

3.3 自从 Sumi 于 1987 年首次报道在纳豆中存在一种具有溶栓作用的酶之后,人们开始对纳豆激酶进行深入系统的研究。研究结果表明,纳豆激酶是一种很有潜力的新型溶栓药物。近年来,我国学者致力于纳豆激酶基因工程方面的研究并取得了一些进展,彭勇等^[7]已经克隆了纳豆激酶成熟肽基因并实现了在大肠杆菌中表达,虽然表达率高达 26%,但表达产物为融合型,以包涵体形式存在,没有测出溶栓活性,而且该基因与文献^[3]报道的核苷酸序列同源性为 93.4%。文献^[8]报道,由于枯草杆菌素 E 在大肠杆菌中表达需要 pro 前导肽序列,表达后可通过自身加工切除 pro 序列,形成有活性的成熟肽。而 NK 与枯草杆菌素 E 具有高度同源性,基于此原因,刘北城等克隆了包括前导肽在内的完全正确的纳豆激酶原基因,并实现了在大肠杆菌中的表达,其表达产物为 38kD 的纳豆激酶原和 28kD 的纳豆激酶,两者的表达产物共占 18%,成熟肽占 50%。与文献^[7,8]相比,本文关于纳豆激酶基因工程的研究从基因的正确性、表达率、表达产物的分离纯化以及活性方面都有一定提高,从而为用基因工程菌生产纳豆激酶奠定了基础。

参考文献:

[1] Sumi H, Hamada H, Tsuchida H, et al. A novel fibrinolytic enzyme in the vegetable cheese natto, a typical and popular soybean food in the Japanese diet [J]. *Experientia*, 1987, 43(20): 1110 - 1111.
[2] Fujita M S, Nomura K C, Hong K G, et al. Purification and characterization of a strong fibrinolytic enzyme in the vegetable cheese NATTO, a popular soybean fermented food in Japan [J]. *Biochem Biophys Res Commun*, 1993, 197(3): 1340 - 1347.
[3] Nakamura TK, Yamagata YH, Ichishima EJ. Nucleotide sequence of the Subtilisin NAT gene from *Bacillus subtilis* (natto) [J]. *Biosci Biotech Biochem*, 1992, 56(11): 1869 - 1871.
[4] Saito HG, MiaraKI. Preparation of transforming deoxyribonucleic acid by phenol treatment [J]. *Biochem Biophys Acta*, 1963, 72: 619 - 629.
[5] 张淑梅, 张云湖, 赵晓祥, 等. 纳豆激酶基因的克隆与表达 [J]. 中国生物化学与分子生物学报, 1999, 15(6): 912 - 915.
[6] 刘北城, 宋后燕. 纳豆激酶基因的克隆及其在枯草杆菌中的表达 [J]. 生物化学与生物物理学报, 2002, 34(4): 338 - 340.
[7] 彭勇, 黄庆, 张正义. 枯草杆菌 DC-2 纳豆激酶基因的克隆及其融合蛋白在 *E. coli* 中的表达 [J]. 中国生物化学与分子生物学报, 2002, 18(5): 559 - 563.
[8] 刘北城, 官孝群, 宋后燕. 纳豆激酶原的基因克隆及在大肠杆菌中的表达 [J]. 上海医科大学报, 1999, 26(6): 401 - 404.

编码锌指蛋白的人类新基因 *TFL76* 的电子克隆

李慧, 李亦平, 慈宏亮

(北京师范大学生物医学研究所, 北京 100875)

摘要:目的: 根据基因同源同功原理电子克隆人类新基因。方法: 利用基于基因识别软件 Genescan 和 EST 拼接的同源基因克隆法得到人类新基因序列 *TFL76*, 再利用生物信息学数据库和软件对其进行功能的预测和分析。结果: *TFL76* 的 cDNA 序列长 2268bp, 开放阅读框编码 677 个氨基酸残基, 含 12 个连续的 C_2H_2 型锌指基序, 其分子量为 76kDa。编码区序列被 4 个内含子分割。染色体定位于 19q13.4。此位点存在很多与胃癌、膀胱癌、乳腺癌等癌症相关的基因。*TFL76* 的 N 末端含有多种蛋白激酶的磷酸化位点和核定位信号。结论: *TFL76* 可能是一个和癌症相关的核转录因子。

