Nucleic Acids Research, Vol. 20, No. 15 ***4001–4007***

cDNA cloning of a novel heterogeneous nuclear ribonucleoprotein gene homologue in *Caenorhabditis elegans* using hamster prion protein cDNA as a hybridization probe

Makoto Iwasaki, Koichi Okumura, Yutaka Kondo, Tatsuro Tanaka and Hisanaga Igarashi*
Shionogi Institute for Medical Science, 2-5-1 Mishima, Settsu-shi, Osaka 566, Japan

Received April 27, 1992; Revised and Accepted July 13, 1992 DDBJ accession no. D10877

ABSTRACT

The mammalian prion protein (PrP^c) is a cellular protein of unknown function, an altered isoform of which (PrP^{sc}) is a component of the infectious particle (prion) thought to be responsible for spongiform encephalopathies in humans and animals. The evolutionary conservation of the PrP gene has been reported in the genomes of many vertebrates as well as certain invertebrates. In the genome of nematode *Caenorhabditis elegans*, the sequence capable of hybridizing with the mammalian PrP cDNA probe has been demonstrated, predicting the presence of the PrP gene homologue in *C.elegans*. In this study, Southern analysis with the hamster PrP cDNA (HaPrP) probe confirmed the previous observation. Moreover, Northern analysis revealed that the sequence is actively transcribed in adult worms. Thus, we screened *C.elegans* cDNA libraries with the HaPrP probe and isolated a cDNA that hybridizes to the same sequence in *C.elegans* that hybridized with the HaPrP probe in the Southern and Northern analyses. The deduced amino acid sequence of this cDNA, however, is substantially homologous with heterogeneous nuclear ribonucleoprotein (hnRNP) core proteins rather than mammalian PrP^c. The hnRNPs contain the glycine-rich domain in the C-terminal half of the molecule, which also seemed to be in PrP^c at the N-terminal half of the molecule. Both of the glycine-rich domains are composed of tracts with high G + C content, indicating that these tracts may due to the hybridizing signals. These results suggest that this cDNA clone is derived from a novel hnRNP gene homologue in *C.elegans* but not from a predicted PrP gene homologue.

INTRODUCTION

Subacute spongiform encephalopathies, such as Creutzfeldt-Jakob disease, Gerstmann-Sträussler Syndrome and kuru in man and scrapie, bovine spongiform encephalopathy, and transmissible mink encephalopathy in animals, appear to be transmissible via infectious agents (1). These agents have yet to be conclusively identified and are usually referred to as unconventional viruses. The overall properties of the agent, which has been designated as a prion (2), differ from those of any known virus or viroid and early on gave rise to speculations that it might be devoid both of nucleic acid and of protein, consist of protein only, be a polysaccharide or a membrane fragment. Recent reports, however, have documented the existence of prion protein (PrP) mRNA in a variety of uninfected tissues (3, 4). PrP genes have also been detected by hybridization in the genomes of many vertebrates, including mouse, rat, rabbit, sheep, goat, and human, as well as in certain invertebrates, such as nematode *Caenorhabditis elegans*, *Drosophila*, and possibly yeast (5). Comparison of hamster(3), mouse (6, 7), rat (8), sheep (9), bovine (10) and human (11, 12) PrP^c amino acid sequences deduced from nucleotide sequences of their PrP genes have revealed extensive homology (approximately 90%). A recent report (13) described the isolation of a cDNA from fractions enriched for the acetylcholine receptor-inducing activity in chicken. This cDNA encodes a chicken protein that is identical to the mouse PrP^c at 33% of its amino acid positions. Although the function of PrP^c is unknown, these findings raise the possibility that the PrP^c serves normally to regulate the chemoreceptor number at the neuromuscular junction.

An exhaustive study of the tiny roundworm *C.elegans* has revealed a wealth of information about development and the brain. In a body composed of very few cells, there are 302 neurons in its nervous system, as opposed to 100 billion or so in man (14). Application of the nematode *C.elegans* as an experimental material for understanding brain function is of great benefit to investigating the biological functions of PrP^c especially in the central nervous system (CNS) and the roles of PrP^c in progressive degeneration of the CNS.

Our Southern analysis with the hamster PrP (HaPrP) cDNA probe supported the previous data (5) that *C.elegans* genomic DNA digested with *Eco*RI had two hybridized bands, suggesting that the PrP gene homologue might be present in *C.elegans*.

* To whom correspondence should be addressed

Northern analysis data demonstrated that it appeared to be actively transcribed. By screening *C.elegans* cDNA libraries by hybridizing with the HaPrP probe, we isolated positive clones and evaluated them in detail. In this study, we concluded that our clones were derived from a novel heterogeneous nuclear ribonucleoprotein (hnRNP) gene homologue in *C.elegans* but not from a predicted PrP gene homologue.

MATERIALS AND METHODS

HaPrP probe fragment

The cDNA insert was excised from pHaPrP (HaPrP cDNA containing plasmid) (3) by cleavage with *Bam*HI, followed by preparative agarose gel electrophoresis. The recovered fragment DNA (2 kb) was labeled with [α-^{32}P] dCTP and served as a probe for hybridization during the Southern and Northern analyses and cDNA cloning. This fragment contains the whole sequence in the probe previously utilized by Westaway and Prusiner (5) and furthermore includes the 3′ non-coding region.

Preparation of genomic DNA and total RNA from *C.elegans*

In this study, the wild-type strain (N2) of *C.elegans* var. Bristol (15) was used. The worms were cultured on agar plates at 20°C. The mixed population of worms mainly contained adults or L4 larvae. They were washed and concentrated in M9 buffer (15). Isolation of genomic DNA from the concentrated worms was performed using the general method (16). For preparation of total RNA, the worms (1 g in wet weight) were suspended in 10 ml of GTC solution (4 M guanidine thyocyanate, 10 mM Tris (pH 7.5), 1 mM EDTA, 0.5%(w/v) Sarcosyl, 1%(v/v) 2-mercaptoethanol), a one-third volume of glass beads was added, and the mixture was vortexed ten times for 30 seconds and centrifuged. The supernatant was passed through a needle (18-gage) ten times, layered over a solution (5.7 M CsCl/10 mM EDTA), and ultracentrifuged (M. Imagawa, personal communication). The RNA pellet was dissolved in distilled water. Poly(A+) RNA was selected by Oligotex-dT30 (Nippon Rosch K.K., Tokyo, Japan).

