

番木瓜 proteinase omega 基因启动子的克隆 及功能初步研究

杨英军^{1,2}, 周 鹏¹

(1 中国热带农业科学院热带生物技术研究所、热带作物生物技术国家重点实验室,
海南 海口 571101; 2 河南科技大学, 河南 洛阳 471003)

摘要: 采用 PCR 技术从番木瓜基因组中克隆了 proteinase omega 基因的部分序列及其 5 侧翼序列。序列分析表明, 克隆到的基因序列与 GenBank 中的序列同源性为 96%, 长 1039 bp 的 5 端侧翼序列在 GenBank 数据库中没有同源片段。预测 5 端侧翼序列有两处基础启动子区域, 转录起始位点 (TSS) 分别为 A, T。在基础启动子区域都存在 TATA-box, 上游发现多处 CAAT-box, G-box, I-box 等顺式作用元件和 AT 富含区。构建了植物表达载体并用基因枪轰击番木瓜的叶组织, GUS 基因瞬间表达结果表明, 该长 1 039 bp 的 5 端侧翼序列具有驱动 GUS 基因在乳管中表达的功能。该启动子的发现对进一步研究启动子的功能和开发番木瓜作为生物反应器具有重要意义。

关键词: 番木瓜; proteinase omega; 启动子; 克隆; 序列分析

中图分类号: Q 75 文献标识码: A 文章编号: 0253-2700 (2005) 05 - 0545 - 07

Cloning and Functional Analysis of a Novel Proteinase Omega Promoter from *Carica papaya*

YANG Ying-Jun^{1,2}, ZHOU Peng^{1**}

(1 National Key Biotechnology Laboratory for Tropical Crops, Institute of Bioscience and Biotechnology, CATAS,
Haikou, 571101, China; 2 Henan University of Science and Technology, Luoyang 471003, China)

Abstract: Proteinase omega was a kind of proteinase in papaya (*Carica papaya* L.) laticifers. A 5 flanking sequence and partial gene sequence of proteinase omega were isolated from the genomic DNA via PCR technology. DNA sequence analysis and homology comparison indicated that it had 96% homology with the proteinase omega gene which had been submitted to GenBank. The two core promoter regions and some upstream regulatory elements in the fragment were analysed using the software of PROMOTER PREDICTION and PLANT CARE. Transcriptional start sites (TSS) were A, T respectively. TATA-box, CAAT-box, G-box, I-box, AT rich regions and other cis-elements were found in the promoter in identical positions common to all promoter sequence regions. Compared with the data in GenBank, the results

基金项目: 国家“863”计划资助课题 (2002AA241141), 国家转基因植物研究与产业化开发 (JY04-B-02) 和中国热带农业科学院基金资助课题 (RKY0419)

通讯作者: Corresponding author. E-mail: zhqtw@public.hk.hi.cn. Tel: 0898-66988564

收稿日期: 2005-03-11, 2005-04-18 接受发表

作者简介: 杨英军 (1968-) 男, 副教授, 在职博士研究生, 主要从事植物生物反应器的研究工作。

showed that a new promoter was obtained, and its GenBank accession number was AY695444. Binary vectors were constructed through fusing 1039 bp 5' flanking region of proteinase omega gene with the GUS gene, Transient GUS expressions were observed in papaya leaves transferred via particle bombardment. GUS activities were detected only in laticiferous cells.

Key words: *Carica papaya*; Proteinase omega; Promoter; Cloning; Sequence analysis

番木瓜 (*Carica papaya* L.) 是著名的热带果蔬, 鲜果外观美丽, 色香味俱佳, 素有“岭南佳果”之称, 在我国南方地区大面积种植。除了食用之外, 番木瓜还具有极其重要的工业应用价值, 含量丰富的木瓜蛋白酶类物质在科研、食品加工、医疗、化妆品、美容等方面有着十分广泛的应用 (夏杏洲和胡雪琼, 2002)。大量研究表明番木瓜是一种集营养、防病、抗癌、保健、美容于一体的有百益而无一害的果之珍品 (夏杏洲和胡雪琼, 2002; 刘因华和曹志红, 2004; Pendzhiev, 2002; Mikhailchik 等, 2004)。

