

一个大鼠心肌缺血-再灌诱导表达上调的新基因 **Mip1** 的克隆和特性分析*

袁 灿 张华莉 刘 瑛 王秋鹏 肖献忠**

(中南大学湘雅医学院病理生理教研室, 长沙 410078)

摘要 采用生物信息学方法和 5' cDNA 末端快速扩增法 (RACE) 技术相结合, 克隆了一个大鼠心肌缺血-再灌诱导表达上调的新基因 Mip1, 并经 RT-PCR 测序及多组织膜 RNA 印迹证实. 生物信息学分析显示, MIP-1 定位于大鼠染色体 1q12 区, 含 5 个外显子和 4 个内含子, 开放阅读框 (ORF) 为 1 827 bp, 编码 608 个氨基酸, 其编码蛋白 N 端含 KRAB 结构域, C 端含 14 个连续的 C2H2 型锌指蛋白结构域, 氨基酸第 277 位至 293 位为双向的核定位信号, 多组织膜 RNA 印迹显示该基因在脑组织表达最高, 其次是心脏, 在其他组织表达较低或无表达. 进一步深入研究该基因的功能具有重要生物学意义.

关键词 生物信息学, cDNA 末端快速扩增法, 基因克隆, Mip1

学科分类号 R36

短时间的心肌缺血-再灌可调动机体的内源性保护机制, 修复受损细胞, 并增强心肌对损伤的抵抗性和耐受性 (即缺血预适应)^[1]. 心肌的内源性保护机制涉及内源性保护物质的释放、膜离子通道的开放、信号转导分子的激活及基因表达改变等多个方面. 其中基因表达改变被认为是心肌内源性保护现象的重要分子基础. 研究表明, 短暂的心肌缺血-再灌可以诱导心肌细胞内多个基因表达上调, 这些基因的表达产物可发挥抗氧化、抗凋亡、促生长等功能, 从而对心肌细胞起到保护作用^[2,3]. 因此, 筛选及克隆短暂心肌缺血-再灌诱导表达上调的基因并探讨其功能, 对阐明心肌的内源性保护分子机制具有重要意义.

本室采用抑制消减杂交技术建立了大鼠心肌缺血-再灌诱导表达上调基因的 cDNA 文库^[4], 并从文库中筛选得到了多个已知基因及可能代表新基因的表达序列标签 (EST) 片段. 从这些 EST 片段出发, 通过生物信息学方法和 5' cDNA 末端快速扩增法 (rapid amplification of cDNA end, RACE) 相结合, 我们成功地克隆了一个大鼠新基因 Mip1, 并经 RT-PCR 测序和多组织膜 RNA 印迹证实. 该基因的 GenBank 登录号为 AY221750.

1 材料和方法

1.1 材料

SMART™ RACE cDNA 扩增试剂盒、Advantage 2 DNA 聚合酶及 Rat MTN Blot 试剂盒购自 Clontech 公司; T4 DNA 聚合酶和 pGEM-T 载体购自 Promega

公司; AMV 逆转录酶购自大连宝生物有限公司; 随机引物标记试剂盒为宝灵曼公司产品; [α ³²P] dCTP 购自北京亚辉公司; DH5 α 菌株为本室保存.

1.2 方法

1.2.1 EST 拼接: 从大鼠心肌缺血-再灌诱导表达上调基因的 cDNA 文库中我们筛选得到了一个长度为 510 bp 的 EST 片段, 与 NCBI 的 GenBank 数据库 (<http://www.ncbi.nlm.nih.gov/>) 进行 Blastn 比对未发现任何同源的已知基因. 采用 Bioecclone 软件 (<http://bioinfo.biosino.org:9090/clone.html>) 进行 EST 拼接, 构建该基因的 EST 重叠群, 再与 NCBI 的 GenBank 数据库进行 Blastn 比对, 仍未发现任何同源的已知基因, 表明该片段代表一新基因, 我们将该基因命名为 Mip1.

