

FRG4 基因全长克隆及生物信息学分析

王仁¹, 姜志胜^{1,2}, 王佐², 赵战芝², 李国华¹, 杨向东²

(南华大学 1. 病理生理学教研室, 2. 心血管病研究所 动脉硬化化学湖南省重点实验室, 湖南省衡阳市 421001)

[关键词] 生物信息学分析; 动脉粥样硬化; 突触相关蛋白; 跨膜区; 功能结构域

[摘要] 为了对新的人突触相关蛋白 FRG4 进行全长克隆及生物信息学分析, 从人胎肝文库 PCR 扩增 FRG4 基因全长 cDNA 序列, 用生物信息学方法对 FRG4 基因进行基因组结构分析、多序列同源性比较、跨膜区段、亲疏水性分析、功能结构域预测等。结果表明 cDNA 文库基础上运用热启动 PCR 获得 FRG4 全长 cDNA 序列, 生物信息学分析显示 FRG4 基因与人类的突触相关蛋白有 99% 同源性; 定位在 X 染色体的短臂 2 区 2 带 2 亚带 2 次亚带; 由 9 个外显子和 8 个内含子组成; 无跨膜区段; 有一结构域 BSD; 该基因编码蛋白可能为一水溶性蛋白。

[中图分类号] Q81

[文献标识码] A

The Full-Length cDNA Cloning and Bioinformatic Analysis of FRG4 Gene

WANG Ren¹, JIANG Zhi-Sheng^{1,2}, WANG Zuo², ZHAO Zhan-Zhi², LI Guo-Hua¹, and YANG Xiang-Dong²

(1. Department of Pathophysiology, 2. Institute of Cardiovascular Disease, Key Laboratory for Atherosclerosis of Hunan Province, University of South China, Hengyang 421001, China)

[KEY WORDS] Bioinformatics Analysis; Atherosclerosis; Synapse Associated Protein; Transmembrane Domains; Functional and Structural Domains

[ABSTRACT] To analyze the basic character of a novel Homosapiens synapse associated protein gene (FRG4) with bioinformatics. FRG4 full-length sequence was obtained by hot-start PCR from human fetal liver library and its genomic constitution, the multi-sequences homology, the transmembrane domains, hydrophilicity and hydrophobicity, as well as its functional and structural domains were analyzed and predicted by bioinformatics. We have successfully attained the full-length cDNA sequence of FRG4, bioinformatic analysis shows FRG4 gene has 99% homology with Homosapiens synapse associated protein 1 (SYAP1), consists of 9 exons and 8 introns and located in Xp22.2. None transmembrane domains but a BSD domain was discovered, and its encoding protein is probably a water-solubility protein.

FRG4 基因是应用抑制消减杂交法 (suppression subtractive hybridization, SSH) 在氧化型低密度脂蛋白 (ox-LDL) 致 U937 细胞泡沫化过程中筛选到的差异表达基因 (Genbank 登录号为: GenBank, Accession BI 502586)^[1,2]。应用热启动 PCR 快速扩增 cDNA 末端技术, 获得 FRG4 基因全长 cDNA 序列 (1 213 bp), 其编码蛋白与人突触相关蛋白 (Homosapiens synapse associated protein 1, SYAP1) 有 99% 同源性。经联机检索, 突触相关蛋白与脂蛋白代谢和泡沫细胞形成、凋亡的关系尚少有文献, 因此, FRG4 基因的发现, 为研究动脉粥样硬化提供了新的思路。

在新基因功能研究中生物信息学技术是一重要工具, 该技术可以改变传统的生物学实验模式, 实现从理论推测出发, 然后再回到实验中去, 进而追踪或验证这些理论假设, 对新基因功能的研究具有很好的指导作用^[3]。本文就是利用生物信息学技术对 FRG4 基因进行相关分析, 为深入展开对该基因的研究提供理论基础。

