Am. J. Hum. Genet. 57:1050–1060, 1995

Molecular Cloning of a Full-Length cDNA for Dentatorubral-Pallidoluysian Atrophy and Regional Expressions of the Expanded Alleles in the CNS

Osamu Onodera, Mutsuo Oyake, Hiroki Takano, Takeshi Ikeuchi, Shuichi Igarashi, and Shoji Tsuji

Department of Neurology, Brain Research Institute, Niigata University, Niigata, Japan

Summary

Dentatorubral-pallidoluysian atrophy (DRPLA) is an autosomal dominant neurodegenerative disorder characterized by genetic anticipation and variable combinations of symptoms including myoclonus, epilepsy, cerebellar ataxia, choreoathetosis, and dementia. Recently, we discovered that DRPLA is caused by unstable expansion of a CAG repeat of a gene on the short arm of chromosome 12. We determined the consensus DRPLA cDNA sequence containing the complete coding region for 1,185 amino acids. The CAG repeat, which is expanded in DRPLA, is located 1,462 bp downstream from the putative methionine initiation codon and encodes a poly-glutamine tract. Although poly-serine and proline tracts exist near the CAG repeats, these poly-serine or proline tracts did not show any polymorphisms, which is in strong contrast to the high heterogeneity in the length of the CAG repeat. Northern blot analysis revealed a 4.7-kb transcript that is widely expressed in various tissues including heart, lung, kidney, placenta, skeletal muscle, and brain. Reverse transcription–PCR analysis revealed that the expanded alleles are transcribed to levels comparable to those of normal alleles. These results indicate that there is no difference in transcriptional efficiency between expanded and normal alleles. Furthermore, mRNA from cerebellar hemispheres of DRPLA patients showed smaller sizes of CAG repeats compared with other regions of the brain, which reflects somatic mosaicism of the expanded alleles of the DRPLA gene.

Introduction

Dentatorubral-pallidoluysian atrophy (DRPLA) is an autosomal dominant neurodegenerative disorder characterized by various combinations of ataxia, myoclonus, epilepsy, choreoathetosis, and dementia (Smith et al. 1958; Naito and Oyanagi 1982). Since the report by Naito and Oyanagi (1982) describing DRPLA in 16 patients of five Japanese families), DRPLA has predominantly been described in Japanese, and the prevalence rate of DRPLA in Japanese has been estimated to be 0.4–0.7/100,000 (Inazaki et al. 1990). The age at onset is quite variable, ranging from the 1st decade to the 6th decade with the mean age at onset being 32.1 years of age (Hirayama et al. 1990; Naito 1990). Recently, we and others identified unstable expansion of a CAG repeat in a gene on the short arm of chromosome 12 as the causative mutation for DRPLA (Koide et al. 1994; Nagafuchi et al. 1994*b*).

To date, five neurodegenerative diseases associated with CAG repeat expansions, spinal and bulbar muscular atrophy (SBMA) (La Spada et al. 1991), Huntington disease (HD) (The Huntington's Disaease Collaborative Research Group 1993), spinocerebellar ataxia type 1 (SCA1) (Orr et al. 1993), DRPLA (Koide et al. 1994; Nagafuchi et al. 1994*b*) and Machado-Joseph disease (MJD) (Kawaguchi et al. 1994) have been reported. There are a number of similarities in these disorders in terms of genetic as well as clinical features: (1) there are considerable heterogeneities in clinical presentations and ages at onset even in a single pedigree; (2) there is a strong inverse correlation between the age at onset and the degree of CAG-repeat expansion; (3) male meiosis seems to be the major source for larger intergenerational increase of CAG-repeat length (Igarashi et al. 1992; La Spada et al. 1992; Andrew et al. 1993; Duyao et al. 1993; The Huntington's Disaease Collaborative Research Group, 1993; Orr et al. 1993; Kawaguchi et al. 1994; Koide et al. 1994; Nagafuchi et al. 1994*b*; Ikeuchi et al. 1995). In light of the similarities between these diseases, there seems to be a common mechanism for selective neuronal cell loss caused by expansions of CAG repeats. The structures of cDNAs for SBMA (Chang et al. 1988), HD (The Huntington's Disease Collaborative Research Group 1993), SCA1 (Banfi et al. 1994), MJD (Kawaguchi et al. 1994) and DRPLA (Nagafuchi et al. 1994*a*) causative genes have been described. Except for the androgen receptor gene, the causative gene for SBMA, the deduced amino acid sequences of the cDNAs

Received June 19, 1995; accepted for publication August 1, 1995.
Address for correspondence and reprints: Shoji Tsuji, Department of Neurology, Brain Research Institute, Niigata University, Asahimachi-dori 1, Niigata, Niigata 951, Japan. E-mail: tsuji@cc.niigata-u.ac.jp

0002-9297/95/5705-0010$02.00

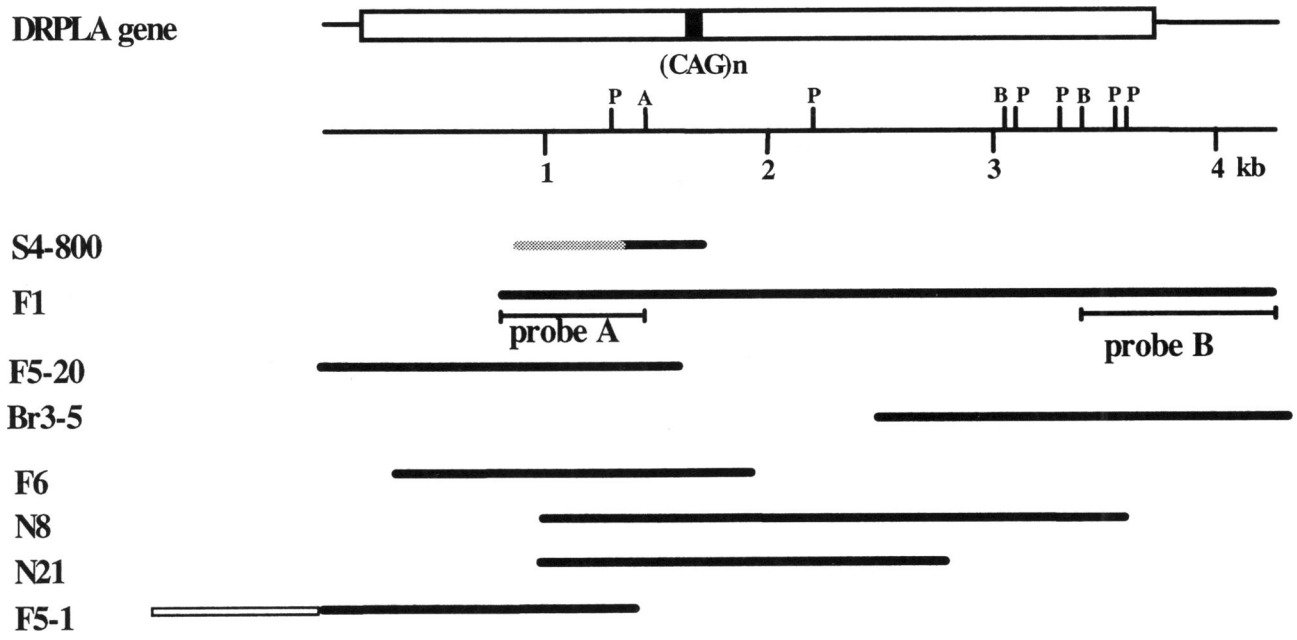


Figure 1 Schematic representation of the organization of DRPLA cDNA. Coding region is indicated by a box. The position of the CAG repeat is indicated by a blackened box. Restriction sites for *Acc*I (A), *Bam*HI (B), and *Pst*I (P) are shown below the box. The size and the position of each cDNA clone are shown below the restriction map. Stippled region in S4-800 indicates chimeric portion. Positions of probes A and B used for further screening of DRPLA cDNA are indicated below F1. The unblackened box at the 5′ side of F5-1 indicates the portion that could not be confirmed by other independent cDNA clones.

for HD, SCA1, MJD, and DRPLA causative genes do not contain any domains suggesting the functions of the gene products (The Huntington's Disease Collaborative Research Group 1993; Banfi et al. 1994; Kawaguchi et al. 1994; Nagafuchi et al. 1994*a*).