番木瓜酶学及相关生物技术研究亦是近年来相当活跃的一个领域 (吴显荣, 1999)。proteinase omega (E.C.3.4.22.30, 也称为 papaya peptidase A) 是木瓜胶乳主要的 4 种巯基蛋白酶 (木瓜蛋白酶、木瓜凝乳蛋白酶、甘氨酸内切酶和 proteinase omega) 中重要的一种, 有助于分解胃肠道中的动物蛋白质, 具有极好的助消化功效。由于纯化较为困难, 使得它的研究晚于其他的蛋白酶类。1988 年, Dubois 等分析了它的多肽链氨基酸组成, 发现与 papaya 的同源性为 69.8% (Dubois 等, 1988), 后来一些学者相继纯化得到了 proteinase omega, 并根据氨基酸序列建立了三维结构模型 (Topham 等, 1991; Sumner 等, 1993), 1998 年进一步研究了酶学动力学 (Yoshihide 等, 1998)。

proteinase omega 蛋白在乳管中特异表达的特性使番木瓜很适合作为生物反应器, 即利用基因工程技术, 构建具有乳管特异表达活性的外源基因转化载体, 将外源基因转入番木瓜并在乳管中特异表达, 从而使番木瓜变为药物生产工厂, 张更林等 (2003) 就番木瓜作为抗结核病口服疫苗作了尝试性工作。目前, 虽然该酶已进行了克隆和转化研究 (Revell 等, 1993), 但基因的表达调控方式的研究尚未见到报道, 因此, 分离、克隆高效表达的、分泌量大的 proteinase omega 基因启动子, 揭示基因的表达及调控的分子机制, 构建高效的表达载体, 提高外源基因表达水平, 对大量表达外源药物蛋白, 开发番木瓜作为一种植物疫苗具有十分重要的意义。

1 材料与方法

1.1 材料及菌种

番木瓜品种为“穗中红” (*Carica papaya* L. 'Sui Zhong Hong') 取自本所温室。大肠杆菌 *Escherichina coli* DH5 为本室保存。

1.2 试剂

克隆载体 pGEM-T-easy、Wizard DNA clear up 回收试剂盒购自 Promega 公司, 限制性内切酶、T₄-DNA 连接酶、Genome Walking Kit 等购自大连宝生物公司, 其他生化试剂为国产分析纯。

1.3 方法

1.3.1 启动子的克隆

1.3.1.1 总 DNA 的提取、PCR 引物的设计与合成 以番木瓜幼嫩的叶片为试材, 采用改良的 CTAB 法

在 GenBank 数据库中没有任何同源片段, 表明该片段为 proteinase omega 基因的 5 端侧翼序列, 并且为首次克隆到。递交 GenBank 数据库, 登录号为 AY695444。

