

用生物信息学方法克隆大鼠核不均一核蛋白 A2/B1 基因

张华莉, 袁 灿, 肖献忠

(中南大学湘雅医学院病理生理学教研室, 湖南省长沙市 410078)

[关键词] 病理学与病理生理学; 心血管内源性保护机制; 生物信息学; 缺血预适应; 电子克隆; 核不均一核蛋白 A2/B1 基因; 表达序列标签

[摘 要] 采用外显子预测、序列匹配、序列拼接等生物信息学方法电子克隆了一个在大鼠心肌缺血预适应中表达上调的表达序列标签所代表的核不均一核蛋白 A2/B1 基因, 用逆转录聚合酶链反应证实了该基因的开放阅读框。它的开放阅读框为 1026 bp, 由 10 个外显子构成, 推测编码 341 个氨基酸。它的编码区与人和小鼠的核不均一核蛋白 A2/B1 基因分别有 92% 和 95% 的一致性, 蛋白质序列的一致性几乎为 100%。这些结果表明核不均一核蛋白 A2/B1 基因在不同物种间高度保守, 它在大鼠心肌缺血预适应中表达上调, 可能在缺血预适应的内源性保护机制中发挥了重要的作用。

[中图分类号] R363

[文献标识码] A

Molecular Cloning of Rat Heterogeneous Nuclear Ribonucleoproteins A2/B1 Gene by Bioinformatics Technique

ZHANG Hua-Li, YUAN Can, and XIAO Xian-Zhong

(Department of Pathophysiology, Xiangya School of Medicine, Central South University, Changsha 410078, China)

[KEY WORDS] Ischemic Preconditioning; In Silicon Cloning; Heterogeneous Nuclear Ribonucleoproteins A2/B1 Gene; Expressed Sequence Tag; Myocardium; Reverse Transcriptase Polymerase Chain Reaction

[ABSTRACT] **Aim** To clone the upregulated expressed gene during rat myocardium ischemic preconditioning. **Methods** New gene was in silicon cloned using exon finding, blast alignment, assembling program and so on. Reverse transcriptase polymerase chain reaction (RT-PCR) assay confirmed the open reading frame (ORF) of new gene. **Results** Rat heterogeneous nuclear ribonucleoproteins (hnRNP) A2/B1 gene was cloned by bioinformatics methods. The product of RT-PCR was 1.1 kb, which was identical to the predicted length. The ORF of rat hnRNPA2/B1 gene was 1 026 bp, which was composed of ten exons and encoded 341 amino acids. The coding sequence showed 92% and 95% identity respectively in 1 026 bp overlap with that of human and mouse, and the predicted polypeptide showed almost 100% identity with that of human and mouse. **Conclusions** The results showed that hnRNPA2/B1 gene was highly conserved among various species. It maybe plays an important role in endogenous protection of ischemic preconditioning.

核不均一核蛋白(heterogeneous nuclear ribonucleoproteins, hnRNP)家族在核不均一 RNA 的转录后加工以及信使 RNA 的包装、转运和翻译中起着重要的作用, 到目前为止已鉴定了 20 余种 hnRNP^[1,2], 其中相对分子质量在 30~40 kDa 的 A1、A2、B1、B2、C1 和 C2 是 hnRNP 的核心蛋白, hnRNP A/B 蛋白是最小的 hnRNP。hnRNP A/B 蛋白的氨基末端有两个 RNA 识别功能域, hnRNP A/B 蛋白通过 RNA 识别功能域与 RNA 结合从而发挥作用^[3,4]。

本室在大鼠心肌缺血预适应动物模型中用抑制消减杂交的方法筛选到一个在缺血预适应时表达上调的表达序列标签(expressed sequence tag, EST) 159^[5]; 本研究采用序列匹配、序列拼接和外显子预测^[6,7]等生物信息学的方法, 电子克隆了该 EST 所代表的基因—大鼠 hnRNP A2/B1 基因, 并用逆转录聚合酶链反应的方法证实了它的开放阅读框。

1 方法

1.1 大鼠核不均一核蛋白 A2/B1 基因的电子克隆

将大鼠心肌缺血一再灌注中表达上调的一个新基因代表 EST(编号 159)用 BioEclone 程序^[8]进行 EST 拼接。将拼接的重叠群用基本局部匹配搜索工具(basic local alignment search tool, blast)^[9]与美国国家生物技术信息中心^[10]的人和小鼠 EST 数据库进