As the first step toward a better understanding of the molecular mechanisms of the selective neuronal degeneration in DRPLA, we have undertaken molecular cloning of full-length cDNA clones of the DRPLA gene. We further analyzed the expression of mutant genes containing expanded CAG repeats as well as the wild-type gene in various regions of the CNS.

Material and Methods

Screening of cDNA Libraries

Approximately 1×10^6 phage clones from an adult human occipital cortex cDNA library (Stratagene) were screened with an oligonucleotide 5′-CACCAGTCTCAACACATCACCATCATCA (Li et al. 1993) end-labeled with [γ32-P]ATP. Hybridization was performed at 58°C for 18 h in 6 × SSC (1 × SSC = 0.5 M NaCl, 15 mM sodium citrate) containing 0.05% sodium pyrophosphate. Finally, filters were washed for 30 min at 58°C in 6 × SSC-0.05% sodium pyrophosphate. For screening of the overlapping clones, we screened 1×10^6 phage clones from a fetal human brain cDNA library (Stratagene) and an adult human occipital cortex cDNA library (Stratagene) using the 843-bp *Eco*RI fragment of S4-800 (fig. 1), 648-bp *Acc*I and *Eco*RI fragment of F1 (probe A, fig. 1), or 851-bp *Bam*HI and *Eco*RI fragment of F1 (probe B, fig. 1) as the probe. The cDNA fragments were labeled using the random priming technique (Feinberg and Vogelstein 1983). Hybridization was performed at 42°C in a solution containing 50% formamide, 5 × SSC, 1 × Denhardt's solution (50 × Denhardt's solution = 1% Ficoll, 1% polyvinylpyrrolidone and 1% bovine serum albumin), 20 mM sodium phosphate (pH 7.0), 10% dextran sulfate and 50 μg/ml denatured salmon sperm DNA (Wahl et al. 1979). After hybridization, filters were washed to a final stringency of 0.1 × SSC 0.1% sodium dodecyl sulfate (SDS) at 60°C. The λ-ZAP II clones were self-excised into pBluescript SK(−) plasmid (Short al. 1988).

Screening for the Mouse Homologue of Human DRPLA Gene

Approximately 1×10^6 phage clones from mouse teratocarcinoma and brain cDNA libraries (Stratagene) were screened with the 648-bp *Acc*I and *Eco*RI fragment of F1 (probe A, fig. 1) or 851-bp *Bam*HI and *Eco*RI fragment of F1 (probe B, fig. 1) as the probe. The cDNA fragments were labeled using the random priming technique (Feinberg and Vogelstein 1983). Hybridization was performed as described above. After hybridization,

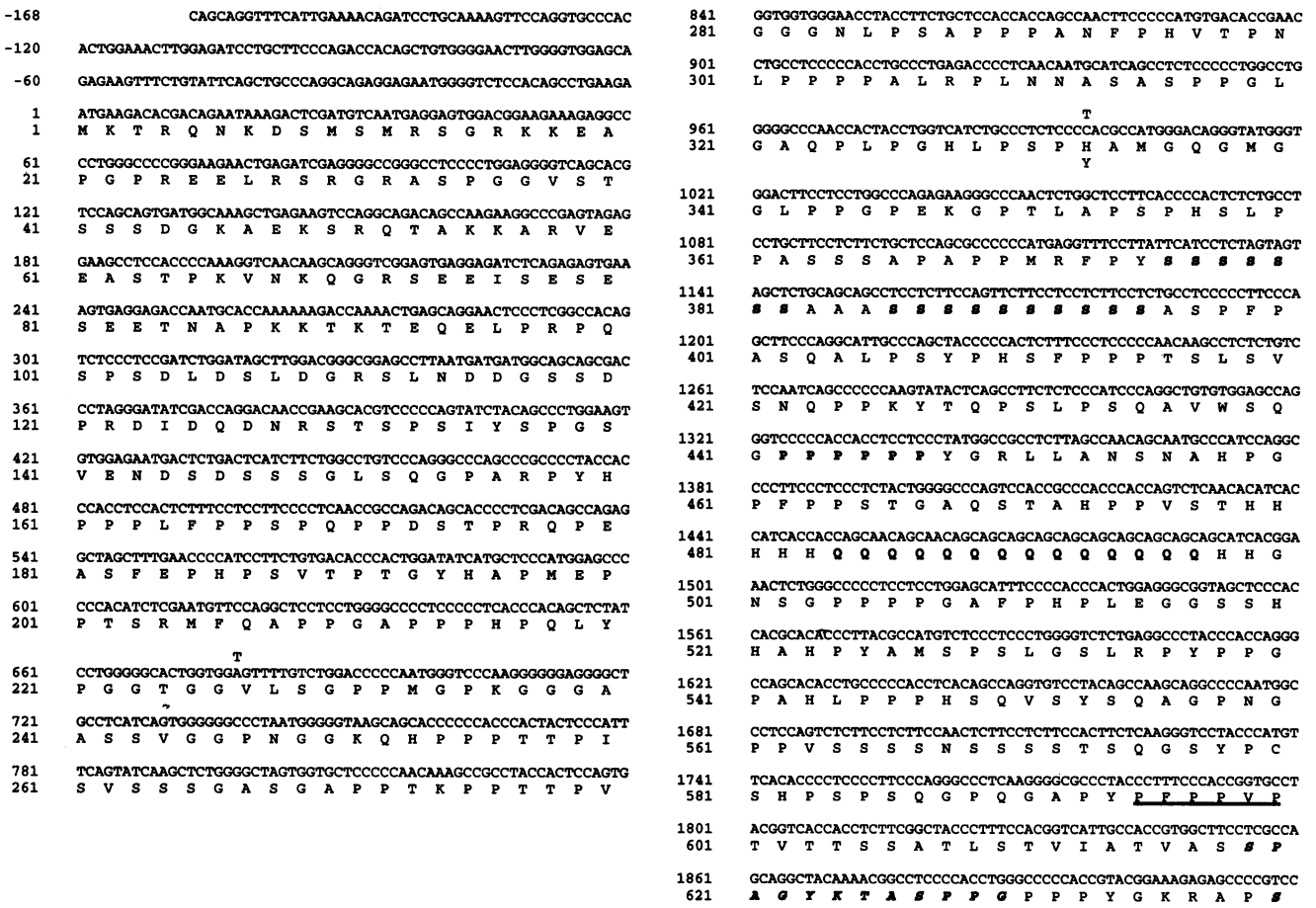


Figure 2 Consensus nucleotide sequence of DRPLA cDNA. The deduced amino acid sequence is shown below the nucleotide sequence. Boldfaced letters indicate homopolymeric amino acid tracts. Italic letters indicate direct repeats of amino acid sequences. Double underline indicates a potential SH3 binding domain. Differences between those reported by Nagafuchi et al. (1994*a*) and ours, in the nucleotide and amino acid sequences, are noted above and below the sequences, respectively. Gaps are represented by hyphens.

filters were washed to a final stringency of 0.1 × SSC 0.1% SDS at 60°C. The cDNA inserts of λ-ZAP II clones were self-excised into pBluescript SK(−) plasmid (Short et al. 1988).

