FEBS Letters 508 (2001) 170–174 FEBS 25419

Molecular cloning of the cDNA coding for *Xenopus laevis* prion protein

Bice Strumbo[a], Severino Ronchi[a], Liana C. Bolis[b], Tatjana Simonic[a,*]

[a] *Dipartimento di Patologia Animale, Igiene e Sanità Pubblica Veterinaria, Sezione di Biochimica e Fisiologia Veterinaria, Università di Milano, via Celoria 10, 20133 Milan, Italy*

[b] *Dipartimento di Scienze Farmacologiche, Laboratorio di Biologia Comparata, Università di Milano, via Balzaretti 9, 20133 Milan, Italy*

Received 1 August 2001; revised 9 October 2001; accepted 9 October 2001

First published online 26 October 2001

Edited by Takashi Gojobori

Abstract Isolation and characterization of the cDNA coding for the 216-residue *Xenopus laevis* prion protein is reported. Existence of this protein in amphibians was suggested by an EST fragment (accession number BG813008), while a conclusive demonstration is presented here. This protein exhibits a higher identity level to avian and turtle prion (more than 44%) than to mammalian prion (about 28%). Although most of the structural motifs common to known prion proteins are conserved in *X. laevis*, the lack of repeats represents a substantial difference. Other features worth noting are the presence of not perfectly conserved hydrophobic stretch, which is considered the prion signature, as well as the complete absence of histidine residues.

Key words: Prion protein; cDNA cloning; Reverse transcription-polymerase chain reaction; *Xenopus laevis*

1. Introduction

Prion protein is a normal cellular protein (PrP^c) exhibiting unique properties, due to its ability of adopting, in particular conditions, an abnormal conformation. This conformational shift changes a physiological protein (PrP^c) to a pathological protein isoform, liable for several neurodegenerative diseases, called transmissible spongiform encephalopathies, described up to now in mammalian species [1–4]. The well known epidemic appearance of such a disease in cattle (bovine spongiform encephalopathy) some years ago and its likely transmission to humans, by feeding infected meals, disclosed new potential problems in food safety [5]. Consequently, also non-mammalian species and especially those of large food-consumption became of interest. Among eatable vertebrate species, besides mammals and birds, fish occupy the most important position and, moreover, for a number of species, fish-breedings have been developed. Therefore, obvious questions have been raised about fish PrP existence as well as infectivity considerations. No sequence similar to mammalian PrP has been found until now in any fish species.

PrP has been well characterized in a number of mammalian species, showing a high degree of primary structure identity and, in some cases, three-dimensional structure has been resolved by nuclear magnetic resonance (NMR) [6]. The first non-mammalian PrP was identified in 1991, by chance, in chicken. This protein conserves all structural motifs present in mammalian PrP, despite a low identity level of the primary structure [7,8]. As far as structural features are concerned, similar comments fit even for turtle PrP sequence, identified several years later [9], representing the first example of such protein in reptiles, which precedes birds and mammals from an evolutionary point of view. During the past 2 years, a number of novel bird PrP cDNAs have been reported, confirming high identity levels among deduced amino acid sequences inside the same zoological class [10]. On the contrary, no PrP sequences have been described in reptile species other than the turtle.

Although it is apparent that troubles are encountered in finding more primitive PrPs, the importance of finding them in lower vertebrates goes beyond a speculative point of view.

Very recently a *Xenopus laevis* expressed sequence appeared in EST database, corresponding to a portion of the first amphibian PrP (GenBank/EMBL/DDBJ accession number BG813008) (Washington University *Xenopus* EST Project). Again, in addition to chicken PrP, an important progress came from a search not specifically addressed.

In this paper the whole cDNA sequence of *X. laevis* PrP is reported. Moreover, both the nucleotide and the deduced amino acid sequences are analyzed in comparison with known PrPs.

2. Materials and methods

2.1. Reverse transcription (RT)-polymerase chain reaction (PCR) and PCR assays

Oligonucleotides used for PCR assays and for sequencing were purchased by MWG Biotech, Germany, and are listed in Table 1.

