Molecular cloning and nucleotide sequence of a cDNA clone coding for rat brain myelin proteolipid

André Dautigny, Patrick M. Alliel, Luc d'Auriol+, Danièle Pham Dinh, Jean-Louis Nussbaum*, Francis Galibert+ and Pierre Jollès

Laboratoire des Protéines, Université de Paris V, 45, rue des Saints-Pères, 75270 Paris Cedex 06, +Laboratoire d'Hématologie Expérimentale, Centre Hayem, Hôpital Saint-Louis, 2 place du Docteur Fournier, 75475 Paris Cedex 10 and **Institut de Neurochimie du CNRS, Laboratoire de Chimie Biologique, 67084 Strasbourg Cedex, France*

Received 12 June 1985

A cDNA library from rat brain was constructed in pBR322 and screened with a 14-mer mixed oligonucleotide probe based on residues 231–235 of bovine proteolipid (PLP). A positive clone was isolated: it contained a 1334-base-pair cDNA insert and was subjected to DNA sequence analysis. The cDNA encoded information for the 276 amino acids of rat PLP. Comparison with bovine PLP sequence showed a complete amino acid sequence homology except for 4 amino acid residues.

Rat brain *Myelin* *Proteolipid* *Oligonucleotide probe* *cDNA sequence*

1. INTRODUCTION

Proteolipid-apoprotein of myelin (PLP) represents a major protein constituent of central nervous myelin [1]. It is highly hydrophobic with a molecular mass of 26 kDa [2]. The complete amino acid sequence of the bovine PLP [2,3] and partial structural data of the rat PLP [4] have been reported. This paper deals with the cloning of the cDNA sequence complementary to rat PLP mRNA. Such a cDNA probe represents a powerful tool for the study of the molecular mechanisms involved in myelogenesis.

2. MATERIALS AND METHODS

2.1. *Isotopes and enzymes*

L-[^{35}S]Methionine, [α-^{32}P]dATP and [γ-^{32}P]ATP were obtained from New England Nuclear and oligo(dT)-cellulose from P-L Biochemicals. Reverse transcriptase was from Life Sciences and all other enzymes were purchased from Boehringer, Mannheim.

2.2. *Synthesis of a mixed oligonucleotide probe*

A mixture of eight 14-mer oligonucleotide probes:

$$3'\ CT^{C}_{T}AA^{G}_{A}GT^{C}_{T}TACTG\ 5'OH$$

corresponding to all possible codons for amino acids 231–235 of the known amino acid sequence of the bovine proteolipid [2,3] was synthesized by the solid-phase phosphite triester method [5] with an automated DNA synthesizer (Biosearch). Oligonucleotides were purified by gel electrophoresis and labeled at their 5'-end with [γ-^{32}P]ATP and polynucleotide kinase [6].

2.3. *Construction of recombinant plasmids and transformation*

Synthesis of double-stranded cDNA was accomplished by the procedure of Wickens et al. [7]. Ligation to $(dC)_n$-tailed *Pst*I-cleaved pBR322 was performed according to [6]. The annealed molecules were used to transform *Escherichia coli* strain C600 [8].

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/85/$3.30

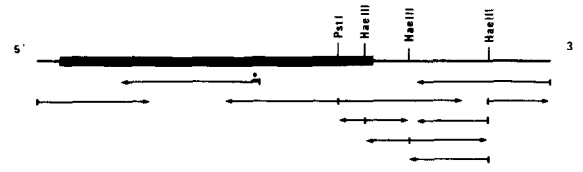

Fig.1. Sequence strategy of the rat PLP cDNA clone. The heavy line identifies the protein coding region, the thin lines denote 5'- and 3'-untranslated regions. Arrows indicate the direction and the extent of sequence determination. * indicates the position of the second primer:

3' CGACACGGACACATGTAAATGAAG 5'

