生物物理学报　第十九卷　第二期　二〇〇三年六月
ACTA BIOPHYSICA SINICA Vol.19 No.2 Jun. 2003

利用序列比较寻找胰岛素基因表达启动区与 DNA 结合蛋白的结合位点

周士新，孙　啸，陆祖宏，谢建明，董献军，徐　伟，王　崎
(东南大学分子与生物分子电子学国家重点实验室, 江苏 南京 210096)

摘要：NDF1、IPF1 和 HNF4 是与胰岛素基因表达有关的 DNA 结合蛋白，通过比较 SWISSPROT 蛋白质数据库中人类、小鼠、大鼠这三种核蛋白氨基酸一级序列、模体和结构域，发现其结构十分相似，根据蛋白质结构和功能的关系，推测这些 DNA 结合蛋白与胰岛素基因结合的核苷酸序列相似；从 GenBank 核酸数据库中获得人类、小鼠、大鼠胰岛素 DNA 序列，用 ClustalW 比较三者 Promoter 区的核苷酸序列，显示有一段核苷酸序列较为相似，同时搜索 TRANSFAC 基因转录数据库中 NDF1、IPF1 和 HNF4 蛋白核苷酸结合位点，发现核酸比对保守的部分序列与 TRANSFAC 数据库中这三个转录因子的 DNA 结合位点一致，另外一些核酸保守序列可能为其他未知 DNA 结合蛋白的结合位点。这种核酸序列比对设计为分子生物学实验寻找和验证胰岛素 DNA 结合蛋白与核苷酸的结合位点提供了简单而实用的方法。

关键词：DNA 结合蛋白(DBP)；　胰岛素基因启动区；　模体和结构域；　结合位点
中图分类号：Q615　　文献标识码：A　　文章编号：1000-6737(2003)02-0161-06

胰岛 β 细胞具有特定的表达成年哺乳动物胰岛素基因的功能，胰岛素基因的表达调控比较复杂，受多种因素控制，其中 DNA 结合蛋白是重要的调节因子，DNA 结合蛋白在基因转录的启动区（promoter）与双链 DNA（dsDNA）结合；在调节人类和啮齿动物（小鼠、大鼠）胰岛素基因表达 DNA 结合蛋白中，有三种蛋白较为明确，即神经分化因子 1（neurogenic differentiation factor 1，NDF1 或 NeuroD）、胰岛素启动因子 1（insulin promoter factor-1，IPF1 或 pancreatic duodenal homeobox-1，PDX1）和肝细胞核因子 4（hepatocyte nuclear factor 4，HNF4）[1~3]，其功能缺陷使胰岛素基因无法正常转录，是二型糖尿病发生的重要遗传因素[4-6]。

由于人类和啮齿类动物（小鼠、大鼠）均存在这 3 种蛋白质因子，本文利用生物信息学序列比对的方法，比较人类、小鼠、大鼠的 NDF1、IPF1（PDX1）和 HNF4 三种蛋白质的氨基酸序列，了解其序列是否相似，同时分析 DNA 结合蛋白的模体(motif)和结构域(domain)特点。通过比较人类、小鼠、大鼠胰岛素基因 Promotor 区的核苷酸序列，试图发现保守的核苷酸位点，预测胰岛素转录启动区中同 DNA 结合蛋白结合的核苷酸序列。

1　材料和方法

1.1　数据来源

人类、小鼠、大鼠的 NDF1、IPF1（PDX1）和 HNF4 三种蛋白质因子的氨基酸序列来源于北京大学生物信息中心镜像 SWISSPROT 蛋白质数据库[7]，人类、小鼠和大鼠编码胰岛素的 DNA 序列来源于欧洲分子生物学实验室的 EMBL[8](网址 http://www.cbi.embl-heidelberg.de)，美国生物技术信息中心的 GenBank [9] (Benson et al., 2001；http://www.ncbi.nlm.nih.gov/Web/Genbank)；DNA 结合蛋白模体来源于 PROSITE 数据库(网址 http://cn.expasy.org/prosite)，结构域来源于 SMART 数据库（网址：smart.embl-heidelberg.de/smart/），调控因子来源于 TRANSFAC 数据库（网址：www.gene-regulation.com）。序列查询和比对数据均为 Fasta

收稿日期：2002-11-26
基金项目：国家自然科学基金“创新群体科学基金”项目(60121101)；国家 863 高科技项目(2002AA231071)
作者简介：周士新，1968 年生，主管医师，博士生，电话:(025) 3795174，E-mail: qizhou98@seu.edu.cn
通讯作者：陆祖宏，教授，E-mail: zhlu@seu.edu.cn