Total RNA was extracted from *X. laevis* liver using the Chomczynski and Sacchi method [11]. 3 μg of total RNA were reverse transcribed using the SuperScript II Reverse Transcriptase (Life Technologies, UK) and different primers, depending on experimental purposes. PCRs were carried out using Platinum Taq DNA polymerase (Life Technologies, UK) in a total volume of 50 μl containing 1× buffer (20 mM Tris–HCl, pH 8.4; 50 mM KCl), 1.5 mM $MgCl_2$, 200 μM dNTPs and 0.4 μM of each primer. All PCR assays were performed following the producer protocol with an annealing temperature of 5°C below the minimum melting temperature and an elongation time ranging from 30 s to 2 min depending on the length of the expected fragment.

The 3′-end of cDNA was isolated using 3′-RACE as described previously [12]. Genomic DNA was extracted from liver of *X. laevis* using a standard method [13]. PCR assays were carried out on 200 ng of DNA template as described above.

2.2. DNA sequence analysis

Sequence analysis of the DNA fragments was performed on both strands directly on purified PCR products by the Taq dye-deoxy ter-

*Corresponding author. Fax: (39)-2-5835 8123.
E-mail address: tatjana.simonic@unimi.it (T. Simonic).

PII: S0014-5793(01)03027-7

Table 1
Oligonucleotides used as primers in PCR assays

Primer	Sequence
EF1	^{13}CTCTGGACTTGTTTAGTCC^{31}
EF2	$^{-99}$CACAGAGGCACAGCACC^{-83}
XF1	^{268}GGTGCTATTGGAGGCTAC^{285}
XF2	^{356}ACTATAACCAGATGCC^{374}
ER1	^{285}GTAGCCTCCAATAGCACC^{268}
XR1	^{457}CTGACATATTGTAGCAGTCC^{438}
XR2	^{726}CGGAACAGTTCAGAATCGG^{708}
FR	TGACTAAGCTTGTCGACG
AdT	CAGGAAACAGCTATGACTAAGCTTGTCGACGT$_{(17)}$

Numbering refers to the cDNA sequence shown in Fig. 1. E prefix indicates primers designed on the EST fragment; X prefix indicates primers designed on our sequence. F, forward; R, reverse; A, anchor.

minator method, using an automated 3100 DNA sequencer (Applera, USA).

The cDNA sequence of *X. laevis* PrP has been deposited in the EMBL database under accession number AJ320186.

2.3. Computer assisted analysis

DNA sequences were analyzed using MT Navigator PPC software (release 1.0.2b3) and alignments were performed with PC-GENE (release 15) and Omiga (release 1.0) software.

Secondary and tertiary structure predictions were performed using the program ProModII at the SWISS-MODEL Automated Protein Modelling Server (http://www.expasy.ch/swissmod/swiss-model.html).

3. Results and discussion

A *X. laevis* expressed sequence (GenBank/EMBL/DDBJ BG813008) was deposited very recently in EST database. This DNA fragment, spanning 538 bp, comprises 151 bp of 5′-UTR and 387 bp of the coding sequence. The encoded peptide (129 amino acids) exhibits some features of known PrPs, for this reason it has been identified as partial sequence belonging to PrP. The above sequence was the starter of this work, aimed to obtain the whole cDNA coding region.

Existence of PrP in amphibians has therefore been suggested by this *X. laevis* EST fragment, while a conclusive demonstration is presented here.

3.1. Isolation and analysis of the whole coding sequence

As first control for the presence of PrP coding sequence in total RNA extracted from liver, using EF1/ER1 primer couple, a 273-bp fragment was amplified, corresponding to positions 13–285 of the cDNA sequence shown in Fig. 1 which will always be used as reference for numbering. This cDNA was completed by 3′-RACE using semi-nested PCR: a same 3′-primer (FR), annealing to the anchor sequence (AdT), was coupled in two consecutive PCR assays with XF1 and then XF2 specific forward oligonucleotides. This procedure led to a product of 412 bp which overlapped the 3′-end of the EST DNA fragment with its 5′-end and comprised the lacking part of the coding sequence (264 bp) as well as a short 3′-UTR. This cDNA segment corresponds to positions 356–748. A further PCR assay was aimed to: (i) encompass the sequence lying between the 3′-portion of the first 273-bp fragment and the 5′-portion of the novel part of cDNA; and (ii) extend our sequence towards the 5′-end of cDNA. For this purpose a forward primer (EF2) was designed within 5′-UTR, on the basis of the EST sequence, while a reverse primer (XR1) was located within the 3′-region of the coding sequence newly isolated. A 556-bp DNA fragment was obtained, corresponding to positions −99 to +457.

