Cloning and Characterization of Rat Cellular Nucleic Acid Binding Protein (CNBP) cDNA

Jun YASUDA, Shoji MASHIYAMA, Reiko MAKINO, Shigekazu OHYAMA, Takao SEKIYA, and Kenshi HAYASHI

Oncogene Division, National Cancer Center Research Institute, 1-1, Tsukiji 5-chome, Chuo-ku, Tokyo 104, Japan and Division of Genome Analysis, Institute of Genetic Information, Kyushu University, Maidashi 3-1-1, Higashi-ku, Fukuoka 812, Japan

(Received 9 February 1995)

Abstract

We cloned and sequenced the cDNAs which code for rat cellular nucleic acid binding protein (CNBP). In-frame insertion/deletion differences were found among the clones at two sites in the open reading frame, suggesting alternative splicing of the message or the presence of multiple genes which code for this protein. The deduced amino acid sequence revealed that one rat CNBP sequence was completely identical to its human counterpart. This striking conservation, together with the fact that homologous genes have been found in various organisms including Schizosaccharomyces pombe, suggests that CNBP plays a basic biological role in eukaryotic cells. The recombinant GST-CNBP fusion protein produced in Escherichia coli bound to a G-rich single-stranded RNA and DNA in a sequence-specific manner.

Key words: CCHC box; zinc finger; GST-CNBP fusion protein; RNA binding

We screened a cDNA expression library for the clone which encoded the protein that binds to the 5′ region of rat c-myc intron 1, using the in situ detection method of Vinson et al. This region is known to be involved in positive and negative transcriptional control of the gene, and transcription pausing (our unpublished results. Also see review). The nucleotide sequence of the binding site had three G-stretches (Fig. 1). One clone, AMAS, was obtained after screening 10^6 phages of a Agt 11 rat liver cDNA expression library (Clontech). The insert was excised as a 1.5-kb EcoRI fragment and subcloned into pBluescript SK(+) (Fig. 1). Both strands of the insert of the obtained plasmid, pMAS, were then sequenced. A polyadenylation signal and poly(A) stretches were found at one end of the insert. A homology search of the nucleotide sequence of the GenBank database revealed a high degree of similarity to human CNBP cDNA. We then screened another λgt 11 rat liver cDNA library (randomly primed, Clontech) using a 316-bp fragment at the 5′-end of the pMAS insert as a hybridization probe. Three clones (AS2, AS3, and AS5) were obtained from among 2.0×10^6 phages. After subcloning to the plasmids as described above, the inserts were sequenced. All three clones carried nearly identical sequences to that of the AMAS insert to the 3′-end, but some discrepancies were found at both ends and in the internal regions among the four clones. The 3′-heterogeneities all consisted of truncation of the longest clone pS2, and were found outside the coding region. This heterogeneity could be explained by the different usage of the polyA signal or oligo dT-priming at internal A-clusters during library construction.

The four clones also showed different 5′-structures. We performed RT-PCR of rat liver total RNA using one primer which had a 5′-end sequence specific to each clone and another primer which had a sequence of a common region. One reaction, using a sequence of pS2 as the 5′ primer, yielded amplification product, indicating that mRNA with the same sequence as the pS2 insert is expressed in rat liver. Primer extension analysis using rat liver total RNA revealed that the mRNA extended an additional 150 bases upstream of the 5′ end of the pMAS insert as a hybridization probe. Three clones (AS2, AS3, and AS5) were obtained from among 2.0×10^6 phages. After subcloning to the plasmids as described above, the inserts were sequenced. All three clones carried sequences that were nearly identical to that of the λMAS insert to the 3′-end, but some discrepancies were found at both ends and in the internal regions among the four clones. The 3′-heterogeneities all consisted of truncation of the longest clone pS2, and were found outside the coding region. This heterogeneity could be explained by the different usage of the polyA signal or oligo dT-priming at internal A-clusters during library construction.

