

BBAEXP 92429

Identification of a developmentally regulated gene in the mouse central nervous system which encodes a novel proline rich protein

Takashi Sazuka ^{a,c}, Yasuhiro Tomooka ^b, Sandeep Kathju ^d, Yoji Ikawa ^{a,c},
Makoto Noda ^f and Sharad Kumar ^f

^a Laboratory of Molecular Oncology and ^b Cell Biology, Institute of Physical and Chemical Research (RIKEN), Tsukuba Life Science Center, Tsukuba (Japan), ^c Institute of Biological Sciences, University of Tsukuba, Tsukuba (Japan), ^d Department of Human Genetics, University of Michigan Medical School, Ann Arbor (USA), ^e Department of Biochemistry, Tokyo Medical and Dental University School of Medicine, Tokyo (Japan) and ^f Department of Viral Oncology, Cancer Institute, Kami-Ikebukuro, Tokyo (Japan)

(Received 21 May 1992)

Key words: CNS development; Proline rich protein; Down-regulation of expression; Gene expression

A full length cDNA whose corresponding mRNA is down-regulated during the mouse embryonic brain development was isolated. The cDNA contains a single long open reading frame which could encode a protein with relative molecular mass of 41 kDa. The predicted gene product contains long stretches of prolines towards the NH₂-terminus, followed by a leucine/proline rich region. The cDNA probe detected a number of mRNA species in Northern blot analysis. The reverse transcriptase-polymerase chain reaction analysis of mRNA from adult mouse tissues indicated that heart and testis expressed this gene (named *NDPP-1*) at relatively high levels, while lower levels of mRNA were detected in a number of other tissues. Expression of *NDPP-1* was also detected in embryonic carcinoma and pheochromocytoma cell lines, but not in fibroblasts. The cDNA hybridized to genomic DNA from several vertebrates species in Southern blot analysis indicating interspecies conservation of this gene. The interesting pattern of expression of the *NDPP-1* gene during mouse brain development and the structure of its putative protein product indicate that this gene may play an important biological role in the development of mouse central nervous system.

Introduction

Formation of mouse central nervous system (CNS) starts at embryonic day 8 (E8) when the neural plate is formed which eventually folds to form the neural tube around E10 [1]. In this process, multipotential neural precursor cells (NPC), derived from the cells in the neural plate differentiate into distinct cell lineages of the neurons and the glial cells and finally give rise to a number of different cell types, including various types of neurons, astrocytes and oligodendrocytes [2–4]. The efforts to understand the mechanisms which bring about the differentiation, development and lineage determination of CNS, have only recently begun to yield interesting information. The recent advances in cell culture techniques of NPC [5] provided the way to study these mechanisms at the molecular level. Isolation and characterization of the unknown genes whose expressions are temporally regulated during the differ-

entiation of the NPC is an important step towards elucidating the mechanism of CNS development. Here we report the isolation of a novel gene *NDPP-1* from neural precursor cDNA library, whose expression is developmentally down-regulated in mouse brain.

Materials and Methods

Cell culture

NPC and mouse embryo fibroblasts (MEF) were prepared and maintained as described before [5,6]. PC12 cell line was maintained in DMEM containing 10% fetal calf serum (FCS) and 5% horse serum. PCC4 (embryonal carcinoma) and NIH/3T3 were maintained in alpha-MEM and DMEM, respectively, supplemented with 10% FCS. To induce differentiation, PC12 cells were treated with 0.1 µg/ml of nerve growth factor (NGF) for indicated period, or 1 mM dibutyl cyclic AMP (db cAMP) and 1 µM phorbol-12-myristate 13-acetate (PMA) for 2 days, while PCC4 cells were treated with 1 µM retinoic acid for 3 days. NGF, db cAMP and PMA all were from Sigma.

Screening of cDNA library

A unidirectional cDNA library was constructed from oligo(dT)-primed poly(A)⁺ RNA from NPC in lambda

Correspondence to: S. Kumar, Department of Viral Oncology, Cancer Institute, 1-17-1, Kami-Ikebukuro, Toshima-ku, Tokyo, Japan 170.

The nucleotide sequence reported in this paper will appear in the DDBJ, EMBL and GenBank databases under accession number D10727.

Uni-Zap™ XR vector (Stratagene) according to the instructions supplied by the manufacturer with some modifications [7]. Plaques were screened by ³²P-labeled HLH oligo DNA probe (see below) according to standard protocols [8]. Hybridizations were carried out at 40°C in a solution containing 6 × SSC (20 × SSC is 3 M NaCl and 0.3 M tri-sodium citrate, pH 7.0), 5 × Denhardt's solution (100 × Denhardt's solution is 2% polyvinylpyrrolidone, 2% bovine serum albumin and 2% ficoll 400), 0.25% SDS, 20 mM sodium phosphate (pH 6.5) and 200 μg/ml sheared and denatured salmon sperm DNA. Filters were washed first at room temperature followed by at 35°C in 3 × SSC, 0.1% SDS prior to autoradiography.

DNA sequencing and computer analysis

Bluescript SK(−) was rescued from CL-1 phage with R408 helper phage (Stratagene). Nested deletions of this plasmid were generated by a ExoIII deletion kit (Pharmacia). Some of the DNA fragments were sub-cloned into M13mp18 and 19 or Bluescript SK(+). Both double-strand sequencing (Bluescript) and single-strand sequencing (M13) were performed using Sequenase (USB) according to the protocols supplied by the manufacturer. Ambiguous regions were sequenced again using 7-deaza sequencing protocol (USB). Computer analyses were performed using the Wisconsin Genetics Computer Group program package. Homology search was performed using FASTA and TFASTA algorithms [9].

Isolation of RNA and Northern blotting

Tissues were dissected and frozen quickly in liquid nitrogen. After homogenization with polytron, poly(A)⁺ RNA were isolated using poly(A)⁺ RNA extraction kit according to the protocol provided by the manufacturer (Fast Track, In Vitrogen). RNA samples were electrophoresed on 1.2% agarose/2.2 M formaldehyde gels, and transferred to Biodyne A nylon membrane (Pall) [8]. Northern hybridizations were carried at 42°C in 50% formamide, 5 × SSPE (20 × SSPE is 3 M NaCl, 0.2 M NaH₂PO₄ and 20 mM EDTA, pH 7.0), 5 × Denhardt's solution, 0.1% SDS and 100 μg/ml sheared and denatured salmon sperm DNA. Final washing was at 60°C in 0.2 × SSC, 0.1% SDS. Exposure, analysis and printing were performed using BAS 2000 Bioimage Analyzer (Fuji). The quantity and the integrity of the RNA samples were monitored by rehybridizing the same blot with a oligo DNA probe for the mouse glyceraldehyde-3-phosphate dehydrogenase (GA-3-PDH) gene [10] under the same conditions.