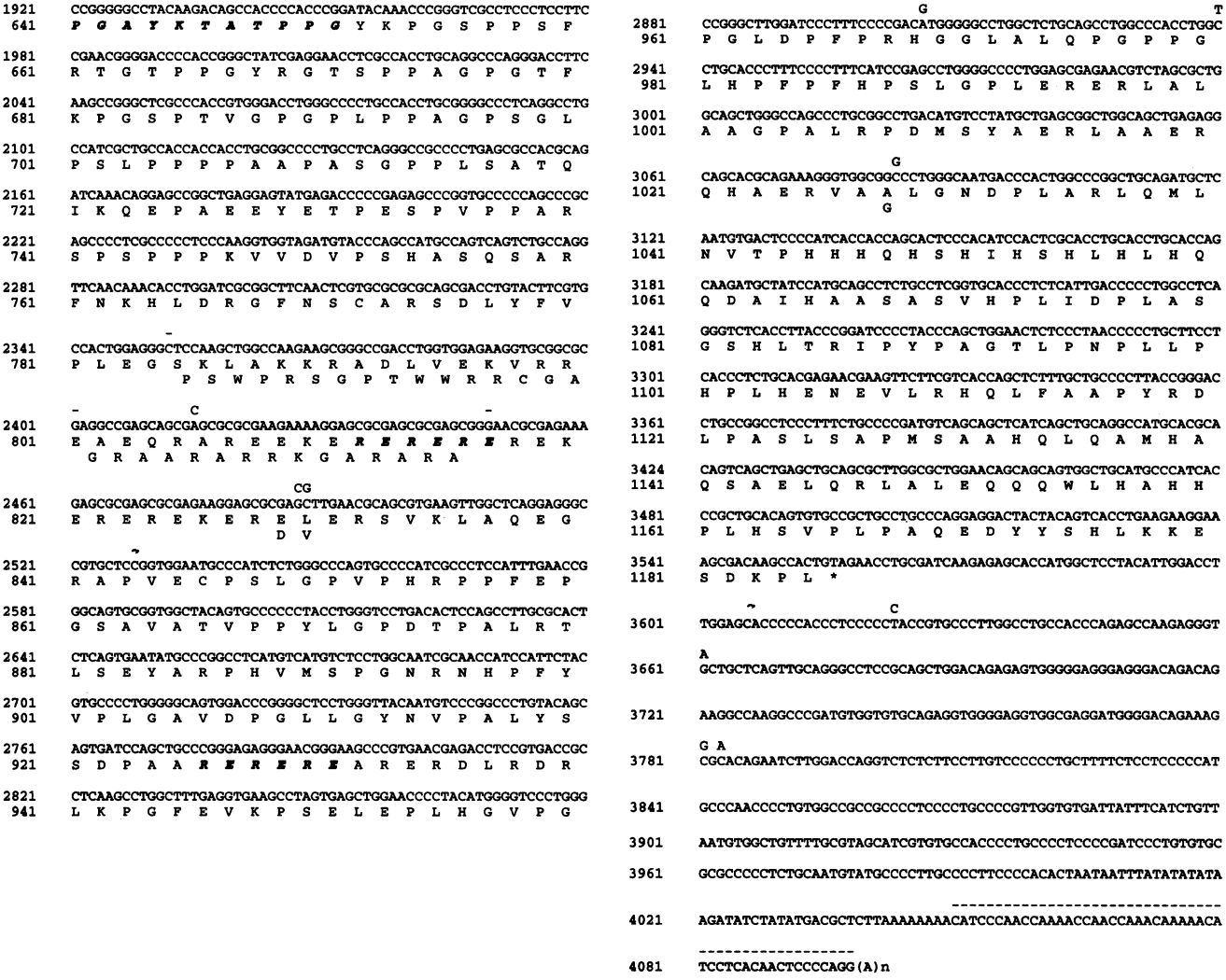
Nucleotide Sequence Analysis

Nucleotide sequences were determined by cycle sequencing or the dideoxynucleotide chain termination method using double-stranded plasmid DNAs as templates with an automated fluorescent DNA sequencer (Pharmacia P-L Biochemicals) (Sanger et al. 1977; Chen and Seeburg 1985; Murray et al. 1989). Every portion of the consensus DRPLA cDNA sequence was confirmed by at least two independent clones.

Database searches were carried out using the BLAST program through GenBank (release 83), European Molecular Biology Laboratory (EMBL) (release 38), DNA Data Bank of Japan (DDBJ) (release 17), SWISS (release 28), and Protein Information Resource (release 40). The nucleotide sequence described in this paper has been deposited in GSDB, DDBJ, EMBL, and National Center for Biotechnology Information nucleotide sequence databases under the accession number of D38529.

Analysis of Polymorphisms of Poly-Proline and Poly-Serine Stretches in DRPLA Gene

The primer sequences used for amplification of poly-serine tracts at amino acid residues 376-382 and 386-395 were PSF: 5′-GCCTCCTGCTTCCTCTTCTGCTC and PSR: 5′-CTTGTTGGGGGAGGGAAAGAGT. The primer sequences used for the amplification of the poly-proline tract at amino acid residues 442-447 were PPF: 5′-CCTTCTCTCCCATCCCAGGCTGTG and PPR: 5′-TGGGCATTGCTGTTGGCTAAGAGG. PCR was conducted in a final volume of 25 μl containing 20 ng of genomic DNA, 250 nmol of each primer, 10 mM Tris HCl, pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$, 0.001% (w/v) gelatin, 10% (v/v) dimethylsulfoxide, 200 mM dATP, 200 mM dCTP, 200 mM TTP, 50 mM dGTP, 150 mM 7-deaza-dGTP and 0.75 U *Taq* DNA polymerase. PCR



amplification consisted of 30 cycles of 2 min denaturation at 94°C, 30 s annealing at 57°C and 1 min extension at 72°C. PCR products were electrophoresed in 5% denaturing polyacrylamide gels and were autoradiographed.

Northern Blot Analysis

Northern blot filters of poly A(+) RNA from various human tissues were obtained from Clonetech. Hybridization was performed as described above using the 851-bp *Bam*HI and *Eco*RI fragment of F1 (probe B, fig. 1) or 2.0-kb human β-actin cDNA obtained from Clonetech as the probe. The radioactivities of the bands were measured using a laser image analyzer (Fuji BAS2000 system) (Amemiya and Miyahara 1988).

Extraction of RNA from Autopsied Tissues of DRPLA Patients

Frozen autopsied tissues from three DRPLA patients were used for the analysis of DRPLA mRNAs. Case N 17-76 is that of a woman with age at onset of 13 years and death at 26 years, who showed a progressive myoclonus epilepsy (PME) phenotype. Case N17-81 is that of a man with a PME phenotype with age at onset of 8 years and death at age 18 years. Case N17-88 is that of a woman with the non-PME phenotype with age at onset at 61 years and death at age 76 years. The following regions of the CNS were available for extraction of RNA: N17-76, frontal cortex and cerebellar hemisphere; N17-81, frontal cortex, frontal white matter, striatum, and cerebellar hemisphere; and N17-88, frontal cortex, temporal cortex, occipital cortex, striatum, cerebellar hemisphere, and dentate nucleus. Total RNA was extracted by the guanidinium thiocyanate procedure (Chirgwin et al. 1979). To avoid potential contamination of genomic DNA in the total RNA preparation, the RNA was treated with RNase-free DNase (Promega) (0.5 U/μg RNA) in 40 mM Tris-HCl, pH 7.9, containing 10 mM NaCl and 6 mM $MgCl_2$ for 15 min at 37°C, and then the RNA was extracted with phenol/chloroform to remove the DNase activity.