[收稿日期] 2003-09-29

[修回日期] 2004-02-16

[基金项目] 国家 973 重点项目(G2000056908)及国家自然科学基金(30170373)资助

[作者简介] 张华莉, 博士, 助理研究员, 主要从事心血管内源性保护的机制研究, 联系电话 0731-2355020, E-mail 为 zh172@yahoo.com.cn。袁灿, 博士, 讲师, 主要从事心血管内源性保护的机制研究。肖献忠, 博士, 教授, 博士研究生导师, 主要从事心血管内源性保护和缺血性休克发病机制及防治研究, 本文通讯作者, 联系电话 0731-2355019, E-mail 为 xianzhongxiao@hotmail.com。

行匹配,确定人和小鼠是否存在与该重叠群高度相似的序列;将小鼠中具有很高相似性的 hnRNP A2/B1 基因的 cDNA 序列与大鼠的 EST 和基因组序列进行 blast 匹配,推测大鼠 hnRNP A2/B1 基因的 cDNA 序列;根据内含子和外显子交界的头尾处符合 GT/AG 规律,用 Spidey 程序^[11]确定大鼠 hnRNP A2/B1 基因的基因组结构。用 GENSCAN^[12]和 MZEF^[13]程序预测大鼠 hnRNP A2/B1 基因的外显子,进一步证实上述结果。用开放阅读框查询程序分析大鼠 hnRNP A2/B1 基因的开放阅读框。用翻译程序将开放阅读框翻译成氨基酸序列。

1.2 逆转录聚合酶链反应证实大鼠核不均一核蛋白 A2/B1 基因的开放阅读框

在推测的大鼠 hnRNP A2/B1 基因 cDNA 序列上设计两条引物 hnRNPA2F 和 hnRNPA2R。按常规方法(即用 Trizol 试剂)抽提大鼠心肌组织的总 RNA,用鸟类成髓细胞瘤病毒逆转录酶(Takara 公司)进行逆转录,以 1 μ L 的逆转录产物作为模板,用引物 hnRNPA2F 和 hnRNPA2R 进行聚合酶链反应。

1.3 大鼠核不均一核蛋白 A2/B1 基因的核苷酸和蛋白质序列匹配

用 blast 匹配大鼠 hnRNP A2/B1 基因的编码区和人 hnRNP A2/B1 基因以及小鼠 hnRNP A2/B1 基因的编码区。用 Clustal W 程序^[14]比较大鼠、小鼠和人 hnRNP A2/B1 基因编码的氨基酸序列。

2 结果

2.1 大鼠核不均一核蛋白 A2/B1 基因的电子克隆

将 EST159 用 BioEclone 程序进行 EST 拼接,得到一个 803bp 的重叠群;用 blast 匹配这个重叠群与小鼠 EST 数据库,得到许多一致性为 90% 的小鼠 EST,这些 EST 属于同一个基因簇,该基因为小鼠 hnRNP A2/B1 基因,推测 159 属于 hnRNP A2/B1 基因的大鼠垂直同源物。将小鼠 hnRNP A2/B1 基因的信使 RNA(GenBank: AF073993)与大鼠 EST 数据库进行 blast 匹配(图 1, Figure 1),除了中间 260 bp 无 EST 匹配以外,均有许多大鼠 EST 与小鼠 hnRNP A2/B1 的 5' 端和 3' 端相匹配。将这些与 AF073993 相匹配的 EST 用 BioEclone 程序进行拼接,得到大鼠 hnRNP A2/B1 基因的 5' 端 cDNA 重叠群和 3' 端 cDNA 重叠群。

将上述的 cDNA 重叠群与大鼠的高通量基因组序列进行 blast 匹配,它与一个基因组序列(GenBank: AC119021.3)完全匹配,说明 AC119021.3 包含

大鼠 hnRNP A2/B1 基因;用 Spidey 程序分析 cDNA 重叠群和 AC119021.3,确定了大鼠 hnRNP A2/B1 基因的 7 个外显子。AF073993 中无大鼠 EST 匹配的 260bp 与 AC119021.3 进行序列匹配,得到 3 个一致性达 95% 以上的片段,说明这 3 个片段是大鼠 hnRNPA2/B1 基因的另外 3 个外显子。根据内含子和外显子交界的头尾处符合 GT/AG 规律和小鼠 A2/B1 基因的基因组结构,我们确定了这 3 个外显子的交界处。我们用 GENSCAN 和 MZEF 程序预测 AC119021.3 的外显子(表 1, Table 1), GENSCAN 预测出 9 个外显子;其中有 3 个外显子的大小不一致;MZEF 预测出 8 个外显子。