RT-PCR

Tissues were quickly dissected and frozen in liquid nitrogen. The frozen tissues were homogenized with polytron in guanidium thiocyanate-containing buffer,

and their total RNAs were isolated by centrifugation through a CsCl cushion [8]. RT-PCR was performed using the total cellular RNA as template according to the protocol provided by the manufacturer (Perkin-Elmer Cetus). In brief, total RNA (1 μg) was annealed with oligo(dT) primer, reverse transcribed at 42°C for 45 min, denatured at 94°C for 5 min, and then cooled to 5°C. PCR reactions were carried out for 20 cycles (each cycle being 95°C, for 1 min and 60°C, for 1 min). One-fifth of each PCR product was electrophoresed through a 1.5% agarose gel, transferred to Biodyne A and hybridized with ³²P-labeled 0.48 kb *Pst*I-*Xho*I fragment of the *NDPP-1* cDNA (see Fig. 1) [8]. To estimate the efficiency of the RT-PCR reaction with each RNA sample, β-actin mRNA was detected under the same conditions. β-Actin primers for PCR were designed from the cDNA sequence [11] (see below for the primers used).

Southern analysis

Genomic Southern analysis was performed using interspecies Zoo-Blot (Clontech) using standard protocols [8]. Filters were hybridized at 42°C in a solution containing 50% formamide, 5 × SSPE, 5 × Denhardt's solution, 0.1% SDS and 100 μg/ml sheared and denatured salmon sperm DNA. Final washing was carried out at 65°C in 0.2 × SSC, 0.1% SDS.

Oligonucleotides

The oligonucleotides used in these experiments were as follows.

- HLH probe, ATGTA(G/C)(G/C)GITGGCIITIII(G/C)AGGATCICACCTTIGG [12].
- GA-3-PDH, ATACGGCCAAATCCGTTACACCGACCTTACCAT.
- NDPP-1A, TCAGTCGTCGTCTCGCAGGAATAG (nucleotide numbers 3978–4001 in Fig. 2).
- NDPP-1B, AGACATGTGTAAATATTAGACC (nucleotide numbers 4303–4280 in Fig. 2).
- actinoligo1, ATGGATGACGATATCGCTGCG (nucleotide numbers 81–101) [11].
- actinoligo2, CTCGGTGAGCAGCACAGGGTG (nucleotide numbers 401–381) [11].

All oligonucleotides were synthesized with Applied Biosystems DNA Synthesizer and purified on OPC columns (Applied Biosystems).

Results

cDNA isolation

Since helix-loop-helix (HLH) containing transcription factors are known to play important role in the development and differentiation of *Drosophila* nervous system [13], we were interested in isolating this family of genes from mouse NPC and study their functions in the development of CNS. Two such genes have been isolated from mouse sympathetic neuronal precursor

cell line using a PCR technique [14]. We used an oligonucleotide probe designed on the basis of HLH region of *myc* and *myoD* proteins [12] to screen $2 \cdot 10^5$ plaques from a cDNA library constructed from NPC [7]. After two cycles of hybridization screening, four positive clones were obtained. Further analysis indicated that two of these clones hybridized strongly to a *myc* gene probe under stringent conditions and were omitted. The remaining two clones CL-1 and CL-2 contained 4.3 kb and 1.7 kb cDNA inserts, respectively. Initial sequence analysis of these clones indicated that both represented novel genes. In the present paper we describe the characterization of CL-1. To our surprise, the sequence of the predicted gene product of CL-1 did not show any possible HLH structures, while the region which possibly hybridized to the HLH probe was mapped as indicated in the Figs. 1 and 2. We have pursued the study of this clone because it encoded a possible protein of unusual structure and the expression of this gene showed an interesting pattern during the development of CNS.

Structure of NDPP-1 cDNA

The nucleotide sequence of CL-1 cDNA is shown in Fig. 2. Total length of this cDNA was 4334 bp including poly(A) tail. The largest open reading frame (ORF) is 1167 bp, which begins from a possible initiation codon (nucleotide position 864–866). The ORF is flanked by 5' and 3' untranslated sequences of 863 bp and 2304 bp respectively, and can encode a possible polypeptide of 389 amino acids. Estimated molecular weight of this protein is 41300. The NH₂-terminal portion (amino acid position 2–52) of this putative protein is extremely rich in prolines. Because of the characteristic structure of its product, we have named this gene *NDPP-1* (NPC derived proline rich protein-1). The NDPP-1 protein contains two hydrophobic regions, one towards NH₂-terminus (amino acid position 30–91), which contains the proline clusters, and another leucine/proline rich region located in the middle of the protein (amino acid position, 163–195). There is a short basic region consisting of four arginines (amino acid position, 273–276). Although there are possible alpha helices near the basic region, these structures are different from typical HLH structure, and do not show significant homologies with other HLH peptide sequences in the database (Swiss Prot). The extreme COOH-terminal region (amino acid position, 348–386) is also capable of forming an alpha helix. Possible glycosylation sites are indicated in Fig. 2. The 3' untranslated region contains a repeat sequence cluster, (GT)₂₂(GC)₆cg(AC)₆N₁₂(GA)₇, (N represents any nucleotide). A polyadenylation signal-like sequence (AATATT) is located towards the end of the sequence (nucleotide position, 4285–4288) while a typical polyadenylation signal (AATAAA) is found just down-

