生物化学与生物物理进展 *Progress in Biochemistry and Biophysics*, 2006, 33(7): 691~698
www.pibb.ac.cn

PIBB

人致肺纤维化相关因子的克隆和生物信息学分析 *

陈晓华 [1,2]** 蔡国平 [1,2]

([1]清华大学生物科学与技术系，北京 100084；[2]清华大学深圳研究生院生命科学学部， 深圳 518055)

摘要 几丁质酶是自然界广泛存在的一类降解几丁质的水解酶类，但是直至近些年才在哺乳动物体内发现存在有几丁质酶样蛋白. 早年曾于矽肺大鼠中纯化出矽诱导的支气管肺泡灌洗蛋白 iSBLP[58]，体外具有促进人胚肺成纤维细胞 2BS 增殖的作用，N端测序显示与哺乳动物几丁质酶蛋白家族成员具有高度同源性. 生物信息学分析表明，来源于人的结肠、肾和胃的几个表达序列标签 (EST) 克隆和大鼠的这一蛋白质序列匹配. 随后成功地从人肾 RNA 样品中克隆到一组 cDNA，其序列及相应的氨基酸序列彼此高度相似，并与 GenBank 中的几个人几丁质酶蛋白高度相似. 和人基因组序列比较，揭示这些分子可能来自于同一基因，为可变剪切的产物.

关键词 几丁质酶，表达序列标签(EST)，cDNA，生物信息学，可变剪切
学科分类号 Q7

几丁质，又称壳聚糖，为 β-1,4-N- 乙酰葡萄糖胺的线性多聚体，是自然界中仅次于纤维素大量存在糖的多聚物. 它是许多物种，如节肢动物，包括甲壳类和昆虫，以及软体动物、线虫和蠕虫的结构组分. 几丁质也发现存在于大部分真菌的细胞壁，含量由不足 1%至高达 40%不等. 几丁质酶 (EC 3.2.1.14)为 β-1,4-N- 乙酰葡萄糖胺内切酶，具有水解几丁质的功能，已确定存在于许多生物体内，包括那些体内不产生几丁质的生物. 在哺乳动物中尚未发现几丁质，但是和细菌、真菌、植物以及无脊椎动物几丁质酶同源的蛋白质在近些年来得到了确认， 如人 HC-gp39[1]、哺乳动物输卵管特异性糖蛋白(OGP)[2~4]、致敏小鼠 T 细胞分泌的 Ym1[5]，但这些蛋白质均缺乏降解几丁质的能力. 目前陆续在哺乳动物体内发现具有酶活性的几丁质酶蛋白， 首先发现的是在 Gaucher 患者血清中浓度升高的 chitotriosidase[6~8]，随后研究人员在小鼠和人体组织克隆出哺乳动物酸性几丁质酶(AMCase)[9]，在牛血清中纯化出由牛肝细胞表达的蛋白 CBPb04[10].

本实验室曾从矽肺大鼠纯化出一特异性单链多肽，命名为 iSPBLP[58]，或 iSPMF-p58，其表观分子质量为 58 ku，p*I* 为 4.5，体外具有促进人胚肺成纤维细胞 2BS 增殖的作用. 其 N 端的 15 个氨基酸，YNLVCYFTNWAQYRP，与和真菌、细菌以及植物几丁质酶同源的哺乳动物几丁质酶样蛋白，如 HC-gp39 和 OGP 具有同源性[11]. 研究显示，哺乳动物几丁质酶样蛋白在细胞对周围环境改变的应答中起着重要作用，参与正常生理反应，以及组织重建和退化相关的病理过程中的细胞 - 细胞级联反应和细胞 - 细胞因子相互作用. 我们推测 iSBLP[58] 可能在矽肺纤维化过程中起着促进成纤维细胞增殖的作用. 在哺乳动物中水解几丁质活性酶类的发现，揭示哺乳动物可能对含有几丁质组分的微生物具有特异的防御作用[9, 10, 12].

我们现已克隆出 iSBLP[58] 的全长 cDNA，并确定其具有水解几丁质的活性(序列号 AY486074 和 AAR28968，相关数据尚未发表). 由于其与小鼠 AMCase 高度同源，p*I* 为 4.5，我们将其命名为大鼠 AMCase. 组织表达检测显示，其 mRNA 在肺泡巨噬细胞以及胃和其他组织表达，预示该基因在体内具有重要功能. 结合生物信息学分析方法，我们从人肾组织总 RNA 样品中克隆到一组序列，这些序列具有一些显著特点，其存在的可能意义在本文中进行了初步探讨.

*国家自然科学基金资助项目(30270515).
** 通讯联系人.
Tel: 010-62777546, E-mail: cxh01@mails.tsinghua.edu.cn
收稿日期：2006-02-19，接受日期：2006-03-30

1 材料和方法

1.1 材料

1.1.1 组织总 RNA 样品购自 Clontech 公司.

