

人脑红蛋白 (NGB) 全长 cDNA 序列的克隆*

王春丽 张成岗**

(军事医学科学院放射医学研究所, 北京 100850)

摘要 人脑红蛋白 (neuroglobin, NGB) 是新发现的神经系统特异的携氧蛋白, 然而其全长 cDNA 序列一直未见报道. 采用电子序列延伸技术和 cDNA 序列末端快速扩增技术 (rapid amplification of cDNA ends, RACE) 研究发现, 人 NGB 全长 cDNA 序列为 1 909 bp, 5' 非编码区为 375 bp, 编码区 (456 bp) 可编码 151 个氨基酸, 3' 非编码区为 1 078 bp, 其中含 27 bp 的 poly (A) (GenBank 接受号: AF422797). 综合采用电子序列延伸技术与 RACE 技术是获得全长 cDNA 序列的有效方法, 为后续的功能研究提供了重要基础.

关键词 脑红蛋白, 电子序列延伸技术, cDNA 末端快速扩增技术

学科分类号 Q786, R852.11

脑红蛋白 (neuroglobin, NGB) 是新近发现的体内第三类重要的携氧蛋白, 能够特异性地向脑组织供氧, 在神经系统氧摄取、氧运输和氧利用等生理过程中起重要作用. 该基因最早由 Burmester 等^[1,2]发现, 随后国内外多家实验室对其予以报道^[3~5]. 然而, 文献中只报道了该基因部分 cDNA 序列, 其 5' 端和 3' 端的序列均属未知, 且由于未能获得 RNA 印迹结果, 也难以判断该基因 mRNA 的大小, 从而, 能否获得 NGB 全长 cDNA 序列已成为进行深入功能研究的瓶颈. cDNA 序列末端快速扩增技术 (rapid amplification of cDNA ends, RACE) 是近年来出现的能够快速获得全长 cDNA 的新方法^[6], 而且已经有较多成功的例子^[7]. 因此, 为了便于对人 NGB 基因进行深入的结构与功能研究, 本文在进行 RNA 印迹分析的基础上, 综合采用电子序列延伸技术、5'-RACE 和 3'-RACE 技术获得了人 NGB 基因的全长 cDNA 序列.

1 材料和方法

1.1 试剂及试剂盒

人正常多组织 RNA 印迹膜和 SMARTTM RACE cDNA Amplification Kit 均购自 ClonTech 公司. MMLV 反转录酶 Superscript II (200 U/μl) 购自 Gibco 公司. LA Taq DNA 聚合酶、pMD18-T Vector 载体为 TaKaRa 公司产品. ProGene DNA Thermal Cycler 型 PCR 仪为 TechNe 公司产品.

1.2 cDNA 序列电子延伸

采用本实验室自行设计的运行于 PC/Linux 环

境的“全自动电子序列延伸系统” (AutoCTG) 进行 cDNA 序列电子延伸分析^[8,9]. 所使用的人 EST 数据库为 2002 年 1 月 20 日创建, 含有 EST 序列 3 929 177 条.

1.3 引物设计与合成

采用 Primer Premier 软件^[10]按照 SMART RACE 试剂盒的要求“23~28 nt、50%~70% GC 比例、 $T_m \geq 65^\circ\text{C}$ ”设计一系列 RACE 引物 (pD 表示下游引物, pU 表示上游引物). 5'-RACE 引物: pD262 (5'-GCAGCATCAATCACAAGCA-3'), pD266 (5'-ACTGGAAGAGGGGCAGCAGGTCA-3'), 检测引物: pU264 (5'-GGATCCGCATGGAGCGCCCGAGCCC-3'), 可与 pD266 扩增进行阳性片段的筛选 (pU264 与 pD266 扩增产物的大小为 138 bp); 3'-RACE 引物: pU264 (序列同前), pU263 (5'-AGCCGCAGCCCTCTGGAACA-3'), 检测用上游引物: pU267 (5'-CCGCTGGAGCACGGCACCCTCCTGT-3'), 检测用下游引物: pD255 (5'-CTCGAGCTCGCCATCCCAGCCTCG-3'). pU267 与 pD255 扩增产物的大小为 403 bp. 这些引物与模板的匹配方式见图 1. 所设计的引物均由上海博亚 (BioAsia) 公司合成. 使用浓度均为 10 μmol/L.