(Pearson)格式。

1.2 方法

1.2.1 人类、小鼠和大鼠 NDF1、IPF1 和 HNF4 蛋白氨基酸序列的两两比对和多重比对

用 ClustalW 程序（CLUSTALW 1.82 Multiple Sequence Alignments）[10] 先用动态规划算法对各序列进行两两比较，获得最优得分值，形成系统发生树（phylogenetic tree）[11]，再依据系统发生树，依次对人类、小鼠和大鼠蛋白质转录调控蛋白的氨基酸序列进行多重比对。ClustalW 算法（网址：http://www.ebi.ac.uk/clustalw）是目前应用得最广泛的多重比对程序。

1.2.2 人类、小鼠和大鼠 NDF1、IPF1 和 HNF4 的模体和结构域分析

分别将人类、小鼠、大鼠的 NDF1、IPF1 和 HNF4 三种蛋白质的氨基酸一级序列输入到 PROSITE ScanProsite 的序列分析框中，查询三种蛋白质的模体[12]；同时也将这三种蛋白质的氨基酸序列输入到 SMART 的序列分析框中，查询结构域[13,14]。

1.2.3 人类、小鼠、大鼠胰岛素基因 Promoter 区的核苷酸序列比较

用 ClustalW 算法，将人类、小鼠、大鼠的胰岛素基因 Promoter 区的核苷酸序列进行三重比对，注明三者序列相同的部分。

1.2.4 调节因子在胰岛素基因启动区与 DNA 的可能结合位点分析

用 ClustalW 程序将 TRANSFAC 数据库中存在的相关调节因子的 DNA 结合序列与人类、小鼠、大鼠的胰岛素基因 Promoter 区的核苷酸序列进行比对，在比对中寻找有无与 NDF1、IPF1 和 HNF4 相同的核苷酸序列，这些相同的核苷酸序列很可能是与蛋白的结合位点，并试图发现是否存在其它调节因子的 DNA 结合位点。

Table 1 The amino acids paired-alignment of NDF1, IPF1 and HNF4 among human, mouse and rat

		NDF1			IPF1			HNF4		
		human	mouse	rat	human	mouse	rat	human	mouse	rat
residues(aa)		356	357	357	283	284	283	465	465	465
scores	human		98	98		87	88		95	96
	mouse	98		99	87		94	95		99
	rat	98	99		88	94		96	99	

Scores were the degrees of similarity by comparing two sequences. The score was 100 if two sequences were all the same

2 结 果

2.1 人类、小鼠和大鼠 NDF1、IPF1 和 HNF4 氨基酸序列的两两比对

ClustalW 程序先将 3 种物种的 NDF1、IPF1 和 HNF4 氨基酸序列进行两两比对，人类与小鼠、大鼠氨基酸长度相同或仅相差 1 个氨基酸；人类和小鼠、大鼠 NDF1 和 HNF4 两两比对相似性得分较高，均大于 94，IPF1 稍低，但均大于 85，若两个序列完全相同，相似性得分为 100，两两比对结果见表 1。

2.2 人类、小鼠、大鼠 NDF1、IPF1 和 HNF4 的模体和结构域

用 ScanProsite 和 SMART 程序分别分析人类、小鼠、大鼠 NDF1、IPF1 和 HNF4 三种蛋白质的氨基酸一级序列，人类、小鼠、大鼠 NDF1、IPF1 和 HNF4 在相同位置分别含有与 DNA 结合的螺旋 -