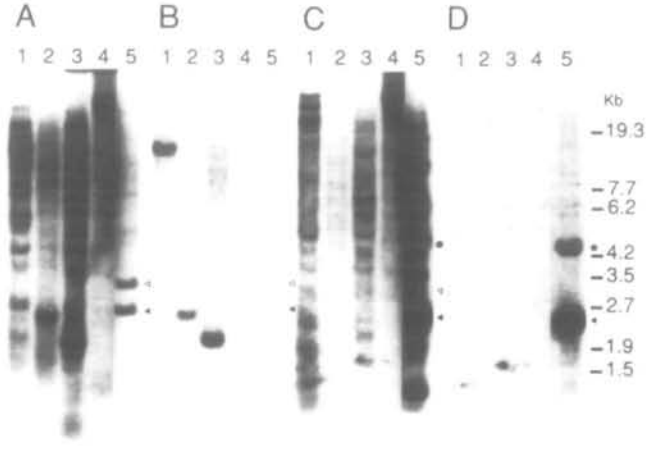

Figure 1. Southern analysis of genomic sequences related to pHaPrP. Human (T98G glioblastoma cell line, ATCC # CRL-1690) (lane 1), mouse (NIH3T3 cell line, ATCC # CRL-1658) (lane 2), rat (Donryu rat placenta tissues) (lane 3), chicken (chicken embryo-fibroblast) (lane 4), and *C.elegans* (lane 5) genomic DNAs were digested with *Eco*RI, electrophoresed through a 0.8%(w/v) agarose gel, transferred to a nylon filter and hybridized with a [^{32}P]-labeled cDNA insert of pHaPrP (A and B) or the C21 clone (C and D). The filter was washed under conditions of low stringency (A and C) or high stringency (B and D) as described in 'Materials and Methods'. The three bands, 2.5 (◀), 3.0 (◁) and 4.5 kb (●), are observed in *C.elegans*. The DNA molecular size markers (λ-*Eco*T141 digest; Takara Shuzoh Co., Ltd., Kyoto, Japan) are shown on the right.

Southern analysis

Genomic DNAs of various vertebrates (10 μg) and *C.elegans* (3 μg) were completely digested with *Eco*RI and electrophoresed on 0.8% (w/v) agarose gels. The DNAs were denatured, reneutralized, and transferred to nylon membranes (Hybond-N+; Amersham Int. plc., Buckinghamshire, UK). Hybridization with the HaPrP probe was performed in a hybridization solution (6×SSC, 5×Denhardt's, 30%(v/v) formamide, 0.5%(w/v) SDS, 0.5 mg/ml heparin) at 42°C overnight. The filter was washed in a solution (0.1×SSC, 0.5%(w/v) SDS) at 42°C or 65°C under conditions of low stringency or high stringency, respectively, followed by exposure to X-ray film (Kodak Co., Tokyo, Japan) with an intensifying screen.

Northern analysis

Total RNAs (10 or 20 μg) or poly(A+) RNAs (2.5 μg) were heat-denatured, electrophoresed on 0.66 M formaldehyde, 1.2%(w/v) agarose gels and transferred to Hybond-N+ nylon membranes. Hybridization and washing were performed under the same procedure as described in the Southern analysis.

C.elegans cDNA library screening and nucleotide sequence analysis

The cDNA libraries inserted in λgt10 and λgt11 (kindly provided by M. Imagawa and H. Yasuda) were used. The NM514 strain for λgt10 and the Y1090 strain for λgt11 were used as host bacteria. The screening method employed was that of Sambrook et al. (17). Hybridization was performed under the low or high stringency condition as described above. The positive clones were rescreened twice more and subcloned into the *Eco*RI site of pBluescript (Stratagene, CA, USA). The cDNAs were sequenced by the Sanger dideoxynucleotide chain-termination method (18) using Sequenase Ver. 2.0 (United States Biochemical, OH, USA).

RESULTS

Confirmation of the sequences in *C.elegans* reacted with the HaPrP probe in Southern analysis

Westaway and Prusiner (5) reported the highly evolutionary conservation of the cellular gene encoding the scrapie prion protein. In the same report, they also predicted the presence of the PrP gene homologue in *C.elegans*. In order to verify this prediction, we first re-examined the experiment as follows. Southern analysis of genomic DNAs under the low stringency condition revealed several strong and faint bands in human, mouse, rat and chicken genomic DNAs (Fig. 1A). Under the same conditions, two strong bands (3.0 and 2.5 kb) were observed in the *C.elegans* genomic DNA. This observation was identical to that described by Westaway and Prusiner (5), confirming the presence of hybridizing signals in *C.elegans* with our HaPrP probe. Under the high stringency condition, a specific band in each mammal, corresponding to the PrP gene, that is, about 15 kb in human, 1.9 kb in rat, and 2.3 kb in mouse, remained (Fig. 1B). In contrast to the mammals, the intensity of every signal in chicken and *C.elegans* was clearly diminished (Fig. 1B, lanes

4 and 5). In other experiments, the intensity of the 2.5 kb band in *C.elegans* tended to be slightly stronger than that of the other 3.0 kb band (data not shown). These data were also identical to that reported by Westaway and Prusiner (5) and supported their prediction.

Identification of transcripts of *C.elegans* derived from the sequences hybridized with the HaPrP probe in Northern analysis

We further examined whether the predicted PrP gene homologue in *C.elegans* would transcribe. As shown in Fig. 2A, Northern analysis revealed four bands (2.2, 1.7, 1.4, 0.7 kb) in both total and poly (A+) RNA from *C.elegans*. This demonstrates that the predicted PrP gene homologue in *C.elegans* might actively transcribe. Thus, to clarify the former prediction, we next attempted to obtain cDNA clones derived from the predicted PrP gene homologue in *C.elegans*.