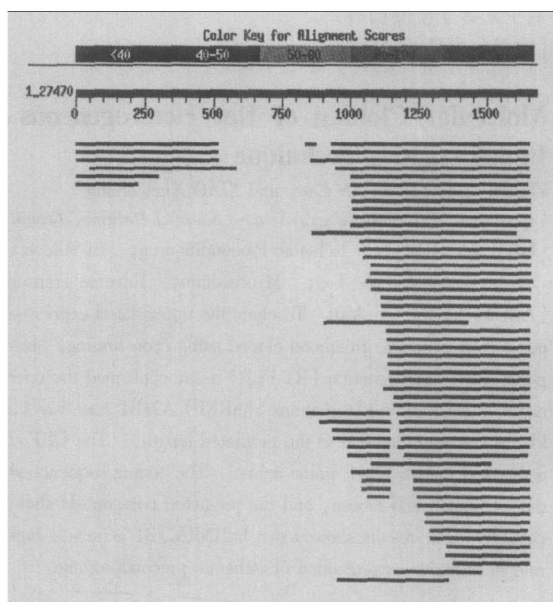


图 1. AF073993 与大鼠 EST 数据库 Blast 匹配的结果
Figure 1. Blast analysis between AF073993 and rat EST database

将 5' 端 cDNA 重叠群、3' 端 cDNA 重叠群和三个新确定的外显子拼接成一个 1 245 bp cDNA 重叠群(图 2, Figure 2)。开放阅读框查询程序分析存在一个 1 026 bp 的开放阅读框,起始密码子邻近的-3 位为鸟嘌呤(G)而 +4 位为鸟嘌呤(G),符合 Kozak 规则,起始密码子上游同一相位有一个终止密码子(图 2, Figure 2)。该开放阅读框由 10 个外显子构成(表 1, Table 1),推测编码 341 个氨基酸。

2.2 大鼠核不均一核蛋白 A2/B1 基因开放阅读框得到证实

用引物 hnRNPA2F 和 hnRNPA2R 进行聚合酶链

TAGAAAAGCGGCGGTCTCGGCTCCAGCGGCAGTAGCAGCGGCGCGGTCGCCGTGTCAGGTGCTCCGTGCGGCGCAGTTTCTTGCCAT
 CGGCTGCTCTCACCACAGCGTCAGGACGAGTCCCGTGGTGTCCGTCGCGGAGGTCTTTCTCTCTCGCTCGGCTGCGGGAATCGAG
 CTGAAGCGACTGAGTCCCGGATGGAGAGAGAAAAGCAACAGTTCGCTAAGCTCTTTATTGGTGGCTTAAGCTTTGAAACCACAGAAGAA
 AGTTTGAGAAAATACTACGAGCAATGGGCAAGCTCACAGACTCTGCTGTTATCCGGGATCCTGCAAGCAAAAAGATCAAGAGGATTTCG
 CTTTGAACCTTCTCATCTATGGCTGAGGTTGATGCTGCCATGGCTGCAAGGCCTCATTCCATTGATGGCAGGGTGGTTGAGCCAAAAC
 GTGCTGTGCAAGAGAGGAGTCTGGGAAACCAGGAGCCCATGTGACTGTGAAGAACTGTTGTTGTTGGAATTAAGGAAGATACTGAG
 GAGCATCACCTTAGAGATTACTTTGAAGAATATGAAAAAATGATACTATTGAAATAATTACTGATAGCCAGTCTGAAAAAAAAGAGG
 CTTTGGCTTTGTTACCTTTGATGACCATGATCCTGTGGATAAGATTGTCTTCAGAAATATCACACCATAAATGGTCACAATGCAGAAG
 TTAGAAAAGCGATTGTCTAGACAAGAAATGCAGGAAGTCCAAAGTCTTAGGAGTGAAGAGGAGGAACTTTGGTTTGGAGATTCTCGA
 GGTGCGGCTGGCAATTTTGGACCAGGACCAGGAAGCAACTTTAGGGGGGCTTCTGATGGATATGGAAGTGGACGTGGATTGGGGATGG
 CTATAATCGGTATGCAAGCAGGACCTGACGCTGCCAATTTTGGAGCTAGCCCTGCTTATGACAGGAAAGAGCAGGATATGCTGCTGGAG
 GACCTGCGATATGCAACCAGGGTGGGGCTACGGAGGTGTTATGACAATATGAGGAGGAAAATTATGGAAGTGGAAATTAACAATGAT
 TTTGGAATTATAACCAGCAACCTTCTAACTACGGTCCAATGAAGAGTGGAAAATTTGGTGGTAGCAGGAACATGGGAGGACCATATGG
 TGGAGGAACTATGCTCTGGAGGAAGTCCGCGGCTATGCTGGAAGGAGCCGATATTGACCTTCTCTACTTACCATGG