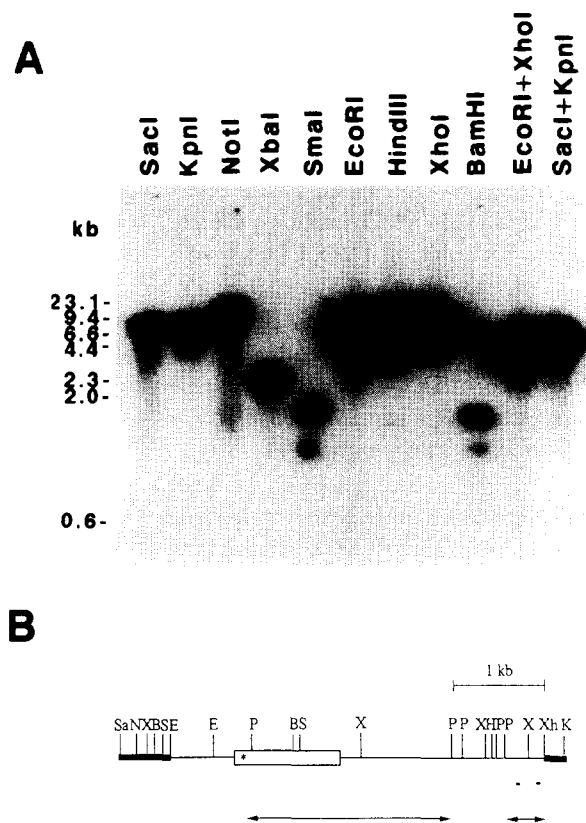


Fig. 1. (A) Hybridization of HLH oligonucleotide probe to the CL-1 clone. Half a μg of plasmid DNA per lane, digested with the indicated restriction enzyme, resolved by 1% agarose gel electrophoresis was transferred to nylon membrane and hybridized to end labeled HLH oligonucleotide probe at 50°C in a solution containing $6 \times \text{SSC}$, $5 \times \text{Denhardt's solution}$, 0.25% SDS, 20 mM sodium phosphate (pH 6.5) and 200 $\mu\text{g/ml}$ of sheared and denatured salmon sperm DNA. Washing was performed in $2 \times \text{SSC}$ at room temperature. (B) Physical map of the CL-1 insert. Filled boxes indicate the flanking regions of the insert derived from pBluescript (SK-) vector, while the open box indicates the open reading frame. Restriction sites; Sa, *SacI*; N, *NotI*; X, *XbaI*; B, *BamHI*; S *SmaI*; E, *EcoRI*; P, *PstI*; H, *HindIII*; Xh, *XhoI*; and K, *KpnI*. The position of asterisk (*) indicates the location of the possible sequence which hybridized to HLH probe shown in (A). The short bars underneath the right-hand side of the map indicate the location of primers used for RT-PCR. The long arrow indicates the 2.1 kb *PstI* fragment used as probe in Northern analyses, while the short arrow indicates the *PstI-XhoI* fragment used for hybridization with RT-PCR products.

stream of ORF (nucleotide position, 2362–2367). The 3' untranslated region also contains seven 'ATTTA' motifs. The same motifs have been found in some mRNAs with short half lives, such as the mRNAs for GMCSF, IL-1, TNF, c-fos and c-myc [15–17], and they are known to bind to a protein which causes destabilization of mRNA [18]. Thus, the presence of these sequences in *NDPP-1* mRNA may indicate that this gene is also post-transcriptionally regulated.