The four clones also showed different 5′-structures. We performed RT-PCR of rat liver total RNA using one primer which had a 5′-end sequence specific to each clone and another primer which had a sequence of a common region. One reaction, using a sequence of pS2 as the 5′ primer, yielded amplification product, indicating that mRNA with the same sequence as the pS2 insert is expressed in rat liver. Primer extension analysis using rat liver total RNA revealed that the mRNA extended an additional 150 bases upstream of the 5′ end of the pS2 insert (data not shown). By 5′ Rapid amplification of cDNA end (5′RACE) using rat liver total RNA as a template, an additional 24 nucleotides upstream of the 5′ end of the pS2 insert were obtained.

The composite nucleotide sequence of the pS2 insert and the 5′RACE product, together with the deduced amino acid sequence of the putative coded protein, are shown in Fig. 2. In clone pS5, a 21-bp in-frame deletion was found at bases 269 to 289 of the composite. The CD-NA's of rat and a reported human homolog showed identical flanking nucleotide sequences for this deletion. The
Rat c-myc exon 1/ intron 1

Figure 1. Schematic diagram of the rat c-myc gene exon1/intron 1 boundary. Arrow indicates the second promoter of rat c-myc gene. Open boxes indicate exons 1 and 2. The thick line indicates a protein binding site we used as a probe for in situ screening. The nucleotide sequence of the binding site is shown beneath the thick line. Three G-stretches are underlined. The sequence of the oligomer which was used as a competitor in the electrophoresis mobility retardation assay is indicated by a box.
Figure 2. The nucleotide and amino acid sequence of rat CNBP cDNAs. The 24-base sequence from the 5'-end was obtained from 5'RACE and used to make a composite with the pS2 insert. The polyadenylation signal is indicated with bold letters (bases 1544 to 1549 of the composite). The deleted amino acid residues in pS5 are italicized. The site of the 3-base insertion found in the pMAS sequence is shown by a small arrow between bases 383 and 384, with a deduced amino acid residue. The 3'-ends of the other three clones are indicated by small arrows with the names of the respective clones. CCHC boxes are boxed, and a glycine-rich region is underlined in the amino acid sequence.
Figure 3. Sequence-specific binding of GST-CNBP fusion proteins to RNA. The probes (see text) are indicated at the top of the panel.

The cell extracts used are indicated at the top of the lanes. Binding conditions were essentially the same as those of Lichtsteiner et al. Sense and antisense RNAs were synthesized by in vitro transcription in the presence of $[^{35}P]UTP$ (Amersham). One microliter (5,000 cpm) of these probes was incubated with each extract at 4°C for 15 min in 10 μl of binding buffer: 15 mM HEPES (pH 7.6), 50 mM KCl, 12% glycerol, 0.05 mM EDTA, 1 mM DTT, 1 mM MgCl₂, 20 units of RNasin (Promega), 0.1 mM PMSF, and 1 μg of heat-treated yeast total RNA. The amount of cell lysate was adjusted by the enzyme activity of GST (0.01 units). The RNA-protein mixture was electrophoresed in 5% native polyacrylamide gel at 4°C. The running buffer was 0.25×TBE. A specific competitor (Fig. 1) was added at 100-fold excess of the probe.

molog. Homology in the 5′-untranslated region between the two species was 74.5%; however, the deduced amino acid sequence of rat CNBP was completely identical to its human homolog. This striking conservation, together with the fact that the homolog has also been found in fission yeast (byr 3) suggests that this protein may have played a crucial role in eukaryotic cells throughout evolution.

Recently, multiple CCHC boxes have also been found in Leishmania HEXBP, Tetrahymena cnjB product, Caenorhabditis glh-1 product and Crithidia fasciculata UMSBP. The CCHC boxes are considered a zinc finger motif for single-stranded nucleic acid binding, and one or two copies of the prototypical CCHC box are present in retroviral nucleocapsid proteins, or gag. The gag proteins are thought to bind to viral RNA and play essential roles in dimerization and packaging of the viral RNA genome. Several mutagenesis experiments using gag have shown that the CCHC box is essential for RNA packaging, which suggests that the box has a specific role in protein-RNA interactions.