Reverse Transcription (RT)–PCR

Ten micrograms of total RNA and 20 pmol of random hexamer primers were dissolved in distilled water in a final volume of 13.5 μl. The solution was heated at 68°C for 5 min, and then the tube was quickly chilled on ice. The solution was adjusted to 50 mM Tris-HCl (pH 8.3), 3 mM $MgCl_2$, 75 mM KCl, and 0.5 mM each dNTP in a final volume of 20 μl and 200 U of recombinant Moloney-murine leukemia virus reverse transcriptase was added. The reaction was continued for 60 min at 37°C. An aliquot (1/25) of each RT reaction mixtures was subjected to PCR. The primer sequences and PCR conditions for amplification of the CAG repeat were the same as those as described elsewhere (Koide et al. 1994; Nagafuchi et al. 1994*b*). PCR was performed in a total volume of 25 μl containing 6.25 pmol of each primers, including 1.25 pmol of the reverse primer labeled with [γ32-P]ATP, 250 μM of each dNTP, 50 mM KCl, 10 mM Tris-Hcl (pH 8.4), 2 mM $MgCl_2$, 2.5 M N,N,N-trimethylglycine and 1.25 U *Taq* DNA polymerase (Takara Syuzo). PCR products were electrophoresed in 5% denaturing polyacrylamide sequencing gels.

Results

Isolation of DRPLA cDNA Clone

We initially screened an adult human occipital cortex cDNA library (Stratagene) using a ^{32}P-labeled oligonucleotide, 5′-CACCAGTCTCAACACATCACCATCATCA, as the probe (Li et al. 1993). Through screening of 1×10^6 clones, we identified one positive clone, S4. We subcloned the 843-bp *Eco*RI fragment of S4 into pBluescript SK(−) (S4-800). To obtain larger cDNA clones, we further screened 1×10^6 clones from a fetal human brain cDNA library (Stratagene) using the S4-800 as the probe and selected four clones (F1, F6, N8, and N21) for the further analysis.

To isolate cDNA clones extending further toward the 5′ and 3′ side, we screened the same libraries using the probes A and B as shown in figure 1. We selected two cDNA clones from the fetal brain library (F5-1 and F5-20) and Br3-5 from a human occipital cortex cDNA library for further analysis (fig. 1).

Nucleotide Sequence of Human DRPLA cDNA

The complete nucleotide sequences of these cDNA clones (F1, F5-1, F5-20, and Br3-5) were determined for both strands, and a consensus nucleotide sequence consisting of 4,267 bp was determined (figs. 1 and 2). To confirm the consensus nucleotide sequence of the DRPLA cDNA, partial nucleotide sequences of N8, F6, and N21 were also determined. Every portion of the consensus sequence of the DRPLA cDNA was confirmed by at least two independent clones. Since the 756-bp segment on the 5′ side of F5-1 could not be confirmed by other clones, we did not include this portion in the DRPLA cDNA. The presence of a poly (A) tail at position 4099 was confirmed by independent cDNA clones, although the consensus sequence for the polyadenylation was not identified (Wickens and Stephenson 1984).

Compared with the nucleotide sequence recently reported by Nagafuchi et al. (1994*a*), there were substantial discrepancies. In particular, there were gaps at nt 2352, 2401, and 2449 in their sequence, which led to different reading frames at amino acid residues at 785–816. The last 49 bp in the DRPLA cDNA were absent in their sequence. In addition, there were four neutral and four missense single base substitutions in the coding region, and four single base substitutions in the noncoding region compared with their sequence.

Since there is an in-frame stop codon 6 bp upstream of the putative ATG initiation codon, the ATG at position 1 is most likely the translation initiation codon, although the flanking sequence of the ATG codon is not in good agreement with Kozak's consensus sequence (Kozak 1987). Assuming the ATG codon as the translation initiation codon, the DRPLA cDNA is predicted to code for 1185 amino acids with a predicted molecular mass of 125 kD, which was named "DRPLAP" (DRPLA protein) (fig. 2).

The CAG repeat, which is expanded in DRPLA patients, is located at position 1462 and is predicted to code for a poly-glutamine tract. Poly-proline and poly-serine tracts were also identified in the deduced amino acid sequence at positions 442–447 (Pro6), 376–382 (Ser7), and 386–395 (Ser10). Analysis of amino acid composition of DRPLAP indicates high percentages of proline (18.8%), serine (12.9%) and alanine (9.1%) residues. A proline-rich motif of XpΦPpXP (X represents a nonconserved residue; Φ represents a hydrophobic residue; and p represents a residue that tends to be proline), which has recently been identified as a ligand for Src homology 3 (SH3) domains (Yu et al. 1994), is located at amino acid residue 595 of DRPLAP (PFPPVP) (fig. 2). A direct amino acid repeat, SPXXYKTAXPPG, is present at positions 619 and 640. The amino acid sequence of RERERE is also present at positions 812 and 926. A hydropathy plot indicates that the protein is generally hydrophilic with no obvious hydrophobic membrane-spanning regions (fig. 3).

Homology searches revealed no significantly homologous sequences except for one human expressed sequence (M78755), which was located at chromosome 1p (Adams et al. 1992). The nucleotide sequence of M78755 shows high homology to human DRPLA cDNA (63.3% identity in nucleotide sequence).

Polymorphism Analysis of a Poly-Glutamine, Poly-Serine, or Poly-Proline Tracts Lengths in DRPLA Gene

There are poly-serine and poly-proline tracts located upstream of the CAG repeat. The poly-serine tract oc-

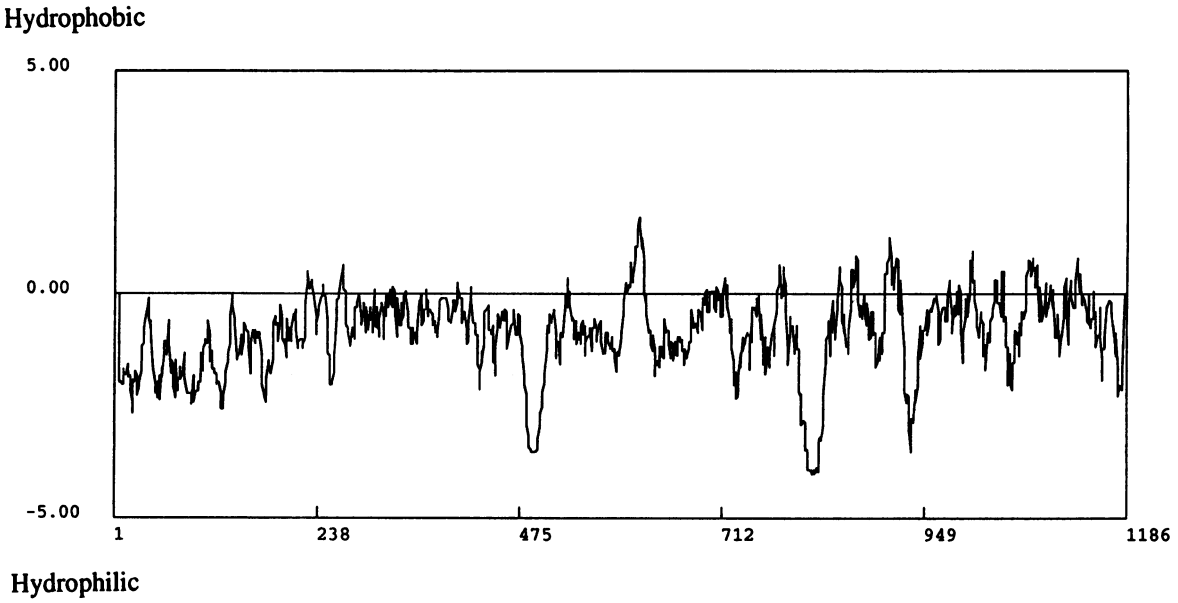


Figure 3 Kyte-Doolittle hydropathy profile of the deduced amino acid sequence of DRPLAP. The hydropathy plots of 11 residues were obtained using the algorithm of Kyte and Doolittle (1982).

curs at position 376 in the deduced amino acid sequence. The 17 serine residues are interrupted by three arginine residues at the eighth position. The six proline residues occur at position 442 in the deduced amino acid sequence. To investigate whether the poly-serine and poly-proline tracts exhibit length polymorphism, we analyzed the lengths of the poly-serine and poly-proline tracts in 20 normal subjects and 20 DRPLA patients. We did not find any length polymorphisms for poly-proline or poly-serine tracts in the normal and the DRPLA patients' chromosomes. These results are in striking contrast to the high heterozygosity of the length of the poly-glutamine tract (ranging in length from 8 to 35 repeat units with 87.1% heterozygosity; 140 normal chromosomes were analyzed).