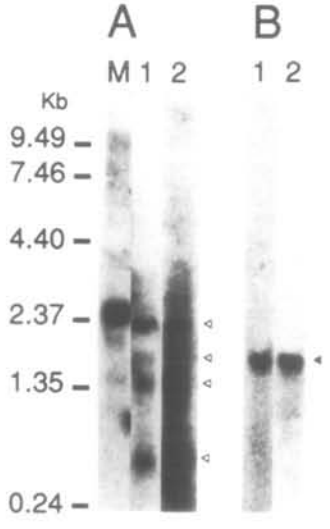

Figure 2. Detection of pHaPrP-related transcripts in mouse and in *C.elegans*. Samples of total (10 μg) (lane1) and poly (A+) RNA (2.5 μg) (lane 2) from *C.elegans*, and total RNA (10 μg) (lane M) from mouse fibroblast cells (NIH3T3) were electrophoresed through a 0.66 M formaldehyde, 1.2%(w/v) agarose gel, transferred to a nylon filter, and hybridized with a [^{32}P]-labeled cDNA insert of pHaPrP (A) or the C21 clone (B). The filter was washed under conditions of low stringency (A) or high stringency (B) as described in 'Materials and Methods'. In *C.elegans*, the four bands (2.2, 1.7, 1.4, 0.7 kb) observed in A (lanes 1 and 2) are indicated with open triangles and the one band (1.7 kb) in B (lanes 1 and 2) is indicated with a closed triangle. NIH3T3 cells (A, lane M) show a prominent band (2.5 kb) as expected. RNA molecular standards (0.24–9.5 kb RNA Ladder, BRL, MD, USA) are shown on the left.

```
C21  AGATTACGGCGGTGGCTGGGGACAGCAAGGAGGCGGCGGTCAAGGCGGAT  110
PRP  ACCCCATGGTGGTGGCTGGGGACAGCCCCATGGTGGTGGCTGGGGACAGC  194

C21  GGGGAGGCCCACAACAGCAGCAAGGAGGTGGTGGCTGGGGACAGCAAGGC  160
PRP  CCCATGGTGGTGGCTGGGGTCAGCCCCATGGTGGTGGCTGGGGTCAAGGA  244

C21  GGCGGCGGTCAAGGCGGATGGGGAGGACCACAGCAACAGCAGCAAGGAGG  210
PRP  GGTGGCACCCACAATCAGTGGAACAAGCCCAGTAAGCCAAAAACCAACAT  294

C21  CTGGGGAGGTCCTCAACAAGGAGGCGGCGGCGGTGGCTGGGGAGGCCAAG  260
PRP  GAAGCACATGGCCGGCGCTGCTGCGGCAGGGGCCGTGGTGGGGGGCCTTG  344

C21  GCCAGCAGCAAGGCGGCTGGGGAGGTCAATCCGGTGCTCAGCAGTGGGCT  310
PRP  GTGGCTACATGCTGGGGAGTGCCATGAGCAGGCCCATGATGCATTTTGGC  394

C21  CATGCTCAAGGAGGAAACAGAAACTATTAAATTAATTCCTTAAGCCCCTC  360
PRP  AATGACTGGGAGGACCGCTACTACCGTGAAAACATGAACCGCTACCCTAA  444
```

Figure 3. Nucleotide sequence homology between cDNAs of the C21 clone and the hamster PrP. Identical sequence repeats are indicated by underlining. Nucleotide sequences not shown did not exhibit significant homology. Numbering of residues is from the 5′ end of the cDNA.

cDNA cloning of *C.elegans* PrP gene homologues and evaluation by Southern and Northern analyses

From the screening of total 6×10^5 clones of *C.elegans* cDNA libraries, we obtained two positive clones for hybridizing with the HaPrP probe under the low stringency condition. A positive clone (designated C21) out of the two clones was selected from the library in the λgt10, because it held the longer insert and the other held the identical but shorter insert. The nucleotide sequence data of the inserted cDNA (651-bp) of the C21 clone (shown in Figs. 3 and 4) indicate 49.0% sequence homology (the GenBank Data base as of 7/1/1992 was evaluated). Almost identical, continuous sequence repeats of 19 base pairs were observed between this clone DNA and the HaPrP cDNA used as a probe (Fig. 3). We think that the hybridization between the positive clone and the HaPrP probe was due to this homology.

Using the inserted cDNA of the C21 clone as a probe (designated the C21 probe), Southern analysis revealed many bands in *C.elegans*, chickens, and mammals under the low stringency condition (Fig. 1C). Under the high stringency condition, two bands (4.5 and 2.5 Kb) in *C.elegans* and no band in others remained (Fig. 1D). Of the two bands in *C.elegans*, one major band, 2.5 kb, corresponds to the former 2.5 kb band which was hybridized with the HaPrP probe as shown in Figs. 1A and 1B. Since this 2.5 kb band was hybridized with the HaPrP probe more strongly than the other 3.0 kb band as described above and also in the paper reported by Westaways and Prusiner (5), the C21 clone obtained here was considered to be derived from the predicted PrP gene homologue in *C.elegans*.

Northern analysis with the C21 probe showed that only one band (1.7 kb) was hybridized in both total and poly (A+) RNA specimens from *C.elegans* (Fig. 2B), revealing that the C21 clone gene actively transcribes in a size of 1.7 kb. This transcript was considered to correspond to one of the four transcripts which were demonstrated to be detected with the HaPrP cDNA probe (Fig. 2A). While the 1.7 kb RNA was clearly not the strongest band, the 2.5 kb DNA band was the stronger one. The discrepancy between these Northern and Southern blots may be explained by the possibility that the level of the 1.7 kb RNA expression in the RNA preparation used here is lower than that of the other three RNAs expression. The other three bands were not detected even under the low stringency condition with the C21 probe (data not shown) and presumably are derived from other distinct genes or from three transcripts produced by alternative splicing of a single primary transcript.