Some features of the 847-nt cDNA, reported in Fig. 1, need to be analyzed. The 5′-UTR exhibits two consecutive ATG at its 3′-end. The first ATG is likely to be part of the 3′-end of 5′-UTR, because it is preceded by three pyrimidines (TTT), as it is known that a pyrimidine at the −3 position is an unfavorable context for translation initiation. Consequently, it has to be supposed that the second ATG with a purine at the same −3 position represents the true, or at least the main translation start codon, considering also that the presence of A and C, immediately after the two ATGs, have the same value for starting translation [14,15]. Furthermore, when compared to the EST fragment, our 5′-UTR sequence exhibits a 19-nt insertion (positions −31 to −14) which represents a perfect direct repeat of an upstream sequence (positions −50 to −33), separated by only one nucleotide. Within the coding region, alternative presence of G or A has been found at position 43, depending on individuals, leading to a codon change. The presence of a very short 3′-UTR (79 nt) is surprising, since, in a number of reported mRNAs coding for PrPs, this region is usually rather long. A putative poly-adenylation (polyA) signal TATAAA [16] is present at positions 702–707, followed, after 23 nt, by 18 adenylates which could represent a partial polyA tail. Nevertheless, it cannot be excluded that the here reported *X. laevis* cDNA is an alternative transcript, generated by the use of the above polyA signal. Several attempts of extending 3′-UTR have been unsuccessful.

A PCR assay, carried out on genomic DNA using EF1/XR2 primer couple, gave a product spanning the same size exhibited by the corresponding cDNA and subsequent sequence analysis confirmed that this 714-bp fragment was not

```
                                  CACAGAGGCACAGCACCAGGACCTGACACCCACATAGCT   -61
TCTCTTTGGCACACTCTATACCCTCACCAACACTCTATACCCTCACCCAGGTTGTTTATG    -1
ATGCCACAAAGTCTCTGGACTTGTTTAGTCCTTATCTCCCTAATATGCACATTGACTGTA    60
 M  P  Q  S  L  W  T  C  L  V  L  I  S  L  I  C  T  L  T  V     20

TCTTCCAAGAAGAGCGGTGGTGGGAAAAGTAAAACTGGAGGATGGAACACAGGGAGCAAC   120
 S  S  K  K  S  G  G  G  K  S  K  T  G  G  W  N  T  G  S  N     40

CGGAACCCCAACTACCCAGGAGGCTACCCAGGGAATACTGGAGGCAGCTGGGGGCAACAA   180
 R  N  P  N  Y  P  G  G  Y  P  G  N  T  G  G  S  W  G  Q  Q     60

CCTTATAATCCTAGCGGTTATAACAAGCAATGGAAACCTCCCAAGTCCAAAACCAACATG   240
 P  Y  N  P  S  G  Y  N  K  Q  W  K  P  P  K  S  K  T  N  M     80

AAGTCGGTGGCCATAGGCGCTGCTGCTGGTGCTATTGGAGGCTACATGCTCGGTAATGCA   300
 K  S  V  A  I  G  A  A  A  G  A  I  G  G  Y  M  L  G  N  A     100

GTGGGTCGTATGAGTTATCAATTCAACAATCCCATGGAGTCCCGTTATTATAACGACTAC   360
 V  G  R  M  S  Y  Q  F  N  N  P  M  E  S  R  Y  Y  N  D  Y     120

TATAACCAGATGCCAAATCGCGTTTACAGGCCTATGTACAGAGGAGAGGAGTACGTGTCA   420
 Y  N  Q  M  P  N  R  V  Y  R  P  M  Y  R  G  E  E  Y  V  S     140

GAGGACAGGTTCGTGAGGGACTGCTACAATATGTCAGTGACAGAGTACATCATAAAGCCG   480
 E  D  R  F  V  R  D  C  Y  N  M  S  V  T  E  Y  I  I  K  P     160

ACTGAAGGAAAGAACAACAGCGAGCTAAACCAGTTGGATACCACGGTAAAGTCCCAAATT   540
 T  E  G  K  N  N  S  E  L  N  Q  L  D  T  T  V  K  S  Q  I     180

ATTCGCGAGATGTGCATCACCGAGTACAGGAGAGGATCGGGATTCAAAGTGCTCTCTAAC   600
 I  R  E  M  C  I  T  E  Y  R  R  G  S  G  F  K  V  L  S  N     200

CCTTGGCTGATCCTTACTATCACTCTCTTTGTTTACTTTGTGATAGAGTGATCAAAGGAA   660
 P  W  L  I  L  T  I  T  L  F  V  Y  F  V  I  E  *              216

ATATTAATAAAAAGGCCAAATGTATGTATATATAGAGAGAGTATAAACCGATTCTGAACT   720
GTTCCGTCTCAAAAAAAAAAAAAAAAA                                    748
```