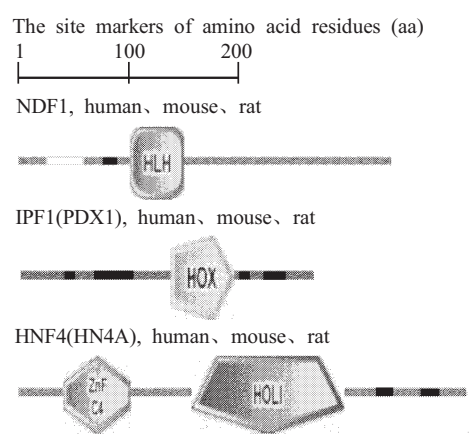

Fig.1 The domains of NDF1, IPF1 and HNF4 among human, mouse and rat. The white line stands for coiled structure of NDF1 proteins, and black ones for segments of low compositional complexity. HLH, HOX, ZnF-C4 and HOLI stand for domains of helix-loop-helix，homeobox domain，zinc finger (C4 type) and ligand binding domain of hormone receptors respectively

```
CLUSTAL W (1.82) multiple sequence alignment
ins1-rat        ---------------------------------------------CTGCAGACTTAGCACT 16
ins1-mouse      -----------------------------------AGTGCTTTCTGCAGACCTAGCACC 24
ins-human       CCGCTGGTCCTGAGGAAGAGGTGCTGACGACCAAGGAGATCTTCCCACAGACCCAGCACC 60
                                                          *  *****  *****

ins1-rat        AGGCAAGTG-TTTGGAAATTACAGCTTCAGCCCCTCTCGCCATCTGCCTACCTACCCCTC 75
ins1-mouse      AGGCAAGTG-TTTGGAAACTGCAGCTTCAGCCCCTCTGGCCATCTGCCTACCCACCCCAC 83
ins-human       AGGGAAATGGTCCGGAAATTGCAGCCTCAGCCCC--CAGCCATCTGCCGACCCCCCCACC 118
                *** ** ** *  ***** * **** ********    ********** ***  ***  *
                   I            II            III              IV
ins1-rat        CTAGAGCCCTTAATGGGCCAAACGGCAAAGTCCAGGGGGCAGAGAGGAGGTGCTTTG-GA 134
ins1-mouse      CTGGAGACCTTAATGGGCCAAACAGCAAAGTCCAGGGGGCAGAGAGGAGGTACTTTG-GA 142
ins-human       CCG----CCCTAATGGGCCAGGCGGCAGGGGTTGACAGGTAGGGGAGATGGGCTCTGAGA 174
                *      ** **********  * ***  *       ** ** *  ** *  ** ** **
                          V             VI
ins1-rat        CTATAAAGCTAGTGGAGACCCAGTAACTCCCAACCCTAAGTGACCAGCTACAATCATAGA 194
ins1-mouse      CTATAAAGCTGGTGGGCATCCAGTAACCCCCAGCCCTTAGTGACCAGCTATAATCAGAGA 202
ins-human       CTATAAAGCCAGCGGGGGCCCAGCAGCCCTCAGCCCTCCAGGACAGGCTGCA-TCAGAAG 233
                *********  * **    **** * * * ** ****    ***  ***  * *** *
                TATA box
ins1-rat        CCATCAGCAAGCAGGTATGTACT-----CTCCTG-----GGTGAGCCCGGTTCCCCCAGC 244
ins1-mouse      CCATCAGCAAGCAGGTATGTACT-----CTCCTC-----TTTGGGCCTGGCTCCCC-AGC 251
ins-human       AGGCCATCAAGCAGGTCTGTTCCAAGGGCCTTTGCGTCAGGTGGGCTCAGGGTTCCAGGG 293
                   ** ********* *** *      *   *        ** **   *    **  *
ins1-rat        CAA---AACTCTAGG----GACTTT-----AGGAAGGATGTGG----GTTCCTCTCTTAC 288
ins1-mouse      CAA---GACTCCAGC----GACTTT-----AGGGAGAATGTGG----GCTCCTCTCTTAC 295
ins-human       TGGCTGGACCCCAGGCCCCAGCTCTGCAGCAGGGAGGACGTGGCTGGGCTCGTGAAGCAT 353
                      ** * **       ** *     *** ** * ****    * ** *     *

ins1-rat        ATGGA-----------------------------------CCTTTTCCTAGCCTCAACCCT 314
ins1-mouse      ATGGA-----------------------------------TCTTTTGCTAGCCTCAACCCT 321
ins-human       GTGGGGGTGAGCCCAGGGGCCCCAAGGCAGGGCACCTGGCCTTCAGCCTGCCTCAGCCCT 413
                 ***                                      ***   *  ****** ****
ins1-rat        GCCTATCTTCCAGGTCATTGT---TCCAACATGGCCCTGTGGATGCGCTTCCTGCCCCTG 371
ins1-mouse      GCCTATCTTTCAGGTCATTGT---TTCAACATGGCCCTGTTGGTGCACTTCCTACCCCTG 378
ins-human       GCCTGTCTCCCAGATCACTGTCCTTCTGCCATGGCCCTGTGGATGCGCCTCCTGCCCCTG 473
                **** ***  *** *** ***   *    ********** * *** * **** ******
```