1.1.2 工具酶和试剂. 各种限制性内切酶、dNTP、TaqDNA 聚合酶和 pfu 聚合酶购自 TaKaRa 公司；RT-PCR kit、质粒 pcDNA3.1D 和转染试剂 lipofectamine™ 2000 购自 Invitrogen 公司；pGEM-T easy 载体购自 Promega 公司；蛋白质印迹 (Western blot)使用的一抗为自制，二抗为 HRP 标记的羊抗兔多克隆抗体，购自 Dako 公司.

1.1.3 PCR 引物和测序. 引物由上海生工生物工程技术服务有限公司合成，测序由上海英俊生物技术有限公司完成.

1.2 方法

1.2.1 RT-PCR 扩增人致肺纤维化相关因子. 使用 NCBI 服务器上的 Blast 检索 GenBank 人 EST 数据库，发现有人的 EST 克隆和大鼠 iSBLP[58] 匹配，如序列号 BI760250、BI517797、BI518221 和 BI761276 等，这些 EST 序列分别来源于人结肠、肾和胃组织. 我们推断，与大鼠 iSBLP[58] 同源的人 cDNA 序列和人 AMCase 高度相似. 使用 oligo(dT)$_{12\sim18}$ 进行逆转录后，引物 HS1，5′ cat gac aaa gct tat tct cct c 3′和 HAS1，5′ agg tca ggt tta tgc cca gtt g 3′用于 PCR，我们从肾组织得到了 PCR 产物，但是从结肠和胃组织未得到任何 PCR 产物. 由肾组织得到的 RT-PCR 产物的序列和基因 TSA1902-L、TSA1902-S (Saito 等提交，序列号为 AB025008 和 AB025009)[13]以及人 AMCase (Boot 等提交，序列号为 AF290004)[9]高度相似，因此我们接着使用表 1 中列出的引物和下游引物 HAS1 以及人肾组织逆转录产物进行 PCR，所得到的序列长度和序列号也在表 1 列出. 为证明这些序列的可靠性，我们使用另外几对引物进行巢式 PCR 扩增这些序列的部分片段，HexF1S、HexF2S、Hex1S 和 HS1 分别与 Ex8R，5′ gat caa caa gcc cag gct g 3′用于第一步扩增，得到的 PCR 产物再分别使用 HexF1S、HexF2S、Hex1S、HS1 和 Ex5R，5′ gag agt ttt cag ctg gct gtt c 3′以及 Ex5F，5′ cat tgg agg ctg gaa ctt c 3′和 Ex8R，5′ cag cct ggg ctt gtt gat c 3′进行第二步扩增.

Table 1 Specific primers used and the length and GenBank accession number of acquired sequence

	Forward primer	Name	Length	GenBank accession number
1	HS1: 5′ catgacaaagcttattctcctc 3′	CH1[1)]		
		CH2[1)]		
		CH3[1)]		
		CH5	1 214	AY911311
		CH8[1)]		AF290004
2	HexF1S: 5′ gaaacctcctcgtctgtgcac 3′	CH4	1 302	AY825504
		CH6[1)]	1 188	AB025009
		CH7[1)]	1 354	AB025008
3	HexF2S: 5′ gaagcctttgtgataaccacaga 3′	CH1[1)]		
		CH2	1 141	AY789445
		CH3[2)]	1 482	AY911310
		CH8[1)]		AF290004
4	Hex1S: 5′ gctttccagtctggtggtga 3′	CH1	1 341	AY789444
		CH8[1)]		AF290004

[1)] The sequence acquired was not the full sequence submitted. [2)] Later we cloned from human kidney library and obtained the sequence as it.

1.2.2 克隆和 RT-PCR 产物测序. 由 **1.2.1** 得到的 RT-PCR 产物与 pGEM-T easy 载体连接，筛选阳性克隆进行测序.

1.2.3 生物信息学分析. 同源性检索、基因组序列查找以及外显子内含子连接的确定，在 http://www.ncbi.nlm.nih.gov/BLAST/ 针对不同的数据库进行.

氨基酸序列预测使用 http://www.ncbi.nlm.nih.gov/gorf/gorf.html 进行；两个序列的比对使用 http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi；多重序列比对在 ftp://ftp.ebi.ac.uk/pub/clustalw/index.html?进行，使用的是 ClustalW 1.8 版本；基因启动子在 http://www.cbs.dtu.dk/services/Promoter/ 进行预测，版本为 Promoter2.0；转录因子结合位点查找使用 http://www.gene-regulation.com/pub/programs/alibaba2/index.html 中的 AliBaba2.1 程序和 http://www.cbrc.jp/research/db/TFSEARCH.html 中的 TFSEARCH 程序.