1 材料与方法

1.1 细胞和主要试剂

人髓系白血病细胞 U937 细胞购自中科院生物化学与细胞生物所; 人胎肝、胎心、胎盘 cDNA 文库 (载体为 λ gt10) 和高保真聚合酶链反应试剂盒购自 Clontech 公司; 克隆载体 pGEM-T easy 载体和连接酶等试剂购自 Promega 公司; 所用 PCR 引物由北京赛百盛公司合成, 其它试剂均为进口或国产分析纯。

1.2 逆转录聚合酶链反应检测 FRG4 表达

ox-LDL 处理 U937 细胞 48 h 建立泡沫细胞模型 (本所成熟的泡沫细胞模型)^[4], 提取细胞总

[收稿日期] 2009-02-15 [修回日期] 2009-06-09

[基金项目] 国家自然科学基金 (30570766, 30470693)、湖南省教育厅重点科研项目 (05A043)、湖南省教育厅基金 (08C747)、动脉硬化化学湖南省高校科技创新团队资助项目 (湘教通 [2008] 244) 和南华大学归国留学人员科研启动基金

[作者简介] 王仁, 讲师, 博士研究生, 研究方向为动脉粥样硬化发病机制, E-mail 为 wangren73@yahoo.com.cn。姜志胜, 教授, 博士研究生导师, 主要研究方向为动脉粥样硬化和缺血性心肌损伤的发病机制及其防治。通讯作者杨向东, 副教授, 硕士研究生导师, 主要从事心血管疾病相关基因克隆及细胞凋亡机制研究。

RNA。取 2 μg 总 RNA,65℃ 变性 3 min,加入 10 × buffer、MgCl₂、Dntp、引物、Rnase inhibitor 及反转录酶,总体积 20 μL,42℃ 保温 90 min,70℃ 10 min 终止反应。PCR 反应体系共 20 μL,94℃ 预变性 5 min,94℃ 变性 1 min,58℃ 退火 1 min,72℃ 延伸 1 min,共 30 个循环,72℃ 延伸,4℃ 保存。设计引物上游为 5'-cttgctctctgccattgctc-3',下游为 5'-cgggccctggatgctcc-3'。RT-PCR 检测 FRG4 的表达,并分别检测人胎肝、胎心和胎盘 cDNA 文库中 FRG4 的表达。

1.3 FRG4 cDNA 全长克隆及开放读码框分析

以 FRG4 的 cDNA 序列设计引物,上游 5'-CCGGGATCCttccggggcttgagcagttgggtt-3',下游 5'-CCGGAATTC gtagaagaggcaagttcaactt-3',上下游引物各含 BamHI 及 EcoRI 酶切位点,选择合适的文库进行扩增,回收热启动 PCR 产物,连接 pGEM-T easy 载体,提取质粒,酶切鉴定后送上海生工测序,软件分析开放读码框(ORF)。

1.4 新基因 FRG4 的生物信息学分析

用 Blastn、Blastp 进行同源性分析和基因的染色体定位;ORF finder 软件分析该基因的开放阅读框;Compute pL/MW 软件分析该基因编码蛋白的氨基酸组成、分子质量和等电点;SAPS 软件分析该蛋白的电荷分布和跨膜区段;SignalP 软件分析该蛋白的信号肽剪切位点;Psort II 分析该蛋白的亚细胞定位;BIOEDIT 软件分析该蛋白的亲疏水性;Domain Database 预测该蛋白的结构域。

2 结果

2.1 逆转录聚合酶链反应检测结果

FRG4 是用抑制消减杂交(SSH)法获得的差异表达基因,根据其 cDNA 序列设计特异引物,RT-PCR 检测结果表明,FRG4 在泡沫细胞中表达增高

(图 1)。以人胎肝、胎心和胎盘 cDNA 文库为模板进行 PCR 扩增,可见 FRG4 在胎肝 cDNA 文库中表达最高(图 2),选用胎肝 cDNA 文库扩增 FRG4 全长 cDNA 序列。