图 2. 大鼠 hnRNP A2/B1 基因的 cDNA 序列 图中大写黑体字母分别为起始密码子和终止密码子
 Figure 2. The cDNA sequence of rat hnRNP A2/B1 gene

表 1. GENSCAN 和 MZEF 预测的基因外显子

Table 1. Predicted exons by GENSCAN and MZEF programs

外显子名称	外显子大小 (bp)	GENSCAN 预测的结果	MZEF 预测的结果
Exon 1	204	—	—
Exon 2	111	111	111
Exon 3	147	147	147
Exon 4	211	207(?)	211
Exon 5	102	97(?)	102
Exon 6	81	96(?)	81
Exon 7	63	63	63
Exon 8	120	120	120
Exon 9	123	123	123
Exon 10	62	62	—

“?”表示预测的大小与实际不符,“—”表示未预测到的外显子

反应(图 3, Figure 3), 扩增的片段约 1.1 kb, 与预测的长度基本一致。产物回收测序发现, 其序列与上述拼接的一致。

2.3 大鼠核不均一核蛋白 A2/B1 基因的核苷酸和蛋白质序列匹配

大鼠 hnRNP A2/B1 基因的编码区与人 hnRNP A2/B1 基因一致性达 92%, 与小鼠 hnRNP A2/B1 基因一致性达 95%。用 Clustal W 程序比较大鼠、小鼠和人 hnRNP A2/B1 基因编码的氨基酸序列(图 4, Figure 4), 结果发现大鼠 hnRNP A2/B1 基因与人 hnRNP A2/B1 基因在 341 个氨基酸有 100% 的一致性, 与小鼠 hnRNP A2/B1 基因在 341 个氨基酸有 99% 的一致性。

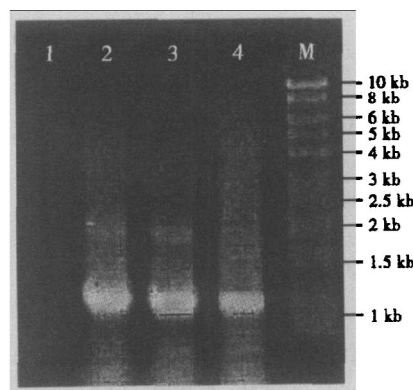


图 3. 逆转录聚合酶链式反应证实大鼠 HnRNP A2/B1 基因的开放阅读框 1 为空白, 2、3 和 4 为大鼠心肌, M 为标准分子质量品。

Figure 3. Reverse transcriptase PCR confirmed the ORF of rat hnRNP A2/B1 gene

3 讨论

我们采用 Blast 匹配、序列拼接、外显子预测、多序列匹配等生物信息学方法获得大鼠 hnRNP A2/B1 基因, 它与人和小鼠 hnRNP A2/B1 基因的核苷酸序列同源性分别为 92% 和 95%, 蛋白质序列的同源性几乎为 100%。以上结果表明 hnRNP A2/B1 基因在不同物种间高度保守, 由此可见它在细胞生命活动中的重要性。

核不均一核蛋白(hnRNP) A2/B1 是 40S 核糖体亚基的主要成分之一, 它不仅能够与端粒序列结合^[15], 还可以与特异性 RNA 片段——hnRNP A2 反

大鼠	MEREKEQFRKLFIGGLSFETTEESLRNYYEQWGKLTDCVVMRDPASKRSRGFGVTFSSM	60
人	MEREKEQFRKLFIGGLSFETTEESLRNYYEQWGKLTDCVVMRDPASKRSRGFGVTFSSM	60
小鼠	MEREKEQFRKLFIGGLSFETTEESLRNYYEQWGKLTDCVVMRDPASKRSRGFGVTFSSM	60