1.2.4 在真核细胞中的瞬时表达和蛋白质印迹检测. 上游引物 HuKFFor，5′ gtc gga tcc cac cat ggc caa gct cat tct 3′；HuKMFor，5′ gtc gga tcc cac cat ggt ttc tac tcc tga gaa c 3′和 HuKSFor，5′ gtc gga tcc cac cat gcg tga agc ttt tga gca 3′分别与下游引物 5′ gtc tct aga ggc cag ttg cag caa tta c 3′用于扩增包含编码区全长的 PCR 产物 ChF、ChM 和 ChS. PCR 产物和质粒 pcDNA3.1D 使用 *Bam*H Ⅰ和 *Xho* Ⅰ消化，然后进行连接转化和阳性克隆筛选，得到 ChF-pcDNA、ChM-pcDNA 和 ChS-pcDNA 阳性质粒，转染真核细胞为 cos-7，质粒 pcDNA3.1D 作为对照同时进行转染. 蛋白质印迹按标准操作进行[14].

2 结　果

2.1 哺乳动物酸性几丁质酶克隆结果与分析

从人肾组织总 RNA 样品中获得多个 RT-PCR 产物(表 1). 产物 CH6、CH7 和 CH8 的序列分别与 TSA1902-S、TSA1902-L[13]和人 AMCase[9]的序列一致；产物 CH1、CH2、CH3、CH4 和 CH5 以及前三者高度相似但是不尽相同. 结果的可靠性采用巢式 PCR 进行了确证.

Blast 分析这些序列显示，人 1 号染色体上的 contig NT_019273 包含这些序列，并且这些序列很可能是同一基因转录的 mRNA 前体的剪切变异体. 通过对基因组的检索还发现，序列号分别为 NM_201653和 AK098814 的另外两个序列也包含于该 contig 中，前者在哮喘患者的肺部表达上调[15]，后者为 Sugano 和 Suzuki 由胃粘膜克隆得到. 这些序列和人类基因组序列的比对还确定了该基因位于人 1 号染色体，并确定了外显子(图 1)，以及外显子 - 内含子接头，如表 2 所示. 除外显子 7 和外显子 8 以及外显子 1 和外显子 1a 之间的内含子，其他供体和受体位点均遵循 GT-AG 剪切规则(表 2).

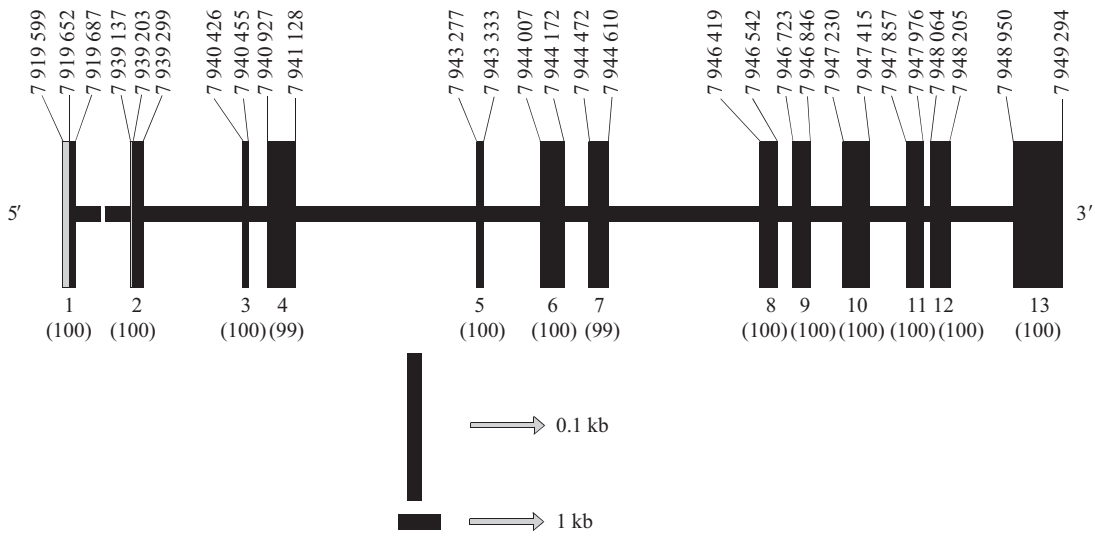

Fig. 1　Exon-intron structure of the deduced gene

Exons and introns are represented by boxes and horizontal lines, respectively. The swarthy areas of the first and second exons represent the variant exons imposed. The start and stop location of each exon in contig NT_019273 is indicated on the top, and the similarity of each exon to corresponding genomic sequence is indicated as % in ( ) below the exon number.