The DNA sequence data from the C21 cDNA clone was so limited that any complete open reading frames (ORF) could not be deduced. Therefore, hybridizing with the C21 probe under the high stringency condition, we rescreened the cDNA libraries and picked several positive clones. From these clones, we selected a positive clone (designated C026) which involved the longest insert, 1431-bp in length. The sequence data of the C026 cDNA revealed that it involves the C21 sequence at the 3′ end of the C026 clone and contains a single long ORF of 1,038 nucleotides, encoding a predicted protein product of 36 kDa (Fig. 4). The size of the single ORF correlated with a 1.7 kb mRNA detected by Northern analysis. The length of the deduced ORF, 346 amino

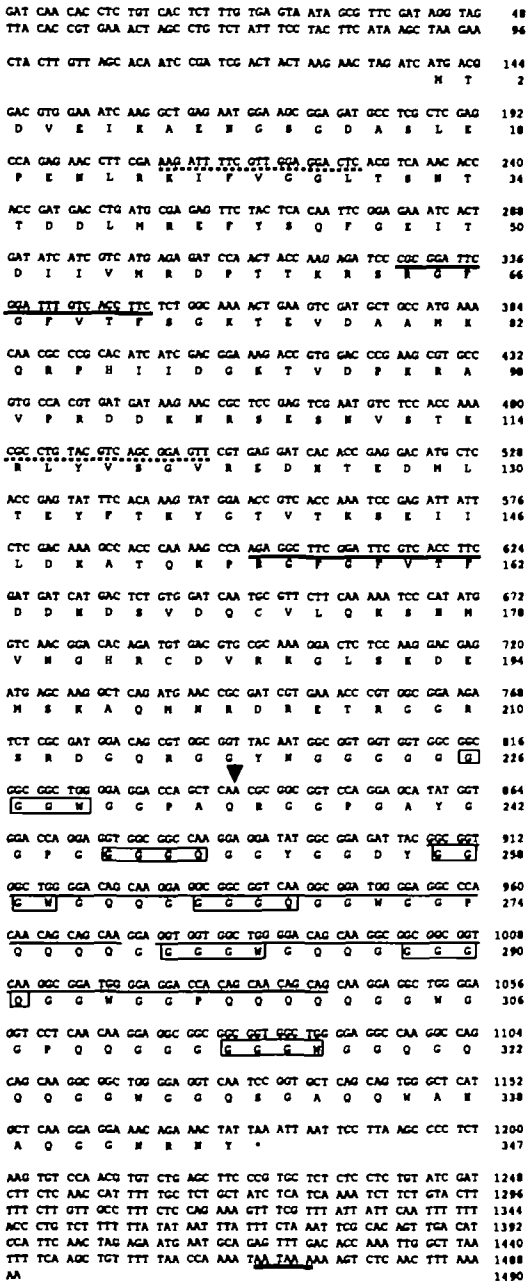

Figure 4. The nucleotide sequence of the C026 and the C21 cDNA inserts and the deduced sequence of the 346 amino acid protein corresponding to the single long open reading frame. The coding strand is shown. The first amino acid residue in the deduced protein sequence corresponds to the first AUG from the 5' end; the last amino acid residue is adjacent to the first in-frame termination codon (UAA). This ATG is probably the true initiation codon, because it is preceded by an in-frame termination codon at nt 133, and its surrounding sequence (ATCATGACG) closely matches the consensus translation initiation site (PuNNATGG) proposed by Kozak (29). In the 3' untranslated region, a putative mRNA processing signal (AATAAA) is indicated with a double underline. At the N-terminus of the deduced protein, an RNA-binding domain consisting of tandemly repeated RNP motifs that contain two highly conserved short segments, referred to as RNP1 and RNP2 (indicated with thick and broken underlines, respectively), is found. GGGW and GGGQ repeats are in boxes and two of the G-Q repeats (22 aa) are underlined. The 5'end of the C21 cDNA is marked with an arrowhead.

acids, was much longer than that of the hamster PrP ORF, 254 amino acids (19). The amino acid sequence comparison of the deduced protein with the hamster PrP^c protein did not give us a significant similarity overall (Fig. 5). On the other hand, another amino acid sequence comparison showed 41% homology between the C-terminal part of this deduced protein and the hamster PrP^c protein at the N-terminus (Fig. 5). This region is characterized as the glycine-rich region that contains three GGGQ and four GGGW repeats as shown in Fig. 4, which also seemed characteristic in the PrP^c sequence, for example, five in human (11) and four in hamster (3). The co-presence of the GGGQ/W repeats in both proteins might imply a similar function in part, although the C-terminus of the putative protein has quite low or no homology with the C-terminus of the mammal PrP^c protein.

The C026 clone is deduced as a novel hnRNP gene homologue in *C.elegans*

Comparison of the deduced C026 protein sequence with sequences in the GenBank Data base revealed homology with the A and B group proteins of human heterogeneous nuclear ribonucleoprotein (hnRNP), A1 and B1 (20, 21), respectively, of rat hnRNP (22), of *Xenopus* hnRNP, XA1a (23), of *Drosophila* hnRNP, Hrb98DE (24) and Hrb87F (25) (Fig. 5), and of grasshopper hnRNP (26) (data not shown) rather than the mammalian PrP^c protein (Fig. 5). The C026 protein is 42−48% identical to these hnRNP members, with identical or conserved amino acid residues in the entire overlapped sequence. In this analysis, conserved amino acids include E and D, K and R, T and S, and A, V, I, L, and F. When the C026 protein is divided into two parts, the N-terminus (amino acid 1 to 195) and the C-terminus (a.a. 196 to 346), the N-terminus shares 56−66% amino acid identity with those of these hnRNP members; whereas, the C-terminus shares only 20−31% identity. Two other features conserved among them are the two RNA-binding domains which contain the two consensus sequences, RNP1 and RNP2, in the N-terminal half of the molecule and a glycine-rich region extending throughout almost the entire sequence in the C-terminal half of the molecule (Figs. 4 and 5). On the other hand, the C026 protein is only 23% identical to the hamster PrP^c. The C026 protein also displays different structural domains as does the mammalian PrP^c, including proline- and glycine-rich repeats in the N-terminal halves of the molecules as well as central and C-terminal hydrophobic regions. This evidence suggests that the C026 clone is derived from a novel hnRNP gene homologue in *C.elegans* but not from a predicted PrP gene homologue.

