

人伴肌动蛋白相关锚定蛋白基因 cDNA 的克隆与功能分析*

袁红丰 王冬梅 李海民 陈琳 王晓 白慈贤 岳文 裴雪涛**

(军事医学科学院输血研究所干细胞研究室, 北京 100850)

摘要 小鼠伴肌动蛋白相关锚定蛋白 (N-RAP) 是已知的与肌纤维生成有关的细胞骨架蛋白, 但是, 与此对应的人类 N-RAP 基因的 cDNA 序列及其功能一直是未知的. 为了明确 N-RAP 基因在人类中所起的作用, 利用生物信息学分析工具和分子生物学技术, 成功地克隆了人 N-RAP 基因. 人 N-RAP 基因 cDNA 序列含有一个 5 088 bp 的开放读码框, 与小鼠 N-RAP 基因有 86% 的相似性. 染色体定位分析表明, 人 N-RAP 基因定位于 10q25~q26 之间, 编码区由 41 个外显子组成, 与 HABP2 和 CASP7 紧密连锁. 电子表达谱分析表明, 肌肉、心脏、脑、脊髓和前列腺有人 N-RAP 基因不同长度的 cDNA 序列. 人 N-RAP 与小鼠 N-RAP 蛋白质序列有 88% 的相似性, 与人 Nebulin 有约 63% 的相似性, 与小鼠 Nebulin 有约 59% 的相似性. 结构域分析发现, N 端为 LIM 结构域, 从第 170 位氨基酸开始, 连续出现 40 个 Nebulin 相关重复序列. RT-PCR 实验表明, 人 N-RAP 基因在成人脑、骨骼肌和心脏组织中都有表达, 在骨髓中不表达. 亚细胞定位研究显示, 人 N-RAP 基因表达于胞浆. 基于序列相似性、结构域分析、表达谱分析和亚细胞定位研究, 可以推测人 N-RAP 与小鼠 N-RAP 同属 Nebulin 家族, 具有类似的功能.

关键词 人 N-RAP 基因, 肌纤维生成, 生物信息学

学科分类号 Q73

小鼠伴肌动蛋白相关锚定蛋白 (Nebulin related anchoring protein, N-RAP) 是 1997 年发现的能结合肌动蛋白的 LIM 蛋白. C 端为 Nebulin 相关重复序列, 可结合肌动蛋白和粘着斑蛋白 (vinculin), N 端为 LIM 结构域, 可结合踝蛋白 (talin), 中间序列为 N-RAP 所特有, 可结合肌肉 LIM 蛋白 (muscle LIM protein, MLP) 和肌动蛋白, 但与肌动蛋白的结合能力比 C 端低 10 倍. 它含 1 175 个氨基酸, 分子质量为 185 ku, 特异表达于骨骼肌和心肌组织, 在 MLP 敲除和 tropomodulin 转基因小鼠模型中的表达都上调, 提示有可能作为扩张性心肌病 (dilated cardiomyopathy, DCM) 的早期标志. N-RAP 被认为在肌纤维生成过程中起作用, 是肌纤维组装的引发中心. 首先, N-RAP 与 β -integrin、talin 和 vinculin 等结合, 形成膜结合复合体. 随后, Z 带蛋白被召集, 肌动蛋白沿着 N-RAP C 端 Nebulin 相关重复序列多聚化, 末端与 Z 带蛋白结合, 形成前肌丝复合体. 最后, N-RAP N 端 LIM 结构域经历翻译后修饰, 前肌丝复合体从膜上释放, 融合形成 Z 带^[1~8].

本实验室从胎儿和成人骨骼肌消减杂交 cDNA 文库中获得了一条新的表达序列标签 (EST), GenBank 接受号为 BQ091951. 利用 BLAST (basic local alignment search tool) 软件进行相似性搜索,

发现该序列与小鼠 N-RAP 有 86% 的相似性, 可能代表小鼠 N-RAP 基因在人类中的同源物. 为了明确 N-RAP 基因在人类中所起的作用, 本研究利用生物信息学分析工具和分子生物学技术, 从成人骨骼肌 cDNA 文库中成功地克隆了人 N-RAP 基因 cDNA 序列, 并对其功能进行了初步的分析.