Table 2 Exon-intron junctions

Exon number	Exon size/bp	3′ exon junction	Splicing donor	Intron size/bp	Splicing acceptor	5′ exon junction	Exon number
1	89	-ccatagtatg	**GT**GAGTGTAA-	19 449	-CTCTATTT**AG**	aagcctttgt-	2
1	89	-atgacaaagc	**GT**GAGTGTAA-	19 517	-GACTGCAA**CC**	aagcctttgt-	2a
2	93	-ctcctcacag	**GT**GGGTTTGT-	1 196	-TTCTATCC**AG**	gtcttgtcct-	3
3	30	-ttgcagctcg	**GT**AAGTCATG-	471	-CTACACAC**AG**	gctctgccta	4
4	202	-tgaaaaataa	**GT**AGGATGAG-	2 148	-TGTTTTAC**AG**	gaacagccag-	5
5	57	-ggactgcccc	**GT**AAGTCTTC-	673	-TCTCCCTC**AG**	tttcactgcc-	6
6	166	-cctggtgcag	**GT**GGGGAAGG-	298	-TATTTTAC**AG**	tgaggaaact-	7
7	139	-tattactgag	**GT**ACATATTCT-	1 807	-TTATTCTG**TA**	gaaatgcgtg-	8
8	125	-aactgtcaca	**GT**GAGTGATG-	180	-CCCACCTC**AG**	gtacctggac-	9
9	124	-cctcaatgtg	**GT**GAGTCCCT-	383	-GACATTGC**AG**	gattatgtca-	10
10	186	-ttactacgag	**GT**ATGTAGAT-	441	-CTTTTGAA**AG**	atctgtacct-	11
11	120	-cgatattaag	**GT**AAGATCAG-	87	-CTTTAAAC**AG**	gctcaatggc-	12
12	142	-cagagtgcaa	**GT**AAGTGACT-	744	-TATGTTTC**AG**	gttgcacggc-	13
13	345						

具有完整 CDS 的分子，CH1、CH2、CH3 和 CH5，其氨基酸序列使用 ORF finder 程序进行了预测. 我们在假设这些分子均可翻译成蛋白质的前提下，对这些氨基酸序列和 CH6、CH7、CH8 和 CH9 的氨基酸序列进行了相互比较，发现这些 cDNA 序列可分为如下 3 类 (图 2 和图 3)：a. CH8，即人 AMCase，含有外显子 4、5、6 和 8 而不含有外显子 7，从外显子 2 开始编码氨基酸，全长为 476 个氨基酸，含有典型的信号肽序列，由 N 端 21 个氨基酸组成，成熟的蛋白质则分泌到细胞外；b. CH1、CH7 和 CH9，不含有外显子 4 和 7，但含有外显子 6，编码氨基酸从外显子 6 开始，氨基酸序列不具有信号肽序列，全长为 368 个氨基酸；c. CH2、CH3、CH5 和 CH6，不含有外显子 6 或含有

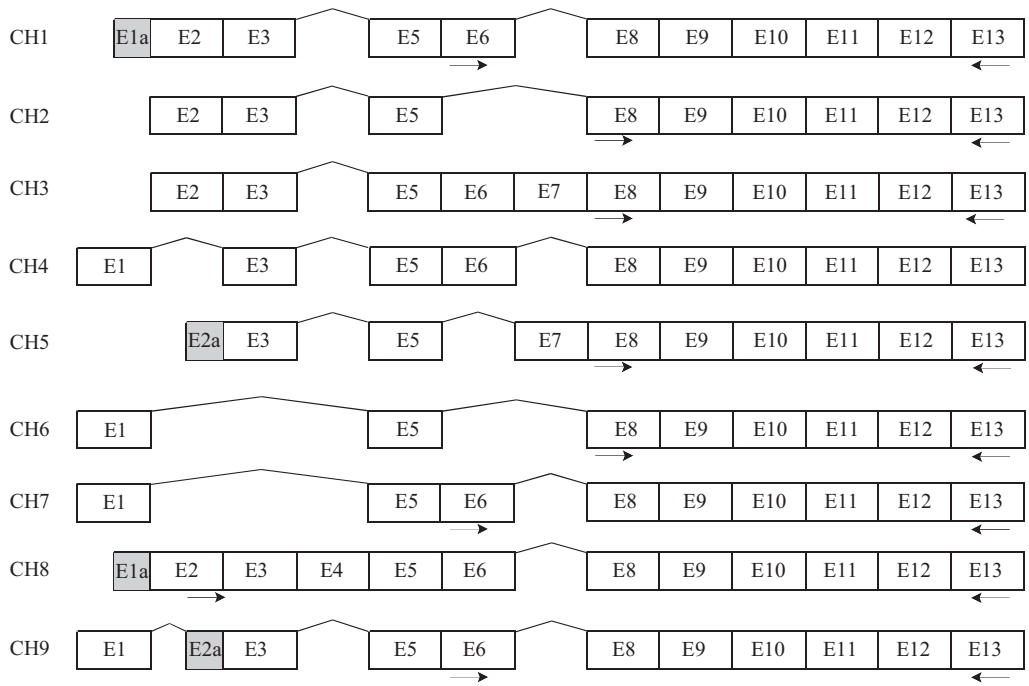

Fig. 2 Schematic representation of the alternatively spliced RNA molecules derived from pre-mRNA molecules (top line) and the protein isoforms predicted to be translated from these splice variants (bottom line)

The coding region of those that have complete CDS are indicated by arrows. Note that (1) CH8, human AMCase, ID AF290004 that includes exon 4, 5, 6, 8 and excludes exon 7 codes amino acids from exon 2; (2) CH1, CH7 and CH9 that exclude exon4 and exon7 but include exon6 code amino acids from exon6; (3) CH2, CH3, CH5 and CH6 that exclude exon 6 or include exon 7 code amino acids from exon8. Their translations are all stop at the same site at exon 13.