Fig. 1. Nucleotide and deduced amino acid sequences of the cDNA coding for *X. laevis* PrP. The first base of the translation start codon is designated +1, the stop codon (TGA) is indicated by an asterisk and the putative poly-adenylation signal (TATAAA) is bold. Ile^{15} and the first base of its codon (black boxed) correspond to the polymorphic site, where, following change of A to G, Val can be alternatively found.

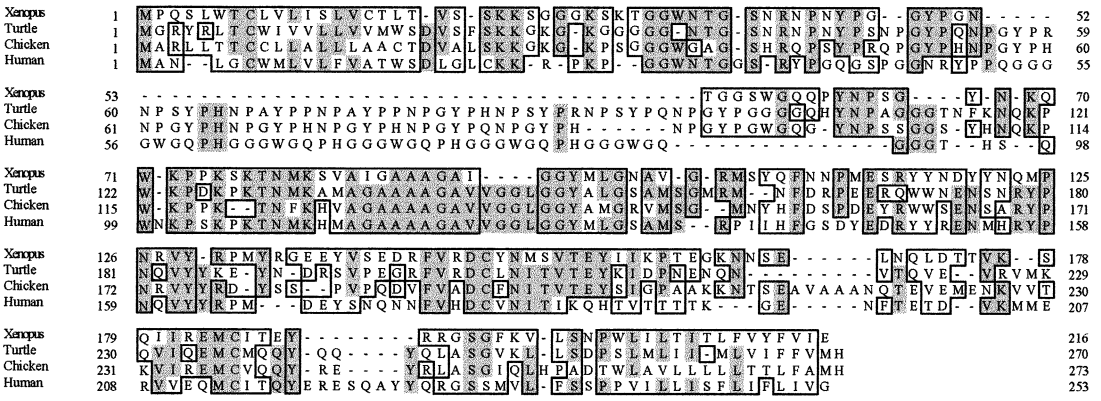

Fig. 2. Comparison of the *X. laevis* prion amino acid sequence with those of representative members of higher vertebrate classes. Conserved amino acids are boxed; gray denotes identical residues.

interrupted by intervening sequences. This preliminary result suggests, according to known PrP gene structures [17] that the coding region and 3′-UTR are part of the same exon. Unfortunately, up to now we have not been able to further extend the sequence at 5′, nevertheless, experiments are in progress to verify the entire gene organization.

3.2. Analysis of the deduced amino acid sequence

A preliminary analysis of some structural features, exhibited by the amino acid sequence deduced from *X. laevis* cDNA (Fig. 1), is useful to allow more documented comments.

In our opinion, as remarked above, precursor protein primary structure starts with Met, corresponding to the second ATG within cDNA sequence. At this N-terminal end, a sequence of 22 amino acids is consistent with an endoplasmic reticulum signal peptide, where the above mentioned presence of G or A in the nucleotide sequence involves a substitution, Ile or Val, at position 15. Occurrence of this conservative change within the signal peptide, lost in the mature protein, suggests that it is likely a polymorphic site. This peptide is followed by a basic region (four lysines out of nine residues) that would represent the N-terminal end of the mature protein. The main difference in comparison with all known PrPs is a complete lack of repeats in this amphibian protein primary structure. Nevertheless, sequences corresponding to re-