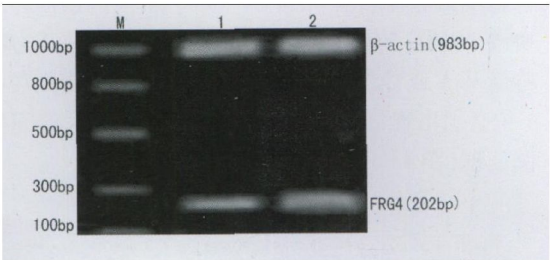


图 1. FRG4 表达检测 M 为 DNA Maker,1 为对照组 U937,2 为处理组泡沫细胞。

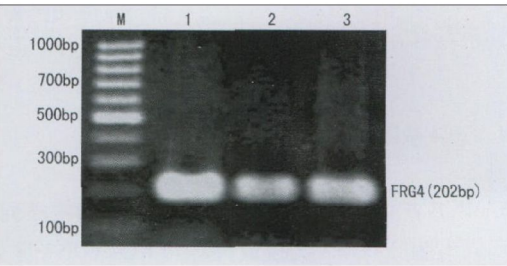


图 2. 人胎肝、胎心和胎盘 cDNA 文库中 FRG4 检测 M 为 Maker,1 为人胎肝文库,2 为人胎心文库,3 为人胎盘文库。

2.2 FRG4 cDNA 全长序列获得及开放读码框分析

从人胎肝 cDNA 文库中提取 DNA 作为模板,根据 FRG4 设计引物进行 PCR 扩增,PCR 产物经琼脂糖凝胶电泳鉴定。随后将 PCR 产物连接于 pGEM-T easy 载体,挑选阳性克隆抽提质粒测序,将测序结果输入 nr 数据库用 Blast 查新、校对并经 Genbank 完整 EST 数据库比对,最终确定 FRG4 cDNA 全长序列为 1 213 bp(图 3)。

```
1 ggcaaccagt gctcgtcg gctctgggg atcgggacg cgccggcgcc cgcgagcgg gATCttccg ggcttgagca gtggttggg cttgcagcag ccggtggcag gcggtgggca
121 gcccaatgga gatgcttacc ccgagcagcc gtccgagacg gtggcagat ctgcggagga ggagctgcag caagcgggag accagagatc cctccaccag gccaaagat tcggcaacta
241 ttatttlaac ttgcatctg ctgccacaaa aagataact gaatcagttg ctgaacacgc acaacaata aagaatccg tagaagaagg aaaaatagat ggcatattg acaagacaa
361 tataggagat tticagagg aacagaaaaa atttttgaa gagcaacata caaagaagtc agaagcagct gtgcccacal ggttgacac taacgatgaa gaaacaatic aacaacaa
481 ttggcctta tcagctgaca agagaaattt ccttcgtgac cctccggctg gcttgcaatt taattcgac ttgatcaga tgaaccctgt ggccttggtc atgctccagg aggatgatc
601 gctaaagcag atgagatttg ccttcgttc taaacttggt aaggaagaag tgttttgag gaactactt taccgctct ccttgattaa gcaatcagcc cagctacagg ccttgctgc
721 ccaacagcag gccgcaggga aggaggagaa gagcaatggc agagagcaag attgcccgt ggcagagcca gtacgacca aaacgccacc cgttgtaate aaatctcagc ttaaaacta
841 agaggatgag gaagaaattt ctactagccc aggtgtttct gattttgca gtgatgcctt cgaatcctgt aacctaaatc aggaagatct aaggaagaa atggagcaac tagtgcttga
961 caaaaagcaa gaggagacag ccgtactgga agaggattct gcagattggg asaaagaact gcagcaggaa cttaagaat atgaagtgtt gacagaaatc gaaaaacgag atgaaaaactg
1081 ggataaggaa atagagaaaa tgcctcaaga ggaaattag ctgttcctga aatagaagaa taatccttaa cagtctgcaa actgacatta aattctagat gtgacaatt actgaaaaa
1201 aaaaaaaaaa aa
```

