Two alleles of a neural protein gene linked to scrapie in sheep

WILFRED GOLDMANN†, NORA HUNTER‡, JAMES D. FOSTER‡, J. MICHAEL SALBAUM*, KONRAD BEYREUTHER*, AND JAMES HOPE‡

*Center for Molecular Biology, University of Heidelberg, Im Neuenheimer Feld 282, D-6900 Heidelberg, Federal Republic of Germany; and ‡Institute for Animal Health, Agricultural and Food Research Council and Medical Research Council, Neuropathogenesis Unit, Ogston Building, West Mains Road, Edinburgh EH9 3JF, United Kingdom

Communicated by D. Carleton Gajdusek, January 4, 1990 (received for review May 1, 1989)

ABSTRACT Sheep are the natural hosts of the pathogens that cause scrapie, an infectious degenerative disease of the central nervous system. Scrapie-associated fibrils [and their major protein, prion protein (PrP)] accumulate in the brains of all species affected by scrapie and related diseases. PrP is encoded by a single gene that is linked to (and may be) the major gene controlling the incubation period of the various strains of scrapie pathogens. To investigate the role of PrP in natural scrapie, we have determined its gene structure and expression in the natural host. We have isolated two sheep genomic DNA clones that encode proteins of 256 amino acids with high homology to the PrPs of other species. Sheep PrPs have an arginine/glutamine polymorphism at position 171 that may be related to the alleles of the scrapie incubation-control gene in this species.

Sheep are the natural hosts of the pathogens that cause scrapie, an infectious degenerative disorder of the central nervous system. The disease is invariably fatal after incubation periods of several months to years and can persist within a flock by spread between flocks or by transmission from ewe to lamb (1). The incubation period of the disease depends primarily on the strain(s) of pathogen and the genotype of the affected animal (2). The expense and handling problems of farm animal experiments has meant that much of the work on the host genetic control and transmissibility of scrapie has been done in rodents. The mouse gene Sinc, with two alleles s7 and p7, is the major gene determining the incubation period of all strains of murine scrapie (2,4), and its sheep homologue Sip (with alleles sA and pA) also appears to control the incidence of natural scrapie in at least some breeds of sheep (5). Paradoxically, a candidate product of these host control genes was discovered in studies on the molecular structure of the scrapie pathogen.

During the subcellular fractionation of scrapie-affected brain, at least some infectivity copurifies with aggregates of a neuronal membrane protein, the prion protein (PrP) (6). These aggregates were first recognized by electron microscopy as structured scrapie-associated fibrils (7,8). They are found in brain extracts (9,10) of all species affected by scrapie and diseases caused by related pathogens, including Creutzfeldt-Jakob disease of man (11,12) and bovine spongiform encephalopathy of cattle (13,14). PrP is a protein of apparent molecular mass 33–35 kDa (15–17). It is glycosylated (18–20) and has two potential N-glycan attachment sites and an N-terminal glycine-rich repetitive sequence (15).

Restriction site polymorphisms in and around the murine PrP gene have been linked to the alleles of Sinc in segregating (21) and congenic (22) mice. In fact, there are two amino acid polymorphisms in PrP in mice with the s7 and p7 alleles (23) and this has led to the proposal that PrP may even be the Sinc gene product. The sheep is the only species other than the mouse in which an incubation time control gene has been defined by responses to experimental infection with scrapie (2), and a similar linkage of the ovine PrP gene to the alleles of Sip has been discovered (24). To investigate the link between PrP and the host gene controlling the incubation period of scrapie, we have sequenced sheep PrP clones, searching for amino acid substitutions that might provide the molecular basis for Sip allelism. In this paper, we report the cloning and sequencing of genomic DNA carrying most of the ovine PrP gene. Two genomic clones encoded predicted proteins of 256 amino acids that differed at codon 171 resulting in an arginine/glutamine polymorphism. These clones also differed in an EcoRI site outside the PrP coding region that has been linked to the alleles of Sip.

MATERIALS AND METHODS

RNA Preparation and Northern Blot Analysis. RNA was prepared from normal brain tissue of Cheviot sheep by the guanidinium isothiocyanate method (25). For purification of poly(A)† RNA, the preparation was followed by oligo(dT) chromatography (26). After electrophoresis in 2.2 M formamide/agarose gels, RNA was transferred to a nylon membrane and hybridized with the 32P-labeled hamster cDNA probe (pEA974 (27)) or the sheep probe (pScr23.4, this paper) in 5× SSPE/50% (vol/vol) formamide/5× Denhardt's solution/denatured fish milt DNA (0.1 mg/ml)/0.5% SDS (1× SSPE = 0.18 M NaCl/10 mM sodium phosphate, pH 7.4/1 mM EDTA; 5× Denhardt's solution = 0.02% polyvinylpyrrolidone/0.02% Ficoll/0.02% bovine serum albumin). Membranes were washed in 1× SSPE at 65°C and exposed to Kodak X-AR film with an intensifying screen at ~70°C (28).

