

草鱼 *CNBP* 的分子特征及其进化分析

陈琼 林刚 王娜 胡成钰*

(南昌大学生命科学学院 南昌 330031)

摘要: 从草鱼 (*Ctenopharyngodon idella*) 肝肾 cDNA 文库中克隆到细胞核酸结合蛋白基因 *CNBP* 的完整开放阅读框序列。分析表明草鱼 *CNBP* 由 163 个氨基酸残基组成, 含有 7 个保守 CCHC 型锌指结构、核定位信号区和 RGG 框, 与其他鱼类的同源性很高。与人及其他脊椎动物的相比, 草鱼细胞核酸结合蛋白在第 3 个锌指中的第 5 个氨基酸残基由 Gly 变成 His, 另外在第 1 锌指和第 2 锌指结构间, 缺失 6~14 个氨基酸残基。虽然如此, 适应性进化分析显示细胞核酸结合蛋白没有经历正达尔文选择 ($\omega \leq 1$), 即这种结构的差异还不足以产生新的功能。这表明 *CNBP* 处于中性进化中。

关键词: 细胞核酸结合蛋白 (*CNBP*); 进化; 草鱼

中图分类号: Q172 文献标识码: A 文章编号: 0250-3263(2008)06-97-06

Molecular Characterization and Evolution of *Ctenopharyngodon idella CNBP*

CHEN Qiong LIN Gang WANG Na HU Cheng-Yu*

(College of Life Science, Nanchang University, Nanchang 330031, China)

Abstract: The open reading frame of cellular nucleic acid binding protein (*CNBP*) was cloned from Grass Carp (*Ctenopharyngodon idella*) cDNA library. The coding region encodes a 163 amino acid polypeptide with the highly conserved general structural organization of seven CCHC zinc finger domains, nuclear located signal and a RGG box which was highly conserved in fish. Compared with human and other vertebrates, the fifth glycine residue at the third zinc finger domain was substituted with histidine and 6–14 amino acid residues were absent between the first and the second zinc finger in the Grass Carp *CNBP*. Nevertheless, the grass carp *CNBP* was not in the positive Darwin selection ($\omega \leq 1$) by adaptive evolution analysis. Therefore, new function was not generated by structural difference of *CNBP*. These indicated that *CNBP* was in the neutral selection.

Key words: Cellular nucleic acid binding protein (*CNBP*); Evolution; Grass Carp (*Ctenopharyngodon idella*)

细胞核酸结合蛋白 (cellular nucleic acid binding protein, *CNBP*) 是一种广泛分布于细胞中的 DNA 结合蛋白。迄今, 已有人 (*Homo sapiens*) 和多种动物的 *CNBP* 基因被克隆^[1-6], *CNBP* 是一个 19 ku 的小分子蛋白, 含有 7 个 CCHC 型 (Cys-X2-Cys-X4-His-X4-Cys) 的锌指结构, 与逆转录核酸结合蛋白非常相似^[7]。在不同物种中, *CNBP* 有较高的同源性, 这种极高的保守性意味着它可能具有非常重要的功能^[8]。近年来, 鱼类 *CNBP* 被相继报道。斑马鱼

(*Danio rerio*) 和银鲫 (*Carassius auratus gibelio*) *CNBP* 都是由 163 个氨基酸残基组成的小分子蛋白质, 其确切功能仍在探索中。

分子进化研究中有效检测正达尔文选择的

基金项目 江西省重点科技攻关项目 (No. 20061B0260301);

* 通讯作者, E-mail: hucy2008@21cn.com;

第一作者介绍 陈琼, 女, 硕士研究生; 研究方向: 水生生物学; E-mail: jonecq2000@sina.com.

收稿日期: 2008-05-23, 修回日期: 2008-08-30

统计学方法是比较核苷酸序列上密码子的错义替换(d_N , 氨基酸变化)与同义替换(d_S , 氨基酸不变)速率。当 d_N/d_S 比率(ω)大于 1, 证明蛋白质受到有益的氨基酸替换, 表现为正选择作用^[9]。分子进化中正选择的出现, 意味着蛋白质在进化中可能衍生出新的结构和功能。基于最大似然法的密码子替换模型可以很好地获得符合数据的真实解释, 已成为在氨基酸水平上检测蛋白适应性进化的一种常规可靠手段^[10]。该方法能较好地分析趋化因子及其受体、转铁蛋白等的适应性进化^[11, 12]。