1 材料和方法

1.1 人 N-RAP cDNA 的克隆

将小鼠 N-RAP 基因全长 cDNA 序列与人类基因组序列作 BLAST 分析 (<http://www.ncbi.nlm.nih.gov/blast>), 发现它与人类 10 号染色体序列有高度相似性, 说明人类确实存在与小鼠 N-RAP 同源的基因. 将小鼠 N-RAP 基因全长 cDNA 与人 EST 库作 BLAST 分析, 发现有多条 EST 序列与小鼠 N-RAP 基因高度相似. 根据对应的基因组序列、EST 序列和我们获得的新 EST 序列, 设计 2 对引物: 上游引物 P1 (ATGAATGTGCAG-

* 国家重点基础研究发展规划项目 (973) (001CB509906) 和国家高技术“863”计划领域重大专项 (2002AA205051) 及主题项目 (2001AA216151).

** 通讯联系人.

Tel: 010-66932203, E-mail: peixt@nic.bmi.ac.cn

收稿日期: 2002-10-14, 接受日期: 2002-12-02

CCCTGTTC)、P2 (TCATCGACCCTCTGCCG) 和下游引物 P3 (AAGACTCAGTGGGCCAAGAA)、P4 (CGAAGATTACGCCAGGACTC), 以人骨骼肌 cDNA 文库 (Clontech 公司) 为模板, 用高保真 DNA 聚合酶 Pyrobest (TaKaRa 公司) 扩增出对应片段. 上游片段扩增采用 Touchdown PCR 技术, 条件如下: 94℃ 预变性 3 min; 94℃ 变性 15 s, 60℃ 退火 15 s, 每经过一个循环退火温度降低 0.5℃, 72℃ 延伸 3 min, 共 25 个循环; 94℃ 变性 15 s, 55℃ 退火 15 s, 72℃ 延伸 3 min, 共 10 个循环; 72℃ 5 min. PCR 结束后, 补加 1 μl rTaq 多聚酶 (TaKaRa 公司) 72℃ 20 min 加尾. 下游片段扩增条件如下: 94℃ 预变性 3 min; 94℃ 变性 15 s, 55℃ 退火 15 s, 72℃ 延伸 3 min, 共 38 个循环; 72℃ 5 min. PCR 结束后, 补加 1 μl rTaq 多聚酶 (TaKaRa 公司) 72℃ 20 min 加尾. 0.7% 琼脂糖凝胶电泳分离 PCR 产物, 采用玻璃奶回收试剂盒 (北京博大泰克生物基因技术有限公司) 从琼脂糖凝胶中回收目的片段. 目的片段以 3:1 摩尔比与 pMD18-T 载体 (TaKaRa 公司) 进行连接. 将连接产物转化大肠杆菌 JM109, 挑选白色菌落进行 PCR 和酶切鉴定, 阳性克隆送上海博亚生物技术有限公司测序. 2 个片段拼接后用 ESTBlast 程序 (<http://www.hgmp.mrc.ac.uk/ESTBlast>) 进行电子延伸, 将延伸后的序列与人类基因组数据库作 BLAST 分析 (<http://www.ncbi.nlm.nih.gov/BLAST>), 确认其可靠性. 根据拼接后的序列, 设计 2 对引物: 外引物 P5 (TCAGCAAAGGA-AAGTGGAAAG)、P6 (GATCTGGGATGGGCTG-G) 和内引物 P7 (GAATTCGGATGAATGTGC-AGCCCTGTTC)、P8 (CTCGAGTCACAACAG-CAGGGCCTTCTTC), 内引物上分别添加 *Eco*R I 和 *Xho* I 酶切位点 (斜体表示), 以便进行后续操作. 以人骨骼肌 cDNA 文库 (Clontech 公司) 为模板, 用 Advantage II DNA 聚合酶 (Clontech 公司) 和外引物 P5、P6 进行第一轮 PCR, 条件如下: 94℃ 预变性 3 min; 94℃ 变性 30 s, 55℃ 退火 30 s, 72℃ 延伸 6 min, 共 20 个循环; 72℃ 10 min. 将 PCR 产物用去离子水稀释 100 倍, 取 1 μl 作模板, 用内引物 P7 和 P8 进行第二轮 PCR, 条件与第一轮 PCR 相同. 电泳分离 PCR 产物, 回收后连入 pMD18-T 载体, 挑选阳性克隆 (pMD18-T/N-RAP) 测序.