Fig.2　The nucleotides multiple alignment of insulin gene promoter among human, mouse and rat. TATA-BOX in the promoter of mouse insulin gene was determined by Wentworth[18] using experiment methods. The nucleotides with asterisks indicate the same sequences in the alignment. The nucleotides with underlines have the similar sequences with the binding sites in TRANSFAC database. The shadow letters "ATG" show the beginning of coding mRNA

环 - 螺旋 （ helix-loop-helix，HLH）[15]、螺旋 - 转角 - 螺旋 （helix-turn-helix，HTH） 同源异型框（homeobox domain）[16]、锌指蛋白 （Zinc-finger）模体和结构域[17]（图 1）。

2.3 人类、小鼠、大鼠胰岛素基因 Promoter 区的核苷酸序列多重比对

小鼠、大鼠和人类胰岛素基因在 GenBank 数据库中的编号分别为 X04725、J00747、J00265。通过对已知小鼠（位于 X04725 的 481~1037 bp）、大鼠胰岛素 DNA 序列 （位于 J00747 的 3841~4390 bp）与人类 DNA 部分序列 （位于 J00265 的 1981~2633 bp）进行比较，发现在距 TATA 框上游 100 多 bp 处，有一段 DNA 序列保守区（转录上游启动区），考虑到结合蛋白与 dsDNA 结合的核苷酸数在 5 个以上，因此将三者 5 个连续或多个仅相差 1~2 个 bp 相同的核苷酸序列考虑为可能与蛋白结合的核苷酸序列（见图 2）。

2.4 调节因子在胰岛素基因 Promoter 区与 DNA 的可能结合位点分析

ClustalW 程序对 TRANSFAC 数据库中相关调节因子的 DNA 结合序列与人类、小鼠、大鼠胰岛素基因 Promoter 区数百个核苷酸的序列比对结果显示，有数段相同序列（相同序列段用罗马数字在图 2 中表示），除发现 NDF1、IPF1 和 HNF4 的 DNA 结合位点外，还发现有 PAX4、增强子（enhancer）GG-Ⅱ 和 IEB1 结合位点相同的序列[19]，提示这些 DNA 位点可能是与 PAX4 和增强子的结合序列。在 DNA 结合蛋白中，HNF4（锌指结构）与 DNA 的结合位点两头较保守；NDF1（bHLH 结构）的 DNA 结合位点含有具有这种结构的共有结合序列 CANNTG（据 PROSITE 数据库，登陆号 PD0C00038，N 为任意脱氧核苷酸）；IPF1（含 HTH 的 homeobox 结构）的 DNA 结合位点具有 TAATG 共有序列。人类、小鼠和大鼠胰岛素基因启动区增强子 GG-Ⅱ 和 IEB1 的 DNA 结合序列很相似（图 2 的 Ⅱ、Ⅲ和表 2）。

Table 2 The possible binding sites analysis of regulative factors in insulin gene promoter

Access number	Name of regulative factors	Binding sites in TRANSFAC databases	Similar sequences of insulin gene promoter (human)	Locations in alignment(Fig.2)
R02708	NDF1	GGAGACATTTG	CCAGCCATCTG	Ⅳ
R02709	IPF1	TCTAATG	CTTAATG	Ⅴ
R09461	HNF4A	GAGGCAGTGggagggcgg agggcGGGGGCCTT	GCAGGGGTTGACAGGT AGGGGAGATGGGCTCT	Ⅵ
R08705	PAX4	TCTGGGAAATGAGGT GGAAAATG	GGGAAATGGTCC GGAAATTG	Ⅰ
R02711	enhancer GG-II	GGAAAT	GGAAAT	Ⅱ
R04457	enhancer IEB1	CTCAGCCCCCAGCCATCT GCCGACCCCCCC	CTCAGCCCCCAGCCATC TGCCGACCCCCCC	Ⅲ

The promoters of human insulin gene with underlines have the similar sequences with the binding sites of regulative factors in TRANSFAC database

3 讨 论

DNA 结合蛋白是一类可与 dsDNA 特异性结合的蛋白质，蛋白质在 DNA 双螺旋大沟和小沟处同核苷酸的主链或碱基相互作用，通过静电吸引、氢键和范德华力连接，使两种分子的构象改变，增加两者之间接触面，产生生物效应[20]。序列特异性结合是由碱基所显示的不同结合模式识别来完成，但蛋白质和 DNA 之间的结合模式非常复杂，不同的蛋白质与 DNA 结合的序列和长度不同；同一蛋白质在执行不同的功能时，识别的序列和长度也有差异；特定蛋白质执行特定的功能时，该蛋白质与 DNA 的结合位点是特异的。在胰岛素的转录起动过程中，NDF1、IPF1(PDX1)和 HNF4 均为 DNA 结合蛋白，是重要的胰岛素转录调节因子，除转录调节外，还具有神经发生、细胞分化、器官发育、基因激活、磷酸化作用等功能[21,22]，在这一特定的基因转录过程中，DNA 结合蛋白与核苷酸的结合