1.2.2 电子克隆: 将以上构建的 EST 重叠群与大鼠的 HTGs (high throughput genomic sequence) 进行 Blastn 比对; 将获得的基因组序列用 GENSCAN 进行外显子预测 (<http://bioweb.pasteur.fr/seqanal/interfaces/genscan.html>), 并拼接预测的外显子; 将拼接的外显子与 EST 重叠群再拼接, 然后分别与 NCBI 的 nr 和 dbEST 数据库进行 Blastn 比对, 确定预测的准确性.

1.2.3 RACE 技术获取基因 5' 末端: 根据以上所获序列设计基因特异性引物 Mip1-5' RACE (5'

* 国家重点基础研究发展规划项目 (973) (G2000056908) 和国家自然科学基金资助项目 (30170373, 30300138).

** 通讯联系人.

Tel: 0731-4805491, E-mail: xianzhongxiao@hotmail.com

收稿日期: 2003-07-28, 接受日期: 2003-09-03

caaggcgatgatatttggcctcggcg 3'), 采用 Clontech 公司的 SMARTTM RACE cDNA Amplification Kit 进行 5' RACE (rapid amplification of cDNA ends) 获取基因的 5' 末端. 具体方法按试剂盒说明书进行. 将 RACE 产物克隆入 pGEM-T 载体, 送上海博亚生物技术有限公司测序. 将所获 RACE 产物序列与预测序列拼接, 获取 Mip1 的基因全长.

1.2.4 RT-PCR 验证所获基因全长: 根据 Mip1 基因的最大开放阅读框 (ORF) 设计扩增引物. 5' 引物: 5' atgctgcagcccgagggaatcaaaat 3'; 3' 引物: 5' ttaggacatttctcgaatgtatactctgatgtc 3'. 引物由上海博亚生物技术有限公司合成. 首先以 2 μ g 大鼠脑组织总 RNA 为模板合成 cDNA 第一链, 再取 5 μ l 第一链合成产物进行 PCR 扩增, 扩增条件为 94 $^{\circ}$ C 30 s, 56 $^{\circ}$ C 30 s, 72 $^{\circ}$ C 150 s, 35 个循环. 将 RT-PCR 产物克隆入 pGEM-T 载体, 送上海博亚生物技术有限公司测序.

1.2.5 多组织膜 RNA 印迹: 采用 Clontech 公司的大鼠多组织膜 RNA 印迹试剂盒确定 Mip1 的组织表达分布及转录本大小.

1.2.6 生物信息学分析 Mip1 的核酸和蛋白质特性: 采用 NCBI 的 ORF Finder 程序分析 Mip1 的基因结构, Blast 程序进行核酸和蛋白质的相似性分

析, Map Viewer 程序观看 Mip1 的染色体定位; 在 ExPASy(<http://us.expasy.org/>)和 SMART(<http://smart.embl-heidelberg.de/>)分析 Mip1 的蛋白质结构域和特性.

2 结 果

2.1 EST 拼接

采用 Bioecclone 软件拼接得到一长度为 1 147 bp 的 EST 重叠群, 该重叠群由以下 EST 拼接而成: AI548984、BF522681、BF394294、BF386074、CB782714.

2.2 电子克隆

将以上构建的 EST 重叠群与大鼠的 HTGs 进行 Blastn 比对获得该基因相应的基因组序列 NW_043352. 将 NW_043352 中包含 EST 重叠群的约 22 kb 序列用 GENSCAN 进行外显子预测, 获得 5 个可能的 外显子. 将 5 个外显子拼接成的 cDNA 片段 (2 055 bp) 与 dbEST 进行 Blastn 比对 (图 1), 发现 cDNA 片段的 3' 端均有许多 EST 匹配, 但 5' 端相似性高的 EST 序列均在 261 bp 处中断, 表明 261 bp 以前的序列预测不对. 经 EST 继续延伸仍无法确定 5' 端全长, 因此决定采用 RACE 技术获取基因的 5' 端.