Comparison of the CAG Repeat and the Flanking Sequences between Mouse and Human DRPLA Genes

To isolate the mouse homologue of the human DRPLA gene, we screened mouse tetracarcinoma and brain cDNA libraries (Stratagene) using the probes A and B as shown in figure 1. We isolated 30 and 25 cDNA clones from the tetracarcinoma and the brain cDNA libraries, respectively. After confirming the restriction maps using *Pst*I and *Bam*HI, we selected two cDNA clones from the fetal brain cDNA library (T5-2 and B8-1) for further analysis. The nucleotide sequences of these two clones flanking the CAG repeat were compared with the sequence of human DRPLA cDNA (fig. 4). Interestingly, the mouse homologue contained only four repeat units, interrupted by CCA between the third and fourth triplets. The nucleotide sequence of the rest of the mouse DRPLA cDNA including the poly-serine stretch, however, showed high homology to that of the human DRPLA cDNA, except for the length of the CAG repeat (84.1% identity in nucleotide sequence and 82.6% identity in deduced amino acid sequence) (fig. 4).

Tissue Expressions of DRPLA Gene

Northern blot analysis of poly A(+) RNA isolated from various human tissues showed that the 4.7-kb mRNA is widely expressed in various tissues including heart, lung, kidney, placenta, skeletal muscle, and brain, and, to a lesser extent, in liver. The expression was higher in fetal tissues, especially in brain. In human adult brain, the transcript is broadly expressed in amygdala, caudate nucleus, corpus callosum, hippocampus, hypothalamus, substantia nigra, subthalamic nucleus, and thalamus, which indicates that the expression of the DRPLA gene is not restricted to the dentatorubral-pallidoluysian system, the most severely affected region in DRPLA (fig. 5).

Regional Expression of Mutant DRPLA Gene Carrying Expanded CAG Repeat in CNS

As the first step to investigate how the mutant DRPLA gene containing an expanded CAG repeat is involved in the pathogenesis of DRPLA, we analyzed regional expressions of the mutant DRPLA genes in the CNS by RT-PCR. The expanded as well as wild-type DRPLA transcripts were observed in the frontal cortex, occipital

A) nucleotide sequence

```
Human   1321   GGTCCCCCACCACCTCCTCCCTATGGCCGCCTCTTAGCCAACAGCAATGCCCATCCAGGC
               ***** ** ** ********************** ******* ****************
Mouse          GGTCCACCTCCTCCTCCTCCCTATGGCCGCCTCTTGGCCAACAACAACACCCATCCAGGC

Human   1381   CCCTTCCCTCCCTCTACTGGGGCCCAGTCCACCGCCCACCCACCAGTCTCAACACATCAC
               ********       *********  ** ** ** ********* *** * * *******
Mouse          CCCTTTCCCT---CTACTGGGGGTCAATCTACAGCCCACCCAGCAGCCCCTACACATCAC

Human   1441   CATCACCACCAGCAACAGCAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCATCACGGA
               ******                      **********  ***** ** ** ***** ***
Mouse          CATCAC--------------------CAGCAGCAGCCACAGCAACAACATCATCATGGA

Human   1501   AACTCTGGGCCCCCTCCTCCTGGAGCATTTCCCCACCCACTGGAGGGCGGT
               ***************** ** ***** * *** ***** ** *** ** **
Mouse          AACTCTGGGCCCCCTCCACCCGGAGCGTATCCTCACCCTCTAGAGAGCAGT
```

B) deduced amino acid sequence

```
Human   441    GPPPPPPYGRLLANSNAHPGPFPPSTGAQSTAHPPVSTHHHHHQQQQQQQQQQQQQQQHHG
               ************** * ******* **  *****   *****       *** *** ***
Mouse          GPPPPPPYGRLLANNNTHPGPFPP-TGGQSTAHPAAPTHHHH-------QQQPQQQHHHG

Human   501    NSGPPPPGAFPHPLEGG
               ********* *****
Mouse          NSGPPPPGAYPHPLESS
```

Figure 4 Comparison of the nucleotide sequence of human DRPLA cDNA with that of mouse homologue cDNA. *A*, Nucleotide sequence. *B*, Deduced amino acid sequence. Asterisks indicate identical sequence. Bold letters in panel B indicate long homopolymeric amino acid stretches.

cortex, white matter, striatum, cerebellar hemisphere, and dentate nucleus (fig. 6). Moreover, the amounts of RT-PCR products from the expanded alleles were comparable to the amount of those derived from the wild-type alleles. As shown in figure 6, there was considerable variation in the size of expanded alleles among various regions of the central nervous system. It should be noted that the cerebellar hemisphere showed the smallest size of expanded alleles. Size distributions of RT-PCR products were similar to those of PCR products of genomic DNA in the corresponding regions of the CNS. In contrast to RT-PCR products of expanded alleles, those of wild-type alleles did not show any differences in size among the various regions of the CNS.

Discussion

In this study we determined the nucleotide sequence of human DRPLA cDNA containing the complete coding region. The nucleotide sequence of a DRPLA cDNA has recently been described by Nagafuchi et al. (1994*a*). Although our nucleotide sequence essentially agrees with theirs, there are disagreements at 15 regions. In addition, the 49 bp at the 3′ noncoding region was absent in their sequence. In particular, there are gaps at nt 2352, 2401, and 2449 in their sequences, which leads to different deduced amino acid residues at 785–817. In the present study, every portion of the consensus nucleotide sequence of DRPLA cDNA was confirmed by at least two independent clones, thereby excluding the possibility of cloning artifacts such as chimeric clones. Moreover, we have cloned a full-length cDNA for the mouse DRPLA gene and compared the human and mouse DRPLA cDNA sequences. The comparison further confirmed our nucleotide sequence and the reading frame as shown in figure 2. The discrepancy is presumably due to errors in the nucleotide sequence published by Nagafuchi et al. (1994*a*). The other single base substitutions in the coding and noncoding regions might represent polymorphisms.