Homology of NDPP-1 gene product with known proteins

Database search (Swiss prot) indicated that NDPP-1 protein shared partial homology with many proteins,

```

1 GAAATTCGGCACGAGAGAACAAGGAGGAACGAAAGGAACAAGGCCAAGGGTACTGTGGGGAAGAGGGAGATGAGGGGGAGAGCAACCAGA 90
91 GACTCCATCTGAATAAAAATTTGCATGAAAATATCATATGGCAGCTGGTTACATTGTTGGATAAATCAAACATGTAAAATAAAATATCT 180
181 TTTAAAAGATAAAAAGGAATAGGTGTATAGACGATGAGAGAATAAAGATAGATAAAGAGTGAGGATGAGTAAGAGACACATTAAGAGAAGTG 270
271 TGATCATGGACTGAACCTTAGGAAAGTGAATCACTGACTCCAGCTCAAACAACCTGCTTTCAGTTTCTGTGAAAGAAGAATGCCACTCAC 360
361 ATTCCAGGTATAGTTAAATGCTAGGTGCATATCGCCTAATTCGATAACCTACTGGAGTGTGTAAACAGGAATGCCGTGTGCATTAGTTATG 450
451 CATAATTTGAATAAAAATTAAGAGCAGCTTTTGCATATGCTTGCCAAGTAGTTTGTTCAGAGAAGCTGATTAATTACATATAAATTAATACTA 540
541 GATGGATTTTCAGAAACACATGAACACACATGCATTTTCTATGGAAATGAATTTCTGCCTAAACTAAAACCTAAACCCTGTAAAGAGAATTTA 630
631 CTGACCTTTCTTCTCCTCTAATCTGAGTCAATGTCGATTTGCCGATTTGCATTTGGAGGATAACTTTGACTCCATGATTCCTGTTTGCCAGTTACTT 720
721 GGTGAGGTCACAGAGAAGAAGTTACAGGATGGAATACATTTGCTCGATGGGGTTGCTCATGCAGGTTGTTAAATCAGGATCGTGCC 810
811 TTATTCAGTGTATTCTCTCTTCTGGCATCACTGTCTCCTCTCTGTCGCCCAAATGCCTCCGTCACCAACAGCACCCAAATGGGTCGCTAG 900
M P P S P T A P (N) G S L D 13
901 ACTCTGTAACATACCCAGTGTCTCCACCGCCTACCTCAGGGCCAGCAGGCCACCTCCGCCGCCACC GCCACC GCCCCGCCACCACCAC 990
14 S V T Y P V S P P P T S G P A A P P P P P P P P P P P P P P P P P P 43
991 CAGCGCTGCCACCGCCGCTGCGCTCCGCTCCGCTCCTCACTCTCACACTGTGGATCACAGGCTTCTCCTCCTCCAGGCACCCCTCTTGCC 1080
44 P L P P P P L P P L A S L S H C G S Q A S P P P G T P L A S 73
1081 CAACTCCCTCATCAAGCCAGTGTCTCCCTTCTCCCTCTGCGAGCTGCCCTGCTGCGGAGACCCCTCTAAATCTGAGCTGGGAG 1170
74 T P S S K P S V L P S P S A A A P A S A E T P L N P E L G D 103
1171 ACTCCTCTGCTTCCGAGCCAGGCTTGCAGGCAGCCTCTCAGCCGGCCGAGTCGCCAACCCACAGGGCTTGTCTTGGGACCACCTGCACC 1260
104 S S A S E P G L Q A A S Q P A E S P T H R A L S W D H L H L 133
1261 TCCGCCACCACCCTCTCCCATCAGGCCCTGCCTACGCCCTCAGCACTTCTCCTCCCCACAGACCCTCCACCACCTCCACTGCCATC 1350
134 R H H P L S H Q A L P T P Q H F L L P Q D P L H H L H C H P 163
1351 CACTGGTCTCTCCACCCTCCACCCTCCACCCTTCTCTAATCAAGCTCCTCCCTCCTCCCCACCTCCTGCCCTCCCTCCCTCC 1440
164 L V L L L L H P L H H L P L F L L K L L P L L P H L P L P S P 193
1441 CGCATCTGGAATTTTCTGGAATCCACGTGAGAAGCAATCGCCTTAACTGGACTTGCAGCTGCAATTCGGGAGCAAACCTTAGGAAAG 1530
194 H L E F S L D P R Q K T I A F (N) W T C S C N C G S K L R K V 223
1531 TGCCCCGGTGGAGGATGGCTCTTCCCAGGTGGAGGAATACTGGGAGTGTGAGCTTGGCCTCATCCAAGCAGACGCTGGGCGTGGGA 1620
224 S R V E D G S F P G G G N T G S V S L A S S K A D A G R G N 253
1621 ATGGACTCTCTCAGGGGTAGTGGCTTAATGGAAGAAATGAGTGGCCTGCTGGCCAGGAGGAGAAGAATGTGAGAGAAGGGATCAA 1710
254 G P L P L G G S G L M E E M S A L L A R R R R I A E K G S T 283
1711 CAATAGAAACAGAAACAAAGGAGACAGAATGAAGATGCAGAGCCTATAACTGCTAAGGCCCATCAACAAGTACACCTGAACCAACCA 1800
284 I E T E Q K E D R N E D A E P I T A K A P S T S T P E P T R 313
1801 GAAAACCTTGGGAAAGAACAAACACAATGAACGGCAGTAAGTCACCTGTCACTCCAGACCCAAATCCACACCTTCATCACGCCAAGTG 1890
314 K P W E R T N T M (N) G S K S P V I S R P K S T P S S Q P S A 343
1891 CCAATGGAGTCCAGACAGAAGGCTTGAATGACAGGCTGAAGCAGGACATTTTAGATGAGATGAGAAAGAGCTGGCAAAGCTGAAGG 1980
344 N G V T E G L D Y D R L K D Y D I L D E M R K E L A K L K E 373
1981 AGGAGCTTATTGACGCAATCAGGCAGGAGCTGAGCAAGTCGAACACTGCATAAAGAAGCAACTACGGAGGGGCAAGGACTTGAATCTGGA 2070
374 E L I D A I R Q E L S K S N T A *
2071 GAAAACAAAATTCCTACAAACAACCTTTAACCCCAAACCTTTAAAGCTGTAAGAAGAAAATGGATACACAGTCAAGGAGGAAAGCCGTCA 2160
2161 ACCTCTGAAAAGCCTCAGACAGTGACTCTGGCGATCAGCTGTCCCCTCAGTGTGCTGCTTTATTCTGTCTGACCTTACCACAGGATGGAG 2250
2251 AATGATATTGGAGTTCCTTAGCAGTACTAACCCGTCAGGCAAGATCACCGTGCATTGAAATATTTTCATGTCTAGATGAGTCTGCACG 2340
2341 TTTTCCATAATCCTATGCTAAATAAGACAGGAAAGGCTCAGAGATTTTTTCCAGAGGGCGCTGATGAAGAATTTGATTTAGTAAGT 2430
2431 TCTGCTGTGCGTTGTAATGTTTCTCAGGTTGTCTTCCATCATATTTGATATTTCCGTAAGACTAAAGATCAAAAGGATGGAACA 2520
2521 AACGGAGTTGATGTGCTCAGAGGGTGGTGTAAATACGAAAGGTCCTAAAATGTTTGTGCTCCACTTGAGGAAATTTAGAAATGATAGGA 2610
2611 AGTCTCTCGGGTTAGCCTTATGCAATTTGTAGTTTAAAACATAAAGTCCGTCAGAAATTAAGAGATTTAGCTTCTCTAAATTTGT 2700
2701 ACAATGCTTTACCAATCTATGACTTCTACATAACCAACAGTGGTCAAATGGAAAACATTTCTATATAGATTTACTGTAGGTTTTCAACC 2790
2791 TTTTTAGACATATGCATGTGGACATTTTTATAATGTAATCAATCACCACAAAGTTAGCTTTTTTAATTACAGACAATAATGCATGTC 2880
2881 ACCTAATATGTAGTGCCCTTTTTCAGGGCTAGTCAGGGAATAATTTGTAGAAATAGGGAAAGTGGGAGGAAAGGAAATCACTTT 2970
2971 TATTTAAAGTTAATTTCTGCAATCTTTTCTCAATTTCTGCATGAATTAATGAAAATATTTTGTGACTTTAATCAGTAACCAT 3060
3061 GTTAACAAAACCAAGTACTTAATCTTTTACATCATGTCTTTCAGCTATTTGATTTTTAACCCAGCAATTTCAACGGTCTGAAACATGATTTCT 3150
3151 GAGCGTAACATGTGAGTCCGCTCACTGTGGACTACGAGTTTGTACTACTGAATTGGCAGTTATTTTACCGCGGCCCCACAGTGTGCACT 3240
3241 GCTTGGGTTACGCTCCTGACTCTGCGGCTGCTCCTTGACCTTTGCTGTGCTAAAATACAAGAAATATCAATTAATGGCAGCAACCAGGGT 3330
3331 CACATAAGTCACTAGGTAAGGGAGACCAATGAGGAGTCTGCAGCTTTCTTTAATAGAATAAAGTGAAGGCTTAAAGTGGTAGGGAGTGT 3420
3421 CTGCTGTACACCCTGAGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 3510
3511 CACACAGGTGAGAGAGAGAGAGAAAGTGATCGAGCTTAAGAGTGTGTAGATAGATATTATGTTCCACCCCTTTATAATTGGCAAGATTA 3600
3601 ATAAGGGAGAAAATTTATATGTTGTAATGTTAGCTTTTTGGAAAATCTAGGAAATCAAAAATTTCTCCAGGCTGTCCATCATGATTTATG 3690
3691 CTGAGTTAATAATGTCCATATTTGCTTTGGAACTTTGGCTATCAAACCTTTACTAAAAAATAGTAGATACATCCACTTGATTGCTTTC 3780
3781 TAGATTCCTATGTCTCAGCTGAAGCAAGCTTGATCAGAGTGTTCCTAAACCCCTGCAGTTTGACAGTAACCCACTGCAGAGCCGTTGT 3870
3871 GAATGAGGGTTTTGTAAAGAGTGTGTTTCTGTAGAACTTTCATCTTTAGATTAAAGTGTGAGGGATGAGCGGTCATAAGTCAATTTCT 3960
3961 GTGGCTTTCTCACCCTCAGTCTGCTCGCAGGAAATAGTAAAGGCATTAAAGGATGCGGCAAGCATGTACGCTGTTCTAAAAGAAGTA 4050
4051 TGCATCTCATTACTGTTGAACGTTTGATTACGCAACTACTTTGATTGTGTACGTGTGTACATCCCTTACAGATCTGATCAAAAATCGCT 4140
4141 TTTCAATTTGTTGCTAATGTTAAATAGTCTGTAGTACGTAGAGTACTGGGCTAGATGGTCCAAGTGAAGACAAAATGCTGAAGTCAAA 4230
4231 TGCTAAGTCTAAACAAGGCTGGTTTATGCATCAAACGTTTGTGCCCTGGCT(AATATT)AACACAATGTCTGTAAAAATGACAAAAA 4320
4321 AAAAAAAAAAAAAA 4334