1.2 人 N-RAP 基因核酸和蛋白质序列分析

核酸序列分析包括开放读码框 (ORF) 查找、相似性搜索、染色体定位、电子表达谱分析, 主要在 NCBI 服务器 (<http://www.ncbi.nlm.nih.gov>) 上进行. 翻译起始位点预测在 NetStart 1.0 Prediction 服务器 (<http://www.cbs.dtu.dk/services/NetStart>) 上进行. 蛋白质序列分析包括分子质量计算、等电点计算、蛋白质翻译、相似性搜索、特征序列模式和结构域搜索、翻译后修饰及拓扑结构预测、一级结构分析、二级结构预测等, 主要在 ExPASy 分子生物学服务器 (<http://cn.expasy.org>) 上进行.

1.3 人 N-RAP 基因在几种成人组织中表达的检测

用 RT-PCR 法. 分别以成人脑、骨髓、骨骼肌和心脏组织的 cDNA 文库 (Clontech 公司) 为模板, 用引物 P5、P6、P7 和 P8 进行巢式 PCR, 扩增人 N-RAP 基因编码区 5 088 bp 的片段. 回收特异性片段后连入 pMD18-T 载体测序证实.

1.4 人 N-RAP 基因表达的亚细胞定位

用 *Eco*R I 和 *Xho* I 酶切 pMD18-T/N-RAP 载体, 回收 N-RAP 编码区, 以相同读码框架连入 pCMV-myc 载体, 构建融合表达载体 pCMV-myc/N-RAP. 用脂质体 Lipofectamine 2000 (Gibco 公司) 介导的方法, 将 pCMV-myc/N-RAP 载体瞬时转染约 80% 汇合的 COS7 细胞, 使 N-RAP 与 c-myc 标签融合表达. 转染 24~48 h 后依次用 4% 多聚甲醛固定细胞 30 min, 用 0.5% Triton X-100 处理 15 min, 用山羊血清封闭 10 min, 用小鼠抗人 c-myc 单克隆抗体 (北京中山生物技术有限公司) 37℃ 孵育 1 h, 用红色荧光 TRITC 标记的山羊抗小鼠 IgG (北京中山生物技术有限公司) 37℃ 孵育 30 min, 用 50 mg/L 4',6-二乙酰基-2-苯基吡啶 (DAPI, Sigma 公司) 复染 15 min, 其间均用 PBS 洗涤 3 次, 每次 5 min. 荧光显微镜下观察红色荧光和蓝色荧光的分布. 以转染 pCMV-myc 载体的 COS7 细胞和未转染基因的 COS7 细胞作为对照.

2 结 果

2.1 人 N-RAP 基因 cDNA 的克隆

用 PCR 方法, 从成人骨骼肌 cDNA 文库中, 分别获得了约 2 kb (上游片段) 和 1.8 kb (下游片段) 的特异条带, 测序后拼接, 发现并没有完整的开放读码框. 为此, 将获得的约 3.7 kb 片段搜索

人 EST 数据库, 进行电子延伸, 得到约 5.4 kb 的电子延伸产物. 据此设计引物 P5、P6、P7 和 P8, 进行巢式 PCR, 从成人骨骼肌 cDNA 文库中获得了约 5.1 kb 的特异条带 (图 1), 测序结果表明我们获得的序列与电子延伸产物只有个别碱基有差异, 其余完全一致. 序列已登录 GenBank, 被命名为人 N-RAP 基因, 接受号为 AY081943.

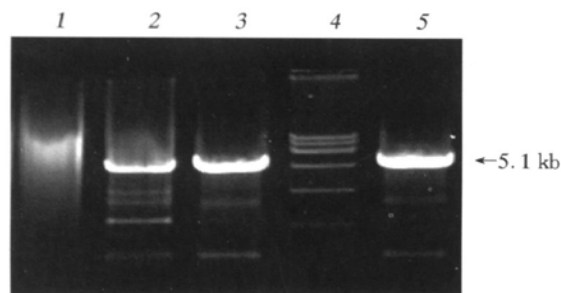


Fig. 1 The expression of human N-RAP in several adult tissues

1: bone marrow; 2: heart; 3: brain; 4: DL15000 marker; 5: skeletal muscle.