关键词: 电子克隆, 锌指蛋白, 核转录因子

中图分类号: R394 文献标识码: A 文章编号: 1004-311X(2003)06-0015-03

In Silicon Cloning of *TFL76*, A Novel Human Gene Encoding a Zinc Finger Protein

LI Hui, LI Yi-ping, CI Hong-liang

(Biomedical Research Institute of Beijing Normal University, Beijing 100875, PRC)

Abstract: Objective: In silicon cloning of novel human genes based on the principal that orthologue genes have the alike function. Methods: The novel human gene *TFL76* by virtue of orthologue gene cloning based on the gene-finding software Genescan and the EST assembly tools, then take use of the bioinformatical databases and softwares to analyze and predict its function. Results: the cDNA length of *TFL76* is 2268b long and a single open reading frame encoding 677 amino acid residues with calculated molecular mass of 76kDa. Analysis of the genome sequence revealed the presence of 4 introns interrupting the coding region. It has many potential protein kinase phosphorylation sites and a nucleus localization signal together with a DNA binding activity conferred by 12 C_2H_2 zinc fingers. *TFL76* is mapped to the chromosome of 19q13.4, a region associated with the stomach cancer, the bladder cancer and the breast cancer. Conclusion: *TFL76* may be a nuclear transcription factor, which might have roles in human cancer.

Key words: In silicon cloning, Zinc finger motif, Nuclear transcription factor

锌指蛋白(zinc finger protein, ZFP)是指含有通过结合 Zn^{2+} 稳定的短的可以自我折叠形成“手指”结构的一类蛋白质,是真核细胞中最普遍存在的一类 DNA 结合蛋白。根据这些 ZFP 锌指结构上的特征,可分为 C_2H_2 型、 C_3H_4 型和 C_4H_4 型等^[1], 其中 C_2H_2 型 ZFP 构成最大的转录因子家族。目前研究的锌指蛋白的主要功能有:参与细胞分化,参与胚胎发育,与多种疾病相关^[2]。

20 世纪 80 年代末期,随着人类基因组计划地开展,基因序列、定位、表达和功能研究等方面都积累了大量数据,并且通过 Internet 网公布于世,人们利用这些数据资源协助克隆人类新基因日益成为人类新基因发现和克隆的主要策略^[3,4]。2000 年 6 月,人类基因组计划中测序工作的基本完成,以及其他包括大鼠、小鼠和斑马鱼在内的许多物种基因组计划地开展,及相关数据库的发布都为我们利用生物信息学的方法克隆新基因带来了新机遇。其中特定基因同源新基因克隆便是一种策略^[5]。2001 年 2 月日本的 RIKEN 公布了 21 076 条小鼠全长 cDNAs 的序列,根据与已知基因的同源性,他们把这些基因分为 9 类,其中“motif-containing protein”这一类代表着编码具有一定结构域蛋白的 573 条小鼠新基因。在此基础上我们通过同源克隆的方法得到了相应的人类新基因。*TFL76* 便是

其中之一。

1 材料与方法

1.1 材料

1.1.1 生物信息学数据库和软件

通过 NCBI(美国国家生物技术信息中心)的 GenBank 数据库实现 DNA 序列查询,NCBI/Blast 软件进行核酸序列的同源性分析,NCBI 的 ORF 查找器对 cDNA 序列进行可读框架分析。

通过 GENESCAN 程序预测基因组序列中可能的外显子。

通过 BDGK(伯克利果蝇基因组计划小组)的启动子数据库检索编码区前可能的启动子位点。

通过 Vector NTI suite 8.0 软件包实现核酸/氨基酸序列的相似性比较,其中的 sim4 程序实现了 cDNA 序列与基因组序列的对齐及显示,Assembly 程序可对 EST 序列进行拼接,Analysis(protein)程序分析氨基酸序列的分子量,等电点等基本性质。