大鼠	AEVDAAMAARPHSIDGRVVEPKRAVAREESGKPCAHVTVKCLFVGGIKEDTEEHHLRDYF	120
人	AEVDAAMAARPHSIDGRVVEPKRAVAREESGKPCAHVTVKCLFVGGIKEDTEEHHLRDYF	120
小鼠	AEVDAAMAARPHSIDGRVVEPKRAVAREESGKPCAHVTVKCLFVGGIKEDTEEHHLRDYF	120

大鼠	EEYGKIDTIEIITDRQSGKKGRCFVTFDDHDPVDKIVLQKYHTINGHNAEVRKALSQRQE	180
人	EEYGKIDTIEIITDRQSGKKGRCFVTFDDHDPVDKIVLQKYHTINGHNAEVRKALSQRQE	180
小鼠	EEYGKIDTIEIITDRQSGKKGRCFVTFDDHDPVDKIVLQKYHTINGHNAEVRKALSQRQE	180

大鼠	MQEVQSSRSGRGNFPGDSRGGGNGFPGPGSNFRGGSDGYGSGRFGDGYNGYGGGGPG	240
人	MQEVQSSRSGRGNFPGDSRGGGNGFPGPGSNFRGGSDGYGSGRFGDGYNGYGGGGPG	240
小鼠	MQEVQSSRSGRGNFPGDSRGGGNGFPGPGSNFRGGSDGYGSGRFGDGYNGYGGGGPG	240

大鼠	GGNFGSPGYGGGRGGYGGGPGYGNQGGYGGYDNYGGGNYGSGNYNDFGNYNQPSN	300
人	GGNFGSPGYGGGRGGYGGGPGYGNQGGYGGYDNYGGGNYGSGNYNDFGNYNQPSN	300
小鼠	GGNFGSPGYGGGRGGYGGGPGYGNQGGYGGYDNYGGGNYGSGTYNDFGNYNQPSN	300

大鼠	YGPMSKSNFGGSRNMGGPYGGGNYGPGSGSGGGYGGRSRY	341
人	YGPMSKSNFGGSRNMGGPYGGGNYGPGSGSGGGYGGRSRY	341
小鼠	YGPMSKSNFGGSRNMGGPYGGGNYGPGSGSGGGYGGRSRY	341

图 4. 大鼠、小鼠和人核不均一核蛋白 A2/B1 基因编码的氨基酸序列匹配 图中“*”表示完全一致的氨基酸残基，“.”表示半保守的氨基酸残基。

Figure 4. Multiple sequences alignment among rat, mouse and human hnRNP A2/B1 gene by Clustal W program

应元件结合,调节含 hnRNP A2 反应元件的 mRNA 的转运^[16],从而参与 mRNA 的合成、转运、翻译和端粒的保护。近年的研究发现在肺癌的早期, hnRNP A2/B1 基因表达明显增高,在癌症和增殖的细胞中与端粒单链 DNA 结合,防止核酸酶对端粒的水解,因此它成为一个早期诊断癌症的生物标记^[17,18]。hnRNP A2/B1 基因在癌症早期的表达上调也说明了基因的转录调节在癌症早期发生了改变。我们通过抑制消减杂交的方法发现 hnRNP A2/B1 基因在心肌缺血预适应时表达上调,并用反向 Northern 印迹加以证实^[5],那么它在缺血预适应中起着怎样的作用? 缺血预适应可增强心肌对其后长时间缺血的耐受性,这种保护机制主要涉及了心肌功能、代谢及基因表达水平的改变^[19]。缺血预适应能快速诱导一组早期基因的表达,它们编码多种转录因子,通过调控其下游基因的表达而对心肌细胞中多种代谢、功能及修复活动进行调节^[20,21]。其次,在遭受短暂冠状动脉阻断及再灌注后,心肌中多种抗氧化酶基因,热休克蛋白基因,低氧诱导因子-1 基因,环加氧酶以及生长因子基因均表达增多^[22-24]。这些表达上调基因的转录调节、mRNA 的转运和翻译可能需要 hnRNP A2/B1 基因的参与。所以 hnRNP A2/B1 基因在缺血