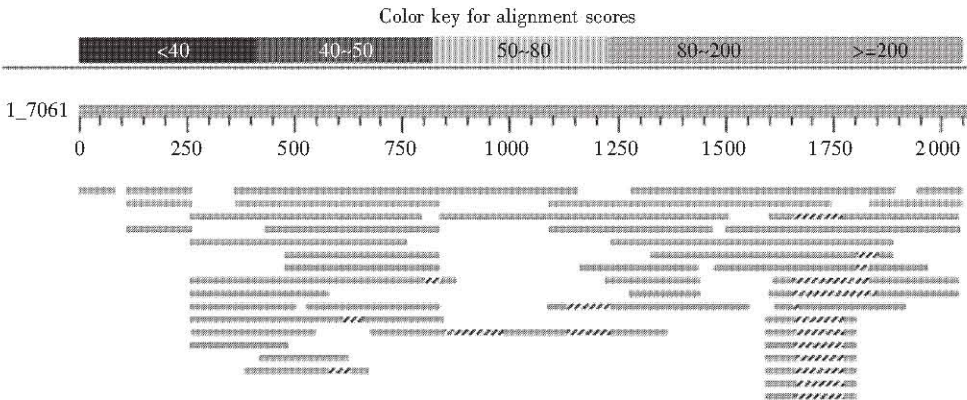


Fig. 1 Alignment of Mip1's cDNA which predicted by GENSCAN in dbEST database

2.3 5' RACE

从距预测的 cDNA 片段 5' 端 423 bp 位置处开始设计引物进行 5' RACE, 获得一约 350 bp 左右的片段 (图 2). 将 PCR 产物纯化后克隆入 pGEM-T 载体, 测序后获得的序列与 GENSCAN 预测序列及 EST 重叠群拼接后, 获得一长度为 3 396 bp 的序列. 将该序列与 dbEST 进行 Blastn 比对完全受 EST 支持 (图 3).



Fig. 2 350 bp fragment was amplified by 5' RACE of Mip1

1: Φ X174/Hae III marker; 2: 5' RACE of Mip1.

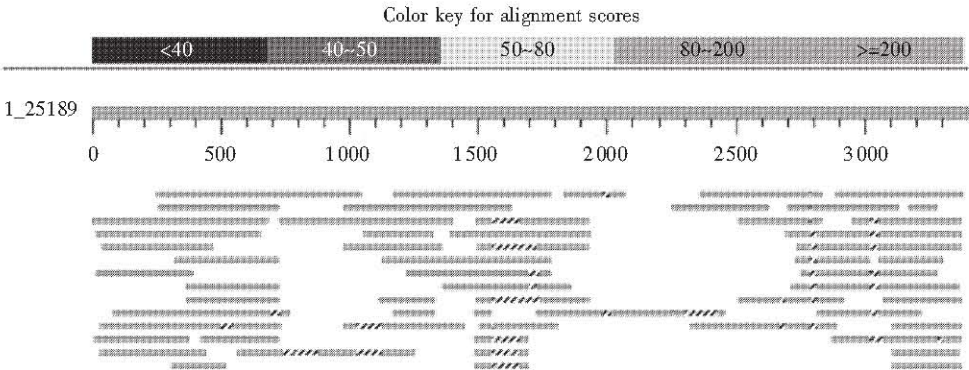


Fig. 3 Alignment of the full length of Mip1's cDNA in dbEST database

2.4 RT-PCR 验证所获基因全长

以大鼠脑组织总 RNA 为模板，RT-PCR 扩增出了 Mip1 的最大开放阅读框架（图 4）。将 PCR 产物克隆入 pGEM-T 载体，送上海博亚生物技术有限公司测序。测序结果与预测结果完全一致。

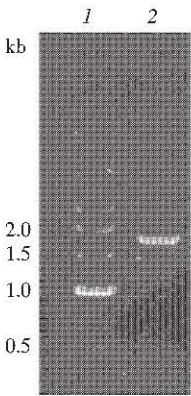


Fig. 4 The ORF of Mip1 was amplified from rat's brain by RT-PCR
1: 1 kb DNA ladder; 2: ORF of Mip1.