本文在分析草鱼(*Ctenopharyngodon idella*) CNBP 结构特征的基础上, 利用密码子替换模型分析 CNBP 分子适应性进化。结果显示, CNBP 没有正选择位点, 即该蛋白处于中性或负进化中。

1 材料与方法

1.1 材料 经 Poly I:C 诱导的草鱼肝、肾混合 cDNA 文库由上海联合基因公司构建。随机对 510 个克隆测序后, 其中编号为 YG002-D06 的克隆(EU709746)经比对确定为 CNBP 开放阅读框序列。

1.2 分析方法

1.2.1 草鱼 CNBP 的分子特征 对草鱼 CNBP 开放阅读框推译和序列相似性检索, 初步确认 CNBP 的相关结构特征。

1.2.2 CNBP 分子适应性进化分析

1.2.2.1 CNBP 系统进化分析 在 GenBank 中检索细胞核酸结合蛋白基因, 经查阅文献确认, 下载氨基酸序列和核苷酸序列。氨基酸序列经 Clustal W 软件(<http://www.ebi.ac.uk/clustalw/>)排序比对分析同源性。再经 EMBOSS win 软件(<http://emboss.sourceforge.net/>)的 tranalign 程序转换成相应的核苷酸序列排布。然后用 PHYLIPY 软件包(<http://evolution.genetics.washington.edu/phylip.html>)中最大似然法(MJ)构建进化树, 自展重复设为 1 000 次。

1.2.2.2 密码子替换模型的参数估计及后验概率检验 参考 Yang 等^[10]的方法对 CNBP 的

适应性进化进行分析。应用 PAML3.15 软件(<http://abacus.gene.ucl.ac.uk/software/paml.html>)的 CODEML 程序中 BranchSite 模型的 Model B, 检测基因的正选择并寻找受到正选择的密码子位点。模型的统计学分析包括两个方面: 第一, 检验基因在进化过程中是否受到正选择。即用似然比检验(likelihood ratio test, LRT)检测序列上是否有一些位点的 $\omega > 1$ 。第二, 进行后验概率检验 ω 估计值大于 1 的位点, 当后验概率很大(> 0.95)时该位点就很可能来自于 $\omega > 1$ 的位点类型, 即可认为该位点处于正选择。而 $\omega = 1$ 和 < 1 分别表示中性进化和负选择作用。

2 结果

2.1 草鱼 CNBP 开放阅读框及其推测的氨基酸序列 草鱼 CNBP 开放阅读框由 492 bp 组成, 编码一个 163 氨基酸残基的蛋白质。该蛋白质的 7 个 CCHC 锌指结构分别位于第 8~21、40~53、60~73、84~97、105~118、123~136、144~157 残基, 其中, 第 25~36 残基(GRGRGRGRGRG)为 RGG 框, 第 76~83 残基(PKKEREQ)为核定位信号区, 第 43~163 残基为核仁组织区(NORs) (图 1)。

2.2 草鱼 CNBP 同源性分析 草鱼 CNBP 与银鲫(AAW8346)和斑马鱼(AA073520)的同源性分别为 98% (160/163 aa) 和 96% (158/163 aa), 可见不同鱼 CNBP 保守性很高。草鱼与银鲫 CNBP 仅在第 9 位(Ser \rightarrow Phe)、第 83 位(Cys \rightarrow Ser)、第 100 位(Ala \rightarrow Gly) 存在差异。草鱼 CNBP 与人(NP003409)的同源性仅为 64% (114/178 aa)。这是因为与其他高等脊椎动物相比, 鱼类 CNBP 有部分氨基酸改变, 其中在第 3 个锌指的第 5 个氨基酸残基由 Gly 变成保守的 His, 另外在第 1 锌指和第 2 锌指结构间, 鱼类 CNBP 缺失 6~14 个氨基酸残基(图 2)。