It is interesting that the CAG repeat in the mouse homologue cDNA is much shorter than that in the human DRPLA cDNA. The mouse DRPLA cDNA contains only three CAG repeat units followed by CAA, which is in striking contrast to human DRPLA cDNA, in which the CAG repeat length is quite variable ranging from 8 to 35 repeat units with a mean unit of 15.2. A stretch of only four CAG repeats is interrupted at the third triplet by CAA. The CAG repeats in mouse homologue genes for other triplet repeat diseases are also shorter than those of human cDNAs. There are two repeats in the mouse androgen receptor gene (He et al. 1990), two and four repeats interrupted by CAA at the third triplet in the mouse huntingtin gene (Lin et al. 1994), eight repeats in the mouse *fmr*-1 gene (Ashley et al. 1992) and five repeats, with two of the five CTG triplets being

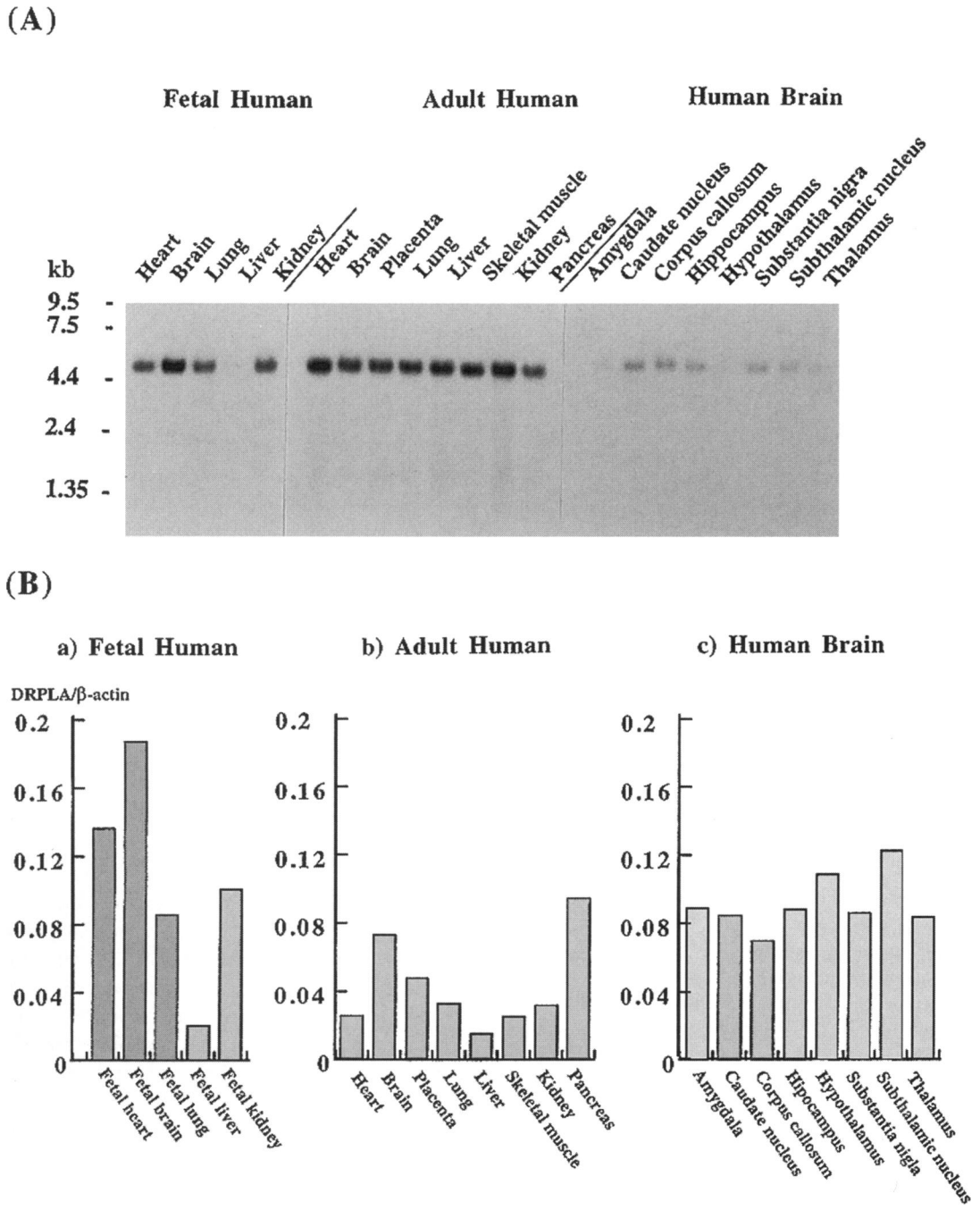


Figure 5 Northern blot analysis of human DRPLA mRNA in various human tissues. *A*, Northern blot analysis of human DRPLA mRNA expression in various human tissues. Each lane contains 2 μg poly A(+) RNA. Hybridization was performed using 851-bp *Bam*HI and *Eco*RI fragment of F1 (probe B; fig. 1). *B*, The ratios of radioactivities of the human DRPLA mRNA to those of β-actin mRNA were measured using a Fuji Bioimaging Analyzer BAS2000.

replaced by CAG in the mouse myotonic dystrophy kinase gene (Jansen et al 1992). Thus, it seems to be a general rule that CAG repeats in mouse homologues are shorter than those in human genes.

As shown in figures 1 and 2, the DRPLA cDNA contains an open reading frame coding for 1,185 amino acids. The CAG repeat codes for a poly-glutamine tract, starting at amino acid residue 488 in the middle of the coding region. The localization of the CAG repeat in the DRPLA gene is in contrast to the localization of CAG repeats near the N-termini in causative genes for other neurodegenerative diseases associated with ex-

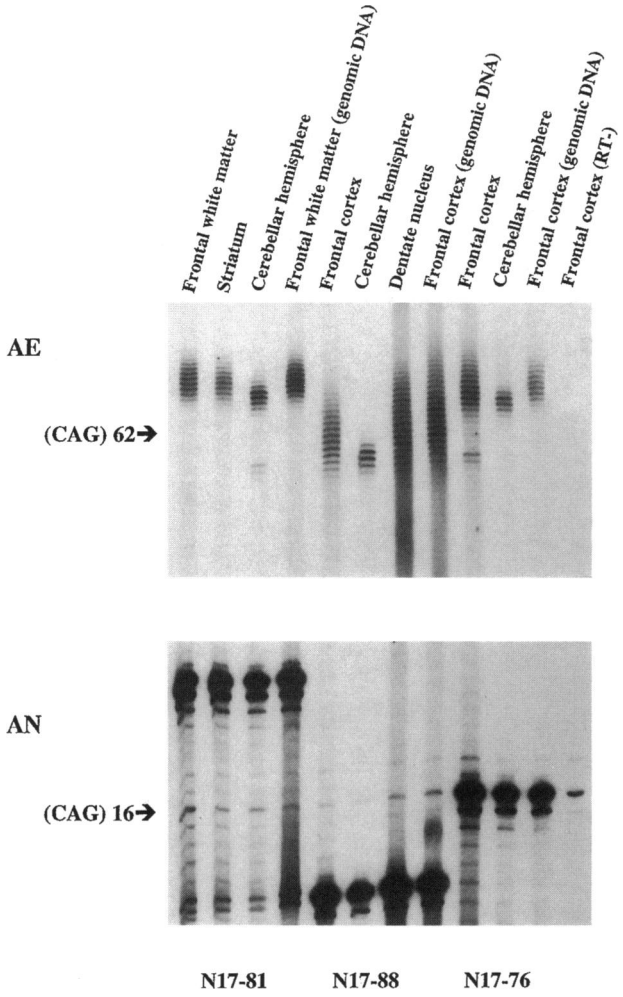


Figure 6 Regional distribution of expanded and wild-type DRPLA gene expression. DRPLA mRNAs derived from expanded as well as normal alleles were analyzed by RT-PCR. DRPLA mRNA of expanded alleles were expressed in all the tissues we analyzed at levels comparable to those of mRNAs derived from normal alleles. RT-PCR products showed similar size distributions to those observed for PCR products derived from genomic DNA. As a negative control, PCR was performed using RNA that was not subjected to RT reaction. AN = normal allele; and AE = expanded allele.

panded CAG repeats; CAG repeats are located at amino acid residue 18 in HD (The Huntington's Disease Collaborative Group 1993), 197 in SCA1 (Banfi et al. 1994) and 58 in SBMA (Chang et al. 1988). In the gene for MJD, which has been most recently found to be associated with an expanded CAG repeat, however, the polyglutamine tract encoded by the CAG repeat is located near the carboxy-terminus (Kawaguchi et al. 1994). Although determination of the functional initiation codon in the DRPLA gene is required, these results indicate that poly-glutamine tracts are not necessarily located near the amino-terminus.