Isolation and Characterization of Sheep PrP DNA Clones. A genomic DNA library was prepared from spleen DNA of a Suffolk sheep by using the λ phage EMBL3 (29). Phage plaques were screened for PrP-related DNA sequences by hybridization to a 470-base-pair Nco I–Sau3AI fragment of plasmid pHaPrP (15), labeled by random priming with [32P]dCTP to a specific activity of 4 × 108 cpm/μg of DNA (30) and hybridized to phage DNA in 5× SSPE/0.1% SDS/5× Denhardt's solution/denatured fish milt DNA (0.01 mg/ml; 108 cpm/ml) for at least 12 hr at 65°C. Final washes of the filters were at 42°C in 0.2× SSPE/0.1% SDS; except in the rescreening, the final wash was done at 60°C in 0.2× SSPE/0.1% SDS. DNA of hybridization-positive phages was amplified, prepared from clear lysates as described (28), and checked for authenticity by Southern blot analyses.

Cloning and Sequencing Procedures. Suitable restriction fragments of the genomic DNA inserts were subcloned into pUC19, M13mp18/mp19, or Bluescript pKS+ (Genofit, Heidelberg) vectors. To facilitate sequencing a set of ordered

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: PrP, prion protein.
†To whom reprint requests should be addressed at the ‡ address.
¶The sequence reported in this paper has been deposited in the GenBank data base (accession no. M31313).
deletions along the ovine genomic insert was generated using exonuclease III and mung bean nuclease (Genofit) (31). Sequencing was done by the dideoxynucleotide termination method (32) using single-stranded DNA of recombinant M13, denatured plasmid DNA templates (33), and T7 DNA polymerase (Sequenase, United States Biochemical) (34). In both alleles, the coding region was sequenced entirely on both strands.

**Nuclease S1 Protection.** The structure of sheep brain PrP mRNA was assessed by hybridization to single-stranded DNA probes derived from ovine genomic clones and nuclease S1 digestion (35). For mapping of the 5' splice junction, a 32P-labeled single-stranded DNA probe was prepared by primer extension to the EcoRI-Sma I region on pSc23.4 and digestion with Pst I. This generated a fragment of 216 nucleotides containing 18 nucleotides of a M13mp18 poly-

---

**Fig. 1.** Alignment of genomic clones and subclones of the ovine PrP gene. The boxed regions indicate the sequenced DNA. □ Coding region; □, 3' untranslated region. E, EcoRI; X, Xba I; H, HindIII; P, Pvu II; S, Sal I. *E* at 3.5 kb indicates the polymorphic EcoRI site that is present in clone 23/21 and absent in clone 32/21.

**FIG. 2.** Sheep PrP DNA and predicted amino acid sequence. The DNA sequence of a Pvu II site in the intron to a Pvu II site downstream of the poly(A) signal site with the predicted PrP amino acid sequence. The sequence of 32/21 between the two Pvu I sites is identical apart from the two indicated nucleotide differences at positions 20 and 583. The corresponding amino acid difference at codon 171 is also shown. The intron/exon border is marked by an arrow; the putative polyadenylation signal and the region of the poly(A) tail attachment are underlined at positions 4004-4014 and 4034-4039, respectively.
linker sequence. Two probes were used to map the 3' end of the ovine PrP mRNA. The short DNA probe was prepared by primer extension. The ovine pScr63.R genomic probe also hybridized efficiently to 4.6-kb PrP mRNA from cattle and goats (Fig. 4).

**RESULTS**

Cloning and Sequencing the Sheep PrP Gene. A genomic DNA library was prepared from spleen DNA of a Suffolk sheep using a λEMBL3 vector and kindly provided by A. J. Clark (AFRC Institute for Animal Physiology and Genetics Research, Edinburgh; ref. 29). The library was screened with a hamster PrP cDNA (plasmid pHaPrP (15)) and two positive clones were isolated. Clone 23/21 with an insert of 14 kilobases (kb) contained EcoRI (4 kb) and HindIII (4.8 kb) restriction fragments whereas the 17-kb insert of clone 32/21 contained EcoRI (7 kb) and HindIII (4.8 kb) fragments. These restriction fragments are similar in size to those detected by Southern blot analyses of total genomic DNA of Suffolk sheep (36) and Cheviot sheep (24). The DNA fragment of clone 23/21 from an EcoRI site (at −0.5 kb, Fig. 1) to a Pvu II site (at 4.2 kb, Fig. 1) was subcloned into pUC Bluescript KS vectors, respectively (plasmids pScr23.4 and bs10, respectively), and sequenced. The insert contained an open reading frame with approximately 83% homology to the hamster PrP cDNA.