图 3. FRG4 全长 cDNA 序列

用 Internet (www. ncbi. nlm. nih. gov/gorf/gorf. html) 中的 ORF finder 软件进行开放阅读框分析, 结果显示, ORF 从第 62 个碱基到第 1 120 个碱基, 编码 352 个氨基酸, 该阅读框起始密码子 ATG 的 -3 位碱基为嘌呤碱基 G, 符合 Kozak 规律; 阅读框起始密码子 ATG 的周围序列部分, 如果 -3 位碱基为嘌呤碱基 A 或 G, 则是有效的翻译起始位点; 否则在 +4 位必须出现 G, 表明该阅读框是一个最大的完整阅读框 (图 4)。

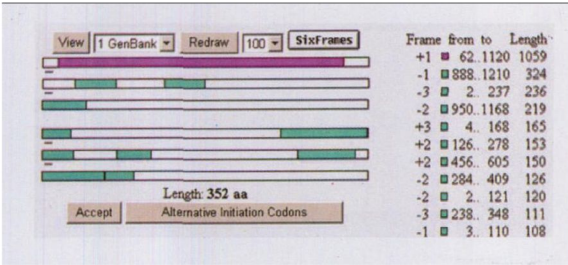


图 4. FRG4 基因的开放阅读框

2.3 PCR 扩增验证 FRG4 基因的开放阅读框序列

以人胎肝 cDNA 为模板, 根据 FRG4 cDNA 序列设计特异的引物, PCR 扩增其 ORF 序列, 产物经 1.0% 琼脂糖凝胶电泳分析, 得到的 PCR 产物大小与分析结果一致 (图 5)。

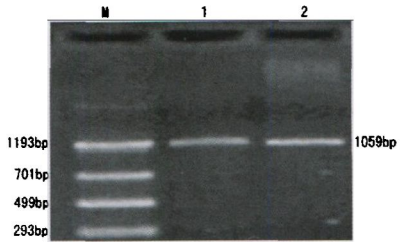


图 5. PCR 扩增验证 FRG4 基因的开放阅读框序列 M 为 DNA marker, 1 和 2 为人胎肝文库中 PCR 扩增新基因 ORF。

2.4 FRG4 生物信息学分析

2.4.1 Blast 及基因组结构分析 运用 Blastn 对新基因 FRG4 进行核酸同源性比较, 结果显示, FRG4 与 SYAP1 同源性达到 99%; 该基因定位在 X 染色体的短臂 2 区 2 带 2 亚带 2 次亚带。进一步分析发现, FRG4 基因由 9 个外显子和 8 个内含子组成。

2.4.2 人 FRG4 蛋白质氨基酸的同源性比较 用 Clustal 软件对人 FRG4 蛋白质氨基酸序列进行多序列同源性比较, 结果显示, FRG4 编码蛋白与人、小鼠和果蝇的氨基酸序列一致率分别为 86%、70% 和 58% (图 6)。



3 讨论

生物信息学(bioinformatics)是一门综合运用生物学、数学、物理学、信息科学及计算机科学等诸多学科的理论方法来研究生物学和生物学过程中的信息流的综合交叉科学。它是当今生命科学和自然科学的重大前沿领域之一,其研究重点主要体现在基因组学(genomics)和蛋白组学(proteomics)两方面,具体说,是从核酸和蛋白质序列出发,分析序列中表达的结构与功能的生物信息。通过它独特的桥梁作用和整合作用,使我们能够从各生物学科中众多分散的观测资料中获得对生物学系统和生物学过程的运作机制的理解,最终达到自由应用于相关实践的目的。

FRG4 是通过抑制消减杂交筛选 ox-LDL 诱导 U937 细胞泡沫化过程中差异表达得到的新基因,RT-PCR 检测人胎肝、胎心和胎盘组织的 cDNA 文库,FRG4 基因在人胎肝 cDNA 文库表达最高,设计酶切位点及引物,将从文库中扩增的全长片段克隆至 pGEM-T easy,建立 FRG4 全长克隆,同时设计引物,以人胎肝 cDNA 文库为模板,PCR 扩增其开放阅读框序列,琼脂糖凝胶电泳检测,其产物片段大小与预期大小一致。