```

Fig. 2. Complete nucleotide sequence of *NDPP-1* cDNA. Sequence was obtained from both strands. The putative protein coding sequence is indicated. The circles show the possible glycosylation sites. The possible sequence which hybridized to HLH probe is double underlined. The ATTTA sequences found in the 3' noncoding region are underlined, while the polyadenylation signal AATAAA and similar sequence AATATT are boxed. Location of two oligonucleotides, used in the RT-PCR studies are also indicated (underlined with arrow-heads to show the orientation).

most of which have the proline rich domains. Some of these homologies extend beyond the proline cluster into the surrounding regions. The proteins which shared relatively high degree of homology with NDPP-1 protein (excluding the ones homologous in proline-rich region alone) include nuclear proteins of EBNA-2 [19] (29.3% in 174 amino acid overlap), EBNA-3 [19] (18.5% in 205 amino acid), circumsporozoite protein [20] (26.0% in 204 amino acids), human breakpoint cluster region protein [21,22] (19.0% in 142 amino acid), RNA replicase polyprotein [23] (24.3% in 185 amino acids), *Drosophila* homeobox protein *cut* [24] (17.6% in 176 amino acids), synapsin IA and IB [25] (29.2% in 147 amino acids), and mouse homeobox protein Hox-2.6 [26] (41.6% in 77 amino acids) (percentage shows identity). The homology search suggested that *NDPP-1* does not belong to any one particular class of genes already described.

Expression of *NDPP-1* during mouse brain development

5 μ g of poly(A)⁺ RNA isolated from embryonic head (E10 and E11) or brain (E13 through adult) at the indicated stages was hybridized to a 2.1 kb *Pst*I fragment of *NDPP-1* on Northern blots. Overall expression of *NDPP-1* in head or brain was consistently low and only faintly detectable in short exposures (e.g., exposure of imaging plates equivalent to an overnight exposure to conventional X-ray films). However, the long exposures showed that the probe hybridized to a number of different mRNA species of various sizes (Fig. 3). We also noted that some of the hybridizing bands were broad and lacked the sharp boundaries. This is not due to the non-specific decomposition of RNA samples because the same blot gave sharp bands when rehybridized with other probes such as GA-3-PDH (Fig. 3). We do not know if these multiple hybridizing bands represent differentially spliced forms of mRNA originated from the same gene or the transcripts from the several closely related genes. At E10 stage one major mRNA species of 4.3 kb and a relatively weaker one of 3.8 kb can be seen among which the 4.3 kb transcript probably represents the *NDPP-1* clone. At E13 stage a larger and prominent band of approximately 5.4 kb, in addition to the 4.3 kb and 3.8 kb species is also observed (Fig. 3). The steady state levels of all these mRNA species are developmentally down-regulated in the brain being only faintly detectable in the adult brain. A very weak but detectable band of approx. 2.3 kb can be seen at stage E10 which also diminishes progressively during development (Fig. 3)

Expression of *NDPP-1* in PC12 and PCC4 cells

Since the level of expression of *NDPP-1* transcripts was developmentally regulated in mouse brain, we examined the possibility that the changes reflect the differentiation of neuronal cells by using two cell lines

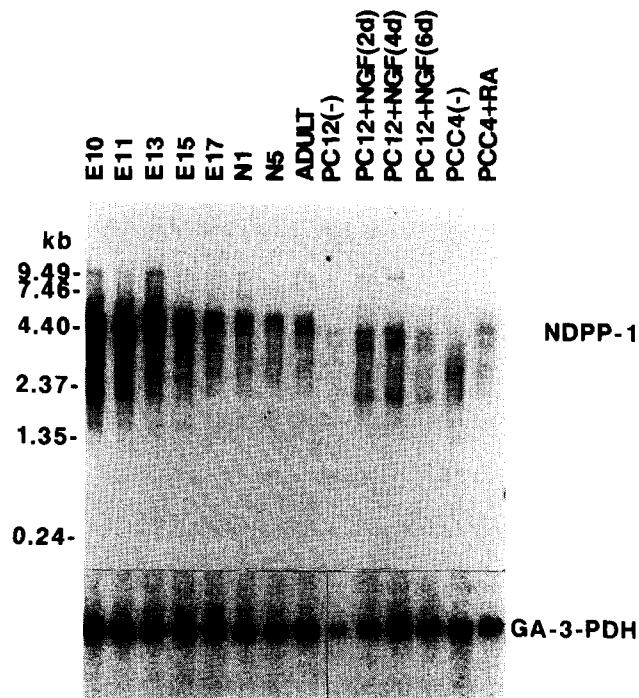


Fig. 3. Northern blot analysis of *NDPP-1* gene expression. Each lane contained 5 μ g of poly(A)⁺ RNA isolated from indicated tissue or cell lines. E, embryonic; N, postnatal; NGF, nerve growth factor; RA, retinoic acid; (-), no treatment. The upper panel shows the hybridization to *NDPP-1* probe, while the lower panel to (GA-3-PDH) which was used as a control to show integrity of RNA.

capable of differentiating into neuronal cells upon induction. One of the cell lines, rat PC12, differentiates into neurons, astrocytes and to a lesser extent into fibroblastoid cells, when induced by NGF [27,28]. PC12 cells were cultured in the absence or presence of 0.1 μ g/ml NGF for 0, 2, 4 or 6 days and 5 μ g of poly(A)⁺ RNA extracted from the cells at each time point was analysed by Northern blot hybridization. The *NDPP-1* probe (2.1 kb, *Pst*I fragment) could detect three faint bands (4.3 kb, 3.8 kb, 2.3 kb) at all time points (Fig. 3). The treatment of PC12 cells with NGF had little or no effects on the intensities of these bands. The other cell line, PCC4, differentiates into many types of cells including neurons after induction by 1 μ M retinoic acid. The expression levels of *NDPP-1* were very low in PCC4 cells and it was difficult to judge if the retinoic acid treatment affected the intensities of the mRNA bands (Fig. 3).