Homopolymeric stretches of glutamine or proline, or proline-rich proteins have been found to occur predominantly in transcription factors (Ross et al. 1993). Recently, Gerber et al. (1994) reported evidence that the length of the poly-glutamine or poly-proline tract directly modulates transcriptional activity. A unique repeat of RERERE, which is present at amino acid residues 812 and 926 in DRPLAP, is frequently found in homeotic proteins with poly-glutamine tracts (S27806 [homeotic protein BarH2] [Higashijima et al. 1992], S03170 [homeobox protein cut] [Blochlinger et al 1988], and A43742 [fsh membrane protein] [Haynes et al. 1989]). It is interesting that the fsh membrane protein has poly-serine tracts as well. These findings raise the possibility that DRPLAP is a transcriptional factor, though this hypothesis must be further tested experimentally.

An interesting proline-rich motif, XpΦPpXP, has been recently described as a ligand for Src homology 3 (SH3) domains found in many intracellular signaling proteins (Yu et al. 1994). Recent investigations have revealed that the SH3 domains provide an effector function to link tyrosine kinase to specific target proteins (Koch et al. 1991). The XpΦPpXP motif is located at amino acid residue 595 of DRPLAP (PFPPVP) (fig. 2), which raises the possibility that DRPLAP acts as an SH3 ligand.

Despite these intriguing structures, the function of the DRPLAP still remains unknown. A homology search through databases did not reveal any domains suggesting the function of the DRPLAP, either. Establishment of expression systems of DRPLAP will be required to elucidate the functions.

In the HD chromosomes, a strong linkage disequilibrium at the polymorphic CCG repeat flanking the CAG repeat has been demonstrated, which suggests the possibility of predisposition of certain chromosomes to expansion of CAG repeats (Andrew et al. 1994; Squitieri et al. 1994). With this background we investigated whether poly-serine and poly-proline stretches located at positions 376–382, 386–395, and 442–447 are polymorphic. Interestingly, we did not find any length polymorphism in these repeats. This result strongly indicates that only the CAG repeats show intrinsic instabilities, with a heterozygosity of 87.1%.

From the northern blot analysis, wide distribution of expression of the DRPLA gene was observed (fig. 5*A* and *B*). Within the CNS, the DRPLA gene is also widely expressed. Thus, the distribution of DRPLAP mRNA expression does not account for the distribution of pathological lesions in DRPLA. These findings are quite similar to those observed for huntingtin (Li et al. 1993; Hoogeveen et al. 1993; Strong et al. 1993) and ataxin 1 (Banfi et al. 1994). Furthermore, high expression of DRPLA mRNA in fetal brain suggests that DRPLAP may play an important role in the development of the CNS.

As the first step toward understanding the molecular mechanisms involved in the pathology caused by the ex-

panded allele, we investigated how the expanded allele is expressed in tissues by RT-PCR analysis. RT-PCR products showed similar patterns to those of PCR products obtained using genomic DNAs that were extracted from the same tissues simultaneously. These results indicate that expanded alleles are transcribed at an efficiency similar to those of wild-type allele. Recently, Yazawa et al. (1995) reported that mutant DRPLAP is present at a level to similar to that of wild-type DRPLAP by the western blotting analysis using polyclonal antibodies raised against synthetic peptide. Taken together, these results indicate that there are no abnormalities in the transcription and translation of the mutant DRPLA gene. In HD and SCA1, the expanded allele has been shown to be expressed as mRNA, but the regional distribution of the mutant mRNA has not been reported (The Huntington's Disease Collaborative Research Group 1993; Banfi et al. 1994). In this study we clearly demonstrated that DRPLA mRNA derived from the expanded alleles are widely expressed in patients' CNSs at levels comparable to those of mRNAs derived from wild-type alleles.

The most striking findings are the somatic mosaicism of DRPLA mRNA expression, and that as shown in figure 6. RT-PCR products derived from cerebellar hemisphere RNA were smaller and showed few bands compared with those derived from cerebral cortex. These differences presumably reflect somatic mosaicism of genomic DNAs in various tissues (H. Takano, O. Onodera, M. Yamada, S. Igarashi, K. Oyanagi, Y. Takiyama, H. Takahashi, et al., unpublished information).

In the present study we described the nucleotide sequence of human DRPLA cDNA containing the complete coding region and regional distribution of DRPLA mRNA of expanded alleles as well as normal alleles of the DRPLA gene. We still have many unanswered questions as to the mechanisms of selective neuronal degeneration in neurodegenerative diseases. The availability of a full-length DRPLA cDNA will be useful for investigating many hypotheses proposed for the potential functions of poly-glutamine tracts. The full-length cDNA will also be useful for creating animal models.

Acknowledgments

This study was supported in part by a Grant-in-Aid for Scientific Research on Priority Areas and a Grant-in-Aid for Creative Basic Research (Human Genome Program) from The Ministry of Education, Science, and Culture Japan, a grant from the Research Committee for Ataxic Diseases, The Ministry of Health and Welfare, Japan, special coordination funds of the Japanese Science and Technology Agency and a grant from the Uehara Memorial Foundation.

References

Adams MD, Dubnick M, Kerlavage AR, Moreno R, Kelley JM, Utterback TR, Nagle JW, et al (1992) Sequence identification of 2,375 human brain genes. Nature 355:632–634

Amemiya Y, Miyahara J (1988) Imaging plate illuminates many fields. Nature 336:89–90

Andrew SE, Goldberg YP, Kremer B, Telenius H, Theilmann J, Adam S, Starr E, et al (1993) The relationship between trinucleotide (CAG) repeat length and clinical features of Huntington disease. Nat Genet 4:398–403

Andrew SE, Goldberg YP, Theilmann J, Zeisler J and Hayden MR (1994) A CCG repeat polymorphism adjacent to the CAG repeat in the Huntington disease gene: implications for diagnostic accuracy and predictive testing. Hum Mol Genet 3:65–67

Ashley CT, Sutcliffe JS, Kunst CB, Leiner HA, Eichler EV, Nelson DL, Watson ST (1992) Human and murine FMR-1: alternative splicing and translational initiation downstream of the CGG-repeat. Nat Genet 4:244–251

Banfi S, Servadio A, Chung M, Kwiatkowski TJ Jr, McCall AE, Duvick LA, Shen Y, et al (1994) Identification and characterization of the gene causing type 1 spinocerebellar ataxia. Nat Genet 7:513–520

Blochlinger K, Bodner R, Jack J, Jan LY, Jan YN (1988) Primary structures and expression of a product from cut, a locus involved in specifying sensory organ identity in Drosophilia. Nature 333:629–635

Chang C, Kokontis J, Liao S (1988) Molecular cloning of human and rat complementary DNA encoding androgen receptors. Science 240:324–326

Chen EY, Seeburg PH (1985) Supercoil sequencing: a fast and simple method for sequencing plasmid DNA. DNA 4:165–170

Chirgwin JM, Przybyla AE, MacDonald RJ, Rutter WJ (1979) Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry 18:5294–5299

Duyao M, Ambrose C, Myers R, Novelletto A, Persichetti F, Folstein S, Ross C, et al (1993) Trinucleotide repeat length instability and age of onset in Huntington's disease. Nat Genet 4:387–392

Feinberg AP, Vogelstein BA (1984) A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Anal Biochem 137:266–267

Gerber HP, Seipel K, Georgiev O, Hefferer M, Martin H, Rusconi S, Schaffner W (1994) Transcriptional activation modulated by homopolymeric glutamine and proline stretches. Science 263:808–811