DISCUSSION

As already mentioned, Westaway and Prusiner (5) have described that three invertebrate DNAs reacted with the mammalian PrP probes in the order nematode ~ *Drosophila* >> yeast, and suggested that the sequences in vertebrates DNA as well as in these invertebrates DNA may arise from authentic PrP genes. In order to verify this suggestion, we decided to isolate the authentic PrP gene from nematode *C.elegans*. Our data on Southern analysis with the HaPrP probe supported the previous data, moreover, our Northern data strongly suggested that the sequence in *C.elegans* was actively transcribed. Thus, we screened *C.elegans* cDNA libraries with the HaPrP probe under the low stringency condition the same as in the Southern and Northern analyses and isolated the C21 clone. This cloned cDNA

```
PRP
C026  MTDVEIKAENGSGDASLEPENLRKIFVGGITSNTTDDLMREFYSQFGEIT   50
HA1           MSKSESPK...Q..L.I...SFE...ESL.SHFE.W.TL.    41
HB1     MEKTLETVPLERKKR.K.QF..L.I...SFE..EESL.NY.E.W.KL.  48
XA1a          MHKSE.PN...Q..L.I...SFE...ESL..HFE.W.TL.    41
D98   NGNSNGHDDDFPQ.SIT...HM..L.I...DYR...ENLKAHFEKW.N.V   58
D87   AEQNDSNGNYDD.EEIT...Q..L.I...DYR....GLKAHFEKW.I.V   51
                          RNP2
PRP
C026  DIIVMRDPTTKRSRGFGFVTFSGKTEVDAAMKQRPHIIDGKTVDPKRAVP  100
HA1   .CV.....N....|.......YATVE......NA...KV..RV.E.....S  91
HB1   .CV.....AS...|.........SMA......AA...S...RV.E.....A  98
XA1a  .CV.....NS...|.......YLSTD......TA...KV..RV.E.....S  91
D98   .VV..K..R....|.....I.Y.HSSMI.E.Q.S...K...RV.E......  108
D87   .VV..K..K....|.....I.Y.QSYMI.N.QNA...K...R..E......  101
                   RNP1
PRP
C026  RDDKNRSESNVSTKRLYVSGVREDHTEDMLTEYFTKYGTVTKSEIILDKA  150
HA1   .E.SQ.PGAHLTV.KIF.G.IK..TE.HH.RD..EQ..KIEVI..MT.RG  141
HB1   .EESGKPGAH.TV.K.F.G.IK..TE.HH.RD..EE..KIDTI...T.RQ  148
XA1a  .E.SS.PGAHLTV.KIF.G.IK..TE..H.R...EQ..KIEVI..MT.RG  141
D98   .Q.IDSPNAGATV.K.F.GALKD..D.QSIRD..QHF.NIVDIN.VI..E  158
D87   .QEIDSPNAGATV.K.F.G.L.D..D.EC.R...KDF.QIVSVN.VS..D  151
                   RNP2
PRP
C026  TQKPRGFGFVTFDDHDSVDQCVLQKSHMVNGHRCDVRKGLSKDEMSKAQM  200
HA1   SG.K...A....|.....KI.I..Y.T....N.E...A...Q..AS.SS  191
HB1   SG.K...G....|....P..KI....Y.TI...NAE...A..RQ..QEV.S 198
XA1a  SG.K...A....E.....KI.I..Y.T..N.NSQ...A...Q..ASVSG  191
D98   .G.K...A..E..|..Y.P..KV....Q.QL..KMV..K.A.P.QNDQQGGG 208
D87   .G.K...A.IE..|..Y.P..KII...T.SIKNKTL..K.AIA.QD.DRQGG 201
          RNP1
PRP           FVAMWTDV.LCKKRPKP...NTGGS.YP.Q.SP..NRYPPQ    52
C026  NRDRETRGGRSRDGQRGGYNGGGGGGGGWGGPAQR-GGPGAYGGPG--GG  247
HA1   SQRGRSGS.NFGG.RG..FG.NDNF.R.GNFSGRG-.FG.SR..G.--Y.  238
HB1   S.SGRGGNFGFG.SRG..G.F.P.P.SNFR.GSDG-Y.S.RGF.D.--YN  245
XA1a  SQRERGGS.NYGSRGGF.NDNF..R..NF..NRGG-..GFGNR.Y.--.D  238
D98   GGRGGPG.RAGGNRGNM.GGNY.NQN..GNWNNGG-NNW.NNR.GN--DN  255
D87   GGG.GGPRAGG.G..GDRGQ....W..QNRQNGGG-NWG..G..G.--F.  248

PRP   .G.TW.QPH......PH...W.QPH..GWG.PH.....-....THNQ.NK  101
C026  GQGGYGGDYGGGWGQQGGGGQG-GWGGPQQQQGGGGWGQQGGGGQGGWGG  296
HA1   .S.DGYNGF.NDGSNF....SY-NDF.NYNN.SSNFGPMK..NFG.RSS.  287
HB1   .Y..GP.GGNF.GSPGY...R.-.Y..GGPGY.NQ.G.YG..YDNY.G.N  294
XA1a  .YN.D.QLWWQPSLLGWNR.Y.-AGQ.GGYGA.Q..GYGG..Q.G.YG.N  287
D98   WGNNSF.GG...G.GY...NNS-WGNNNPWDN.N..GNFG...NNWNN..  304
D87   NS..NF.GGQ..GSGGWNQQG.-SG...WNN....NG.WN....G..Y..  297

PRP   .SKPKTNMKHMAGAAAA.AVV..L.GYML.SAMSRPMMHFGNDWEDRYYR  151
C026  PQQQQQGGWGGPQQGGGGGGWGGQGQQQGGWGGQSGAQQWAHAQGGNRNY  346
HA1   .YGGGGQYFAKPRNQ..Y..SSSSSSYGS.RRF                  320
HB1   YGSGNYNDF.NYN.QPSNY.PMKS.NFG.SRNMGGPYGGGNYGP..SGGS  344
XA1a  GGYDGYN.G.SGFS.S..NFGSSG.YNDF.NYNSQSSSNFGPMK...YGG  337
D98   NDFGGYQQNY.GGPQR...NFNNNRM.PYQG..GFK.GGGNQGNY.GN.Q  354
D87   GNSNGSW.GN.GGG.....FGNEYQ.SY..GPQRNSNFGNNRPAPYSQGG  347
```