* 国家自然科学基金资助项目 (30100049, 39900041, 39900074) 与军事医学科学院科技创新研究启动基金 (0102001, 9905105) 部分资助项目.

** 通讯联系人.

Tel: 010-66931590, E-mail: zhangcg@nic.bmi.ac.cn

收稿日期: 2002-04-15, 接受日期: 2002-06-23

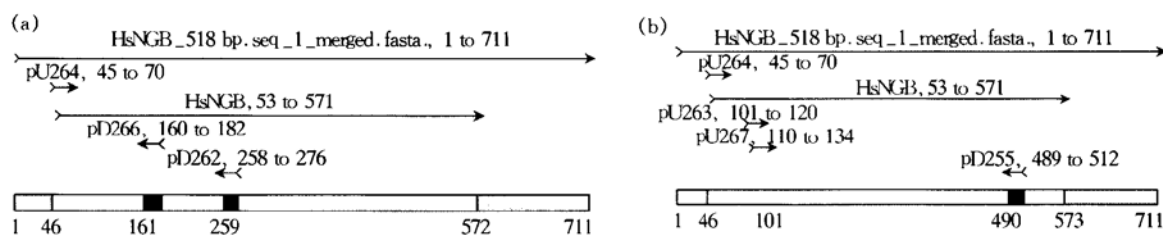


Fig 1 Diagram shows the location and direction of RACE primers in the known cDNA sequence (711 bp) of human NGB (HsNGB)

(a) 5'-RACE primers; (b) 3'-RACE primers. Top sequence: the 711 bp sequence of electronic elongated sequence for HsNGB. The third line of HsNGB is the published sequence of 518 bp submitted by Burmester T (Nature, 2000). Created by using the SequencherTM software (version 4.0.5 for Windows).

1.4 RNA 印迹分析

用引物对 pU267 和 pD255 从人胎脑总 mRNA 反转录的 cDNA 文库中扩增出目的基因片段 (403 bp), 1% 琼脂糖凝胶电泳回收纯化后, 用随机引物标记试剂盒 (Promega 公司产品) 标记探针. RNA 印迹过程严格按照操作流程进行. 洗膜后压片于 -70℃, 2 d 后显影.

1.5 5'-RACE

按照试剂盒的操作流程, 参考文献 [11] 进行. 首先进行反转录反应, 制备 5'-RACE-Ready cDNA. 随后, 采用通用引物混合物 UPM (此时作为上游引物) 和 pD262 进行首轮 PCR 扩增. 用 1% 琼脂糖凝胶电泳检测. 如果扩增产物电泳条带明显, 则回收鉴定并克隆后测序. 如电泳未发现扩增条带, 则将 PCR 产物稀释 10 倍, 用巢式引物对 (NUP+ pD266) 继续进行 PCR 扩增, 直至获得特异性条带. 所有扩增条带均使用内引物对 pU264+ pD266 进行 PCR 鉴定 (图 1a), PCR 条件同上. 通过鉴定的原始片段 (NUP+ pD266 的扩增产物) 用 1% 琼脂糖凝胶电泳回收, 克隆至 pMD 18-T 载体并用 377XL 测序仪 (PE 公司) 测序.

1.6 3'-RACE

按照试剂盒的操作流程进行. 首先进行反转录反应, 制备 3'-RACE-Ready cDNA. 随后, 采用通用 pU264 和引物混合物 UPM (此时作为下游引物) 进行首轮 PCR 扩增. 用 1% 琼脂糖凝胶电泳检测. 如果扩增产物电泳条带明显, 则回收鉴定并克隆后测序. 如电泳未发现扩增条带, 则将 PCR 产物稀释 10 倍, 用巢式引物对 (pU263+ NUP) 进行 PCR 扩增, 直至获得特异性条带. 所有扩增条带均使用内引物对 pU267+ pD255 进行 PCR 鉴定 (图 1b), PCR 条件同上. 通过鉴定的原始片段 (pU263+ NUP 的扩增产物) 用 1% 琼脂糖凝胶电泳回收, 克隆至 pMD 18-T

载体并用 377XL 测序仪 (PE 公司) 测序.