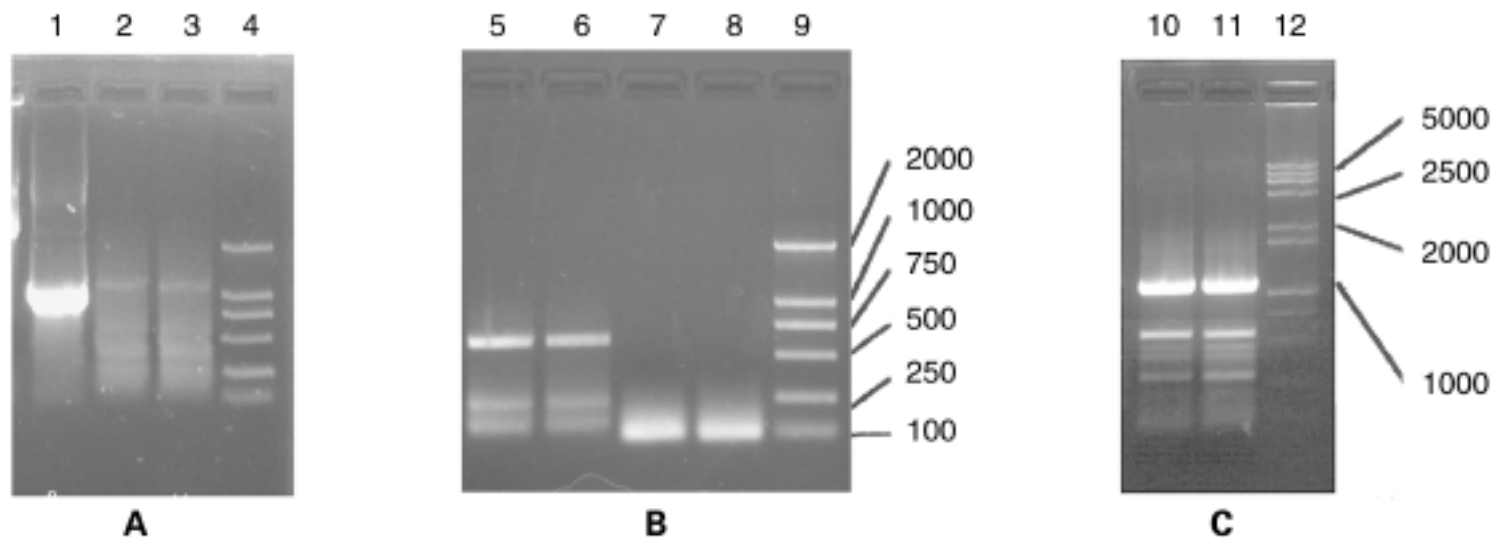


图 1 番木瓜基因组 PCR 扩增结果

A. PCR 扩增获得 5B-700; B. PCR 扩增获得 1B-600 产物; C. 以番木瓜基因组 PCR 扩增得到 proteinase omega 基因及其 5 端侧翼序列; 1. 泳道为扩增得到的 5B-700; 5, 6. 泳道为扩增得到的 1B-600 产物; 2, 3, 7, 8. 泳道为巢式 PCR 第一次扩增的结果; 10, 11. 泳道为扩增得到的 proteinase omega 基因及其 5 端侧翼序列; 4, 9, 12. 泳道为分子量标准 (分别为 DL2000, DL2000, DL15000+2000)

Fig. 1 Agarose gel electrophoresis of PCR products amplified from the papaya genomic DNA

A. The nested PCR products amplified from the genomic DNA digested with *Xba*I; B. The nested PCR products amplified from the genomic DNA digested with *Eco*RI; C. The gene and promoter fragment amplified from the papaya genomic DNA
1, 5, 6 lane: The second PCR products, 5B-700 and 1B-600 fragments; 2, 3, 7, 8 lane: The first PCR products; 10, 11 lane: The gene and promoter fragment, 1100 bp; 4, 9, 12 lane: Marker (DL2000, DL2000, DL15000+2000)

采用启动子预测软件 plant CARE 对该启动子序列进行分析, 结果表明, 在翻译起始密码子上游存在着多个可能的 TATA-box (TATA-box 的一致序列为 TATA (A T) ATA) 和 CAAT-box。再利用 Berkeley Drosophila Genome Project 软件分析, 发现全序列共有两个基础启动子区, 分别位于 905~955 bp, 947~997 bp 区 (图 2 的单划线区), 可能性分别为 0.79 和 0.66, 在这两个基础启动子区的前端分别具有启动子的另一个关键的启动子元件 CAAT-box。进一步对 5 上游远端调控序列进行分析, 还发现多个 G-box, I-box, AT1-motif 等光响应元件, 其中 G-box 序列是广泛存在于植物基因启动子区的一类顺式调控元件, 是 bZIP 蛋白与核 DNA 相结合的位点, 在基因的转录调节中占着十分重要的地位 (Harter 等, 1994; Kircher 等, 1999)。另外还发现许多明显的逆境应答调控元件, 如 ERE 是乙烯响应元件, ABRE 是 ABA 响应元件, HSE 是热激响应元件, WUN-motif 是伤害响应元件等重要的顺式作用元件。在此调控序列 830~838 bp 的负链中还含有一个的 AT 丰富序列 (TAAAtatt), AT 富含区广泛存在于植物启动子中, 能与植物细胞核内存在的 AT 丰富序列结合蛋白作用, 对基因的表达可能起调控作用 (Holk 等, 1996)。