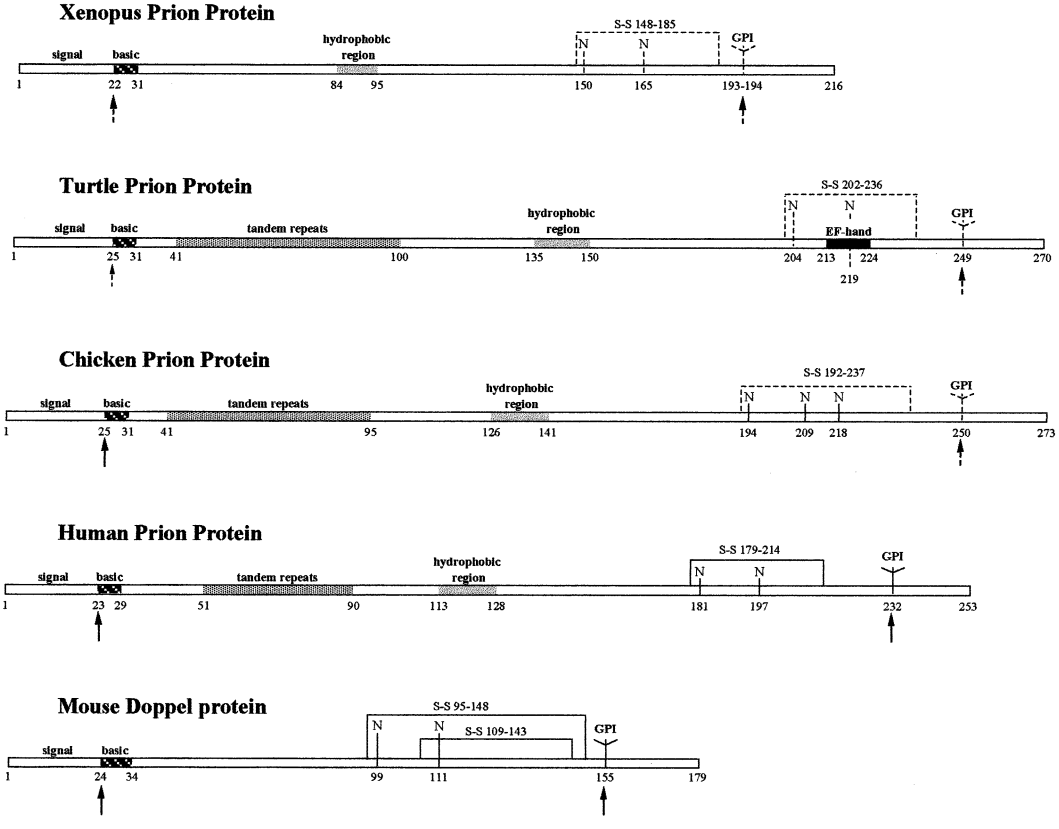

Fig. 3. Schematic representation of relevant prion and Doppel protein features. Numbers refer to positions in amino acid sequences. Boldface arrows indicate sites of proteolytic cleavage; disulfide bridges are marked by a line connecting the two involved Cys. N indicates glycosylation sites; GPI indicates GPI anchor. Dotted lines indicate putative sites.

gions immediately before and after repeats are present also in *X. laevis* or, better, show considerable conservation, especially when compared to turtle and avian PrP (Fig. 2). A second significant difference concerns the hydrophobic stretch, considered a PrP signature, which is not perfectly conserved in *X. laevis*, where, along a sequence of 16 residues, identical in all known PrPs, there is a deletion of four consecutive amino acids and some substitutions (Fig. 2).

The C-terminal portion retains all relevant features present in other PrPs: two cysteines (positions 148 and 185) corresponding to the same residues, conserved in all species, forming usually an intramolecular disulfide bond; two putative *N*-glycosylation sites, the first (Asn^{150}) corresponds exactly to the first modification site in all PrPs, the second (Asn^{165}), due to surrounding context, appears to be equivalent to the extra glycosylation site present in avian PrPs rather than to the second common site [8,18]. Finally, a peptide of 22–23 amino acids at C-terminal end might be lost in the mature protein because of the predicted presence of a glycosyl phosphatidylinositol (GPI) anchor site, likely at position 193 or 194, where a dipeptide Ser–Gly is present, corresponding to the same dipeptide in turtle and chicken. In this last case, Gly^{249} is thought to be the probable target for GPI addition [8]. A schematic representation of PrP structural motifs, compared among different vertebrate classes, is shown in Fig. 3.