1.7 序列组装

将已知序列、电子延伸的序列和通过 5'-RACE 及 3'-RACE 获得的序列片段用 SequencherTM 软件进行序列组装, 结合测序峰图进行序列校正, 并标记出接头引物部分 (UPM 和 NUP), 最后获得拼接好的序列.

1.8 全长序列鉴定

根据实验所获得的 HsNGB cDNA 序列分别在其 5' 和 3' 端设计一对引物 (pU313: 5'-TTCCCAggCCACCATAgCggCTggG-3' 和 pD314: 5'-AgAgAggCACACAgAACCAgTTTAT-3'), 从人胎脑总 RNA 经 Oligo (dT) 引导反转录形成的 cDNA 中扩增 NGB 基因的全长 cDNA 序列. PCR 条件同前. 所扩增的片段用 1% 琼脂糖凝胶电泳回收, 克隆至 pMD 18-T 载体并用 377XL 测序仪 (PE 公司) 全长测序进行鉴定.

2 结果

2.1 RNA 印迹实验

采用 RNA 印迹检测发现, 人 NGB 只在正常成人脑中含有一个约 1.9 kb 的转录本 (图 2).

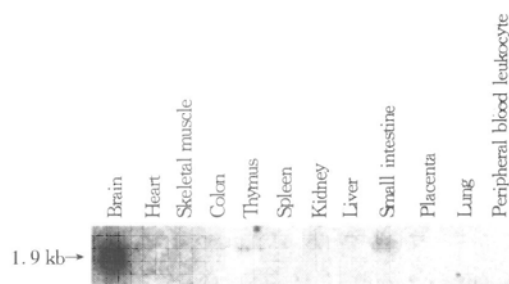


Fig 2 Northern blot analysis of HsNGB transcription in 12 normal human tissues

A specific band of 1.9 kb was only detected in normal human brain.

聚核苷酸成分 SMART II Oligonucleotide (5'-AAGCAGTGGTAACAACGCAGAGTACGC GGG-3') 3'端的“GGG”将通过和“3'-CCC (N)-5'”匹配作为反转录反应的模板,从而使反转录反应继续进行,直至反转录到 SMART II Oligonucleotide 的5'端.此即为所谓的“模板跳转机制”(switching mechanism at 5' end of RNA transcript, SMART).此方法可保证反转录所获得的 cDNA 序列能够最大限度地到达 mRNA 的5'端.同时,巢式引物 NUP 的引物序列 (5'-AAGCAGTGGTAACAACGCAGAGT-3') 又是 SMART II Oligonucleotide 5'端的23个碱基,引物混合物 UPM 则由两条引物组成 (5'-CTAATACGACTCACTATAGGGCAAGCAGTGGTAACAACGCAGAGT-3'、5'-CTAATACGACTCACTATAGGGG-3'),其中长引物3'端的23 bp (划线部分) 又与 NUP 的序列一致.因此,在反转

录反应完成后, cDNA 的5'端就具备了引物 UPM 和 NUP 的结合位点,用户只需根据已有 cDNA 的序列片段设计合适的基因特定的下游引物 (reverse gene specific primer, R-GSP), 并使用 UPM + R-GSP或 NUP+ R-GSP 启动 PCR 反应,从而即可扩增未知基因的5'端序列.对于3'-RACE,情况则比较简单.当进行反转录反应时,所使用的引物3'-RACE cDNA synthesis primer (5'-AAGCAGTGGTAACAACGCAGAGTAC (T)₃₀N-₁N-3') 5'端的23 bp 即为 NUP 的结合位点,从而将作为下游引物的 UPM 和 NUP 的结合位点引入到了反转录所生成的 cDNA 序列中,用户只需根据已有 cDNA 的序列片段设计合适的基因特定的上游引物 (forward GSP, F-GSP), 并使用 F-GSP+ UPM 或 F-GSP+ NUP 启动 PCR 反应,即可扩增未知基因的3'端序列^[12].