2.3 启动子驱动 GUS 基因在番木瓜叶组织中的瞬时表达

以含有全长启动子序列的质粒为模板, 用引物 QDZ1、QDZ2 扩增获得了基因 5 端启动子序列约 1 039 bp PCR 产物 (图 3), 回收后分别与同样经过酶切的载体 pCAMBIA1301 大片段连接, 并将重组质粒命名为 pCAMBIA1039。

```

1  gaattccggc acggcggcgg ttcgcttcag atcggctatt ttcagccgaa tttggctgat tttggatcaa
71  tccctgcgag aaAtatgatt gcaggcgtgc cttgatgatg tagatataat aataacttgt atggcccaga
141 tgagaataat tctctgaatg taactAttg tgtaattatt tatggttatt gattttcaca ttaataatag
211 ctgttttggg aatttcattt tgaggattaa tttcactggt tccttttggg tctttccttt atttactata
281 attattataa ttgttgaaat cttctagaaa ttagaaattt tattatataat cTatatatta tgaaaattaa
351 ccacaaaaat ttatatattt attaatattg cttatattgt aacatttttt aataaatata ttgaTttatt
421 tagataatgt cattataatt cgattttatt gatggcatca actcataaaa aaactgttaa tatggagtat
491 atagatcggt ttattatatt taaaggcaat gttttttttt ttttaatttt tatattgaaa actttttGTT
561 TCTATGCAAT TTTTGCTCCC AGTAAATAAT TGTGAAAAGA CTATTCTAAA GTTATATCTT ATACATACCA
    S22
631 CGTGGAACAA ATTTAGTTTG ATAAATTTAA AACAGTGTA TTAANAATCAA ATTTTTTCTA AGATTTAATA
701 ATAAAATTA TTAANAAGTAA CATCTACTAA TATTAAACT AACACATGGC AAATATATAA TTATTACAAC
771 TCTTACACCT CGCGTCAAT TTTATTAACC AAATTAATCT TTTGTTGAAA TTCACTATT TGCATAATTA
841 AATATTTTAT CGAAACTAAA ACTTTTAAAT TAACAAGCGT GAATGAGAAG AAATTTTTTA TTTATTATA
911 TTTTTTTATA AAAAAACAA AGCTTATTGG AAACACAAGC TAGCTAGTAT ATAAACAAC ATGATCCTCC
981 CATTCTCTT CATCCCTTCA TACATTTGAG AACTAAAAA GTATATAGCT AGTGTCACAA TGGCTATGAC
1051 ATGTTCAATT TCAAAGTTGC TCTTTGTTGC AATATGTCTC TTTGTTTATA TGAGTGTGTC CTTTGGTGTG
1121 TTTTCTATTG TGGGGTATTG GCAAGATGAC TTGACATCTA CTGAGAGGCT TATTCAGCTA TTTGAATCGT
1191 GGATGTTGAA GCacaataaaa ttttacgaga atgttgatga gaaactttac agatttgaga tttttaagga
    s2
1261 caaattgacg agacaaacaa aaagaataac agttattggc ttgtctaaac tacgattaaa cgagtttgct
1331 gatttaagca atgatgaatt caatgagaag tatgttggtt cccttattga tgcaacaatt gaacaatcct
1401 atgatgaaga gtttattaat gaagatattg taaatttggc cgagaatgtc gattggagaa aaaaaggagc
1471 tgtcactccc gtaagacatc agggttcatg cggta
    P2

```

图 2 番木瓜 proteinase omega 酶基因的启动子序列及其分析

下划线为预测的基础启动子序列，较大的黑体字母为预测的转录起始位点，阴影部分为基因编码区，斜体部分为设计的引物序列，方框内为 TATA-box，CAAT-box 元件和翻译起始密码子 ATG

Fig. 2 The sequence of proteinase omega promoter and its analysis of *Carica papaya* L.