预适应的内源性保护机制中可能发挥了重要的作用。

【参考文献】

- [1] Burd CG, Dreyfuss G. Conserved structures and diversity of functions of RNA-binding proteins. *Science*, 1994, 265 (5172): 615-621
- [2] Mattaj JW. RNA recognition: a family matter? *Cell*, 1993, 73 (5): 837-840
- [3] Kumar A, Williams KR, Szes W. Purification and domain structure of core hnRNP proteins A1 and A2 and their relationship to single-stranded DNA-binding proteins. *J Biol Chem*, 1986, 261 (24): 11 266-273
- [4] Shyu AB, Wilkinson MF. The double lives of shuttling mRNA binding proteins. *Cell*, 2000, 102 (2): 135-138
- [5] 袁 灿, 吕青兰, 陈广文, 刘 瑛, 王尧玲, 刘海军, 等. 大鼠心肌缺血预适应诱导表达上调基因的筛选及鉴定. *生物化学与生物物理进展*, 2003, 30 (3): 390-394
- [6] Burge C, Karlin S. Prediction of complete gene structures in human genomic DNA. *J Mol Biol*, 1997, 268 (1): 78-94
- [7] Zhang MQ. Identification of protein coding regions in the human genome based on quadratic discriminant Analysis. *Proc Natl Acad Sci USA*, 1997, 94 (2): 565-568
- [8] BioEclone. 进行 EST 序列拼接的软件,网址是 <http://www.biosino.org/pages/lab.htm>
- [9] Blast. 进行两个核苷酸或氨基酸序列匹配的软件,网址是 <http://www.ncbi.nlm.nih.gov/BLAST/>
- [10] 美国国家生物技术信息中心(National Center for Biotechnology Information, NCBI). <http://www.ncbi.nlm.nih.gov>
- [11] Spidey. 根据 cDNA 序列和基因组序列确定基因的外显子和内含子的软件,网址是 <http://www.ncbi.nlm.nih.gov/IEB/Research/Ostell/Spidey/>
- [12] GENSCAN. 在基因组 DNA 中进行外显子预测的软件,网址是 <http://genes.mit.edu/GENSCAN.html>
- [13] MZEF. 在基因组 DNA 中进行外显子预测的软件,网址是 <http://aragon.cshl.org/genefinder/human.htm>
- [14] Clustal W. 进行多个核苷酸或氨基酸序列匹配的软件,网址是 <http://>

- www.ebi.ac.uk/clustalw/index.html
- [15] Ishikawa F, Matunis MJ, Dreyfuss G, Cech TR. Nuclear proteins that bind the pre-mRNA 3' splice site sequence r(UUAG/G) and the human telomeric DNA sequence d(TTAGGG)_n. *Mol Cell Biol*, 1993, 13(7): 4 301-310
- [16] Ainger K, Avossa D, Diana AS, Barry C, Barbarese E, Carson JH. Transport and localization elements in myelin basic protein mRNA. *J Cell Biol*, 1997, 138 (5): 1 077-087
- [17] Zhou J, Mulshine JL, Unsworth EJ, Scott FM, Avis IM, Vos MD, et al. Purification and characterization of a protein that permits early detection of lung cancer. Identification of heterogeneous nuclear ribonucleoprotein-A2/B1 as the antigen for monoclonal antibody 703D4. *J Biol Chem*, 1996, 271 (18): 10 760-766
- [18] Zhou J, Nong L, Wloch M, Cantor A, Mulshine JL, Tockman MS. Expression of early lung cancer detection marker: hnRNP-A2/B1 and its relation to microsatellite alteration in non-small cell lung cancer. *Lung Cancer*, 2001, 34 (3): 341-350
- [19] 周伏文, 李元建. 缺血预适应的血管内皮保护作用及机制. *中国动脉硬化杂志*, 1999, 7(4): 366-369
- [20] Das DK, Maulik N, Morara I. Gene expression in acute myocardial stress. Induction by hypoxia, ischemia, reperfusion, hyperthermia and oxidative stress. *J Mol Cell Cardiol*, 1995, 27 (1): 181-193
- [21] Nishizawa J, Nakai A, Higashi T, Tanabe M, Nomoto S, Matsuda K, Ban T, Nagata K. Reperfusion causes significant activation of heat shock transcription factor 1 in ischemic rat heart. *Circulation*, 1996, 94 (9): 2 185-192
- [22] Deindl E, Schaper W. Gene expression after short periods of coronary occlusion. *Mol Cell Biochem*, 1998, 186 (1-2): 43-51
- [23] Tyagi P, Tayal G. Ischemic preconditioning of myocardium. *Acta Pharmacol Sin*, 2002, 23 (10): 865-870
- [24] 吕青兰, 袁灿, 张华莉, 陈广文, 王尧玲, 邓恭华, 等. 用 cDNA 芯片检测大鼠心肌缺血预适应后基因表达谱的改变. *中国动脉硬化杂志*, 2003, 11 (3): 189-193
- (此文编辑 胡必利)