1	TTCCAGGCCACCATAGCGGCTGGCGGAGGGAGCGCGCCTTGCTGGCCTGGAGGGGGCGGGGGC	66
67	CGTGGCGGCTTTAAAGCGCCAGCCAGGCGTGGCGGGTGGGGCGGCTCTGGCGGCTGGCGGGCG	132
133	CAGGGCGCAGCGGCCAAGCGGGTCCCCGGAAGCACAGCTGGGGTGTCTCCACCTACGACTGGCCG	198
199	CGCGCCTTTCTCTCCCGCGCCAGGGAAGGAGCGGCTGCGGCCCCCGCGGGCGGAGGCACGGGGG	264
265	GCGTACGAGGGGCGAGGGGACCGCTCGCGGAGGAGATGGCGCGGCACGTGCGGTGACGGCACCC	330
331	GAGCCCTGAGGGTCCAGCCCCGCGCTCCGCGTCCCCGGGACAGCATGGAGCGCCCGGAGCCCCGAG	396
1	M E R P E P E	7
397	CTGATCCGGCAGAGCTGGCGGGCAGTGAGCCGACGCCCCGCTGGAGCACGGCACCGTCTGTTTGCC	462
8	L I R Q S W R A V S R S P L E H G T V L F A	29
463	AGGCTGTTTGCCCTGGAGCCTGACCTGCTGCCCTCTTCCAGTACAACCTGCCGCGAGTTCTCCAGC	528
30	R L F A L E P D L L P L F Q Y N C R Q F S S	51
529	CCAGAGGACTGTCTCTCCTCGCCTGAGTTCTGGACCACATCAGGAAGGTGATGCTCGTGATTGAT	594
52	P E D C L S S P E F L D H I R K V M L V I D	73
595	GCTCAGTGACCAATGTGGAAGACCTGTCTACTGGAGGATACCTTGCCAGCCTGGCGAGGAAG	660
74	A A V T N V E D L S S L E E Y L A S L G R K	95
661	CACCGGGCAGTGGGTGTGAAGCTCAGCTCCTTCTCGACAGTGGGTGAGTCTCTGTCTACATGCTG	726
96	H R A V G V K L S S F S T V G E S L L Y M L	117
727	GAGAAGTGTCTGGGCCCTGCCTTACACAGCCACACGGGCTGCCTGGAGCCAACCTACGGGGCC	792
118	E K C L G P A F T P A T R A A W S Q L Y G A	139
793	GTAGTGCAGGCCATGAGTCGAGGCTGGGATGGCGAGTAAGAGGCGACCCCGCCCGGAGCCCCCAT	858
140	V V Q A M S R G W D G E *	151
859	CCATCTGTGTCTGTCTGTTGGCCTGTATCTGTTGTAGCCAGGCTCCCCAAGCTTCCCTGCATCTT	924
925	GGTCTGTGTCCTTGGCCCACTGGAGAGGTGATGGGGCAGGGCTGGGTCTCAGTATCCTAGAGT	990
991	CCAGCTGCAGAAAGGATGGCTTTTCTCCAGGAAGGGGCTTCTGGGTGTCCCTCATCCCCAGTAG	1056
1057	CCTCTTTCTTGCGTTTCTTTTACCTTTTTTGGCACTCCCTCTGACCCCGGATGAGTGTTTTGGT	1122
1123	GGCAGAGGTGGGATGAGCTGGAAGGTATGGAGGTGGGAGAGGATGGGGCTCTTCTGTCTGTCTG	1188
1189	CTTCTTCAGGTGAGTGCAGGCCAAGGCGGGGTGAGATGGCTGAGCTTCCAGCGCCTTCTGTCTG	1254
1255	CCTGCCCAGTCCCTTCACTGCTTCTTCCGCCCCAAGATGGCTTGCTTTTCAAAATAAAGAGAAAGA	1320
1321	GCAGCTTACGCTTCTTGGTGAATCCAGGCAGTGGGAGCAGAATCAGAACTGCCAGGGAAGGGA	1386
1387	AGGGGACCTGGGTCTCAATGGGTCTCATTTGAGTCTCGCGGGCTGTGCAGATGCCCTGACAGAGT	1452
1453	CGTTTCTCTTGGCGGCATTCCCTTCCCTCATTCAGCACTTCTGCTGGGAACCTCCCTGACTATTC	1518
1519	CGTGTGTCAGGAACCCAGCTAGCTGGCCAGGTGGGAGGGGCTGGGGACCGCCAGGAAGGAGGG	1584
1585	GCTACTTCATCCAGAGAGACCCGAGTTCCCCAGGCCCTTCATCACCACCCGCTCCCGCAGGAGA	1650
1651	GAGTCTTACCTCCCTTGGCCCTCCTTCTGGCTCAGCCTGCAGCGACTGTGAGGCCACAGCTCCTC	1716
1717	AGATTCACTGCCCCGTGTGTGCCAGTACTCAGGCAGCTGGAGAGAAGGAAGGCAGCAGCAGAGGC	1782
1783	CCCCGCCCTCACCCAGCCATCTGCACTTGTACCATTGTCTGTGCTGACTGTGGTCTATAAAT	1848
	poly(A) signal	
1849	TCATGAGAAATAAACTGGTTCTGTGCTCTCTAAAAATAAAAAAAAAAAAAAAAAAAAAA	1909