3.3. Some comments on X. laevis PrP

The *X. laevis* protein comprises 216 residues, corresponding to a calculated molecular mass of 24 344 Da, and is therefore 15–20% shorter with respect to known PrPs, mainly due to lack of repeats region. On the whole, it shows a higher degree of identity to avian and turtle (more than 44%) than to mammalian PrPs (about 28%). As far as Doppel is concerned [19], identity level is very low (12–13%), supporting the opinion that this *X. laevis* protein is the amphibian version of PrP, in spite of the fact that absence of repeats and presence of an imperfect hydrophobic stretch (PrP signature) suggest that it could be lying between prion and Doppel.

Observations that this novel PrP is completely devoid of histidine residues, as well as that His content of PrP decreases with lowering of evolutionary level, are noteworthy, but of unknown meaning. Histidines represent about 4% of amino acid residues in mammalian and avian PrP, decrease to 1.5% in turtle and are absent in *X. laevis*. Presence of these residues in the repeats region of mammalian PrP has been reported to be responsible of Cu^{2+} binding and this behavior has been considered the basis for some functional hypotheses [20,21]. On the contrary, such ability seems rather questionable for avian and turtle PrP, suggesting that copper binding could not be the primary function of the PrP [9,22]. Absence of both repeats and histidines in *X. laevis* protein could support the above hypothesis. A second role for His, in particular for the His preceding the hydrophobic stretch (corresponding to His^{111} of the human sequence in Fig. 2) and conserved both in mammals and birds, as crucial residue involved in conformational changes of PrP, has been reported [23]. Moreover, very recent studies, carried out using the prion peptide PrP 106–126 (following numbering of human protein), showed that Cu^{2+} and/or Zn^{2+} binding to the peptide is critical for its aggregation and neurotoxicity, and that His^{111} is directly involved in this binding [24]. Since the corresponding His is present neither in turtle, nor in *X. laevis* PrP, questions arise about PrP conformational behavior and neurotoxicity in these lower vertebrates. The predicted three-dimensional structure of the *X. laevis* PrP C-terminal region (Fig. 4) was computed by a comparative protein modelling software [25]. Unlike the two short antiparallel β-strands, which are absent in this novel PrP, the three α-helices, resolved by NMR in PrP of some mammalian species [6], appear to be well conserved in the amphibian protein. However, three residues downstream the first α-helix (H1), a couple of amino acids are predicted to retain the same secondary structure (H1a), confirming that this region has remarkable tendency to assume this structure.

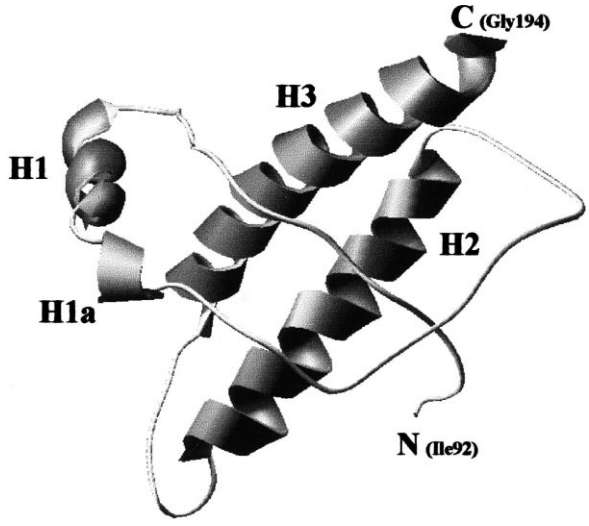

Fig. 4. Ribbon diagram of the predicted secondary and tertiary structures of the C-terminal region (residues 92–194) of *X. laevis* PrP. The three α-helices (H1, H2 and H3) as well as the extra short tract (H1a), downstream of the first one, are indicated. This structure was produced using the SWISS-MODEL Automated Protein Modelling Server and is based upon the coordinates of human (1QLZA.pdb), bovine (1DX0A.pdb), syrian hamster (1B10A.pdb) and mouse (1AG2.pdb) PrPs. The diagram was produced using the SwissPDB 3.7b2 software.

Isolation and characterization of the first whole coding sequence for an amphibian PrP is here reported. Although further work, already in progress, is needed to verify mRNA size, gene structure as well as the possible presence of prion-like proteins, analysis of both the *X. laevis* cDNA sequence and the deduced protein primary structure led to a further approach towards PrP origin. This result gives new support to hypothesis that ancestors of PrP have to be found also in the most primitive vertebrates.