采用 PROSITE 数据库进行蛋白质功能位点分析,SMART 服务器进行结构域的确,Signal P 服务器预测信号肽剪切位点,联网至清华大学生物信息学院亚细胞定位数据库对蛋白质序列进行亚细胞定位分析。

1.2 方法

1.2.1 *TFL76* 的电子克隆流程

1)从网址(<http://www.gsc.richcn.go.jp/e/FANTOM>)下载类别为“motif-containing protein”的小鼠新基因;

2)用 BLAST 的方法在人类基因组中搜寻与小鼠新基因具有同源序列的片段^[6];

收稿日期: 2003-06-05, 修回日期: 2003-07-16

作者简介: 李慧(1978-),女,硕士生,专业方向:分子生物学, E-mail: huisp@163.com
李亦平(1954-),男,博士,教授,博士生导师,从事骨细胞胚胎发育、肿瘤机理研究,发表论文 30 余篇。

3) 下载人类基因组中与小鼠新基因具有同源序列的片段,用 GeneScan 分析这一片段中可能包含的人类新基因候选基因;

4) 将 Genescan 推测的外显子拼接成 cDNA 序列;

5) 用拼接所得到的 cDNA 序列与人类 EST 比对,确定这个基因是真基因还是假基因。如果这条它与人类 EST 具有相同的序列,说明它是一个真基因或者其片段,并且把它与这些 EST 序列利用 Vector NTI suite 中的 assembly 程序拼接,重复这一过程,得到人类候选新基因。如果它不与任何人类 EST 同源,则可能是一个假基因,被暂时淘汰。如果过一段时间后,该基因与 NCBI 中新加入的 EST 同源,则认为它是真基因。把人类候选基因与非冗余序列库(blastnr)进行比较,淘汰已知功能的人类基因,则得到人类新基因。

1.2.2 将所得到的新基因与对应的鼠基因进行核苷酸序列的比对。

1.2.3 TFL76 的核酸序列分析

利用上述数据库及其软件进行开放阅读框分析,并把核苷酸序列翻译成氨基酸序列,基因组定位,染色体定位,基因启动子位点分析等。

1.2.4 氨基酸序列分析 利用上述数据库与软件进行:

1) 基本信息分析包括分子量、等电点、前导肽、蛋白质亚细胞定位分析等。

2) 基于结构功能域数据库的蛋白质功能预测。

2 结果

2.1 电子克隆得到的 TFL76 的核苷酸序列与氨基酸序列

核苷酸序列长 2268bp, 编码区序列长为 2031bp, 编码 677aa 长度的氨基酸序列, 起始密码子的 + 4 位为鸟嘌呤 G, 符合 KOZAK 规则, 如图 1。

2.2 TFL76 与其他序列的比对结果

与同源鼠核酸序列比对结果为: 同源性较高的区域为 1589bp, 此部分的同源性达 82%, 氨基酸序列同源性达到 70%, 说明克隆到的是鼠的同源基因。和人类已知蛋白序列同源比分析结果为和 z180 同源性达到 39.4% 和 z1328 同源性达到 34.8% 和 z135 同源性达到 34.1% 表明是一种新型的锌指蛋白。

2.3 TFL76 的基因组定位和染色体定位

用 vector NTI suite 8.0 的 sim4 程序将 TFL76 定位于基因组上, 结果显示 TFL76 为一个包含 5 个外显子的基因, 如图 2。其外显子/内含子边界处基本符合 GT-AG 规则, 如表 1。TFL76 定位于染色体的 19q13.4 位。

表 1 TFL76 的基因组结构表

Exon number and size (nt)	5' splicing site	Intron number and size (nt)	3' splicing site
1 (40)	GATCCCATCA	1 (13406)	aggetcaggeATGGCGGCCT
2 (132)	GCCCGGGGTG	2 (2848)	gatttagt agTTGTGGAATG
3 (164)	GAGGACTGGG	3 (3572)	gccgcccc agCTTGGCGCGG
4 (24)	GCCGCCAACCC	4 (21240)	ctttttgc agTTGGGATATG
5 (1634)			