是特异的。

Glick 等[23]通过实验研究表明大鼠的 NDF1、IPF1 和 HNF4 三种蛋白质与胰岛素转录启动区（位于转录起始 TATA 框的前 100 多个 bp）的核苷酸结合。比较人类、小鼠和大鼠这三种 DNA 结合蛋白的氨基酸一级结构，发现三者之间很相似（两两比对得分均大于 85 分），特别是 NDF1 和 HNF4，人类和小鼠、大鼠极为相似（两两比对得分大于 94 分），显示人和鼠调节胰岛素基因转录的氨基酸序列保守，且三物种相同的调节因子具有相同的 DNA 结合模体和结构域，说明三物种相同的蛋白质与 DNA 结合的位点（核苷酸序列）应相似，推测这三种物种在转录启动区应具有一段相似的 DNA 序列，ClustalW 多重比对结果显示在胰岛素基因编码区前人类、小鼠、大鼠确有一段同源序列。

一般说来，不同的哺乳动物的转录起动区的核苷酸序列差异较大，通过核苷酸序列比对的方法分析人类、小鼠、大鼠胰岛素基因转录启动区的 DNA 序列，结果发现存在数段相同的 DNA 序列，其中有一部分很可能是与蛋白质结合的序列。除了 NDF1、IPF1 和 HNF4 三个转录因子外，在比对保守序列中还发现存在 PAX4、增强子 GG-Ⅱ和 IEB1 结合位点，推测 PAX4 蛋白、增强子 GG-Ⅱ和 IEB1 参与胰岛素基因转录的调控。用常规的分子生物学方法测定 DNA 结合蛋白的核苷酸结合位点是非常复杂的，本文通过比较不同哺乳动物胰岛素基因的 Promoter 区 DNA 片段，发现保守的核苷酸序列，推测 DNA 结合蛋白的结合位点，为实验寻找和验证胰岛素 DNA 结合蛋白与核苷酸的结合位点提供了简单而实用的方法。由于在胰岛素基因启动区中人和鼠还有一些相同序列，其中某些核酸序列是否为未知 DNA 结合蛋白的结合位点，有待进一步研究。

参考文献：

[1] Miyachi T, Maruyama H, Kitamura T, et al. Structure and regulation of the human NeuroD (BETA2/BHF1) gene [J]. *Brain Res Mol Brain Res*, 1999,69:223~231.

[2] Inoue H, Riggs AC, Tanizawa Y, et al. Isolation, characterization, and chromosomal mapping of the human insulin promoter factor 1(IPF-1) gene[J]. *Diabetes*, 1996,45:789~794.

[3] Kritis AA, Argyrokastritis A, Moschonas NK, et al. Isolation and characterization of a third isoform of human hepatocyte nuclear factor 4[J]. *Gene*, 1996,173:275~280.

[4] Malecki MT, Jhala US, Antonellis A, et al. Mutations in NEUROD1 are associated with the development of type 2 diabetes mellitus[J]. *Nat Genet*, 1999,23:323~328.

[5] Hani EH, Stoffers DA, Chevre JC, et al. Defective mutations in the insulin promoter factor-1 (IPF-1) gene in late-onset type 2 diabetes mellitus[J]. *J Clin Invest*, 1999,104:R41~R48.

[6] Moller AM, Urhammer SA, Dalgaard LT, et al. Studies of the genetic variability of the coding region of the hepatocyte nuclear factor-4alpha in Caucasians with maturity onset NIDDM[J]. *Diabetologia*, 1997,40:980~983.

[7] Gasteiger E, Jung E, Bairoch A. SWISS-PROT: Connecting biological knowledge via a protein database [J]. *Curr Issues Mol Biol*, 2001,3:47~55.

[8] Stoesser G, Baker W, van den Broek A, et al. The EMBL Nucleotide Sequence Database[J]. *Nucleic Acids Res*, 2001, 29:17~21.