2.2 人 N-RAP 基因核酸和蛋白质序列分析

人 N-RAP 基因含有一个 5 088 bp (166~5 253 bp) 的开放读码框, 第 85~87 位有同相位的终止密码子 TGA. 结合小鼠 N-RAP 序列和 NetStart 1.0 预测结果, 推测第 166 位开始的 ATG 可能是翻译起始位点. 核酸序列相似性搜索发现, 人 N-RAP 基因与小鼠 N-RAP 基因有 86% 的相似

性. 搜索人类基因组数据库, 发现人 N-RAP 基因定位于 10q25~q26 之间, 编码区由 41 个外显子组成, 与 HABP2 和 CASP7 紧密连锁, 3' 端已与 HABP2 负链重叠. 根据小鼠 N-RAP 基因序列和染色体定位结果, 可以确定我们获得的 cDNA 序列代表了人 N-RAP 基因全长. 电子表达谱分析发现, 不同长度的人 N-RAP 基因 cDNA 序列来源于肌肉、心脏、脑、脊髓和前列腺等组织. 5 088 bp 的 ORF 可翻译成含 1 695 个氨基酸的蛋白质, 预测分子质量为 193 ku, 等电点 pI 为 9.7, 无疏水区, 为可溶性蛋白, 无信号肽和线粒体定位信号. 151~165 位 (KSLGEEYTEDYEQPR) 可能是 PEST 区. 第 425 位和第 1 182 位的 Thr 可能是 O 型糖基化位点, 可能存在多个 Ser、Thr 和 Tyr 磷酸化位点. 蛋白质序列相似性搜索表明, 人 N-RAP 与小鼠 N-RAP 有 88% 的相似性, 与人 Nebulin 有约 63% 的相似性, 与小鼠 Nebulin 有约 59% 的相似性. 利用 SMART 服务器分析结构域, 发现 N 端 5~57 位为 LIM 结构域, 从第 170 位开始, 连续出现 40 个 Nebulin 相关重复序列 (NEBU, 图 2). LIM 结构域富含半胱氨酸, 与 Zn 离子结合, 不与 DNA 结合, 可能作为蛋白质相互作用的界面, 而 Nebulin 相关重复序列则与肌动蛋白结合, 作为肌动蛋白多聚化的模板. 基于序列相似性和结构域分析, 可以推测人 N-RAP 与小鼠 N-RAP 同属 Nebulin 家族, 它们具有类似的功能.



Fig. 2 The domain organization of human and mice N-RAP

2.3 人 N-RAP 基因在几种成人组织中表达的检测

以我们可以获得的成人脑、骨髓、骨骼肌和心脏组织的 cDNA 文库 (Clontech 公司) 为模板, 用引物 P5、P6、P7 和 P8 进行巢式 PCR, 扩增人 N-RAP 基因编码区 5 088 bp 的片段. 图 1 显示, 除骨髓组织外, 从脑、骨骼肌和心脏组织中均扩增到 5.1 kb 片段. 回收该片段, 连入 pMD18-T 载体测序, 结果证实扩增到的特异片段为人 N-RAP 基因

相应片段. 实验结果表明, 成人脑、骨骼肌和心脏组织中均表达 N-RAP 基因. 这与电子表达谱分析的结果是一致的.

2.4 人 N-RAP 基因表达的亚细胞定位

荧光显微镜下观察, 转染 pCMV-myc 载体的 COS7 细胞, 红色荧光在整个细胞均匀分布, 未转染基因的细胞则没有荧光, 而转染 pCMV-myc/N-RAP 载体的细胞, 红色荧光分布于胞浆 (图 3).

DAPI 与 DNA 特异结合, 在紫外光激发下发蓝色荧光, 可区分胞浆和胞核. 实验表明, 人 N-RAP

基因表达于胞浆. 这与其可能行使参与肌纤维生成的功能相吻合.