RT-PCR analysis of *NDPP-1* expression

Northern blot analysis of *NDPP-1* expression showed consistently low levels of transcripts in embryonic brain and cell lines. For this reason, we used quantitative RT-PCR technique to analyse *NDPP-1* mRNA. Total RNA was transcribed by reverse transcriptase to generate cDNA, and the cDNA was used as a template for

PCR. A pair of primers (NDPP-1A, NDPP-1B, see Materials and Methods), derived from the 3' untranslated region of *NDPP-1* were used. The expected size of PCR product is 326 bp. The conditions for PCR reactions were adjusted in such a way that the amount of the product would be proportional to the RNA input. To confirm that the RT-PCR products were correctly and specifically amplified from *NDPP-1* mRNA, we examined the size of the products by Southern blot analysis using *Pst*I-*Xho*I fragment as a probe (see Fig. 1). The analysis showed that NIH/3T3, MEF, and PC12 cells expressed *NDPP-1* at low levels (Fig. 4). Since PC12 cells were derived from rat, it is possible that the apparent low-level expression reflects the differences in the *NDPP-1* sequence between mouse and rat, which would result in poorer amplification during the earlier steps of PCR reaction. In the EC and the PCC4 cells, the products were readily visible. Consistent with earlier observation (Fig. 3), treatment of PC12 and PCC4 cells with differentiation inducers failed to show any profound effect on the levels of *NDPP-1* mRNA. We could eventually detect positive bands in all samples when RT-PCR were carried out for 35 cycles (data not shown), suggesting that all the

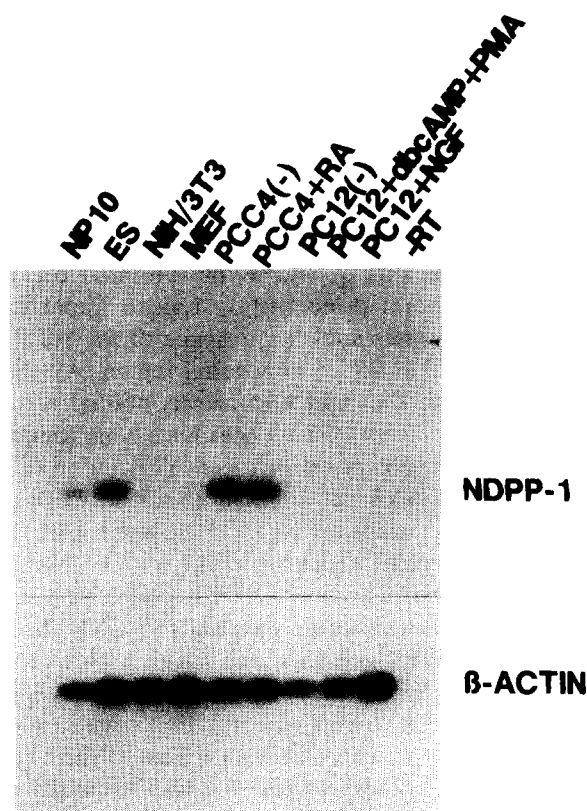


Fig. 4. Detection of *NDPP-1* expression in several cell lines by RT-PCR analysis. PCR products were electrophoresed on two identical gels, transferred to nylon membranes and hybridized to *NDPP-1* and β -actin probes as described in the Materials and Methods. The expected size of both *NDPP-1* and β -actin PCR products were 326 bp.

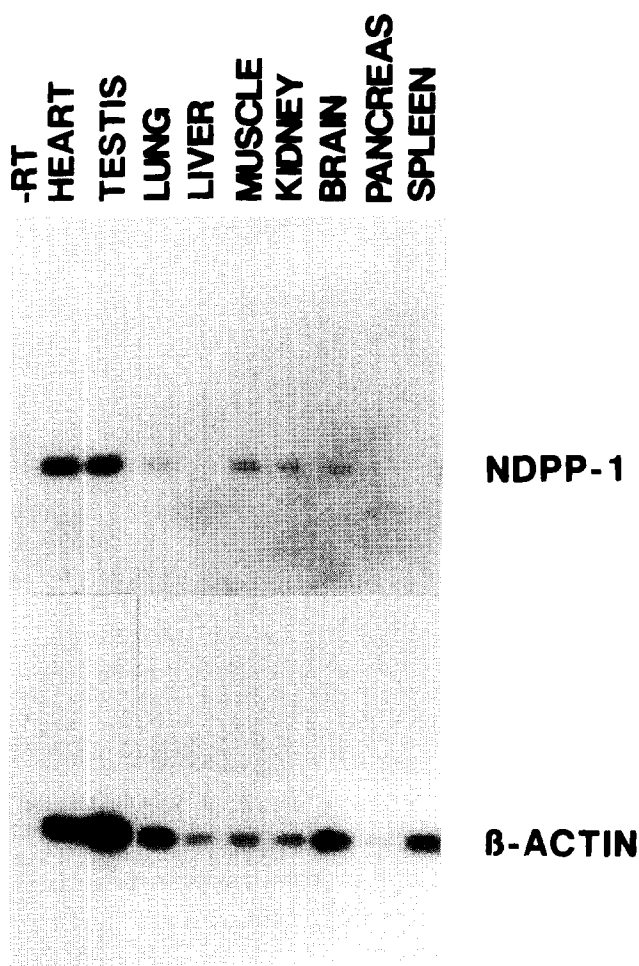


Fig. 5. Detection of *NDPP-1* expression in the adult mouse tissues. For details see Materials and Methods and legend to Fig. 5.

cell lines we studied expressed *NDPP-1*, albeit at very low levels. Only very weak signals were obtained when poly(A)⁺ RNA from adult mouse tissues was hybridized to the *NDPP-1* probe on Northern blots (data not shown). We therefore used the RT-PCR technique to examine the tissue distribution of the *NDPP-1* transcripts (Fig. 5). After Southern blot analysis of RT-PCR products, *NDPP-1* expression was detected in a number of tissues, more prominently in heart and testis and less so in lung, skeletal muscle, kidney, pancreas and brain.