Acknowledgements: This work was supported by Ministero dell'Università e della Ricerca Scientifica e Tecnologica (MURST 60%). We thank Lorenzo Sangiorgio for his valuable technical support.

References

[1] Prusiner, S.B. (1982) Science 216, 136–144.
[2] Prusiner, S.B. (1991) Science 252, 1515–1522.
[3] Prusiner, S.B. (1997) Science 278, 245–251.
[4] Collinge, J. (2001) Annu. Rev. Neurosci. 24, 519–550.
[5] Lasmézas, C.I., Fournier, J.G., Nouvel, V., Boe, H., Marcé, D., Lamoury, F., Kopp, N., Hauw, J.J., Ironside, J., Bruce, M., Dormont, D. and Deslys, J.P. (2001) Proc. Natl. Acad. Sci. USA 98, 4142–4147.
[6] Wüthrich, K. and Riek, R. (2001) Adv. Protein Chem. 57, 55–82.

[7] Harris, D.A., Falls, D.L., Johnson, F.A. and Fishbach, G.D. (1991) Proc. Natl. Acad. Sci. USA 88, 7664–7668.
[8] Gabriel, J.M., Oesch, B., Kretzschmar, H., Scott, M. and Prusiner, S.B. (1992) Proc. Natl. Acad. Sci. USA 89, 9097–9101.
[9] Simonic, T., Duga, S., Strumbo, B., Asselta, R., Ceciliani, F. and Ronchi, S. (2000) FEBS Lett. 469, 33–38.
[10] Wopfner, F., Weidenhöfer, G., Schneider, R., von Brunn, A., Glich, S., Schwarz, T.F., Werner, T. and Schätzl, H.M. (1999) J. Mol. Biol. 289, 1163–1178.
[11] Chomczynski, P. and Sacchi, N. (1987) Anal. Biochem. 162, 156–159.
[12] Simonic, T., Duga, S., Negri, A., Tedeschi, G., Malcovati, M., Tenchini, M.L. and Ronchi, S. (1997) Biochem. J. 322, 729–735.
[13] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd edn., Cold Spring Harbor Laboratory, Cold Spring, NY.
[14] Kozak, M. (1986) Cell 44, 283–292.
[15] Kozak, M. (1997) EMBO J. 16, 2482–2492.
[16] Guo, Z. and Sherman, F. (1995) Mol. Cell Biol. 15, 5983–5990.
[17] Lee, I.Y., Westaway, D., Smit, A.F.A., Wang, K., Seto, J., Chen, L., Acharya, C., Ankener, M., Baskin, D., Cooper, C., Yao, H., Prusiner, S.B. and Hood, L.E. (1998) Genome Res. 8, 1022–1037.
[18] Rudd, P.M., Wormald, M.R., Wing, D.R., Prusiner, S.B. and Dwek, R.A. (2001) Biochemistry 40, 3759–3766.
[19] Silverman, G.L., Qin, K., Moore, R.C., Yang, Y., Mastrangelo, P., Tremblay, P., Prusiner, S.B., Cohen, F.E. and Westaway, D. (2000) J. Biol. Chem. 275, 26834–26841.
[20] Shiraishi, N., Ohta, Y. and Nishikimi, M. (2000) Biochem. Biophys. Res. Commun. 267, 398–402.
[21] Wong, B.S., Vénien-Bryan, C., Williamson, R.A., Burton, D.R., Gambetti, P., Sy, M.S., Brown, D.R. and Jones, I.M. (2000) Biochem. Biophys. Res. Commun. 276, 1217–1224.
[22] Marcotte, E.M. and Eisenberg, D. (1999) Biochemistry 38, 667–676.
[23] Salmona, M., Malesani, P., De Gioia, L., Gorla, S., Bruschi, M., Molinari, A., Della Vedova, F., Pedrotti, B., Marrari, M.A., Awan, T., Bugiani, O. and Tagliavini, F. (1999) Biochem. J. 342, 207–214.
[24] Jobling, M.F., Huang, X., Stewart, L.R., Barnham, K.J., Curtain, C., Volitakis, I., Perugini, M., White, A.R., Cherny, R.A., Masters, C.L., Barrow, C.J., Collins, S.J., Bush, A.I. and Cappai, R. (2001) Biochemistry 40, 8073–8084.
[25] Guex, N., Diemand, A. and Peitsch, M.C. (1999) Trends Biochem. Sci. 24, 364–367.