Fig. 5 Nucleotide sequence of cDNA encoding HsNGB

First line, nucleotide sequence of the cDNA for HsNGB; second line, the deduced amino acid sequence of HsNGB. The stop codon of translation is designated by an asterisk. The putative poly-adenylation signal is underlined.

生物信息学在核酸和蛋白质序列分析中的应用越来越广泛,代表 cDNA 序列片段的表达序列标签 (expressed sequence tag, EST) 序列数据库中记录的数量也在急剧增加。因此,利用一定的生物信息学算法使用这些 EST 数据对 cDNA 序列进行电子序列延伸,也是目前获得较长甚至全长 cDNA 序列的策略之一。本文中即采用了我们自行开发的运行于 PC/Linux 环境的“全自动电子序列延伸系统 (AutoCTG)”^[8]进行辅助序列延伸,对原始的种子序列在 5' 端和 3' 端均进行了有效延伸,并基于此进一步设计更合理的 RACE PCR 引物,也为本研究的成功奠定了基础。目前我们正在开发 AutoCTG 系统的网络版本,不久将为其他同行提供服务。

作者在以前的文献中曾讨论过 RACE 技术应用过程中的一些问题,包括: Marathon RACE 和 SMART RACE 的区别; RACE 中的引物设计问题; 扩增产物异常 (扩增条带、扩增条带出现 smear 现象或者扩增条带不唯一); 反转录反应条件的优化以及对全部阳性克隆进行测序以获得所有可能的阳性序列片段等^[11]。此处就关于试剂盒的选用问题再探讨如下。ClonTech 公司所推出的 SMART RACE 推荐使用其 SMART RACE 试剂盒 (产品编号: K1811-1), 并推荐使用 Advantage 2 Polymerase Mix。然而,该试剂盒价格相对较高。我们以前所采用的 Promega 公司的 T-载体的价格也相对较高。本文采用 TaKaRa 公司的产品——LA Taq DNA 聚合酶及其 PCR 体系,以及 TaKaRa 公司 pMD18-T Vector 载体,也获得了良好的结果,在价格上具有一定的优势。进一步分析可发现,SMART RACE 试剂盒的所有成分包括反转录时所需的 SMART II Oligonucleotide、5'-RACE cDNA Synthesis Primer、3'-RACE cDNA Synthesis Primer 以及 RACE PCR 时所需的 Universal Primer Mix (UPM)、Nested Universal Primer (NUP) 等均可自行合成,而这一点可为经常需要大量从事 RACE 实验的科研人员节省较多的科研经费。