2.5 多组织膜 RNA 印迹结果

多组织膜 RNA 印迹结果显示，Mip1 在脑和心表达较高，肝、睪丸、肾、骨骼肌有弱表达，脾和肺中未检测到表达（图 5）。Mip1 只有一个转录本，分子质量大小约为 3.4 kb 左右，与我们所获得的 cDNA 大小相符。

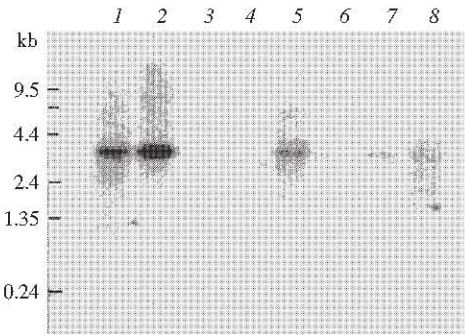


Fig. 5 Multiple tissue Northern blot result of Mip1
1: heart; 2: brain; 3: spleen; 4: lung; 5: liver; 6: skeletal muscle; 7: kidney; 8: testis.

2.6 Mip1 的核酸与蛋白质特性的生物信息学分析

采用 ORF Finder 软件分析 3 396 bp Mip1 的 cDNA 全长的最大开放阅读框为 1 827 bp，起始密码子 ATG 位于 120 bp 处，终止密码子位于 1 946 bp 处，起始密码子前同一相位存在一终止密码子 TAG，3 343 bp 处有加尾信号“ AATAAA ”，polyA 尾出现于 3 382 bp 位置，显示其为一完整基因（图 6）。采用 Spidey 软件分析该基因含 5 个外显子，4 个内含子（图 7）。Map Viewer 显示 Mip1 定位于大鼠染色体 1q12 区（图 8）。Mip1 编码 608 个氨基酸，Blastn 比对显示其与人类 R31155_1 蛋白有 70% 的同源性。Expasy 和 SMART 分析显示 Mip1 的编码蛋白 N 端含 KRAB 结构域，C 端含 14 个连续的 C2H2 型锌指蛋白结构域（图 9），氨基酸第 277 位至 293 位为双向的核定位信号。

