-repeat paralogues encoding a novel class of proline-rich proteins in *Oryza sativa*. *J Experimental Botany*, 2006, **57** (11): 2887~2897
- 21 Wu H M, Zou J, May B, *et al.* A tobacco gene family for flower cell wall proteins with a proline-rich domain and a cysteine-rich domain. *Proc Natl Acad Sci USA*, 1993, **90** (14): 6829~6833
- 22 Thomas P, Lee M M, Schiefelbein J. Molecular identification of proline-rich protein genes induced during root formation in grape (*Vitis vinifera* L.) stem cuttings. *Plant Cell Environment*, 2003, **26** (9): 1497~1504
- 23 Chen C G, Mau S L, Clarke A E. Nucleotide sequence and style specific expression of a novel proline-rich protein gene from *Nicotiana glauca*. *Plant Mol Biol*, 1993, **21** (2): 391~395
- 24 Chen C G, Cornish E C, Clarke A E. Specific expression of an extensin-like gene in the style of *Nicotiana glauca*. *The Plant Cell*, 1992, **4** (9): 1053~1062
- 25 Cheung A Y, May B, Kawata E E, *et al.* Characterization of cDNAs for stylar transmitting tissue-specific proline-rich proteins in tobacco. *Plant J*, 1993, **3** (1): 151~160
- 26 Hong J C, Nagao R T, Key J L. Developmentally regulated expression of soybean proline-rich cell wall protein genes. *The Plant Cell*, 1989, **1** (9): 937~943
- 27 Wyatt R E, Nagao R T, Key J L. Patterns of soybean proline-rich protein gene expression. *The Plant cell*, 1992, **4** (1): 99~110
- 28 Schmidt J S, Lindstrom J T, Vodkin L O. Genetic length polymorphisms create size variation in proline-rich proteins of the cell wall. *Plant J*, 1994, **6** (2): 177~186
- 29 John M E, Keller G. Characterization of mRNA for a proline-rich protein of cotton fiber. *Plant Physiol*, 1995, **108** (2): 669~676
- 30 Orford S J, Timmis J N. Abundant mRNAs specific to the developing cotton fibre. *Theor Appl Genet*, 1997, **94** (6~7): 909~918
- 31 Tan H, Creech R G, Jenkins J N, *et al.* Cloning and expression analysis of two cotton (*Gossypium hirsutum* L.) genes encoding cell wall proline-rich proteins. *DNA Seq*, 2001, **12** (5~6): 367~380
- 32 Feng J X, Ji S J, Shi Y H, *et al.* Analysis of five differentially expressed gene families in fast elongating cotton fiber. *Acta Biochim Biophys Sin*, 2004, **36** (1): 51~56
- 33 Ruan Y L, Llewellyn D J, Furbank R T. Suppression of sucrose synthase gene expression represses cotton fiber cell initiation, elongation, and seed development. *The Plant Cell*, 2003, **15** (4): 952~964
- 34 Arpat A B, Waugh M, Sullivan J P, *et al.* Functional genomics of cell elongation in developing cotton fibers. *Plant Mol Biol*, 2004, **54** (6): 911~929
- 35 Sun Y, Fokar M, Asami T, *et al.* Characterization of the brassinosteroid insensitive 1 genes of cotton. *Plant Mol Biol*, 2004, **54** (2): 221~232
- 36 Ji S J, Lu Y C, Feng J X, *et al.* Isolation and analyses of genes preferentially expressed during early cotton fiber development by subtractive PCR and cDNA array. *Nucleic Acids Res*, 2003, **31** (10): 2534~2543
- 37 Wang S, Wang J W, Yu N, *et al.* Control of plant trichome development by a cotton fiber MYB gene. *The Plant Cell*, 2004, **16** (9): 2323~2334
- 38 Wang H Y, Yu Y, Chen Z L, *et al.* Functional characterization of *Gossypium hirsutum* profilin 1 gene (*GhPFN1*) in tobacco suspension cells. *Planta*, 2005, **222** (4): 594~603
- 39 Wu Y, Machado A C, White R G, *et al.* Expression profiling identifies genes expressed early during lint fibre initiation in cotton. *The Plant Cell Physiol*, 2006, **47** (1): 107~127
- 40 Shi Y H, Zhu S W, Mao X Z, *et al.* Transcriptome profiling, molecular biological, and physiological studies reveal a major role for ethylene in cotton fiber cell elongation. *The Plant Cell*, 2006, **18** (3): 651~664

Molecular Characterization and Expression Analysis of Five Novel Genes Encoding Proline-rich Proteins in Cotton (*Gossypium hirsutum*) *

XU Wen-Liang**, HUANG Geng-Qing**, WANG Xiu-Lan, WANG Hong, LI Xue-Bao***

(College of Life Sciences, Huazhong Normal University, Wuhan 430079, China)

Abstract Proline-rich cell wall proteins are widely spread in plants and are believed to function by modeling the architecture of the cell wall surrounding specific cell types. Five genes encoding proline-rich proteins were isolated from cotton cDNA libraries. The most common characteristic of these proteins is the abundant proline residues that occur in repeating motifs of at least two consecutive Pros. Based on amino acid composition, repetitive motifs and domain organization, the five members can be divided into two subgroups: one group similar to common PRPs including GhPRP3, GhPRP6, GhPRP5 and GhPRP4 was composed of two domains, an N-terminal hydrophobic domain (or signal peptide) followed by a proline-rich domain containing different proline-rich repetitive motifs; the other group different from common PRPs including GhPRPL lies in it contains an N-terminal hydrophilic domain, eight repetitive copies of pentapeptide (similar to PPKKE) lies in the C-terminal domain. Expression studies of the six *GhPRPs* have been carried out by quantitative realtime RT-PCR. The results showed that *GhPRP3* and *GhPRP5* were preferentially expressed in 10 dpa fiber, little transcripts was detected in other tissues examined. *GhPRPL* highly expressed in cotyledons, whereas smaller or negligible amounts of its transcripts were detected in other tissues. The remaining two genes, *GhPRP4* and *GhPRP6*, were expressed in all the tissues analysed, but their transcript level is different. *GhPRP4* mRNA is most abundant in hypocotyls, and then in anther, while *GhPRP6* expressed highly in fiber, and then in 10 dpa ovule. Furthermore, the results showed that the fiber-specific *GhPRP3* and *GhPRP5* were also developmentally regulated, suggesting that the genes may play important roles during cotton fiber development.

Key words proline-rich protein, gene expression pattern, fiber specific, *Gossypium hirsutum*

*This work was supported by grants from The National Natural Science Foundation of China (30400022), The National Basic Research Program of China (2004CB117304) and Chenguang Project for Young Scholars in Wuhan City(20065004116-26).

** The first two authors contribute equally to this work.

***Corresponding author . Tel/Fax: 86-27-67862443, E-mail: xbli@mail.ccnu.edu.cn

Received: November 3, 2006 Accepted: January 16, 2007