The subclone derived from the second positive clone, 32/21 (plasmid pScr63.R), also contained an open reading frame. Sequencing between the two flanking Pst I sites (Fig. 2) revealed only two differences: at nucleotide 20 (cytidine in 23/21 and adenosine in 32/21) and at nucleotide 583 (guanine in 23/21 and adenosine in 32/21). The nucleotide-583 difference is particularly interesting because it leads to an amino acid difference (codon 171) in the predicted PrP protein sequences: arginine in 23/21 and glutamine in 32/21. The deduced amino acid sequences of these ovine PrPs have 88–90% homology with the PrPs from rodents and man.

**Determination of an Intron/Exon Border.** Rodent PrP genes contain an intron whose 3' border is 10 base pairs upstream of the translation initiation codon (17, 23). Ten bases upstream from the putative translation start codon of ovine PrP, there is a sequence, TCTTATTTTGGAGA (nucleotides 48–62), with strong homology to the consensus sequence (T/C)_n(N/C)TAGG for 3' splice junctions (37). Additionally, the motif CTGAC at nucleotides 139–143 is identical to the consensus sequence/5' splice acceptor site (38). To identify the intron/exon S1 protection analyses. Sheep brain poly(A)⁺ RNA was hybridized with a genomic DNA probe of 216 bases (Fig. 3) and subsequently digested with nuclease S1. Fragments of 138–140 bases were protected (Fig. 3). The band of 138 bases corresponds to a splicing event at nucleotide 62 at the first AG dinucleotide after the proposed lariat acceptor site. The bands of 139 and 140 bases might be due to incomplete digestion of the RNA-DNA hybrid by nuclelease S1. The DNA upstream of the protected region (sequence not shown) has a low G + C content of 40% with no homologies to the G + C-rich promoter region of the hamster gene (17) or other promoter structures (39). These results indicate that the ovine PrP gene, like its rodent homologues, has an intron/exon border located 10 base pairs upstream of the coding region.

**PrP mRNA in Natural Hosts.** Ovine and rodent PrP mRNAs vary greatly in size (Fig. 4 and ref. 40). By using a hamster PrP cDNA clone [pEA974 (27)] or an ovine PrP genomic clone (pScr23.4) in Northern blot analysis of brain poly(A)⁺ RNA, we estimated the size of ovine PrP mRNA to be 4.6 kb, compared to the 2.5-kb mouse and hamster PrP mRNA (Fig. 4). The ovine PrP genomic probe also hybridized efficiently to 4.6-kb PrP mRNAs from cattle and goats (Fig. 4).

**Determination of the Polyadenylation Site of the Ovine PrP mRNA.** The sequences 5'ATAAA (nucleotides 4009–4014) and ATTTAA (nucleotides 4033–4038) (Fig. 2) resemble the consensus sequence AATAAA, which precedes the polyadenylation site of most eukaryotic mRNAs. To determine the polyadenylation site of the ovine PrP mRNA, we carried out nuclease S1 protection analyses of total RNA from sheep brain by using DNA fragments (insert 23/21) spanning a region 3.4–4.4 kb downstream from the predicted polyadenylation site.

**Fig. 3.** Nuclease S1 protection analysis of sheep brain poly(A)⁺ RNA to determine the position of the intron/exon boundary. (A) Autoradiograph of the nuclease S1 protection analysis of ovine RNA with a DNA probe derived from clone pScr23.4. A 216-nucleotide DNA probe was used for nuclease S1 protection analysis. This DNA probe from position 2 to position 199 contains an additional 18-nucleotide sequence of M13mp18 (lane 1). DNA fragments of 138–140 nucleotides were protected in this analysis (lane 2). The reaction mixtures containing adenine, cytidine, guanine, and thymine of a DNA sequence were used as size markers (lanes 3–6, respectively). (B) Diagram of the region analyzed by nuclease S1 protection. Numbers refer to the PrP sequence shown in Fig. 2. **<c>Fig. 3</c>** Coding region. The test fragment and the protected fragments of the nuclease S1 analysis are aligned.

**Fig. 4.** Northern blot analysis of brain poly(A)⁺ RNA from various species. Lanes: 1, hamster; 2, mouse (C57BL, Sinc⁰); 3, mouse (IM, Sinc⁰); 4 and 5, sheep; 6, cow; 7, goat. Each lane contains approximately 5 μg of brain poly(A)⁺ RNA. Hybridizations were carried out with a hamster cDNA probe (lanes 1–4) or an ovine genomic DNA probe (lanes 5–7). Exposure times were as follows. Lanes: 1, 24 hr; 2–4, 72 hr; 5–7, 24 hr.
intron/exon border (Fig. 5), consistent with the apparent mRNA size (Fig. 4). All assays indicate that poly(A) addition occurs some 20–25 base pairs downstream from the sequence TATAAA. From the Southern blot analyses of the insert of 32/21 with ovine 3' PrP mRNA-specific DNA probes, we conclude that this genomic clone is also capable of encoding the whole 3.2-kb 3' untranslated mRNA region (data not shown).