2.4 基因启动子位点分析

在编码序列前的 2000bp 内查找可能的启动子区域, 结果存在 13 个启动子, 可能性高达 57% 到 95% 的有 9 个, 可能性为 51% 到 56% 的有 4 个, 并且存在有两个 TATA, 如图 3。

2.5 氨基酸序列分析

分子量为 76.0 kDa, 等电点为 9.16, 没有前导肽区段, 是个非分泌型蛋白, 亚细胞定位结果为细胞核, 1 个可能的 cAMP- and cGMP- 依赖的蛋白激酶磷酸化位点, 3 个可能的 PKC 磷酸化位点, 2 个可能的酪氨酸激酶磷酸化位点, 9 个可能的酪蛋白磷酸化位点。

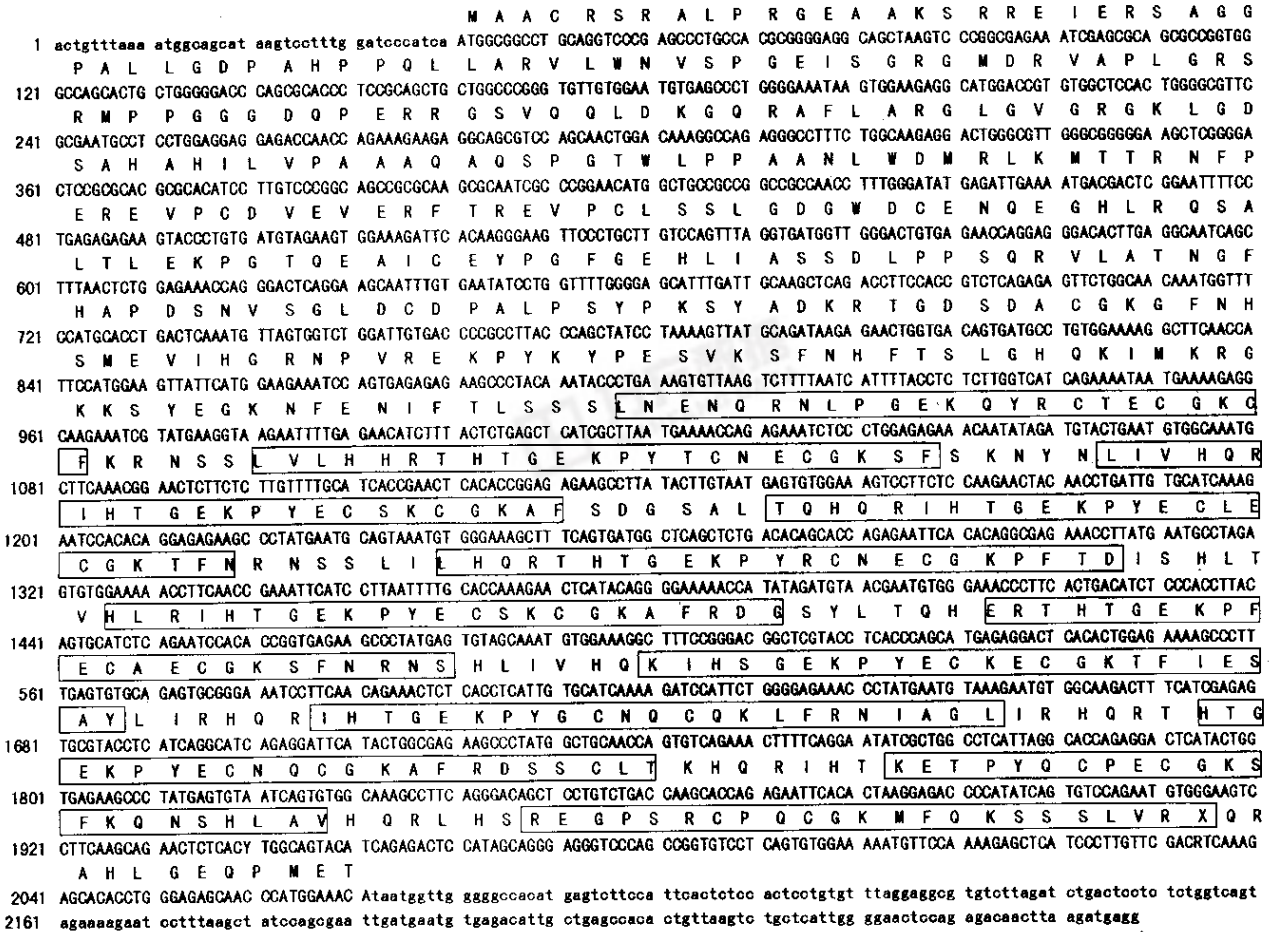


图 1 TFL76 的核苷酸及其编码的氨基酸序列, 编码区用大写, 非编码区用小写, 对应的氨基酸序列在核苷酸序列之上, 黑框处表示 C₂H₂ 结构域

Start	End	Score	Promoter Sequence
175	225	0.86	aataaataaataaagaaggcttacttacagactcttctgctggtgacct
213	263	0.65	tgctggtggcctgaaaaaaggagctactgtgagctctacagctgcaag
418	468	0.59	ggtgcatgactataatcccagctactcaagaggctgaagcaggagaattg
681	731	0.53	tctggagatcctaacaacagaggcctagttaagctatgccagactcttga
742	792	0.95	actgagataataaatacacgtggcttccccacaggcggagcccgcagt
1038	1088	0.89	ctaaaatacaaaaaatagcagggtgtggtgatgcccctatagtccca
1098	1148	0.55	aggctgaggcaaaaaattgcttgaaccaggaggcggaggttgcagtga
1229	1279	0.93	tgttacagttctttaaagcagcgtctctggaatttcttctctgatgtt
1362	1412	0.79	gttacagctcataaatgcaatcggaccacaagagcaagcagcagcaaga