生物信息学软件及数据库资源对 FRG4 进行初步的功能分析,Blast 分析发现 FRG4 与人类突触相关蛋白 SYAP1 有高度同源性,达到 99%,基因组分析结果发现,FRG4 基因由 9 个外显子和 8 个内含子组成,定位于 X 染色体的短臂 2 区 2 带 2 亚带 2 次亚带。另外,对该蛋白的理化性质进行分析表明,该蛋白为分子质量 40 kDa、等电点 4.45 的一略酸性的水溶性蛋白质,亚细胞定位于细胞浆中,无跨膜区段,蛋白序列中无信号肽剪切位点,该蛋白序列中有一功能域 BSD,包含一特异性基元(FW-motif),它是真核转录因子家族的成员。氨基酸序列进行多序列同源性比较,发现与人、小鼠和果蝇的氨基酸序列一致率分别为 86%、70% 和 58%。该基因编码的蛋白

可能是一转录因子,其功能尚需进一步的研究。

目前研究发现,突触相关蛋白(family of synapse-associated proteins, SAP)家族主要分布在兴奋性或抑制性突触的突触前膜或后膜。SAP 的功能与突触膜上的谷氨酸酯受体(glutamate receptor)和钾离子通道(voltage-gated K^+ channels)的定位和功能有关。SAP 在神经系统中重要的作用,它能介导神经递质的释放,促进神经的生长、发育及修复,也与大脑的记忆功能有密切的联系。SAP 基因的突变或表达改变,可能引起一系列的神经精神疾病如 Parkinsonism 病、癫痫病甚至精神分裂症等。FRG4 与人类突触相关蛋白 SYAP1 有 99% 同源性,说明 FRG4 可能是 SAP 家族新成员。当前关于 SAP 功能研究主要集中于神经系统,对其与动脉粥样硬化等心血管疾病的关系尚无文献报道^[5-9]。FRG4 全长 cDNA 克隆的建立及其基本性质的分析,为进一步研究 FRG4 基因的功能作用奠定了基础。

[参考文献]

- [1] 杨向东,王抒,唐蔚青,等. 人单核细胞泡沫化敏感候选基因的筛选[J]. 中国动脉硬化杂志, 2002, 10 (3): 195-198.
- [2] 王仁,杨向东,刘俊文,等. 血管壁细胞突触相关蛋白的表达[J]. 中国动脉硬化杂志, 2004, 12 (1): 128-131.
- [3] 张文红,刘新平,王瑛,等. Ndr2 基因及 NDRG2 蛋白的生物信息学分析[J]. 第四军医大学学报, 2003, 24 (17): 208-214.
- [4] 李全忠,杨永宗,易光辉,等. U937 泡沫细胞模型的建立[J]. 中国动脉硬化杂志, 1997, 7 (2): 152-154.
- [5] Riley DE, Krieger JN. Embryonic nervous system genes predominate in searches for dinucleotide simple sequence repeats flanked by conserved sequences [J]. *Gene*, 2009, 429 (1-2): 74-79.
- [6] Sato J, Shimazu D, Yamamoto N, et al. An association analysis of synapse-associated protein 97 (SAP97) gene in schizophrenia [J]. *J Neural Transm*, 2008, 115 (9): 1355-365.
- [7] Abi-Char J, El-Haou S, Balse E, et al. The anchoring protein SAP97 retains Kv1.5 channels in the plasma membrane of cardiac myocytes [J]. *Am J Physiol Heart Circ Physiol*, 2008, 294 (4): H1851-861.
- [8] Lu Z, Je HS, Young P, et al. Regulation of synaptic growth and maturation by a synapse-associated E3 ubiquitin ligase at the neuromuscular junction [J]. *J Cell Biol*, 2007, 177 (6): 1077-089.
- [9] Marcello E, Gardoni F, Mauceri D, et al. Synapse-associated protein-97 mediates alpha-secretase ADAM10 trafficking and promotes its activity [J]. *J Neurosci*, 2007, 27 (7): 1682-691.

(此文编辑 许雪梅)