Figure 5. Amino acid sequence comparison of the deduced C026 protein with hamster PrP^c (PRP) (3) and hnRNPs from different organisms belonging to the hnRNP family [the human A1 (HA1) (20) and B1 (HB1) (21), the *Xenopus* XA1a, (XA1a) (23), the *Drosophila* Hrb98DE, (D98) (24) and Hrb87F, (D87) (25)]. Amino acids that are identical between C026 and others are dotted. Gaps are indicated by dashes. Conserved residues within the RNA-binding regions (RNP1 and RNP2) are boxed. The amino acid numbers on the right refer to individual proteins. The hnRNPs are divided into two parts between the residues 195 and 196 of the C026 protein by a vertical line, the N-terminus and the C-terminus.

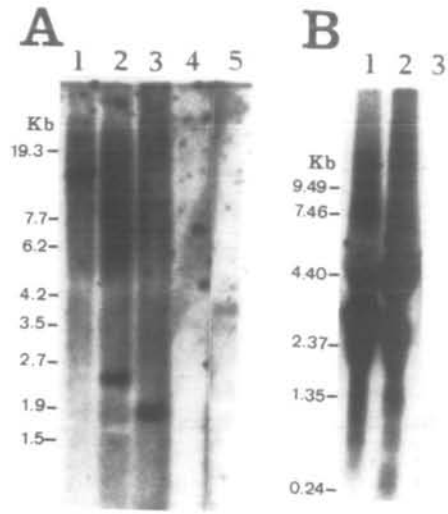

Figure 6. Southern (A) and Northern (A) analyses using the PrP-2 probe containing the 251−733 nt cDNA fragment of the HaPrP cDNA. A: Human (lane 1), mouse (lane 2), rat (lane 3), chicken (lane 4), and *C.elegans* (lane 5) genomic DNAs digested with *Eco*RI were used for Southern analysis using the [^{32}P]-labeled PrP-2 probe under the low stringency condition. In human, mouse and rat, expected bands 15, 2.3, and 1.9 kb, respectively, and unexpected faint bands or smears are seen. In chicken, no band is seen. One faint band, 3.6 kb, is only observed in *C.elegans*. The DNA molecular size markers (λ-*Eco*T14I digest) are shown on the left. B: Samples of total RNA (20 μg) from human (T98G) (lane 1), mouse (NIH3T3) (lane 2) and *C.elegans* (lane 3) were used for Northern analysis using the [^{32}P]-labeled PrP-2 probe under the low stringency condition. In T98G cells and NIH3T3 cells, expected bands 3.0 and 2.5 kb, respectively, and unexpected bands are seen. In *C.elegans*, only one faint band (1.0 kb) is observed. RNA molecular standards (0.24−9.5 kb RNA Ladder) are shown on the left.

Table 1. Statistical significance (Z value) of protein similarity scores between the C026 predicted product and hnRNPs and HaPrP.

Library sequence matched		Similarity score		Z value	
Protein	Species	I	O	I	O
HA1	Human	632	672	77.2	82.3
HB1	Human	623	684	76.0	83.8
XA1a	*Xenopus*	585	702	71.1	86.1
D98	*Drosophila*	526	655	63.5	80.1
D87	*Drosophila*	623	774	76.0	95.4
PRP	Hamster	81	139	6.4	13.9

The statistical significance of each comparison of the two sequences was estimated by the RDF program of Lipman and Pearson (30), which generates the Z value =[(similarity score − mean of random scores(=31.0)]/[standard deviation of random scores (=7.79)], (Ktup=1). I or O indicates the initial or optimized alignment score, respectively. A Z value > 4 is believed to be biologically significant. Other abbreviations for proteins are as described in legend for Fig. 5.

was able to hybridize one of the two bands in *C.elegans* genomic DNA, which was shown to be strongly reacted by the HaPrP probe in Southern analysis. Moreover, this cDNA also reacted with one of the four bands in *C.elegans* mRNA, which were shown to be reacted by the HaPrP probe in Northern analysis. From the evaluation in Southern and Northern analyses, we conclude that this *C.elegans* cDNA obtained here is derived from the sequence previously reacted with the mammalian PrP probe.