1704	1754	0.57	ttggtgtgtttacaaccttgagctagatcacagagtgcctgactggtgat
1764	1814	0.53	ttagctagacataaaggttctccaagtcctactagactcaggagcccag
1800	1850	0.54	actcaggagcccagctggcttcaccagtgatcccgcacaggggccgcag
1927	1977	0.82	gagcaggggggcagcgtcgtcggggaggctccggcccgcagagcccag

图3 启动子位点分析

2.6 TFL76的保守结构域

使用 SMART 服务器分析 TFL76 保守结构域, 结果为含有 12 个连续的 C₂H₂ 型的锌指基序, 如图 4。



图2 TFL76的基因结构图



图4 TFL76的保守结构域

2.7 19q13.4位的癌基因列表

通过查询 Abin 公司提供的 19q13.4 位的癌基因列表, 发现此位点含有与胃癌、家族性神经性瘤、膀胱癌、乳腺癌相关的基因, 它们为 PTPRH、FWT2、LENG4 等。

3 讨论

新基因搜寻和蛋白质功能分析是人类基因组图谱公诸于世后的研究热点, 电子克隆技术以数学为核心, 以计算机和互联网为工具, 利用现有的表达序列标签 EST 和生物信息数据

库, 可以加速对人类基因组未知功能新基因的发掘, 为人类功能基因组学与蛋白质组学研究提供新的线索和基础^[1]。本研究采用基于基因识别软件(如 Genescan)和 EST 拼接的同源基因克隆法, 为电子克隆基因提供了一种新的思路。利用该方法克隆到的新基因 TFL76 未得到其全长序列, 因为没有找到 polyA 位点, 我们可以通过设计引物, 先用 RT-PCR 检测该基因的表达情况, 然后再证实并且得到其 cDNA 序列全长。通过从核酸和蛋白质水平 TFL76 进行了较为全方位的预测, 结果为同时含有和 DNA 结合的位点以及与其它蛋白质的作用位点, 所以可以推断它为转录因子。是否真正具有转录因子活性有待进一步实验证明。TFL76 所在的染色体位点 19q13.4 存在许多癌症相关基因, 为继续探索 TFL76 提供了更加浓厚的兴趣。