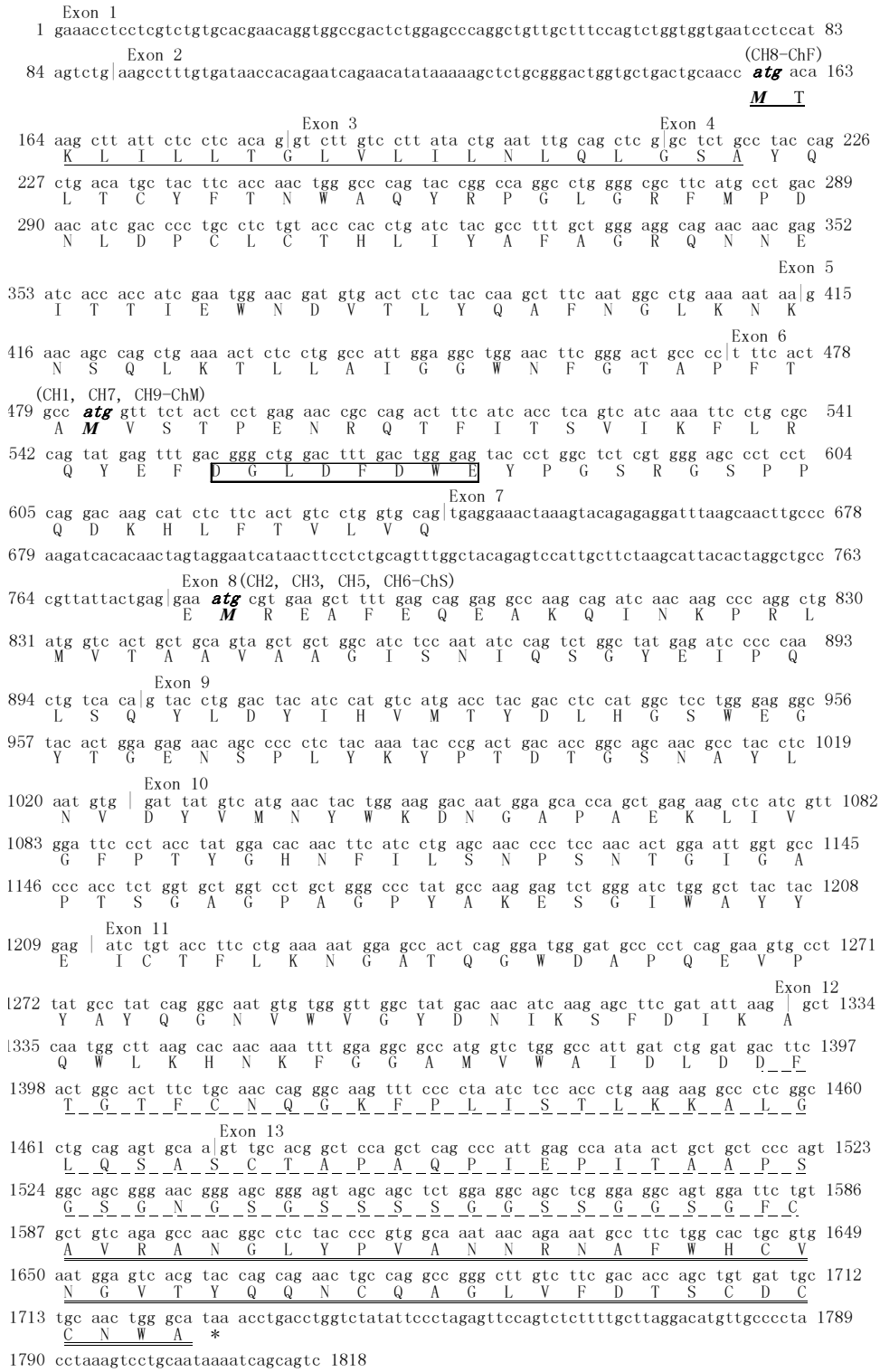

Fig. 3 The cDNA and amino acid sequences of the CH1-3, 5-9 that have complete CDS, which reflect presence of splice variants

The positions of splice sites are indicated by vertical lines and were elucidated by reference to homo sapiens chromosome 1 genomic contig NT_019273.17 (Hs1_19429). The bold italic fonts represent positions of the respective transcript start sites. The signal peptide, the catalytic center for chitinase activity, the hinge reqion and the chitin binding domain are indicated by single underline, box line, dashed underline and double underline respectively. Their amino acid sequences are same from C-terminal with different N-amino acid initiation. All these protein molecules consist of N-terminal region and C-terminal acid chitin-binding domain of 45 amino acids long and 65 amino acids long hinge region in-between, the chitin-binding domain and the hinge region are same to each other, the catalytic regions are same from their C-terminal with different N-terminal amino acid initiation site。Their translations are all stop at the same site at exon 13.