A comparison of the nucleotide and amino acid sequences between the cDNA clone and the HaPrP cDNA was carried out. The maximum homology of the compatible sequence at the nucleotide level is 52%. There are almost identical, continuous sequence repeats of 19 bases in their sequences (see Fig. 3). The cDNA clone appears to be due to hybridization of this very GC-rich region of the HaPrP cDNA. Westaway and Prusiner described that their HaPrP cDNA probe containing the 5′*Bam*HI-*Taq*I cDNA fragment (1−909 nt (5)) hybridized to the two bands, which seem to be the same bands seen by us from Southern analysis in *C.elegans* DNA (Fig. 1A, lane 5). To clarify the region responsible for this hybridization reaction, we removed the 5′ region including these repeats and the 3′ non-coding region from the HaPrP cDNA and used it as a new HaPrP probe, designated PrP-2, which contains the 251−733 nt cDNA fragment of the HaPrP cDNA, for the Southern and Northern analyses. As shown in Fig. 6A, the PrP-2 probe did not hybridize to the previous two bands which reacted with the HaPrP probe in Southern analysis shown in Fig. 1A, but weakly hybridized

to a new 3.6 kb band instead. Moreover, in Northern analysis (Fig. 6B), the PrP-2 probe did not hybridize to the previous four bands which reacted with the HaPrP probe shown in Fig. 2A, but weakly hybridized to a new 1.0 kb band instead. These data indicate that the sequence at the 5′ half region from 1 to 270 nt in the HaPrP cDNA is responsible for hybridizing between the *C.elegans* sequences and the HaPrP cDNA under the low stringency condition. Therefore, the continuous repeats are involved in this 5′ half region of the HaPrP cDNA and probably resulted in the cross-hybridization between them.

Although the C026 clone isolated from a *C.elegans* cDNA library using the C21 probe under the high stringency condition has a full length (346 amino acids) of deduced protein sequence, it shows quite low homology with the mammalian PrP^c; whereas, it shows quite high homology with vertebrate A and B hnRNPs as well as invertebrate hnRNPs (Fig. 5). The statistical evaluation (Table 1) indicates that this deduced 346-amino acid C026 gene product displays extremely more significant similarity to various members of the hnRNP family than to the mammalian PrP gene product. In eukaryotic cells, nascent RNA transcribed by RNA polymerase II is present in the nucleus as RNA-protein complexes, termed the hnRNP complexes. Considerable experimental evidence points to a role of hnRNP complexes in the post transcriptional processing of hnRNA and particularly in the splicing reaction (20, 24). Proteins A and B are members of two related families of basic proteins that share common antigenic determinants and extensive homologies in the primary structure. The vertebrate A and B hnRNPs and the invertebrate hnRNPs may form a distinct subfamily within the larger family of related RNA binding proteins. So far, from *C.elegans*, genomic or cDNA clones for the sequences related to hnRNPs are not reported. The hnRNPs have consensus sequences, RNP1 and RNP2, for RNA-binding domains at the N-terminus. Identical repeats are also found in the N-terminus of the deduced amino acid sequence of the C026 clone. This region of the C026 protein is considered to be an RNA binding domain. The sequences of the C026 protein and the hnRNPs are highly homologous at the N-terminus rather than at the C-terminus. By comparing the amino acid composition of the first 195 amino acids in the C026 protein with those of hnRNPs, it is obvious that the C-terminal half of this protein (residues 196−346) represents an extremely glycine-rich (48%) domain that is quite analogous to that (38−52% glycine) in the C-terminal halves of the hnRNPs (Fig. 5). While the exact sequences are poorly conserved between the C-terminal halves of the different hnRNPs, the compositions are very similar; this concept is also seen between the C026 protein and the hnRNPs. In contrast to this C026 protein and hnRNPs, the N-terminus of the mammalian PrP^c protein (approximately residues 29−94) represents the analogous glycine-rich (approximately 47%) domain (3, 6−12). A striking feature of the mammalian PrP^c protein is that 13% of the residues are in four direct tandem repeats of eight amino acids, i.e., the octapeptide WGQPHGGS/G, from positions 57 to 88. In addition, remnants of this sequence are found on either side of the four repeats. The suggested secondary structure of these repeats is composed exclusively of random coils and β-turns (6). Analysis of the primary structure of the glycine-rich domain of hnRNP revealed that it also contains 16 repeating oligopeptide units and is structurally flexible due to the high content of residues with strong β-turn and random coil prediction (22). However, the consensus sequence in the repeats is absolutely different from that of the previous repeats seen in the mammalian PrP^c protein. Taken together, these findings strongly suggest that this cDNA isolated here was derived from an hnRNP gene homologue but not from the predicted PrP gene homologue in *C.elegans*. This suggestion raises the possibility that the sequences detected in other invertebrates with the mammalian PrP cDNA probe arise from hnRNP genes but not from authentic PrP genes. From the unpublished results by Westaway and Mirenda (27), they failed to clone a homologous PrP gene from *Drosophila* but the clones obtained are predicted to encode threonine-rich proteins unrelated to PrP, which also support this possibility.

The other sequences in *C.elegans*, which correspond to the 3.0 kb band and the three 2.2, 1.4 and 0.7 kb transcripts loosely reacted with the mammalian PrP cDNA probe in Southern (Fig. 1A) and Northern analyses (Fig. 2A), respectively, are considered to differ from an authentic PrP gene, because the sequences did not hybridize with the PrP-2 probe as previously described. However, it would be interesting to examine whether these sequences are related to those of the hnRNPs. Moreover, the other sequences in *C.elegans*, which correspond to the 3.6 kb band and the 1.0 kb band weakly reacted with the PrP-2 probe in Southern and Northern analyses (Fig. 6), respectively, are also considered to differ from an authentic PrP gene. This is why the sequences did not hybridize with the HaPrP probe. Taken together, our findings raise the possibility that animals lower than vertebrates, at least nematode, might not possess authentic PrP gene homologues. The following evidences may support this assumption: the unexpectedly low homology (33%) between the murine PrP^c and the chicken prion-like protein recently cloned (13), i.e., this is extraordinarily lower than the homology between the human ubiquitously expressed protein Oct-1 and its counterpart in chicken (an overall 96% amino-acid sequence identity) (28), and the failure to detect PrP sequences in *Xenopus* (5). We think that PrP genes might be unique in vertebrates, especially mammals. This possibility has already been suggested by Oesch et al. (27) based on unpublished results by Westaway and Mirenda previously described. A recent report (29) that mice engineered to lack the gene for PrP^c show no detectable abnormalities in behavior or development up to at least seven months of age, suggesting a possibility that the function of the missing PrP^c within the cell is assumed by related or different protein(s), or that the function is redundant. Hence, we cannot rule out the possibility that the sequences in vertebrates as well as in *C.elegans*, which weakly reacted with the PrP-2 probe (Fig. 6), are predicted to encode proteins not exactly but partially related to PrP^c.