参考文献:

- [1] Ladomery M, Delleire G. Multifunctional zinc finger proteins in development and disease [J]. *Ann Hum Genet* 2002 Nov 66(6): 331-342.
- [2] Chen W, Liang Y, Deng W, et al. The zinc-finger protein CNBP is required for forebrain formation in the mouse [J]. *Development* 2003 Apr. 1, 130(7): 1367-1379.
- [3] Venter J C, Adams M D, Myers E W, et al. The sequence of the human genome [J]. *Science* 2001 291(5507): 1304-1351.
- [4] Lander ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome [J]. *Nature* 2001 409(6822): 860-921.
- [5] Minoru Kanehisa, Peer Bork. Bioinformatics in the post-sequence era [J]. *Nat Genet* 2003 33: 305-310.
- [6] 李铁石, 常智杰, 傅新元, 等. 基于 Internet 网生物信息资源特定基因同源新基因克隆策略 [J]. *生命科学* 2002, 14(4): 234-237.

原生质体融合子代的筛选和鉴定

邱雁临, 曾莹, 胡赛阳, 夏服宝

(湖北工学院生物工程系, 湖北 武汉 430068)

摘要: 介绍了绿色木霉 N₆ 和黑曲霉 856 原生质体融合子代的研究。经传代、发酵、筛选, 从 11 株初筛的融合子中得到了 3 株纤维素酶活高且稳定的菌株 AT23、AT16、AT34, 其 CMC 酶活分别为亲本绿色木霉 N₆ 的 2.2 倍、1.4 倍、1.2 倍。并对其进行了制霉菌素抗性试验和可溶性蛋白质凝胶电泳分析鉴定。试验结果证明了 AT23、AT16、AT34 是基因发生了重组的融合子且具有杂种优势。

关键词: 融合子, 传代, 发酵, 筛选, 鉴定

中图分类号: Q813.2 文献标识码: A 文章编号: 1004-311X(2003)06-0017-03

Screening and Identifying of Protoplasts Fusion Progeny

QIU Yan-lin, ZENG Ying, HU Sai-yang, XIA Fu-bao

(Bioengineering department of Hubei Polytechnic University, Wuhan 430068, PRC)

Abstract: This paper introduced the study of protoplast fusion progeny between *Trichoderma viride* N₆ and *Aspergillus niger* 856. By generating fermenting and screening 3 strains AT23, AT16, AT34 with high cellulase activity and stable were obtained from 11 strains primary screening fusant. The value of their CMCase activity is 2.2, 1.4, 1.2 times respectively of that of N₆. By anti-nystain ability testing and soluble proteins SDS-PAGE analyzing identity. The results proved that the 3 strains are gene combination fusion progeny possessing hybridization domination.

Key words: Fusant, Generating, Ferment, Screening, Identify

原生质体融合得到的融合子, 在传代中可能不稳定, 会分离成某一亲本类型。故必须将得到的融合子经过反复传代和发酵, 筛选出稳定融合子并对其进行鉴定。考虑到育种目的首先是要得到高酶活菌株, 故先对融合子进行酶活筛选, 再对高

酶活菌株进行鉴定。本鉴定包括两方面: (1) 制霉菌素的抗性测定, 如果是融合子, 则子代将获得了致死亲本 N₆ 的部分性状, 有可能具有抗高浓度制霉菌素的能力, 且酶活越高, 抗制霉菌素能力越强。(2) 可溶性蛋白质的凝胶电泳, 蛋白质是基因编码的产物, 蛋白质的变化即反映了基因的不同。普遍认为, 可溶性蛋白质凝胶电泳分析是微生物的化学分类的有效手段^[1,2]; Giynn 等提出在标准条件下, 菌体蛋白质电泳图谱可重复性, 李