gcaggcgccggggcctctgtgctggcagctctgtggcagctgtgagacggagacgctgTA GggagatagctctcggtctcagctctgaatgtggaccgtcaaaagctggaggaaagaATGcc
tgcaagcccagggaatacaaaatccaaggcaccagtgacttttgggacttagccatctactctacagaggaatgggaafgctgagctccaatcagaaaggattgtacgaagatgtcatgttggag
aactatcacaaatctgtctcagttgtgttgcctgcccaggccaaatatcatgacctgttggagaaaggaaagcaccatggatggtggagccatcaagaaagcgcgggggtccctgaattagggtcta
aggatgagaccaagaaagtaccaccaagtcagttaacaatctggccaagcatctgtaaagaaccagattcttcacaaacaaagggtccacagagaaggcaaaacataaagaatgcagtcacca
gaaaaaacagaaggggcattcagcgcaagaaatctttaaagtgtatagctgtgggaaaacctttttgcgaagcttatctcttaaacctcatcaaggctttcacactggagagaggctctatgagtcagctact
tgtggacaagttttcaggcagatcttaicctcattctccaccagagagttcacactcagaataaagctatgaatgtgataaggtgtgggacatttttaataaaagttactctgatgattcatagaagaagtc
taatggaaaaggaaaattttcacatgagaagacttcagatcatgcccattccctcagtcacaccataataatcacgctattgacagcattcaccagtgcagaaaatgtgggaaagctttcagccggatgtca
tccctttacttataagaagattcacacagaagaagaatacaaaatacagtgcatgtggagagggcttcaaaaagaagccagtcctgtgacataaaaagaattgtattgggaagaaaactcatgaaaata
aggccttaattcaaggtctcaggcagagaacttaccagtcagaaaaccttttactgtgagaaatgcaggaagtccttttagagatctcagccttaattgtcttcacagagagctcacacttcagggaaccc
ctacaatgtgataaggtcagaaagacttcggacgactttcaaccttattctgcacctcagaattcatagtgaggagaaagcagtttaagtgcacaaggtgtgagaaggatgcaatgcctttcatcctttat
tcagcacaagaaaatccataaaaggaaaagaactatcagagtgcaaggatgtggaaagatgtttgtgtgaatgaaaaaccttaaaagtacatctcaactccattcggaggagaacctttcaagtg
aataagtgtgttaaagtttttgcctgcatgctattcttagtgaacatcaaaagaattcactctggagagaagccctatcagtgtagaaatgtgggaaagccttcagtcaccggatctcttccacacataa
gagaattcacagcgaggacagaccttacgaatgtgacctgtgtgggaaagccttcagccagagtgctcatcttgcctcagcatgaaagaattcacactggagagaagccttacgcgtgcaaaatgtgaa
aagtccttactcagcgcatatccctacttcatgagaagaagtcactctgggagaagcccttcagagtgtaatgaatgtgggaaagcattcagtagtgctcagacattatccgacatcagcgaagtc
tcttcagagaaccatfatgaatgtgcaaatgtgtaaggcatatagccggagctcatctctgattcgacatcagagatatacttcggaggaaatgtctCTAAGgctgtcgtctatatctgcaaaagctaaaa
aggacatcaaaagaaggttcagagacatgaggaatgtgtgtaagtgtcttcttactatgatgacgtatgtcccaataaacccatcgtaagtgtgaagcgtctgagacaaacagatatttaataacctgacctac
caaatgtgatcatttagcagcacagtgtaagtgtggagtcctcgttctcctgactgtatgctcttctggaagccatggcctgatcccgctccagcattgtgagagtgatctcgtgaccatccg
cctaggtcacttttacccaatctcagttgaatggatattgtctctgtactatgtaatgctaaaaaacataaaatgggaactatctgcctatacatattttgtgtatgtggggcggtataatattctcctttgaa
agccaagaaacaaaagtagttgtgtgattttgaagtgaagatctgaatcagctgaggatgtctcagaattttggatcaggggtttgtgaatgaagggtgtttttcaatgtgcaatttctatgtgtgagctttgag
cctgagttgaggtcccaggaaaatagagtattaaagttaacaggcaaacaggcaaaagtaggggtgcctcttctcctgacagatctgtgttggggagagaacaaaagagtgagcttttaggtgctagaaa
tagtagctacagtggtgtggaagtagtcaataccacaaaagataccctgagaaaatttttaagacaagctgggggctgtggcagagatactgaaaatgaggaacttcagagcagcagagaagct
attttttactttttacttgaaaataaactgtctatatttatataagtaaacagtggtttcatgcaataattccagtgctgtacaaaaaatactcttgagatacatgtttgggttttcatggcttgatattgtgccccc
caaaactcatatgttgaaagccctaacccttcaataggtgataattttgaaatgagaaatgaagtcatgatgtggatggccttttgatacgctttgatcagacacacaagaaggctttgtctctctgtcccat
gcaagaacacagcgaggccatctgcaaacagggaiaaacccctaccagaacctgaccagtttggtactctcactgtgacttcagcttataagcgaacacaggaacctagtttctctgaagtcagcct
gtctgtgatactgtgttatagcaccctaaactgattatgtgtggaagaactcctaggttaaggtaactggctgttttcttctgtctcacaacacgttaatgattttttgtctcctttccagccctctgtgaa
ctatttaaAATAAAGtctgaattctttgtgtgattaggtcatcacAAAAAAAAAAAAA

Fig. 6 Characterization of Mip1's cDNA

The start condon and end condon and polyA signal were marked with thick black.