外显子 7，编码氨基酸从外显子 8 开始，其氨基酸序列不具有信号肽序列，全长为 315 个氨基酸. 这些 cDNA 的编码均终止于外显子 13 的同一位点处. 这些蛋白质序列由 C 端起是相同的，只是 N 端氨基酸起始位点不同. 保守结构域检索[16]显示，这些蛋白质均具有 N 端酶催化活性部位和 C 端由 45 个氨基酸组成的结合几丁质的结构域，两者之间由 65 个氨基酸组成的铰链区相隔，结合几丁质的结构域和铰链区相互一致，酶催化活性部位从 C 端起相同，只是 N 端氨基酸起始位点不同，如图 3 所示. 核酸和氨基酸序列的比较揭示，尽管它们的 N 端所使用的外显子不同，但是编码框却未发生移位.

包含外显子 1 和外显子 1 前 2 kb 的序列用于启动子分析，结果显示，未发现启动子，同一序列用于转录因子结合位点检索，在外显子 1 上游未发现典型的 TATA 和 CAAT 盒结构.

2.2 在真核细胞中的瞬时表达

我们构建了 3 种表达载体，转染 cos-7 细胞后进行蛋白质印迹检测，显示重组蛋白表达方式有 3 种：a. 引物 HuKFFor 和下游引物构建 ChF-pcDNA 的重组蛋白分泌到培养基中，分子质量约为 55 ku；b. 引物 HuKMFor 和下游引物构建 ChM-pcDNA 的重组蛋白只在细胞内表达，不分泌到胞外，分子质量约为 45 ku；c. 引物 HuKSFor 和下游引物构建 ChS-pcDNA 的重组蛋白只在细胞内表达，不分泌到细胞外，分子质量约为 40 ku (图 4).

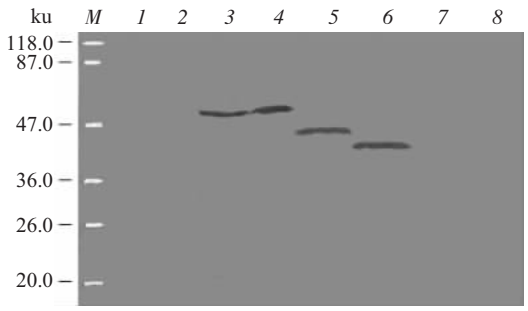

Fig. 4 Western blot analysis of recombinant protein

M: Molecular mass marker of protein, the molecular mass as indicated at the left; *1*: Supernatant of cos-7 transfected with pcDNA3.1D; *2*: Cell lysis of cos-7 transfected with pcDNA3.1D; *3*: Supernatant of cos-7 transfected with ChL-pcDNA; *4*: Cell lysis of cos-7 transfected with ChL-pcDNA; *5*: Cell lysis of cos-7 transfected with ChM-pcDNA; *6*: Cell lysis of cos-7 transfected with ChS-pcDNA; *7*: Supernatant of cos-7 transfected with ChM-pcDNA; *8*: Supernatant of cos-7 transfected with ChS-pcDNA.

3 讨 论

结合生物信息学分析方法，我们从人肾组织总 RNA 样品中克隆到一组序列，根据分析，这组序列极有可能来自于同一个基因，由于可变剪切而形成不同的剪切变异体，这些剪切变异体组成性使用该基因中的 3′端外显子，而 5′端外显子使用情况变化较大 (图 2). 分析显示，除外显子 7 和外显子 8 以及外显子 1 和外显子 1a 之间的内含子，其他供体和受体位点均遵循 GT-AG 剪切规则(表 2). 目前在大部分真核动物基因中发现的剪切大部分都符合经典的 GT-AG 规律，但是也存在少数与 GT-AG 不同的剪切，如在外显子 7 和 8 之间的内含子剪切为 GT-TA，外显子 1 和 2a 之间剪切为 GT-CC，在我们得到的这些 cDNA 序列中，外显子 1 和 2 使用情况较其他 5′端外显子更为复杂，究竟是什么因素导致这些不同的剪切方式，还有待于更深入地进行分析和研究.