The gene code region, primers sequence, the core promoters predicted by the software of promoter prediction, the predicted transcription start sites (TSS), start code ATG and TATA-box, CAAT-box, are indicated by underlines, shadow, italic letters, single underline, black and large letters, and frame respectively

采用基因枪轰击法将植物表达载体 pCAMBIA1039、对照载体 pCAMBIA1301 和不用 DNA 包被的金粉作为空白对照转入番木瓜叶片组织，轰击后暗培养 2 d，经 X-Gluc 染色、脱色后，在显微镜下观察 GUS 组织化学染色情况。结果表明，启动子片段能驱动 GUS 基因

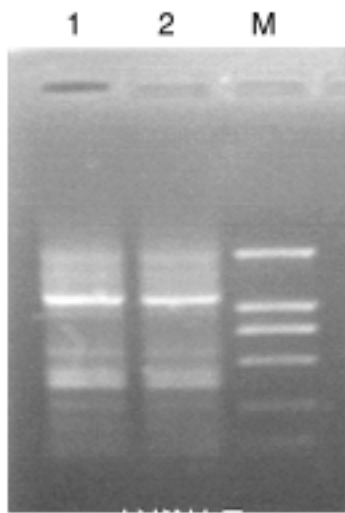


图3 启动子片段的扩增
Fig. 3 The amplification of promoter .1,2 .PCR products
M marker (DL2000)

的表达, 在轰击部位均出现较多的蓝色斑点 (图 4: B), 但 GUS 染色部位有一定差异, 主要分布在叶脉维管组织中, 在叶肉组织染色较少, 而 35S 驱动的 GUS 基因表达在叶脉、叶肉组织中没有较明显的差异 (图 4: A), 因此初步推断该 1 039 bp 的启动子可能具有乳管特异性。

3 讨论

植物是继细胞培养、动物乳腺组织之后做为另一种新型的生物反应器而日益受到人们青睐的, 植物启动子作为最主要的一种转录调节因子, 对于外源基因的表达非常重要。

本研究从“穗中红”番木瓜基因组中扩增获得了一个具有启动子特征的特异序列, 经克隆、测序及序列同源性搜索后证实, 其序列的 3 端与 proteinase omega 基因编码区同源性为 96%, 而 5 端没有同源片段, 是一个新的 5 端调控序列, 递交 GenBank 数据库, 登录号 AY695444。