2.3 CNBP 分子适应性进化分析 以不同物种的 CNBP 及其同工型、变异体蛋白构建系统发育树(图 3)。大鼠(*Rattus norvegicus*)、小鼠(*Mus musculus*) CNBP 及其同工型蛋白 1 和蛋白

```

34  ATGGACATGAGTAOCAGCGAGTGCCTCTGGATGTGGGCGCTCGGACACTGGATCAAGAACTGTCTTAAOCTGGCGCGGTCTGGCAGG
1   M D M S T S E C S G C G R S G H W I K N C P N A G R G R G R
124  GGCGGGGACGAGGAAAGATCTGTTCTGTTAOCGCTGTGGACAGCAGGGOCACATTGOCAGAGACTGTGAACAGACAGAGGATGCATCG
31  G R Q R G K D L F C Y R C G E Q G H I A R D C E Q T E D A C
214  TACAATTGCCACAGGACTGGOCATATTTCCAGAGACTGTAAGAGCCOAAAAAGAAAGAGACCAGTCTTGCTATAACTGGCCAAAAGCT
61  Y N C H R S G H I S R D C K E P K K E R E Q C C Y N C G K A
304  GGTCACGTGGCGGACTGTGAOCCOCCOACGAGCAGAAGTCTACTCTCGCGGGGTTGCGACACATCCAGAAAATTATGTGACAAAG
91  G H V A R D C D H A N E Q K C Y S C G G F G H I Q K L C D K
394  GTGAAATGTTACCGGTGTGGTGTGAGATTGCTCACCTGGCAGTGCAGTGCAGCAAGGCCACTGAGGTAACTGCTACAACCTGGCGAAGAAC
121  V K C Y R C G E I G H V A V Q C S K A T E V N C Y N C G K T
484  GGTCACTGGCGAGAGACTGCACCATTGAAGCGTCTGCGTAA
151  G H L A R E C T I E A S A *

```

图 1 草鱼 *CNBP* 开放阅读框及其推测的氨基酸

Fig. 1 Nucleotide sequence and deduced amino acid sequence of Grass Carp *CNBP*

下划线标出 7 个 CCHC 型锌指结构, 虚下划线标出 RGG 框, 实线框标出核定位信号区。

Underlines denote seven CCHC zinc finger domains, broken underline denote RGG box, and nuclear localization signal denoted by real line frames.

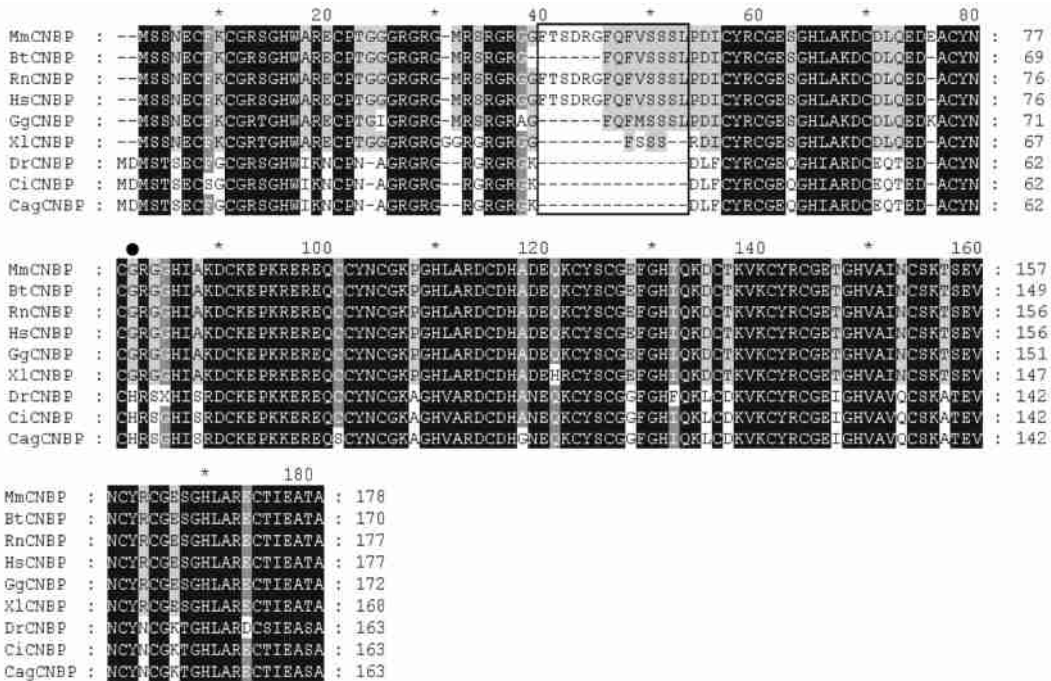


图 2 *CNBP* 同源性分析

Fig. 2 The homology of *CNBP*

黑色阴影表示完全保守的氨基酸位点; 实线框标出第一锌指与第二锌指之间的缺失部分; 黑圆点表示与其他高等脊椎动物相比, 鱼类由 Gly 变成保守的 His; XlCNBP、HsCNBP、RnCNBP、MmCNBP、BtCNBP、GgCNBP、DrCNBP、CiCNBP 和 CagCNBP 分别代表非洲爪蟾 (NP_001084082)、人 (NP_003409)、大鼠 (BAA08212)、小鼠 (AAH58723)、牛 (NP_001029396)、鸡 (NP_990238)、斑马鱼 (AA073520)、草鱼 (EU 709746)、银鲫 (AAW82446) 的 *CNBP*。