Haynes SR, Mozer BA, Bhatia-Dey N, Dawid IB (1989) The Drosophila fsh locus, a maternal effect homeotic gene, encodes apparent membrane proteins. Dev Biol 134:246–257

He WW, Fischer LM, Sun S, Bilhartz DJ, Zhu X, Young CYF, Kelley DB, et al (1990) Molecular cloning of androgen receptors from divergent species with a polymerase chain reaction technique: complete cDNA sequence of the mouse androgen receptor and isolation of androgen receptor cDNA probes. Biochem Biophys Res Commun 171:697–704

Higashijima S, Kojima T, Michiue T, Ishimaru S, Emori Y, Saigo K (1992) Dual Bar homeo box genes of Drosophilia required in two photoreceptor cells, R1 and R6, and primary pigment cells for normal eye development. Genes Dev 6:50–60

Hirayama K, Takayanagi T, Yanagisawa N (1990) Ataxic diseases (spinocerebellar degenerations and Shy-Drager syn-

drome): epidemiologic study. Ann Rep Res Comm Ataxic Dis, pp 46–54

Hoogeveen AT, Willemsen R, Meyer N, de Rooij EK, Roos AC, Bvan Ommen GJ, Galjaard H (1993) Characterization and localization of the Huntington disease gene product. Hum Mol Genet 2:2069–2073

Huntington's Disease Collaborative Research Group, The (1993) A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. Cell 72:971–983

Igarashi S, Tanno Y, Onodera O, Yamazaki M, Sato S, Ishikawa A, Miyatani N, et al (1992) Strong correlation between the number of CAG repeats in androgen receptor genes and the clinical onset of features of spinal and bulbar muscular atrophy. Neurology 42:2300–2302

Ikeuchi T, Koide R, Tanaka H, Onodera O, Igarashi S, Takahashi H, Kondo R, et al (1995) Dentatorubral-pallidoluysian atrophy (DRPLA): clinical features are closely related to unstable expansion of trinucleotide (CAG) repeat. Ann Neurol 37:769–775

Inazaki G, Kumagai K, Naito H (1990) Dentatorubral–pallidoluysian atrophy (DRPLA): its distribution in Japan and prevalence rate in Niigata. Seishin-Igaku 32:1135–1138

Jansen G, Mahadevan M, Amemiya C, Wormskamp N, Segers B, Hendriks W, O'Hoy K, et al (1992) Characterization of the myotonic dystrophy region predicts multiple protein isoform-encoding mRNAs. Nat Genet 1:261–266

Kawaguchi Y, Okamaoto T, Taniwaki M, Aizawa M, Inoue M, Katayama S, Kawakami H, et al (1994) CAG expansions in a novel gene for Machado-Joseph disease at chromosome 14q32.1. Nat Genet 8:221–228

Koch CA, Anderson D, Moran MF, Ellis C, Pawson T (1991) SH2 and SH3 Domains: elements that control interactions of cytoplasmic signaling proteins. Science 252:668–674

Koide R, Ikeuchi T, Onodera O, Tanaka H, Igarashi S, Endo K, Takahashi H, et al (1994) Unstable expansion of CAG repeat in hereditary dentatorubral-pallidoluysian atrophy (DRPLA). Nat Genet 6:9–13

Kozak M (1987) An analysis of 5′-noncoding sequences from 699 vertebrate messenger RNAs. Nucleic Acids Res 15:8125–8148

Kyte J, Doolittle RF (1982) J Mol Biol 157:105–132

La Spada AR, Roling DB, Harding AE, Warner CL, Spiegel R, Petrusweicz IH, Yee WC, et al (1992) Meiotic stability and genotype-phenotype correlation of the trinucleotide repeat in X-linked spinal and bulbar muscular atrophy. Nat Genet 2:301–304

La Spada AR, Wilson EM, Lubahn DB, Harding AE, Fischbeck H (1991) Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. Nature 352:77–79

Li S-H, McInnis MG, Margolis RL, Antonarakis SE, Ross CA (1993) Novel triplet repeat containing genes in human brain: cloning, expression, and length polymorphisms. Genomics 16:572–579

Li S-H, Schilling G, Young WS III, Li XJ, Margolis RL, Stine OC, Wagster MV, et al (1993) Huntington's disease gene (IT 15) is widely expressed in human and rat tissues. Neuron 11:985–993

Lin B, Nasir J, MacDonald H, Hutchinson G, Graham RK, Rommens JM, Hayden MR (1994) Sequence of the murine Huntington disease gene: evidence for conservation, and polymorphism in a triplet (CCG) repeat alternate splicing. Hum Mol Genet 3:85–92

Murray V (1989) Improved double-strand DNA sequencing using the linear polymerase chain reaction. Nucleic Acids Res 17:8889

Nagafuchi S, Yanagisawa H, Ohsaki E, Shirayama T, Tadokoro K, Inoue T, Yamada M (1994*a*) Structure and expression of the gene responsible for the triplet repeat disorders, dentatorubral and pallidoluysian atrophy (DRPLA). Nat Genet 8:177–182

Nagafuchi S, Yanagisawa H, Sato K, Shirayama T, Ohsaki E, Bundo M, Takeda T, et al (1994*b*) Expansion of an unstable CAG trinucleotide on chromosome 12p in dentatorubral and pallidoluysian atrophy. Nat Genet 6:14–18

Naito H (1990) Progressive myoclonus epilepsy and DRPLA. Adv Neurol Sci 34:56–67

Naito H, Oyanagi S (1982) Familial myoclonus epilepsy and choreoathetosis: hereditary dentatorubral-pallidoluysian atrophy. Neurology 32:798–807

Orr HT, Chung M, Banfi S, Kwiatkowski TJ Jr, Servadio J, Beaudet AL, McCall AE, et al (1993) Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1. Nat Genet 4:221–226

Ross CA, McInnis MG, Margolis RL, Li S-H (1993) Genes with triplet repeats: candidate mediators of neuropsychiatric disorders. Trends Neurosci 16:254–260

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467

Short JM, Fernandez JM, Sorge JA, Huse WD (1988) λ ZAP: a bacteriophage λ expression vector with in vivo excision properties. Nucleic Acids Res 16:7583–7600

Smith JK, Gonda VE, Malamud N (1958) Unusual form of cerebellar ataxia: combined dentato-rubral and pallido-luysian degeneration. Neurology 138:2605–2609

Squitieri F, Andrew SE, Goldberg YP, Kremer B, Spence N, Zeisler J, Nichol K, et al (1994) DNA haplotype analysis of Huntington disease reveals clues to the origins and mechanisms of CAG expansion and reasons for geographic variations of prevalence. Hum Mol Genet 3:2103–2114

Strong TV, Tagle DA, Valdes JM, Elmer LW, Boehm K, Swaroop M, Kaatz KW, et al (1993) Widespread expression of the human and rat Huntington's disease gene in brain and nonneural tissues. Nat Genet 5:259–265

Wahl GM, Stern M, Stark GR (1979) Efficient transfer of large DNA fragments from agarose gels to diazobenzyloxymethyl-paper and rapid hybridization by using dextran sulfate. Proc Natl Acad Sci USA 76:3683–3687

Wickens M, Stephenson P (1984) Role of the conserved AAUAAA sequence: four AAUAAA point mutants prevent messenger RNA 3′ end formation. Science 226:1045–1051

Yazawa I, Nukina N, Hashida H, Goto J, Yamada M, Kanazawa I (1995) Abnormal gene product identified in hereditary dentatorubral-pallidoluysian atrophy (DRPLA) brain. Nat Genet 10:99–103

Yu H, Chen JK, Feng S, Daligarno DC, Brauer AW, Schreiber SL (1994) Structural basis for the binding of proline-rich peptides to SH3 domains. Cell 76:933–945