1 747 376 1 730 220

Fig. 7 Alignment of Mip1's cDNA in rat genome database by Spidey software indicate that there are 5 exons and 4 introns

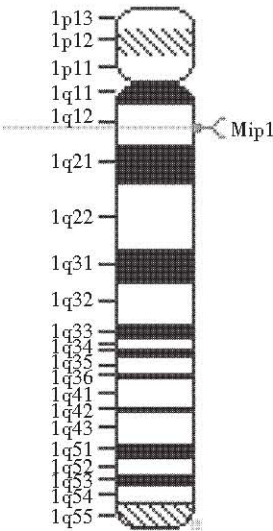


Fig. 8 Mip1 was located in chromosome 1q12 by Map viewer in NCBI

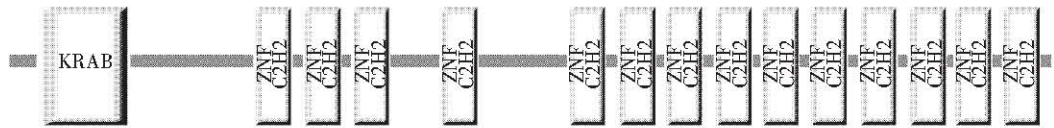


Fig. 9 SMART analysis indicate that the N terminal of Mip1 protein has a KRAB domain, and the C terminal has 14 successive C2H2 type zinc finger domain

3 讨 论

随着多种模式生物的全基因组测序陆续完成^[5,6], 基因组及 cDNA 数据库得到进一步完善. 而计算机分析软件的开发应用, 使人们可能通过与基因组及 cDNA 数据库进行简单的序列比对、拼接及软件预测而获得一个新基因的全长, 即所谓的电子克隆^[7]. 与传统的基因克隆方法相比, 电子克隆法大大加快了新基因的发现速度, 提高了工作效率. 然而, 基因组及 cDNA 数据库并没有囊括所有基因的全部信息, 各种分析预测软件也各有其缺陷及限制, 因此电子克隆法并非万全之策, 许多基因的克隆还需结合实验工作, 并最终必须获得实验验证. 本文从一个大鼠心肌缺血预适应诱导表达上调的 EST 片段出发, 首先通过生物信息学方法获得该基因尽量长的 cDNA 片段, 然后结合 5'RACE 技术获取基因全长, 并采用 RT-PCR 和多组织膜 RNA 印迹加以证实, 由此既快速又可靠地克隆了一个新基因 Mip1.

生物信息学分析显示, Mip1 编码蛋白 C 端含 14 个连续的 C2H2 型锌指结构域, 属典型的 C2H2 型锌指蛋白. 具有锌指结构的蛋白质统称为锌指蛋白 (zinc finger protein, ZFP)^[8]. 锌指结构是蛋白质与 DNA 双螺旋结构之间结合的一种形式, 是最为典型的转录因子标志性结构之一^[9]. 根据组成锌指结构的氨基酸序列, 锌指蛋白可分为 C2H2 型、C3H 型、C4 型和环指型. 其中 C2H2 型是最主要的一种类型^[10], 它们的锌指结构由包括两个半胱氨酸残基和两个组氨酸残基的 21 个氨基酸残基组成, 即 $CX_2C-X_{12}-HX_3H$, 其中 X 代表任何氨基酸. 约有 1/3 的 C2H2 型锌指蛋白含有 KRAB (Kruppel-associated box) 型结构域, 该结构域被认为起转录抑制作用. 而 Mip1 蛋白的 N 端正含有 KRAB 结构域, 因此推测 Mip1 可能是一个具有转录抑制作用的转录因子. 并且生物信息学分析还显

示, Mip1 编码蛋白的氨基酸第 277 位至 293 位为双向的核定位信号, 这进一步增强了其为一核转录因子的可能性.