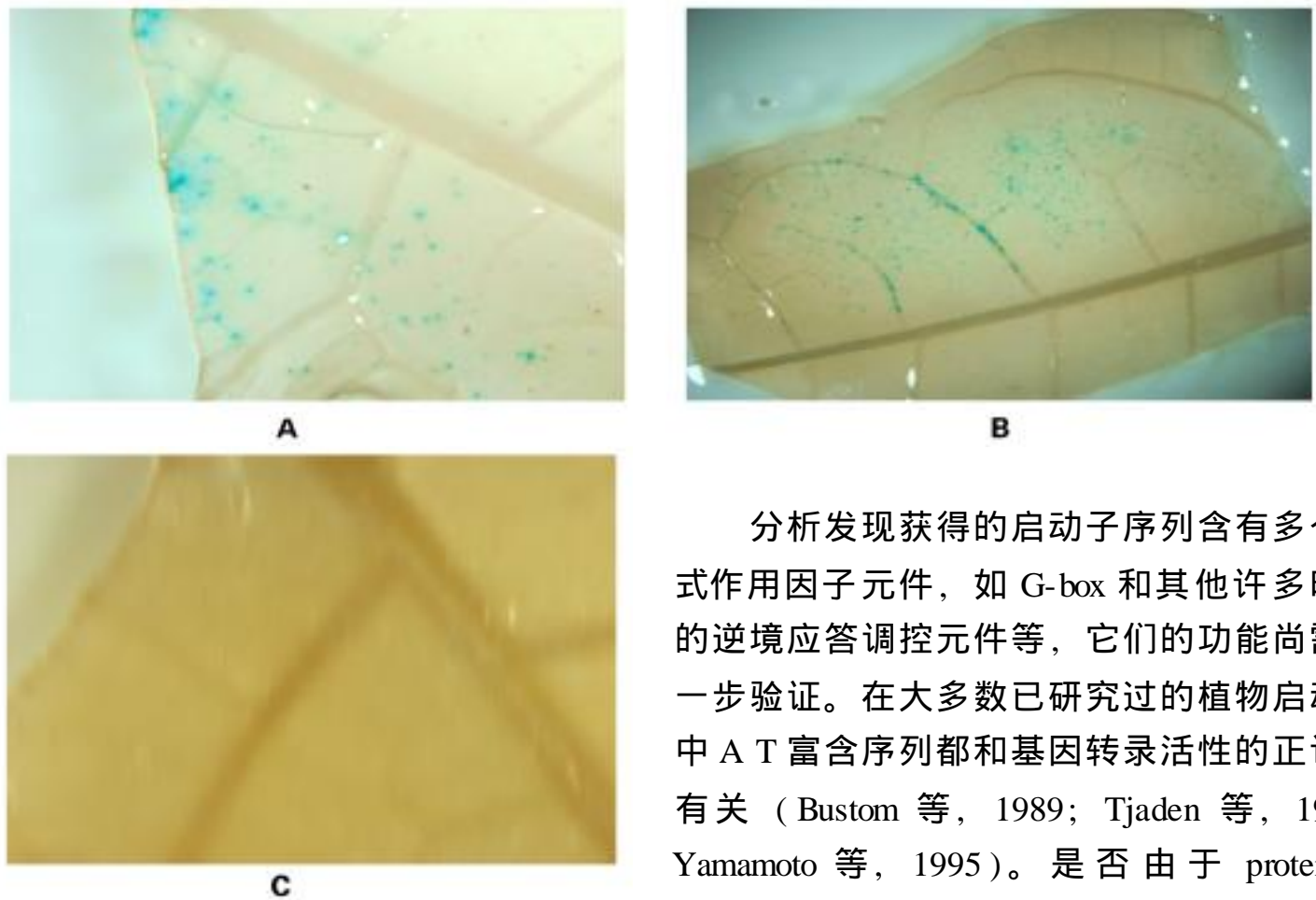


图4 启动子驱动 gus 基因在番木瓜叶片中的瞬时表达
A. 植物表达载体; B. pCAMBIA1301 的瞬时表达;
C. 空白对照

Fig. 4 GUS transient expression in bombarded papaya leaves
A. driven by CaMV 35S promoter; B. driven by laticiferous specific promoter; C. control

分析发现获得的启动子序列含有多个顺式作用因子元件, 如 G-box 和其他许多明显的逆境应答调控元件等, 它们的功能尚需进一步验证。在大多数已研究过的植物启动子中 A T 富含序列都和基因转录活性的正调控有关 (Bustom 等, 1989; Tjaden 等, 1994; Yamamoto 等, 1995)。是否由于 proteinase omega 基因启动子 830 ~ 838 bp 的负链中含有 A T 富含序列, 使得该启动子片段具有调控基因表达的功能, 还有待进一步的探讨。

proteinase omega 基因的启动子功能的瞬时表达研究中, 在 5 端调控序列调控下的 GUS 表达水平 (表达载体 pCAMBIA1039) 较

对照 pCAMBIA1301 中的 35S 启动子驱动的 GUS 基因表达为高, 而非转化组织中无 GUS 表达。观察发现 GUS 染色主要分布在叶脉的组织中, 初步推断该启动子可能具有乳管特异

性。分析启动子片段,发现 pCAMBIA1039 包括所预测的顺式作用元件和两个基础启动子序列,GUS 基因的较高表达水平及乳管特异性是否为在两个基础启动子的协同驱动下的结果,尚需进一步研究。

本研究的下一步工作是分离更长的 5 启动子区域的上游序列,并对该序列进行系列缺失,构建包含不同长度的启动子缺失片段的植物表达载体,通过转基因研究,确定增强子区和启动子活性区等,并在此基础上构建高效表达载体,驱动有重要免疫功能的外源蛋白的高效表达,开发番木瓜使其成为一种植物疫苗。