Black shadow indicated the fully conservative amino acids; Real line frames indicated the deletion amino acid sites between the first and the second Zn finger; Compared with other vertebrates, a substitution of a His residue in place of Gly in fish denoted by black dot; XlCNBP, HsCNBP, RnCNBP,

MmCNBP, BtCNBP, GgCNBP, DrCNBP, CiCNBP and CagCNBP represent CNBP from *Xenopus laevis* (NP_001084082), *Homo sapiens* (NP_003409), *Rattus norvegicus* (BAA08212), *Mus musculus* (AAH58723), *Bos Taurus* (NP_001029396), *Gallus gallus* (NP_990238), *Danio rerio* (AA073520), *Gnopharyngodon idella* (EU709746), *Carassius auratus gibelio* (AAW 82446), respectively.

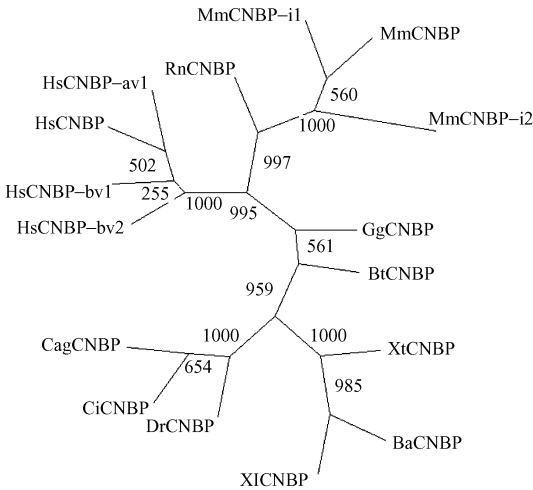


图3 CNBP 系统进化树
Fig. 3 Phylogeny of CNBP

物种以及相关的 GenBank 登录号为: 小鼠 MmCNBP-i1 (NP_038521), 小鼠 MmCNBP-i2 (NP_001103215), 人 HsCNBP-av1 (AAY89856), 人 HsCNBP-bv1 (AAY96754), 人 HsCNBP-bv2 (AAY96755), 非洲爪蟾 XiCNBP (AAI22021), 蝗蚱 BaCNBP (AAD33937), 其余物种名称见图 2; 分枝上的数字为 bootstrap 1 000 次的支持率, 代表分歧度。

The species and GenBank accession numbers of corresponding sequences are *Mus musculus* CNBP isoform 1 (NP_038521), *Mus musculus* CNBP isoform 2 (NP_001103215), *Homo sapiens* CNBP alpha variant 1 (AAY89856), *Homo sapiens* CNBP beta variant 1 (AAY96754), *Homo sapiens* CNBP beta variant 2 (AAY96755), *Xenopus tropicalis* CNBP (AAI22021), *Bifio arenarum* CNBP (AAD33937), respectively. Other species see fig 2. The numerals on branches the support rate of bootstrap 1 000 times, indicate the difference degrees.

2 (MnCNBP-i1, MnCNBP-i2) 位于同一分枝上。人 CNBP 与其 α 变异体蛋白 1、 β 变异体蛋白 1、 β 变异体蛋白 2 (HsCNBP-av1, HsCNBP-bv1, HsCNBP-bv2) 独立位于另一相邻分枝上。草鱼和银鲫的 CNBP 分化较晚, 与斑马鱼的位于同一分枝上。因此, 系统进化树能很好反映不同物种间 CNBP 的亲缘关系, 可用作适应性进化分析的一个数据源。

在 PHYLIPY 树文件基础上, 用 CODEML 程序估计密码子替换模型的参数 (表 1)。从计算的结果来看, 没有得到有大于 1 的 ω 值, 因此, CNBP 位点没有受到正选择作用, 而是处于中性进化或负进化状态。其中, 模型 M3 的参数 $p_0 = 0.5897$, 模型 M8 的参数 $p_0 = 1.0000$, 表示受到选择约束的位点的比例, 说明 CNBP 大多数位点受到功能或结构上的选择约束以及序列在进化上的高度保守性。