Comparison of the PrP mRNAs of Various Species. Ovine PrP mRNA contains a 3.2-kb structure between its stop codon at position 840–842 and the polyadenylation site around nucleotide 4039 (Fig. 2). This 3' untranslated region is more than twice the size of the corresponding region in hamster, mouse, and human PrP mRNA. A comparison of the ovine sequence with the rodent sequence revealed three large insertions. Insertion D (about 0.5 kb) and insertion F (about 1.3 kb) seem to be sheep specific, whereas insertion B (0.25 kb) is also found in the human mRNA (Fig. 6).

DISCUSSION

Homology of Gene Structure and the mRNA. The hamster PrP gene has two exons: a 5' exon of 56–82 nucleotides preceded by a G+C-rich region, probably the transcription promoter, and a 3' exon of 1960 nucleotides that encodes the complete protein coding region of the gene. These exons are linked by a 10-kb intron (17) and this gene organization is conserved in mice (23). Our data indicate a similar structure in sheep with PrP encoded by only one exon. We detected an intron/exon boundary just on the 5' side of the start codon and, although this is consistent with an upstream exon, we have no direct data on the size of the intron or the 5' exon. The estimated size of the ovine PrP mRNA is 4.6 kb (Fig. 4) and the distance from the 3' intron/exon border to the polyadenylation site of the ovine PrP gene is about 4 kb. Hence approximately 0.6 kb of the PrP mRNA sequence remains undetermined and may be made up of the poly(A) tail and a 5' noncoding exon of the ovine PrP gene, similar to that found in the hamster gene (17). The highly conserved sequence (80–90% identical) of all known PrP mRNA coding regions is striking. The 3' untranslated regions of these mRNAs are also conserved in hamster, man, and sheep, although the much larger sheep sequence has several insertions. Because of their similar size, it is likely that the bovine and caprine PrP mRNAs also have this unusually large 3' untranslated region.

Protein Structure and Homologies. The primary structure of PrP is highly conserved in man (42), hamster (17), mouse (23, 43), and sheep (this paper). From the genomic DNA, we can deduce the sequence of a 256-amino acid protein with an overall homology of 88–90% with PrPs of other species.

The sheep PrP has a predicted 24-amino acid signal peptide (44), two possible sites of asparagine-linked glycosylation (at residues 184 and 200) (45), and an extremely hydrophobic C-terminal sequence. Its highly basic N-terminal region precedes a series of keratin-like glycine-rich segments that are conserved in PrPs across species. The amino acid inserted between residues 9 and 10 is predicted to be a glycine residue, which makes it identical to the bovine sequence (14). This insertion creates a Pro-Gly-Gly-Gly- sequence identical to a repetitive sequence in the tau protein (46), another protein implicated in degenerating processes of the central nervous system.

During biosynthesis and intracellular transport of PrP to the cell surface, the protein is posttranslationally modified. The signal peptide is cleaved off (16) and a phosphatidylinositol-glycolipid is attached to the new C-terminal amino acid created by removal of a hydrophobic tail (47, 48), and we have some evidence that this is also the case for the ovine PrP (J.H., unpublished data). The hydrophobic C-terminal region of the ovine molecule is highly conserved and shows only one substitution when sequences of PrPs from various species are compared (Fig. 6).
The PrP gene has been linked to the major gene controlling the incubation period (Sip) of the various strains of scrapie pathogens in mice and sheep. We have shown by gene cloning that its protein and gene structure are highly conserved in the natural host of scrapie. It is possible that the PrP polymorphism we have detected in the Suffolk sheep is related to the alleles of the sip gene.

We thank Dr. A. J. Clark and his colleagues at the Agricultural and Food Research Council Institute for Animal Physiology and Genetics Research Station, Edinburgh, for the sheep genomic DNA library and Drs. N. Robakis (New York) and B. Oesch (Zürich) for the hamster PrP cDNA clones, pEA974 and pHaPrP. This work was supported by funding through the Bundesministerium Forschung und Technologie and Sonderforschungsbereich to K.B. and travel costs were met in part from a North Atlantic Treaty Organization collaborative research grant (86/0112) awarded to J.H. and K.B.


CONCLUSION

The PrP gene has been linked to the major gene controlling the incubation period (Sip) of the various strains of scrapie pathogens in mice and sheep. We have shown by gene cloning that its protein and gene structure are highly conserved in the natural host of scrapie. It is possible that the PrP polymorphism we have detected in the Suffolk sheep is related to the alleles of the sip gene.