尽管存在这些剪切变异体，但是由这些剪切变异体翻译出蛋白质则主要有 3 种，本文标记为 ChF、ChM 和 ChS，这些蛋白质的氨基酸序列由 C 端起是相同的，只是 N 端氨基酸起始位点不同. 保守结构域检索[16]显示，这些蛋白质均具有 N 端酶催化活性部位和 C 端由 45 个氨基酸组成的结合几丁质的结构域，两者之间由 65 个氨基酸组成的铰链区相隔，结合几丁质的结构域和铰链区相互一致，酶催化活性部位从 C 端起相同，只是 N 端氨基酸起始位点不同. 核酸和氨基酸序列的比较揭示，尽管它们的 N 端所使用的外显子不同，但是编码框却未发生移位 (图 2 和图 3). ChF 和 ChM 的 N 端含有不同几丁质酶酶活性必需的催化活性中心的保守活性位点残基(DXXDXDXE)，而 ChS 缺乏成熟 ChL 蛋白的 N 端 100 个氨基酸 /ChM 的 N 端 53 个氨基酸，不具备这一活性位点(图 3). 虽然 ChF 和 ChM 都具有酶催化活性中心的保守活性位点 (图 3)，但 ChM 蛋白缺乏成熟 ChF 蛋白 N 端 87 个氨基酸，这种序列的不同有可能导致这两者酶活性存在差异. 根据这三者的氨基酸序列，可以推测 ChF 蛋白是分泌到胞外的，而 ChM 和 ChS 则在细胞内执行功能，瞬时表达产物蛋白质分析也证实了这一点. 由此可以推测这三者在生物学功能上也极可能各具特点.

Saito 等[13]采用原位杂交证实编码 TSA1902 (CH6 和 CH7) 的基因位于染色体 1p13. 和基因组序

列比对显示本文的结果与之一致. 针对外显子上游序列进行的启动子分析尚不能确定启动子区域. 我们通过生物信息学分析，进而克隆得到了这些cDNA 序列，目前尚有很多问题亟需解决，例如：如果这些分子是同一初始转录体的不同剪切方式的产物，那么有哪些因子参与可变剪切？目前发现，越来越多的剪切变异体，其中一些已翻译成蛋白质执行精确的生物学功能[17~19]，而有一些则通过mRNA 的 nonsense-mediated decay 调节转录表达水平[20, 21]，CH8，即人 AMCase 已确定以蛋白质形式执行功能，CH9，很可能也是以蛋白质形式执行功能. 但是，我们得到的这些分子是否翻译成蛋白质执行生理学功能或参与病理过程，还是不翻译成蛋白质而只是以 RNA 形式参与表达调节或其他细胞活动，以及这些分子由哪类细胞表达，分子的亚细胞定位等等，均需要进行进一步的细胞和功能研究.

如果这些分子翻译成蛋白质，那么参与转录和翻译的调节因子在不同组织和不同时期应有所不同. 蛋白质产物具有相同的 C 端几丁质结合结构域以及铰链区，但是不同的 N 端氨基酸起始位点，有可能导致不同全长的蛋白质在细胞和机体内具有不同的稳定性和几丁质酶催化活性，三维晶体结构研究可能来回答这些问题.

氨基酸序列起始位点的多样性也预示它们在机体内具有不同的功能. 我们观察到，在人肾组织 CH1～8 具有不同水平的表达以及 CH1 和 CH3 在矽肺患者肺泡巨噬细胞的表达 (数据未发表)；Saito 等[13]检测到 TSA1902 (CH6 和 CH7) mRNA 在肺组织特异性表达，但是我们未能从人肺组织 cDNA (购自 Clontech 公司)克隆到该基因. CH8，即人 AMCase 在胃组织大量表达，可能参与胃肠道食物的消化，而它在肺组织较低水平表达表明它有可能参与免疫防御[9]. CH9，在哮喘患者肺部表达大大上调，被证实参与 Th2 细胞介导的哮喘[15]. 而我们在矽肺患者肺泡巨噬细胞观察到 CH1 和 CH3 的表达，显示这些分子可能在体内还具有免疫防御以外的其他重要功能.

有助于了解这些分子的生物学功能以及在矽肺和其他疾病中的作用.

参 考 文 献

1 Hakala B E, White C, Recklies A D. Human cartilage gp-39, a major secretory product of articular chondrocytes and synovial lining cells, is a mammalian member of a chitinase protein family. J Biol Chem, 1993, **268** (34): 25803～25810

2 Sendai Y, Abe H, Kikuchi M, *et al*. Purification and molecular cloning of bovine oviduct-specific glycoprotein. Biol Reprod, 1994, **50** (4): 927～934

3 Sendai Y, Komiya H, Suzuki K, *et al*. Molecular cloning and characterization of a mouse oviduct-specific glycoprotein. Biol Reprod, 1995, **53** (2): 285～294

4 Suzuki K, Sendai Y, Onuma T, *et al*. Molecular characterization of a hamster oviductspecific glycoprotein. Biol Reprod, 1995, **53** (2): 345～354

5 Guo L, Johnson R S, Schuh J C. Biochemical characterization of endogenously formed eosinophilic crystals in the lungs of mice. J Biol Chem, 2000, **275** (11): 8032～8037

6 Hollak C E, van Weely S, van Oers M H, *et al*. Marked elevation of plasma chitotriosidase activity, a novel hallmark of Gaucher disease. J Clin Invest, 1994, **93** (3): 1288～1292