3 讨论

CNBP 为典型的 DNA 结合蛋白, 其 7 个重复 CCHC 型锌指占了 CNBP 很大的部分, 其中, 第 1 与第 2 锌指间有 18~34 个氨基酸残基间隔, 而第 2~7 锌指间隔很短。目前, 有关 CNBP 与 DNA 的结合有一些报道。Aaron 认为 5 至 6 个独立的锌指在一起识别和结合 15~18 bp 长

表 1 CNBP 正选择检验参数估计

Table 1 Estimates of parameters in CNBP by positive selective test

模型 Model	自由参数数目 P	似然值对数 lnL(l)	两倍的似然值对数差 2Δl	参数估计 Estimates of parameters	正选择位点 Positively selected sites
M0	1	-1 954.269 0	12.261 8*	$\omega = 0.022$	无
M3	5	-1 948.138 1		$p_0 = 0.5897, p_1 = 0.1427, p_2 = 0.2676$ $\omega_0 = 0.0028, \omega_1 = 0.0532, \omega_2 = 0.0532$	无
M7	2	-1 948.172 6	0.023 2	$p = 0.1909, q = 3.8311$	无
M8	4	-1 948.161 0		$p_0 = 1.0000, p = 0.3687, q = 11.8193$ $p_1 = 0.0000, \omega = 1.0000$	无

P 是模型的自由参数数目; lnL(l) 是最大似然值的对数; 2Δl 是 M3 与 M0 及 M7 与 M8 间 lnL(l) 之差的二倍; * 号表示似然比检验显著; p_0 表示受到选择约束的位点比例; p_1 表示可能受到正选择位点比例; p 和 q 是 M7 和 M8 中 β 分布 B(p, q) 的参数; ω 指示正选择位点的参数。

P is the number of freedom parameters in the model; lnL(l) is the log likelihood; 2Δl is twice the log likelihood difference (M3&M0, M7&M8); * : significant at 5% level; p_0 and p_1 indicated the proportion of sites under selective constraint and under positive selection, respectively; Parameters p and q for the beta distribution B(p, q) are given under M7 and M8; ω indicated the parameters of positively selected sites.

度的 DNA^[13]。所以, CNBP 第 2~7 锌指连接在一起, 可能作为一个整体特异性识别 DNA, 从而调节基因转录。与高等动物不同, 鱼类 CNBP 第 1 与第 2 锌指间隔较小, 其锌指如何与 DNA 结合还有待进一步研究。研究表明, 不同的锌指可能存在一些功能的差异。例如, McGrath 等用人 CNBP 锌指分别代替人 HIV-1 的逆转录核酸结合蛋白的 N 末端锌指序列, 结果包含人 CNBP 第 5 个锌指的 HIV-1 突变体的复制受阻。而其他 6 个锌指的 HIV-1 突变体均能支持 HIV-1 的复制^[14]。这至少可以说明第 5 个锌指与其他锌指存在一定的差异。

虽然在 CNBP 第 3 与第 4 锌指间存在 1 个核定位信号区, 但对该蛋白质在细胞中的分布还有争议。例如 Warden 等发现鼠 CNBP 存在于细胞质、内质网中^[5]。同样, Western blot 也表明 CNBP 可能在蟾蜍 (*Bufo arenarum*) 卵母细胞的细胞质中^[6]。而 Pablo 等发现 CNBP 在动物的不同发育期有差异性表达, 即在胚胎发育早期分布于细胞质, 原肠胚形成后聚集于细胞核中^[5]。同样, Liu 等的结果表明, CNBP 广泛分布在银鲫卵母细胞的细胞质、细胞核、卵黄发生前和中期的滤泡上皮中^[6]。因此, CNBP 可能在细胞质和细胞核中均有分布, 并各自完成独特的生物学功能。在细胞质中, CNBP 能通过 RGG 框(精氨酸甘氨酸簇)与 RNA 相互作用进而对其进行调节。如 CNBP 能特异性结合核糖体蛋白 L4 mRNA 的 5'UTR 的典型嘧啶区, 参与调节核糖体 mRNA 的翻译^[7]。另一方面, 在细胞核中, CNBP 特异性与单链 DNA 结合, 调节基因转录。如人 CNBP 能与甾醇类调节元件结合, 刺激相关基因的转录, 启动红细胞的终末分化^[8]。此外, CNBP 还可作为核因子参与调节中胚层的表达^[9]。