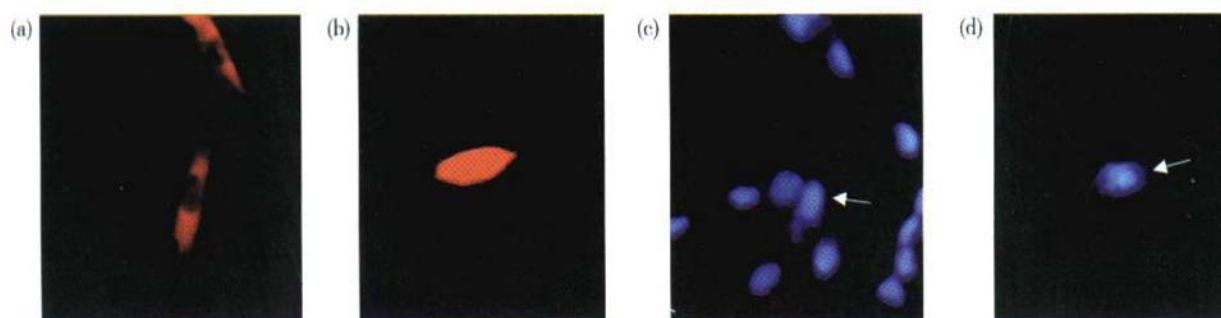


Fig. 3 Subcellular localization of human N-RAP gene

Cellular localization of ϵ -myc fusion protein displayed by immunostaining of COS7 cells transfected with pCMV-myc/N-RAP (a) and pCMV-myc (b). The nucleus of the corresponding COS7 cell transfected with pCMV-myc/N-RAP (c) and pCMV-myc (d) respectively was shown by arrow ($\times 200$).

3 讨 论

生物信息学是应用计算机和信息技术搜集、储存、分析、整理各种生物医学信息, 并予以管理和利用的一门新兴科学. 利用生物信息学手段, 可以显著加快基因克隆与功能研究的进程. 本研究从我们获得的一条新 EST 序列 (接受号为 BQ091951) 出发, 利用生物信息学手段, 仅用几次 PCR 反应, 就成功地克隆到长达 5 417 bp 的人 N-RAP 序列. 相比而言, 利用 RACE 技术和文库筛选方法则要费时费力得多. 这里, 我们所用的克隆策略是基于新发现的人 EST 序列在小鼠中有对应的同源序列这一前提的, 如果没有对应的同源序列, 我们可以利用现在几乎完成的人类基因组计划和大量的 EST 序列获取全长基因. 策略是: 下载人 EST 序列对应的基因组序列, 用基因发现软件如 Genscan 分析可能对应的基因, 将此基因序列与人 EST 数据库作 BLAST 分析, 寻找对应的 EST 序列进行电子延伸, 根据两端的 EST 序列设计引物, 用 PCR 方法扩增即可. 我们用此方法得到的电子延伸产物, 与实际序列基本一致, 只有中间少数外显子有出入, 但这并不影响用 PCR 进行克隆.

生物信息学分析结果显示, 我们获得的序列与小鼠 N-RAP 具有类似的核酸序列 (86%)、氨基酸序列 (88%) 以及结构域. 我们的实验结果和已有的文献报道^[8]表明, 它们具有相似的表达谱. 因此推测人 N-RAP 基因是小鼠 N-RAP 的同源基因, 可能具有类似的功能. 我们发现的一个不同之处在于, 人 N-RAP 在成人脑中也有表达, 电子表达谱分析也表明脑、脊髓和前列腺中都有表达, 而 Luo

等^[8]的实验证实小鼠 N-RAP 特异表达于骨骼肌和心肌组织中.

基因在细胞内表达部位的确定可为基因的功能研究提供丰富的信息. 比如, 细胞膜上表达的基因可能是受体和离子通道, 细胞核内表达的基因可能是转录调节分子. 在没有特异性抗体的前提下, 研究者常用标签 (tag) 来跟踪基因的表达. pCMV-myc 载体中含有 13 个氨基酸的 ϵ -myc 标签, 其后为多克隆位点, 可插入目的基因进行融合表达. 由于 ϵ -myc 标签很小, 融合蛋白在细胞内的定位由目的蛋白来决定. 这样, 用抗 ϵ -myc 标签的抗体就可跟踪目的基因在细胞内的表达. 我们用此方法确定人 N-RAP 基因主要表达于细胞浆, 这与其可能参与肌纤维生成的功能是相吻合的. 人 N-RAP 基因与小鼠 N-RAP 基因是否具有类似的功能, 进一步的研究正在进行中.