The in vivo function of CNBP is unclear. Initially, CNBP was believed to bind to sterol-regulatory element (SRE) in the HMG-CoA reductase gene and negatively control its expression. However, at this time, the protein does not seem to be directly involved in sterol homoeostasis, and is believed to play an important role in cell survival under oxysterol-induced cell death of lymphoid cells. Our results suggest that CNBP binds to single-stranded RNA in a sequence-specific manner, although the biological role of this protein is unclear. The relationship between c-myc gene expression and CNBP also remains unknown. Single-stranded nucleic acid may be involved in several steps of gene expression, e.g., in transcription, mRNA maturation and transport, or translation. Recently, a cDNA of single-stranded DNA binding protein (FBP) was isolated as a regulator of c-myc expression which interacts at the far upstream element of the human c-myc gene. However, there are no apparent homologous regions between FBP and CNBP. The unique structural characteristics of CNBP, which has no apparent functional domains other than the zinc fingers, suggest that the mechanism of any kind of action of this protein may involve interaction with single-stranded nucleic acids or other proteins via fingers.

Acknowledgments: This research was supported in part by Grants-in-Aid from the Ministry of Health and Welfare for the Comprehensive 10-Year Strategy for Cancer Control, from the Ministry of Education, Science and Culture of Japan and from the Special Coordination Fund of the Science and Technolog Agency of Japan.

References

1. Hayashi, K., Makino, R., Kawamura, H., Arisawa, A., and Yoneda, K. 1987, Characterization of rat c-myc and
encoding a single-stranded DNA-binding protein containing nine "CCHC" zinc finger motifs, J. Biol. Chem., 268, 13994–14002.

13. Taylor, F. M. and Martindale, D. W. 1993, Retroviral-type zinc fingers and glycine-rich repeats in a protein encoded by cnjB, a Tetrahymena gene active during meiosis, Nucleic Acids Res., 21, 4610–4614.

14. Roussel, D. L. and Bennett, K. L. 1993, glh-1: A germline putative RNA helicase from Caenorhabditis has four zinc fingers, Proc. Natl. Acad. Sci. USA, 90, 9300–9304.

15. Abeliovich, H., Tzfat, Y., and Shlomai, J. 1993, A trypanosomal CCHC-type zinc finger protein which binds the conserved universal sequence of kinetoplast DNA minicircles: isolation and analysis of the complete cDNA from Crithidia fasciculata, Mol. Cell. Biol., 13, 7766–7773.

16. Summers, M. F. 1991, Zinc finger motif for single-stranded nucleic acids? Investigations by nuclear magnetic resonance, J. Cell. Biochem., 45, 41–48.

17. Dupraz, P., Oertle, S., Meric, C., Damay, P., and Spahr, P. F. 1990, Point mutations in the proximal Cys-His box of Rous sarcoma virus nucleocapsid protein, Proc. Natl. Acad. Sci. USA, 87, 6403–6407.

18. Meric, C., Gouilloud, E., and Spahr, P. F. 1988, Mutations in Rous sarcoma virus nucleocapsid protein p12 (NC): deletions of Cys-His boxes, J. Virol, 62, 3328–3333.

19. Ayala-Torres, S., Johnson, B H., and Thompson, E. B. 1994, Oxysterol sensitive and resistant lymphoid cells: correlation with regulation of cellular nucleic acid binding protein mRNA, J. Steroid Biochem. Mol. Biol., 48, 307–315.

20. Duncan, R., Bazar, L., Michelotti, G. et al. 1994, A sequence-specific, single-strand binding protein activates the far upstream element of c-myc and defines a new DNA-binding motif, Genes Dev., 8, 465–480.

21. Lichtsteiner, S., Wuarin, J., and Schibler, U. 1987, The interplay of DNA-binding proteins on the promoter of mouse albumin gene, Cell, 51, 963–973.