Note added

While this paper was under review, we found a report that, in a survey of expressed genes in *C.elegans*, a cDNA clone corresponds to hnRNP (31).

ACKNOWLEDGMENTS

We thank Drs. Bruno Oesch (Univ. California, USA) and Yasumi Ohshima (Kyushu Univ., Japan) for providing the pHaPrP clone and the N2 strain of *C.elegans*, respectively, and Drs. Masayoshi Imagawa (Tokyo Univ., Japan) and Hachiro Yasuda (Toyohashi Univ., Japan) for the *C.elegans* cDNA libraries. We also thank Dr. Y.Hinuma, the Director of the Shionogi Institute for Medical Science, for his constant support and helpful discussions.

REFERENCES

1. Weissmann, C. (1991) Nature **352**, 679–683.
2. Prusiner, S. B. (1982) Science **216**, 136–144.
3. Oesch, B., Westaway, D., Walchli, M., McKinley, M.P., Kent, S.B.H., Aebersold, R., Barry, R.A., Tempst, P., Teplow, D.B., Hood, L.E., Prusiner, S.B. and Weissmann, C. (1985) Cell **40**, 735–746.
4. Chesebro, B., Race, R., Wehrly, K., Nishio, J., Bloom, M., David, L., Bergstrom, S., Robbins, K., Mayer, L., Keith, J. M. Garon, C. and Haase, A. (1985) Nature **315**, 331–333.
5. Westaway, D. and Prusiner, S. B. (1986) Nucleic Acids Res. **14**, 2035–2044.
6. Locht, C., Chesebro, B., Race, R. and Keith, J.M. (1986) Proc. Natl. Acad. Sci. USA **83**, 6372–6376.
7. Westaway, D., Goodman, P.A., Mirenda, C.A., McKinley, M.P., Carlson, G A. and Prusiner, S.B. (1987) Cell **51**, 651–662.
8. Liao, Y-C., Tokes, Z , Lim, E., Lackey, A., Woo, C.H., Button, J.D. and Clawson, G.A. (1987) Lab. Invest. **57**, 370–374.
9. Goldmann, W., Hunter, N , Foster, J.D., Salbaum, J.M., Beyreuther, K. and Hope, J. (1990) Proc. Natl. Acad. Sci. USA **87**, 2476–2480.
10. Goldmann, W., Hunter, N., Martin, T., Dawson, M. and Hope, J. (1991) J. Gen. Virol. **72**, 201–204.
11. Kretzschmar, H.A., Stowring, L.E., Westaway, D., Stubblebine, W.H., Prusiner, S.B. and Dearmond, S.J. (1986) DNA **5**, 315–324.
12. Liao, Y.-C.J., Lebo, R.V., Clawson, G.A. and Smuckler, E.A. (1986) Science **233**, 364–367.
13. Harris, D.A., Falls, D.L., Johnson, F.A. and Fischbach, G.D. (1991) Proc. Natl. Acad. Sci. USA **88**, 7664–7668.
14. Roberts, L. (1990) Science **248**, 1310–1313.
15. Brenner, S (1974) Genetics **77**, 71–94.
16. Sulston, J. E. and Brenner, S. (1974) Genetics **77**, 95–104.
17. Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, pp. 9.16–19.
18. Sanger, F., Nicklen, S. and Coulsen, A. R. (1977) Proc. Natl. Acad. Sci. USA **74**, 5463–5467.
19. Basler, K., Oesch, B , Scott, M., Westaway, D., Walchli, M., Groth, D.F., MiKinley, M.P., Prusiner, S.B. and Weissmann, C. (1986) Cell **46**, 417–428.
20. Buvoli, M., Boamonti, G., Tsoulfas, P., Bassi, M. T., Ghetti, A., Riva, S. and Morandi, C. (1988) Nucleic asids Res. **16**, 3751–3770.
21. Burd, C. G., Swanson, M. S., Gorlach, M. and Dreyfuss, G. (1989) Proc. Natl. Acad. Sci. USA **86**, 9788–9792.
22. Cobianchi, F., SenGupta, D. N., Zmudzka, B. Z. and Wilson, S. H. (1986) J. Bio. Chem. **261**, 3536–3543.
23. Kay, B. K., Sawhney, R. K. and Wilson, S. H. (1990) Proc. Natl. Acad. Sci. USA **87**, 1367–1371.
24. Haynes, S. R., Rebbert, M. L., Mozer, B. A., Forquignon, F. and Dawid, I. B. (1987) Proc. Natl. Acad. Sci. USA **84**, 1819–1823.
25. Haynes, S. R., Johnson, D., Raychaudhuri, G. and Beyer, A. L. (1991) Nucleic Acids Res. **19**, 25–31.
26. Ball, E. E., Rehm, E. J. and Goodman, C. S. (1991) Nucleic Acids Res. **19**, 397.
27. Oesch, B., Westaway, D. and Prusiner, S.B. (1991) Current Topics in Microbiology and Immunology (Ed. B. Chesebro) **172**, 109–124.
28. Petryniak, B., Staudt, L.M., Postema, C.E., McCormack, W.T. and Thompson, C.B. (1990) Proc. Natl. Acad. Sci. USA **87**, 1099–1103.
29. Kozak, M. (1989) J. Cell Biol. **108**, 229–241.
30. Lipman, D.J. and Pearson, W.R. (1985) Science **227**, 1435–1441.
31. Waterston, R., Martin, C., Craxton, M., Huynh, C., Coulson, A., Hillier, L., Durbin, R., Green, P., Shownkeen, R., Halloran, N., Metzstein, M., Hawkins, T., Wilson, R., Berks, M., Du, Z., Thomas, K., Thierry-Mieg, J. and Sulston, J. (1992) nature genetics **1**, 114–123.