参 考 文 献

- 1 Carroll S L, Herrera A H, Horowitz R. Targeting and functional role of N-RAP, a Nebulin-related LIM protein, during myofibril assembly in cultured chick cardiomyocytes. *J Cell Sci*, 2001, **114** (Pt 23): 4229~4238
- 2 Zhang J Q, Elzey B, Williams G, *et al.* Ultrastructural and biochemical localization of N-RAP at the interface between myofibrils and intercalated disks in the mouse heart. *Biochemistry*, 2001, **40** (49): 14898~14906
- 3 Ehler E, Horowitz R, Zuppinger C, *et al.* Alterations at the intercalated disk associated with the absence of muscle LIM protein. *J Cell Biol*, 2001, **153** (4): 763~772
- 4 Carroll S L, Horowitz R. Myofibrillogenesis and formation of cell contacts mediate the localization of N-RAP in cultured chick cardiomyocytes. *Cell Motil Cytoskeleton*, 2000, **47** (1): 63~76
- 5 Herrera A H, Elzey B, Law D J, *et al.* Terminal regions of mouse Nebulin: sequence analysis and complementary localization with N-RAP. *Cell Motil Cytoskeleton*, 2000, **45** (3): 211~222

- 6 Luo G, Herrera A H, Horowitz R. Molecular interactions of N-RAP, a Nebulin-related protein of striated muscle myotendon junctions and intercalated disks. *Biochemistry*, 1999, **38** (19): 6135~ 6143
- 7 Luo G, Leroy E, Kozak C A, *et al.* Mapping of the gene (NRAP) encoding N-RAP in the mouse and human genomes. *Genomics*, 1997, **45** (1): 229~ 232
- 8 Luo G, Zhang J Q, Nguyen T P, *et al.* Complete cDNA sequence and tissue localization of N-RAP, a novel Nebulin-related protein of striated muscle. *Cell Motil Cytoskeleton*, 1997, **38** (1): 75~ 90

Cloning and Function Analysis of a Novel Gene: Human Nebulin-related Anchoring Protein^{*}

YUAN Hong-Feng, WANG Dong-Mei, LI Hai-Min, CHEN Lin,
WANG Xiao, BAI Ci-Xian, YUE Wen, PEI Xue-Tao^{**}

(*Department of Stem Cell Biology, Beijing Institute of Transfusion Medicine, Beijing 100850, China*)

Abstract Nebulin-related anchoring protein (N-RAP) is a 185 ku actin-binding LIM protein recently discovered in skeletal and cardiac muscle tissues in mouse. It is proposed that N-RAP serves as a link between the terminal actin of the myofibril and the protein complexes at the cell membrane and thus as an organizing center in the initial phase of myofibril assembly. But in human, the sequence and function of N-RAP remain unknown. By using bioinformatics tools, the full length of human N-RAP cDNA, which contains a 5 088 bp ORF, encodes a protein of 1 695 amino acid residues, is successfully cloned. Human N-RAP is mapped to the genomic region between HABP2 and CASP7 at chromosome 10q25~ q26, consisting of 41 exons and 40 introns. Homology searches with the deduced 1 695 amino acid protein sequence reveal human N-RAP shares 88% similarity with mice N-RAP, 63% with human Nebulin and 59% with mice Nebulin. Corresponding EST sequences are found in muscle, heart, spinal cord and prostate tissue. The predicted protein contains LIM domain (5~ 57), which binds two zinc ions, does not bind DNA, seems to act as interface for protein-protein interaction, and Nebulin repeats, tandem arrays of which are known to bind actin. RT-PCR reveals human N-RAP is expressed in adult muscle, heart and brain tissue, not in bone marrow. In addition, subcellular location study shows human N-RAP is expressed in cytoplasm. These results demonstrate that just like mice N-RAP, human N-RAP is proposed to be crucial for myofibrillogenesis.

Key words human N-RAP, myofibrillogenesis, bioinformatics

^{*} This work was supported by grants from The Special Funds for Major State Basic Research of China (001CB509906) and State 863 High Technology R & D Project of China (2002AA205051 and 2001AA216151).

^{**} Corresponding author. Tel: 86-10-66932203, E-mail: peixt@nic.bmi.ac.cn

Received: October 14, 2002 Accepted: December 2, 2002