Interspecies conservation of the NDPP-1 gene

To assess if the *NDPP-1* gene is structurally conserved across species, genomic DNA from several organisms were hybridized on Southern blot to the *NDPP-1* probe under stringent conditions (Fig. 6). The results indicated a high degree of conservation of *NDPP-1* among mouse (15.3 kb, 13.5 kb), rat (17.4 kb, 12.9 kb, 11.7 kb, 1.2 kb) and human (14.7 kb, 3.9 kb, 2.4 kb). Weaker but detectable hybridization was observed with monkey, dog, cow, rabbit and chicken genomic DNA. The intensities of the hybridizing bands proba-

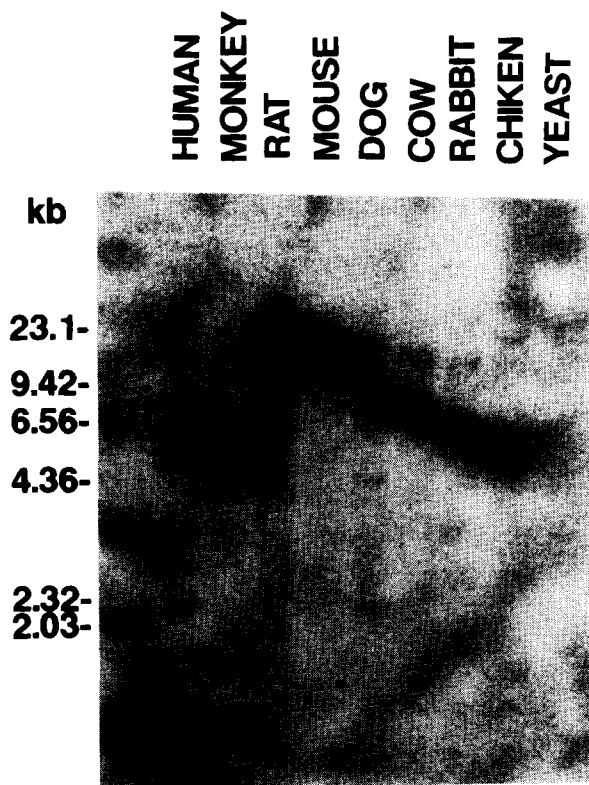


Fig. 6. Inter-species Southern-blot analysis of *NDPP-1*. Commercially available inter-species Zooblot (Clontech), containing *Eco*RI digested genomic DNA from several species, was hybridized with *NDPP-1* probe under stringent conditions as described in the Materials and Methods.

bly represent the varying degrees of homology between the *NDPP-1* sequence of mouse and other vertebrates. No hybridization to yeast DNA was observed under these conditions (Fig. 6).

Discussion

In order to elucidate the mechanisms which bring about the differentiation and development of the CNS derived multipotent precursor cells [5], we have been trying to isolate the genes which are predominantly expressed in these cells (or embryonic brain). To achieve this, we have used two different strategies; one is to isolate the mouse homologues of genes known to play role in *Drosophila* neurogenesis (e.g., see Refs. 13, 14) and the other to isolate cDNA clones by subtracting the NPC library with postnatal and adult brain mRNA [7]. In the first approach, we have tried to isolate the HLH class of genes which play important role in the *Drosophila* neurogenesis [13]. The *NDPP-1* gene described in this paper, had partial homology with the HLH oligonucleotide probe we had used, but did not contain any HLH structure in its predicted protein. However, further studies with the expression of *NDPP-1*, reported in this paper suggested that this

gene might play an important biological role in the development and differentiation of CNS.

Probably the most interesting feature of *NDPP-1* gene is the progressive down-regulation of its expression during the CNS development. The multiple species of transcripts we detect on the Northern blots, very likely represent either the differentially spliced forms of mRNA originating from a single gene or, the presence of closely related species. There are however two lines of support for the first prediction. First, the Southern blot analysis of mouse genomic DNA shows two equal intensity bands, ruling out the presence of several related genes. The second line of evidence comes from our recent isolation of a partial cDNA (3.6 kb in length) which contains identical structure to *NDPP-1* cDNA except the 117 bp sequence at its 5' end. The changed sequence at the 5' end results in the elongation of the reading frame (open at the 5' terminus). It appears therefore likely that this and other mRNA species hybridizing to *NDPP-1* probe, represent alternately spliced forms transcribed from a single gene. The presence of two possible polyadenylation signals present in the 3' noncoding region of *NDPP-1* may also contribute to the heterogeneity observed in the sizes of mRNA. In addition to developing CNS, the *NDPP-1* transcripts can also be detected in certain adult tissues. We do not yet know however, whether the levels of *NDPP-1* mRNA are developmentally regulated in these tissues. To ascertain this we are currently carrying out in situ mRNA analysis using the embryo sections.

Besides the regulated pattern of *NDPP-1* expression, the cDNA has several characteristic features, like the presence of long 5' and 3' untranslated regions, presence of several di-nucleotide repeats capable of forming palindromes and the presence of the ATTTA sequence motifs in the 3' noncoding region. All these features point to the tight post-transcriptional regulation of *NDPP-1*. Repeat sequences have been found in the regulatory region of rat *myc* gene [29] and in the 3' noncoding region of rat calmodulin-sensitive plasma membrane Ca^{2+} transporting ATPase gene [30]. The tri-nucleotide repeat sequence CTG is also found in the 3' noncoding region of cDNA encoded by a gene implicated in the myotonic dystrophy [31–33]. In the myotonic dystrophy patients, the CTG triplet repeat undergoes expansion, thereby causing possible genetic instability [31–33]. Analogous situation occurs in the fragile X-syndrome where a CGG repeat at the 5' end of a gene undergoes expansion [34]. Although the mechanistic basis of the genetic instability in these syndromes remain unclear, these examples clearly indicate that the nucleotide repeat sequences may play a role in the phenomena. It may be relevant to point out here that the subcloning of certain regions of *NDPP-1* cDNA, specially the ones containing the di-nucleotide

repeats, have been increasingly difficult. Also amplification of certain regions of cDNA by PCR have been unsuccessful. We speculate that the repeat sequences present in the *NDPP-1* probably cause the genetic instability and technical difficulties we observe.

The role of the ATTTA motifs present in the 3' noncoding regions of a number of short-lived regulatory gene transcripts, in destabilising the mRNA is well documented [15–18] and as seven of these are present in the 3' noncoding region of the *NDPP-1*, its transcript would be expected to have short half-life. This and the other structural features discussed above, all clearly point to the stringent control of *NDPP-1* gene.

At the level of protein, *NDPP-1* can potentially encode a peptide of approx. 41 kDa, which has a proline rich NH₂-terminus followed by a leucine/proline rich middle region. The search for proteins with homology to putative *NDPP-1* gene product resulted in the extraction of a large number of peptide data base entries, majority of which contained proline rich regions. Interestingly, some of these proteins are involved in neuronal differentiation. For example, *Drosophila* homeobox protein *cut* (absence of which causes the transformation from external sensory organs into chordotonal organs) has proline clusters locate in the COOH-terminal [24]; mouse homeobox protein Hox-2.6 (which is expressed in CNS at the early embryonic stage) has proline clusters in the middle of the protein [26]; synapsins IA and IB (which are peripheral membrane proteins of synaptic vesicles which mediate their attachment to the cytoskeleton) contain proline rich region in the both NH₂- and COOH-terminus [25].