与其他动物和人的 CNBP 相比, 鱼类的第 3 锌指的第 5 个氨基酸残基发生替代 (Gly⁻His), 同时, 在第 1 与第 2 锌指间也比高等脊椎动物的少 6~14 个氨基酸, 其中富含丝氨酸。这些结构差异有什么意义还不清楚。但至少适应性进化分析表明 CNBP 没有受到正选择作用

($\omega \leq 1$), 即这种结构的差异还不足以产生新的功能。这表明不同物种 CNBP 在进化上的高度保守性和功能上的同一性。

参 考 文 献

- [1] Lusis A J, Rajavashisth T B, Klisak I, *et al.* Mapping of the gene for CNBP, a finger protein, to human chromosome 3q13.3-q24. *Genomics*, 1990, **8**(2): 411~414.
- [2] Dominicis A D, Lotti F, Pierandrea Amaldi P, *et al.* cDNA cloning and developmental expression of cellular nucleic acid binding protein (CNBP) gene in *Xenopus laevis*. *Gene*, 2000, **241**(1): 35~43.
- [3] Ruble D M, Foster D N. Molecular cloning and characterization of a highly conserved chicken cellular nucleic acid binding protein cDNA. *Gene*, 1998, **218**(1-2): 95~101.
- [4] Yasuda J, Mashiyama S, Makino R, *et al.* Cloning and characterization of rat cellular nucleic acid binding protein (CNBP) cDNA. *DNA Res*, 1995, **2**(1): 45~49.
- [5] Pablo A, Sebastián C, Lombardo V A, *et al.* Zebrafish cellular nucleic acid binding protein: gene structure and developmental behaviour. *Gene*, 2004, **337**: 151~161.
- [6] Liu J X, Gui J F. Expression pattern and developmental behaviour of cellular nucleic acid binding protein (CNBP) during folliculogenesis and oogenesis in fish. *Gene*, 2005, **356**: 181~192.
- [7] Berg J M. Potential metal-binding domains in nucleic acid binding proteins. *Science*, 1986, **232**(4749): 485~487.
- [8] Flink I L, Morkin E. Organization of the gene encoding cellular nucleic acid binding protein. *Gene*, 1995, **163**(2-3): 279~282.
- [9] Anisimova M, Bielawski J P, Yang Z. Accuracy and power of the likelihood ratio tests in detecting adaptive molecular evolution. *Mol Biol Evol*, 2001, **18**: 1585~1592.
- [10] Yang Z, Nielsen R. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol Biol Evol*, 2000, **17**(1): 32~43.
- [11] 陶敏, 樊棠怀, 徐立中等. 趋化因子及其受体家族分子适应性进化分析. *遗传*, 2007, **29**(12): 1519~1524.
- [12] 樊棠怀, 陶敏, 徐立中等. 转铁蛋白正选择位点的检测及分析. *南昌大学学报(理科版)*, 2007, **31**(3): 298~301.
- [13] Aaron K. Zinc finger peptides for the regulation of gene expression. *J Mol Biol*, 1999, **293**(2): 215~218.
- [14] McGrath C F, Buckman J S, Gagliardi T D, *et al.* Human cellular nucleic acid binding protein Zn²⁺ fingers support

- replication of human immunodeficiency virus type 1 when they are substituted in the nucleocapsid protein. *J Virol*, 2003, **77** (15): 8 524~ 8 531.
- [15] Warden C H, Krisans S K, Purcell Huynh D. Mouse cellular nucleic acid binding proteins: a highly conserved family identified by genetic mapping and sequencing. *Genomics*, 1994, **24**(1): 14~ 19.
- [16] Calcaterra N B, Palatnik J F, Bustos D M, *et al.* Identification of mRNA binding proteins during development: characterization of *Bufo arenarum* cellular nucleic acid binding protein. *Dev Growth Differ*, 1999, **41**(2): 183~ 191.
- [17] Pellizzoni L, Lotti F, Maras B, *et al.* Cellular nucleic acid binding protein binds a conserved region of the 5' UTR of *Xenopus laevis* ribosomal protein mRNAs. *J Mol Biol*, 1997, **267**(2): 264~ 275.
- [18] 李翔, 周建平, 杨锋等. 人细胞核酸结合蛋白 cDNA 克隆及其表达分析. *基础医学与临床*, 2005, **25**(11): 1 004~ 1 008.
- [19] Amas P, Cabada M O, Calcaterra N B. Primary structure and developmental expression of *Bufo arenarum* cellular nucleic acid binding protein: Changes in subcellular localization during early embryogenesis. *Dev Growth Differ*, 2001, **43**(1): 13~ 23.