[9] Benson DA, Karsch-Mizrachi I, Lipman DJ, et al. GenBank [J]. *Nucleic Acids Res*, 2001,28:15~18.

[10] Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice [J]. *Nucleic Acids Res*, 1994,22:4673~4680.

[11] Mount DW. 生物信息学：序列与基因组分析(影印版)[M]. 北京：科学出版社, 2000. 329~331.

[12] Falquet L, Pagni M, Bucher P, et al. The PROSITE database, its status in 2002[J]. *Nucleic Acids Res*, 2002,30:235~238.

[13] Schultz J, Milpetz F, Bork P, et al. SMART, a simple modular architecture research tool: identification of signaling domains[J]. *Proc Natl Acad Sci USA*, 1998,95:5857~5864.

[14] Ivica L, Leo G, Nicholas JD, et al. Recent improvements to the SMART domain-based sequence annotation resource [J]. *Nucleic Acids Research*, 2002,30:242~244.

[15] Yokoyama M, Nishi Y, Miyamoto Y, et al. Molecular cloning of a human neuroD from a neuroblastoma cell line specifically expressed in the fetal brain and adult cerebellum [J]. *Brain Res Mol Brain Res*, 1996,42:135~139.

[16] Stoffel M, Stein R, Wright CV, et al. Localization of human homeodomain transcription factor insulin promoter factor 1 (IPF1) to chromosome band 13q12.1[J]. *Genomics*, 1995,28: 125~126.

[17] Price JA, Fossey SC, Sale MM, et al. Analysis of the HNF4

alpha gene in Caucasian type Ⅱ diabetic nephropathic patients[J]. *Diabetologia*, 2000;43:364~372.

[18] Wentworth BM, Schaefer IM, Villa-Komaroff L, et al. Characterization of the two nonallelic genes encoding mouse preproinsulin[J]. *J Mol Evol*, 1986,23:305~312.

[19] Sheau YS, Christine MMS, Tsai MJ. Molecular characterization of the rat insulin enhancer-binding complex 3b2[J]. *J Biol Chem*, 1995,270:21503~21508.

[20] Twyman RW 著. 陈淳, 徐沁译. 高级分子生物学要义[M]. 北京：科学出版社, 2000. 217~222.

[21] Arava Y, Adamsky K, Ezerzer C, et al. Specific gene expression in pancreatic β-cells: cloning and characterization of differentially expressed genes[J]. *Diabetes*, 1999,48:552~556.

[22] Arava Y, Adamsky K, Belleli A, et al. Differential expression of the protein kinase a regulatory subunit (R1α) in pancreatic endocrine cells[J]. *FEBS Lett*, 1998,425:24~28.

[23] Glick E, Leshkowitz D, Walker MD. Transcription factor β_2 acts cooperatively with E2A and PDX1 to activate the insulin gene promoter[J]. *J Biol Chem*, 2000,275:2199~2204.

TO FIND NUCLEOTIDE BINDING SITES OF DBPS BY COMPARING THE PROMOTER SEQUENCES OF INSULIN GENE

ZHOU Shi-xin, SUN Xiao, LU Zu-hong, XIE Jian-ming, DONG Xian-jun, XU Wei, WANG Qi

(National Laboratory of Molecular and Biomolecular Electronics, Southeast University, Jiangsu Nanjing 210096, China)

Abstract: There were several DNA binding proteins (DBPs) related to insulin gene expression, such as NDF1, IPF1 and HNF4. The structures of three DBPs was homologous by comparing the amino acid sequences, motifs and domains of human, mouse and rat in SWISSPROT protein database. According to the relationships between structure and function of protein, it was supposed that the nucleotide binding sites of DBPs were similar in the promoters of insulin gene. The DNA sequences of human, mouse and rat were obtained in GenBank nuclear database. The ClustalW program of multiple alignment was used to compare nucleotide sequences of the three mammals. It showed that there were some similar sequences in the promoters of insulin gene. In the meanwhile, the nucleotide binding sites of NDF1, IPF1 and HNF4 were searched in the TRANSFAC gene regulation database. The conservative nucleotide sequences were found between the promoters alignment of insulin gene and the nucleotide binding sites of DBPs in the TRANSFAC database. Some other conservative nucleotide sequences in the alignment were perhaps the binding sites of unknown proteins. This device of sequences alignment offered a simple and applicable method to find and confirm the nucleotide binding sites of DBPs in the molecular biology experiments.

Key Words: DNA binding protein(DBP); Promoters of insulin gene; Motif and domain; Binding site