总之,在目前大量新基因被发现并急需对其进行全长 cDNA 克隆的后基因组计划时代,电子序列延伸技术和 RACE 技术将成为相关领域研究的主角。

参 考 文 献

- 1 Burmester T, Weich B, Reinhardt S, *et al.* A vertebrate globin expressed in the brain. *Nature*, 2000, **407** (6803): 520~ 523
- 2 Moens L, Dewilde S. Globins in the brain. *Nature*, 2000, **407** (6803): 461~ 462
- 3 Sun Y J, Jin K L, Mao X O, *et al.* Neuroglobin is up-regulated by and protects neurons from hypoxic ischemic injury. *Proc Natl Acad Sci USA*, 2001, **98** (26): 15306~ 15311
- 4 张成岗, 李 林, 邓美玉, 等. 大鼠脑红蛋白基因编码区的克隆、多态性分析及该基因组织表达谱研究. *遗传学报*, 2001, **28** (11): 997~ 1001
Zhang C G, Li L, Deng M Y, *et al.* *Acta Genetica Sinica*, 2001, **28** (11): 997~ 1001
- 5 张成岗, 李 林, 邓美玉, 等. 大鼠脑红蛋白 (NGB) 的原核表达、抗体制备及其细胞分布. *中国生物化学与分子生物学报*, 2002, **18** (1): 80~ 84
Zhang C G, Li L, Deng M Y, *et al.* *Chin J Biochem Mol Biol*, 2002, **18** (1): 80~ 84
- 6 Zhang Y, Frohman M A. Using rapid amplification of cDNA ends (RACE) to obtain full-length cDNAs. *Methods Mol Biol*, 1997, **69**: 61~ 87
- 7 Fang C, Mkrtchian S, Ingelman Sundberg M. Combination of direct DNA sequencing with degenerate primer-mediated PCR and 5'-/3'-RACE to screen novel cDNA sequences. *Biotechniques*, 1997, **23** (1): 52~ 58
- 8 张成岗, 孙焕东, 欧阳曙光, 等. 基于 PC/Linux 的核酸序列电子延伸系统的构建及其应用. *遗传*, 2002, **24** (1): 50~ 54
Zhang C G, Sun H D, Ouyang S G, *et al.* *Hereditas*, 2002, **24** (1): 50~ 54
- 9 张成岗, 贺福初. *生物信息学方法与实践*. 北京: 科学出版社, 2002. 249~ 257
Zhang C G, He F C. *Bioinformatics: Method and Practice*. Beijing: Science Press, 2002. 249~ 257
- 10 Singh V K, Mangalam A K, Dwivedi S. Primer premier: program for design of degenerate primers from a protein sequence. *BioTechniques*, 1998, **24** (2): 318~ 319
- 11 邢桂春, 张成岗, 魏汉东, 等. 采用 RACE 技术获得全长人新基因 MAGE-D1. *中国生物化学与分子生物学报*, 2001, **17** (2): 203~ 208
Xing G C, Zhang C G, Wei H D, *et al.* *Chin J Biochem Mol Biol*, 2001, **17** (2): 203~ 208
- 12 Matz M, Shagin D, Bogdanova E, *et al.* Amplification of cDNA ends based on template-switching effect and step-out PCR. *Nucleic Acids Res*, 1999, **27** (6): 1558~ 1560

Full-length cDNA Cloning of Human Neuroglobin by Using Electronic Sequence Elongation Technique and RACE Technique*

WANG Chun-Li, ZHANG Cheng-Gang**

(Department of Genomics and Proteomics, Beijing Institute of Radiation Medicine, Beijing 100850, China)

Abstract Human neuroglobin (NGB) is a newly-discovered gene functioned specifically in the oxygen supply and oxygen consumption in the brain. However, the full-length cDNA sequence of NGB was not reported yet. Using the *in silicon* sequence elongation technique, 5'-RACE and 3'-RACE technique, the full-length cDNA sequence of human neuroglobin was successfully obtained (GenBank accession number AF422797). The cDNA sequence of human NGB is 1 909 bp in size and capable of encoding a protein of 151 amino acids. The 5'-untranslated region is 375 bp, while the 3'-untranslated region is 1 078 bp including 27 bp of poly (A) sequence. Results demonstrated that the RACE technique coupled with electronic sequence elongation technique is useful to obtain unknown sequence of 5'-terminal or 3'-terminal part of a new gene.

Key words neuroglobin, electronic sequence elongation technique, rapid amplification of cDNA ends (RACE)

* This work was supported by grants from The National Natural Sciences Foundation of China (30100049, 39900041, 39900074) and Initiative Foundation for Scientific and Technological Innovation of Academic Military Medical Science (0102001, 9905105).

** Corresponding author. Tel: 86-10-66931590, E-mail: zhangcg@nic.bmi.ac.cn

Received: April 15, 2002 Accepted: June 23, 2002