〔参 考 文 献〕

- 刘因华,曹志红,2004.抗艾滋病药物的发展近况[J].云南中医中药杂志,25(4):44—45
- 傅荣昭,孙勇如,贾士荣主编,1994.植物遗传转化技术手册[M].北京:中国科学技术出版社,131—132
- Busto MM, Guiltinan MJ, Jordano J, *et al*, 1989. Regulation of α -glucuronidase expression in transgenic tobacco plants by an AT-rich, *cis*-acting sequence found upstream of a French bean α -phaseolin gene [J]. *The Plant Cell*, 1 (9): 839—853
- Dubois T, Kleinschmidt T, Schnek AG, *et al*, 1988. The thiol proteinases from the latex of *Carica papaya* L. II. The primary structure of proteinase omega [J]. *Biol Chem Hoppe Seyler Aug*, 369 (8): 741—54
- Harter K, Kircher S, Frohnmeyer H, *et al*, 1994. Light-regulated modification and nuclear translocation of cytosolic G-box binding factors in parsley [J]. *Plant Cell*, 6: 545—559
- Holk A, Kaldenhoff R, Richter G, 1996. Regulation of an embryogenic carrot gene (DC2.15) and identification of its active promoter sites [J]. *Plant Mol Biol*, 31 (6): 1153—1181
- Jefferson A, 1987. Assaying chimeric genes in plants: the GUS gene fusion system [J]. *Plant Mol Biol Rep*, 5: 387—405
- Kircher S, Wellmer F, Nick P, *et al*, 1999. Nuclear import of the parsley bZIP transcription factor CPRF2 is regulated by phytochrome photoreceptors [J]. *J Cell Biol*, 144: 201—211
- Mikhailchik EV, Ivanova AV, Anurov MV, *et al*, 2004. Korkina. Wound-healing effect of papaya-based preparation in experimental thermal [J]. *Trauma Bulletin of Experimental Biology and Medicine*, 137 (6): 560—562
- Pendzhiev AM, 2002. Proteolytic enzymes of papaya: medicinal applications [J]. *Pharm Chem J*, 36 (6): 315—317
- Revell DF, Cummings NJ, Baker KC, *et al*, 1993. Nucleotide sequence and expression in *Escherichia coli* of cDNAs encoding papaya proteinase omega from *Carica papaya* [J]. *Gene*, 7 (2): 221—225
- Sumner IG, Vaughan A, Eisenthal R, *et al*, 1993. Kinetic analysis of papain proteinase omega [J]. *Biochem Biophys Acta*, 1164 (3): 243—51
- Tjaden G, Coruzzi GM, 1994. A novel AT-rich DNA binding protein that combines and HMG I-like DNA binding domain with a putative transcription domain [J]. *The Plant Cell*, 6 (1): 107—118
- Topham CM, Salih E, Frarazao C, *et al*, 1991. Structure-function relationships in the cysteine proteinases actinidin, papain and papaya proteinase omega. Three-dimensional structure of papaya proteinase omega deduced by knowledge-based modelling and active-centre characteristics determined by two-hydronic-state reactivity probe kinetics and kinetics of catalysis [J]. *Biochem J*, 280 (Pt 1): 79—92
- Wu XR (吴显荣), 1999. Papaya enzymes and their functions in cosmetic [J]. *Trop Crop Sci Tech (热带作物科技)*, (4): 69—70
- Xia XZ (夏杏洲), Hu XQ (胡雪琼), 2002. The development and its industrialization utilization of papaya resources [J]. *J Trop Crop Agric Sci (热带农业科学)*, 22 (4): 72—75
- Yamamoto S, Nishihara M, Monkawa H, 1995. Promoter analysis of seeds storage protein genes from *Canavalia gladiata* D. C [J]. *Plant Mol Biol*, 27: 729—741
- Yoshihide Ikeuchi, Nikolaos A, Katerelos, *et al*, 1998. Goodenough. The enhancing of a cysteine proteinase activity at acidic pH by protein engineering, the role of glutamic 50 in the enzyme mechanism of caricain [J]. *FEBS Letters*, 437: 91—96
- Zhang GL (张更林), Zhou P (周鹏), Guo AP (郭安平), *et al*, 2003. An initial study of transgenic *Carica papaya* used as a kind of vaccine for anti-*Tuberculosis* [J]. *Acta Bot Yunnan (云南植物研究)*, 25 (2): 223—229