多组织膜 RNA 印迹结果显示, Mip1 在心、脑等耗氧量高的组织中表达量最高, 而其他组织表达较少或无表达, 结合其在心肌缺血预适应时应表达上调, 因此 Mip1 可能与心肌缺血缺氧时的功能代谢调节有关. 进一步深入研究 Mip1 在心肌缺血-再灌时的时空表达模式、亚细胞定位、其上下游调控机制及相互作用蛋白, 探讨其在心肌缺血-再灌时所发挥的功能, 可能为揭示心肌的内源性保护分子机制提供新的思路.

参 考 文 献

- 1 Kloner R A, Jennings R B. Consequences of brief ischemia: stunning, preconditioning, and their clinical implications: part 2. *Circulation*, 2001, **104** (25): 3158 ~ 3167
- 2 Deindl E, Schaper W. Gene expression after short periods of coronary occlusion. *Mol Cell Biochem*, 1998, **186** (1 ~ 2): 43 ~ 51
- 3 Depre C, Tomlinson J E, Kudej R K, *et al.* Gene program for cardiac cell survival induced by transient ischemia in conscious pigs. *Proc Natl Acad Sci USA*, 2001, **98** (16): 9336 ~ 9341
- 4 袁 灿, 吕青兰, 肖献忠, 等. 大鼠心肌缺血预适应诱导表达上调基因的筛选和鉴定. *生物化学与生物物理进展*, 2003, **30** (3): 390 ~ 394
- 5 Yuan C, Lü Q L, Yuan K Y, *et al.* *Prog Biochem Biophys*, 2003, **30** (3): 390 ~ 394
- 6 Venter J C, Adams M D, Myers E W, *et al.* The sequence of the human genome. *Science*, 2001, **291** (5507): 1304 ~ 1351
- 7 Waterston R H, Lindblad-Toh K, Birney E. Initial sequencing and comparative analysis of the mouse genome. *Nature*, 2002, **420** (6915): 520 ~ 562
- 7 Wheeler R B, Sharp J D, Schultz R A, *et al.* *In silico* cloning: the identification of genes within the CLN6 critical region. *Am J Hum Genet*, 2001, **69** (4): 1573 Suppl
- 8 Laity J H, Lee B M, Wright P E. Zinc finger proteins: new insights into structural and functional diversity. *Curr Opin Struct Biol*, 2001, **11** (1): 39 ~ 46
- 9 Tan S, Richmond T J. Eukaryotic transcription factors. *Curr Opin Struct Biol*, 1998, **8** (1): 41 ~ 48
- 10 Wolfe S A, Nekhudova L, Pabo C O. DNA recognition by Cys2His2 zinc finger proteins. *Annu Rev Biophys Biomol Struct*, 2000, **29**: 183 ~ 212

Cloning and Characterization of a New Gene Mip1 Up-regulated During Myocardial Ischemia-reperfusion *

YUAN Can, ZHANG Hua-Li, LIU Ying, WANG Qiu-Peng, XIAO Xian-Zhong **

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

Abstract Using bioinformatics analysis and 5'RACE technology, a new gene Mip1 which is up-regulated during rat myocardial ischemia-reperfusion had been cloned. The full length of Mip1 was proved by RT-PCR sequencing and multiple tissue Northern blot. Bioinformatics analysis indicate that Mip1 is located in chromosome 1q12, including 5 exons and 4 introns. The full-length cDNA has an open reading frame of 1 827 bp, which encodes 608 amino acids. There is a KRAB domain at the N terminal of the hypothetical protein and 14 successive C2H2 type zinc finger domain at the C terminal of the protein. There is a bipartite nuclear targeting sequence from amino acid 277 to 193. Mip1 is expressed abundantly in brain and heart and seldom in other tissues. It is worthwhile to further study the biological functions of Mip1.

Key words bioinformatics, rapid amplification of cDNA end (RACE), gene cloning, Mip1

* This work was supported by grants from The Special Funds for Major State Basic Research of China (G200056908) and The National Natural Sciences Foundation of China (30170373).

** Corresponding author. Tel: 86-731-4805491, E-mail: xianzhongxiao@hotmail.com

Received: July 28, 2003 Accepted: September 3, 2003