7 Renkema G H, Boot R G, Muijsers A O, *et al*. Purification and characterization of human chitotriosidase, a novel member of the chitinase family of proteins. J Biol Chem, 1995, **270** (5): 2198～2202

8 Boot R G, Renkema G H, Strijland A, *et al*. Cloning of a cDNA encoding chitotriosidase, a human chitinase produced by macrophages. J Biol Chem, 1995, **270** (44): 26252～26256

9 Boot R G, Blommaart E F C, Swart E, *et al*. Identification of a novel acidic mammalian chitinase distinct from chitotriosidase. J Biol Chem, 2001, **276** (9): 6770～6778

10 Suzuki M, Morimatsu M, Yamashita T, *et al*. A novel serum chitinase that is expressed in bovine liver. FEBS Lett, 2001, **506** (2): 127～130

11 Guoping C, Fan P, Jingxi S, *et al*. Purification and characterization of a silica-induced bronchoalveolar lavage protein eith fibroblast growth-promoting activity. J Cell Biochem, 1995, **67** (2): 257～264

12 van Eijk M, van Roomen C P, Renkema G H, *et al*. Characterization of human phagocyte-derived chitotriosidase, a component of innate immunity. Int Immunol, 2005, **17** (11): 1505～1512

13 Saito A, Ozaki K, Fujiwara T, *et al*. Isolation and mapping of a human lung-specific gene, TSA1902, encoding a novel chitinase family member. Gene, 1999, **239** (2): 325～331

14 萨姆布鲁克 J, 弗里奇 E F, 曼尼阿蒂斯 T. 金冬雁, 黎孟枫, 等译. 分子克隆实验指南. 第二版. 北京: 科学出版社, 2002. 891～898
Sambrook J, Fritsch E F, Maniatis T. Translated by Jin D Y, Li M F, *et al*. Molecular Cloning: A Laboratory Manual. 2nd. Beijing: Science Press, 2002. 891～898

15 Zhu Z, Zheng T, Homer R J, *et al*. Acidic mammalian chitinase in asthmatic Th2 inflammation and IL-13 pathway activation. Science, 2004, **304** (5677): 1678～1682

16 Marchler-Bauer A, Bryant S H. CD-Search: protein domain annotations on the fly. Nucleic Acids Res, 2004, **32** (Web Server issue): W327～331

17 Miki T, Bottaro D P, Fleming T P, *et al*. Determination of ligand-binding specificity by alternative splicing: two distinct growth factor receptors encoded by a single gene. Proc Natl Acad

Sci USA, 1992, **89** (1): 246～250

18 Martin M M, Willardson B M, Burton G F, *et al*. Human angiotensin Ⅱ type 1 receptor isoforms encoded by messenger RNA splice variants are functionally distinct. Mol Endocrinol, 2001, **15** (2): 281～293

19 Marden J H, Fitzhugh G H, Girgenrath M, *et al*. Alternative splicing, muscle contraction and intraspecific variation: associations between troponin T transcripts, Ca(2+) sensitivity and the force and power output of dragonfly flight muscles during oscillatory contraction. J Exp Biol, 2001, **204** (Pt 20): 3457～3470

20 Maquat L E. Nonsense-mediated mRNA decay. Curr Biol, 2002, **12** (6): R196～R197

21 Lewis B P, Green R E, Brenner S E. Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. Proc Natl Acad Sci USA, 2003, **100** (1): 189～192

Cloning and Bioinformatics Analysis of Human Lung Fibrosis-inducing Factors*

CHEN Xiao-Hua[1,2)**], CAI Guo-Ping[1,2)]

([1)] *Department of Biological Science and Biotechnology, Tsinghua University, Beijing* 100084, China;

[2)] *Life Science Division, Graduate School at Shenzhen, Tsinghua University, Shenzhen* 518055, China)

Abstract Chitinases are ubiquitous chitin-fragmenting hydrolases, however until a few years ago that chitinase like proteins have been found in mammalian. A silica-induced bronchoalveolar lavage protein (iSBLP[58]) with fibroblast growth promoting activity in silicotic rat, which had high sequence homology with members of the mammalian chitinase protein family has been previously purified and characterized. Bioinformatics analysis showed that several human EST clones from pooled colon, kidney or stomach matched the rat protein sequence. Thereafter clone from human kidney RNA samples with several pairs of primers was managed and a set of sequences was obtained, whose cDNA and amino acid sequences have high similarity with each other and several human chitinases in GenBank. Comparison with human genome sequence suggests that these molecules may be from variant transcripts of same a Pre-mRNA. Here the characterization of these sequences is reported.

Key words chitinase, EST, cDNA, bioinformatics, alternative splicing

---

*This work was supported by a grant from The National Natural Science Foundation of China (30270515).

**Corresponding author . Tel: 86-10-62777546, E-mail: cxh01@mails.tsinghua.edu.cn

Received: February 19, 2006 Accepted: March 30, 2006