The data presented in this report on the *NDPP-1* cDNA, its putative protein structure and developmentally regulated expression point to some function in development and/or differentiation of CNS. Further characterization of the expression, protein structure and biological activity of the gene product will shed light on the possible physiological role of this gene.

Acknowledgements

This work was supported by grants from RIKEN and Science and Technology Agency of Japan (STA). S. Kumar was the recipient of a STA fellowship while S. Kathju was the holder of a vacation scholarship funded by STA. Both S. Kumar and M. Noda were previously at RIKEN. We thank H. Kitayama for the synthesis of oligonucleotides.

References

1 Spemann, H. (1938) *Embryonic Development and Induction*, Yale University Press, New Haven.

- 2 Fujita, S. (1963) *J. Comp. Neurol.* 120, 37–42.
- 3 Jacobson, M. (1978) *Developmental Neurobiology*, Plenum Press, New York.
- 4 McKay, R.D.G. (1989) *Cell* 58, 815–821.
- 5 Kitani, H., Shiurba, R., Sakakura, T. and Tomooka, Y. (1991) *In Vitro Cell. Dev. Biol.* 27A, 615–624.
- 6 Wadhwa, R., Ikawa, Y. and Sugimoto, Y. (1991) *Biochem. Biophys. Res. Commun.* 178, 269–275.
- 7 Kumar, S., Tomooka, Y. and Noda, M. (1992) *Biochem. Biophys. Res. Commun.* 185, 1155–1161.
- 8 Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor University Press, Cold Spring Harbor.
- 9 Pearson, W.R. and Lipman, D.J. (1988) *Proc. Natl. Acad. Sci. USA* 85, 2444–2448.
- 10 Sabath, D., Broome, H.E. and Prystowsky, M.B. (1990) *Gene* 91, 185–191.
- 11 Tokunaga, K., Taniguchi, H., Yoda, K., Shimizu, M. and Sakiyama, S. (1986) *Nucleic Acids Res.* 14, 2829–2829.
- 12 Benezra, R., Davis, R.L., Lockshon, D., Turner, D.L. and Weintraub, H. (1990) *Cell* 61, 49–59.
- 13 Vaessin, H., Caudy, M., Bier, E., Jan, L.-Y. and Jan, Y.-N. (1990) *Cold Spring Harbor Symposia on Quantitative Biology*, Volume LV, pp. 239–245.
- 14 Johnson, J.E., Birren, S.J. and Anderson, D.J. (1990) *Nature* 346, 858–861.
- 15 Brawerman, G. (1987) *Cell* 48, 5–6.
- 16 Mullner, E.W. and Kuhn, L.C. (1988) *Cell* 53, 815–825.
- 17 Wilson, T. and Treisman, R. (1988) *Nature* 336, 396–399.
- 18 Vakalopoulou, E., Schaack, J. and Shenk, T. (1991) *Mol. Cell. Biol.* 11, 3355–3364.
- 19 Baer, R., Bankier, A.T., Biggin, M.D., Deininger, P.L., Farrell, P.J., Gibson, T.J., Hatfull, G., Hudson, G.S., Satchwell, S.C., Seguin, C., Tuffnell, P.S. and Barrell, B.G. (1984) *Nature* 310, 207–211.
- 20 Eichinger, D.J., Arnot, D.E., Tam, J.P., Nussenzweig, V. and Enea, V. (1986) *Mol. Cell. Biol.* 6, 3965–3972.
- 21 Hariharan, I.K. and Adams, J.M. (1987) *EMBO J.* 6, 115–119.
- 22 Mes-Masson, A.M., McLaughlin, J., Daley, G.Q., Paskind, M. and Witte, O.N. (1986) *Proc. Natl. Acad. Sci. USA* 83, 9768–9772.
- 23 Keese, P., Mackenzie, A. and Gibbe, A. (1989) *Virology* 172, 536–549.
- 24 Blochlinger, K., Bodmer, R., Jack, J., Jan, L.-Y. and Jan, Y.-N. (1988) *Nature* 333, 629–635.
- 25 Südhof, T.C. (1990) *J. Biol. Chem.* 265, 7849–7852.
- 26 Graham, A., Papalopulu, N., Lorimer, J., Mcvey, J.H., Tuddenham, E.G.D. and Krumlauf, R. (1988) *Genes Dev.* 2, 1424–1438.
- 27 Greene, L.A. and Tischler, A.S. (1976) *Proc. Natl. Acad. Sci. USA* 73, 2424–2428.
- 28 Sugimoto, Y., Noda, M., Kitayama, H. and Ikawa, Y. (1988) *J. Biol. Chem.* 263, 12102–12108.
- 29 Hayashi, K., Makino, R., Kawamura, H., Arisawa, A. and Yoneda, K. (1987) *Nucleic Acids Res.* 15, 6419–6436.
- 30 Greeb, J. and Shull, G.E. (1989) *J. Biol. Chem.* 264, 18569–18576.
- 31 Brook, J.D., McCurrach, H.G., Buckler, A.J., Church, D., Aburatani, H., Hunter, K., Stanton, V.P., Thirion, J.-P., Hudson, T., Sohn, R., Zemelmann, B., Snell, R.G., Rundle, S.A., Crow, S., Davies, J., Shelbourne, P., Buxton, J., Jones, C., Juvonen, V., Johnson, K., Harper, P.S., Shaw, D.J. and Housman, D.E. (1992) *Cell* 68, 799–808.
- 32 Fu, Y.-H., Pizzuti, A., Fenwick, R.G. Jr., King, J., Pizzuti, A., Fenwick Jr., R.G., King, J., Rajnarayan, S., Dunne, P.W., Dubel, J., Nasser, G.A., Ashizawa, T., de Jong, P., Wieringa, B., Korneluk, R., Perryman, M.B., Epstein, H.F. and Caskey, C.T. (1992) *Science* 255, 1256–1258.

- 33 Mahadevan, M., Tsilfidis, C., Sabourin, L., Shulter, G., Amemiya, C., Jansen, G., Neville, C., Narange, M., Barcelo, J., O'Hoy, K., Leblond, S., Earle-Macdonald, J., De Jong, P.J., Wieringa, B. and Korneluk, R.G. (1992) *Science* 255, 1253–1256.
- 34 Fu, Y.-H., Kuhl, D.P.A., Pizzuti, A., Pieretti, M., Sutcliffe, J., Richards, S., Verkerk, A.J.M.H., Holden, J.J.A., Fenwick, R.G. Jr., Warren, S.T., Oostra, B.A., Nelson, D.L. and Caskey, C.T. (1991) *Cell* 67, 1047–1058.