Only the positions of relevant restriction sites are indicated.

```
                                                                                CCG AAA AAA AGA CTA GCC AGC AGC TGC AAT TGG AGC CAG AGT GCC AAA GAC ATG

GGT TTG TTA GAG TGC TGT GCT AGA TGT CTG GTA GGG GCC CCC TTT GCT TCC CTG GTG GCC ACT GGA TTG TGT TTC TTT GGA GTG GCA TTG
Gly Leu Leu Glu Cys Cys Ala Arg Cys Leu Val Gly Ala Pro Phe Ala Ser Leu Val Ala Thr Gly Leu Cys Phe Phe Gly Val Ala Leu
                                10                                  20                                  30

TTC TGT GGA TGT GGA CAT GAA GCT CTC ACT GGC ACA GAA AAG TTA ATT GAG ACC TAT TTC TCC AAA AAC TAC CAG GAC TAT GAG TAT CTC
Phe Cys Gly Cys Gly His Glu Ala Leu Thr Gly Thr Glu Lys Leu Ile Glu Thr Tyr Phe Ser Lys Asn Tyr Gln Asp Tyr Glu Tyr Leu
                                40                                  50                                  60

                                                                                                                    *
ATT AAT GTG ATT CAT GCT TTC CAG TAT GTC ATC TAT GGA ACT GCC TCT TTC TTC TTC CTT TAT GGG GCC CTC CTG CTG GCC GAG GGC TTC
Ile Asn Val Ile His Ala Phe Gln Tyr Val Ile Tyr Gly Thr Ala Ser Phe Phe Phe Leu Tyr Gly Ala Leu Leu Leu Ala Glu Gly Phe
                                70                                  80                                  90

TAC ACC ACC GGC GCT GTC AGG CAG ATC TTT GGC GAC TAC AAG ACC ACC ATC TGC GGC AAG GGC CTG AGC GCA ACG GTA ACA GGG GGC CAG
Tyr Thr Thr Gly Ala Val Arg Gln Ile Phe Gly Asp Tyr Lys Thr Thr Ile Cys Gly Lys Gly Leu Ser Ala Thr Val Thr Gly Gly Gln
                               100                                 110                                 120

AAG GGG AGG GGT TCC AGA GGC CAA CAT CAA GCT CAT TCT TTG GAG CGG GTG TGT CAT TGT TTG GGA AAA TGG CTA GGA CAT CCC GAC AAG
Lys Gly Arg Gly Ser Arg Gly Gln His Gln Ala His Ser Leu Glu Arg Val Cys His Cys Leu Gly Lys Trp Leu Gly His Pro Asp Lys
                               130                                 140                                 150

TTT GTG GGC ATC ACC TAT GCC CTG ACT GTT GTA TGG CTC CTG GTG TTT GCC TGC TCT GCT GTG CCT GTG TAC ATT TAC TTC AAT ACC TGG
Phe Val Gly Ile Thr Tyr Ala Leu Thr Val Val Trp Leu Leu Val Phe Ala Cys Ser Ala Val Pro Val Tyr Ile Tyr Phe Asn Thr Trp
                               160                                 170                                 180

                    *                                                   *
ACC ACC TGC CAG TCT ATT GCC TTC CCT AGC AAG ACC TCT GCC AGT ATA GGC AGT CTC TGC GCT GAT GCC AGA ATG TAT GGT GTT CTC CCA
Thr Thr Cys Gln Ser Ile Ala Phe Pro Ser Lys Thr Ser Ala Ser Ile Gly Ser Leu Cys Ala Asp Ala Arg Met Tyr Gly Val Leu Pro
                               190                                 200                                 210

TGG AAT GCT TTT CCT GGC AAG GTT TGT GGC TCC AAC CTT CTG TCC ATC TGC AAA ACA GCC GAG TTC CAA ATG ACC TTC CAC CTG TTT ATT
Trp Asn Ala Phe Pro Gly Lys Val Cys Gly Ser Asn Leu Leu Ser Ile Cys Lys Thr Ala Glu Phe Gln Met Thr Phe His Leu Phe Ile
                               220                                 230                                 240

                                        *
GCT GCA TTT GTG GGG GCT GCA GCC ACA CTA GTT TCC CTG CTC ACC TTC ATG ATT GCT GCC ACT TAC AAC TTT GCC GTC CTT AAA CTC ATG
Ala Ala Phe Val Gly Ala Ala Ala Thr Leu Val Ser Leu Leu Thr Phe Met Ile Ala Ala Thr Tyr Asn Phe Ala Val Leu Lys Leu Met
                               250                                 260                                 270

GGC CGA GGC ACC AAG TTC TGA TCT CCC ATA GAA ATC CCC CTT TGT CTA ATA GCG AGG CTC TAA CCA CAC AGC CTA TAG TGT TGT GTC TCC
Gly Arg Gly Thr Lys Phe
                    276

TGT CTT AAC TCT GCC TTT GCC ACT GAT TGG CCC TCT TCT TAC TTG ATG AGT ATA ACA AGA AAG GAG AGT ATT GCA GTG ATT AAT CTC TCT

CTG TGG ACT CTC CCT CTT ATG TAC CTC TTT CAG TCA TTT TGC TCC ACA GTG GGC TCC TGC TAG AAA TGG GGA ATA CCT GAG AAG GTG ATT

CCC CCG CTG CAA GTC GCA GAG GAA TGA AAG CTC TAA TTG AGT TTG CAA GCA TCT CCT GAA GGC CAG GAT GTG CTT CCT TCT CAA AGG GCA

CTT CCA TTG AGG AAA ACA AAG TGG AAA GAA AGA TTC TCA GGT AGA AAG CAG GAA TGT CCT TGG TCT CCT TGC CAT CAG TAG GAG TCA AAT

ATA TTC TCT TTG ATG CAC AA
```

Fig.2. Nucleotide sequence of the P-23 insert and its predicted amino acid sequence. The amino acid residues are numbered beginning with the first residue of PLP. The overline indicates the region corresponding to the 14-mer mixed oligonucleotide probe. Boxes denote the initiation codon and the stop codon. *, amino acids differing between rat PLP (this work) and bovine PLP [2,3].

2.4. *Screening and preparation of recombinant plasmids*

Bacterial clones were plated and transferred to nitrocellulose filters [9]. The filters were prehybridized at 65°C for 4 h in: 6 × NET, 5 × Denhardt's solution, 0.5% Nonidet P-40, 100 μg/ml base-hydrolyzed yeast RNA [10]. Hybridization was performed overnight in the prehybridization solution containing 5′-end-^{32}P-labeled oligonucleotide mixture (2 × 10^6 cpm/ml).

Filters were washed at 42°C in 6 × SSC and subjected to autoradiography with an intensifying screen. Plasmid DNA from positive clones was digested with *Pst*I restriction endonuclease and analyzed on agarose gel. After transfer to nitrocellulose membrane [11] DNA was hybridized with the same 5′-labeled probe and washed as described above [10].

2.5. *DNA sequence analysis*

Sequence analysis was performed by the dideoxy method [12]. The primers used for sequencing were a universal primer purchased from Amersham and a specific 24-mer primer synthesized as described above.

3. RESULTS

3.1. *Construction and identification of PLP cDNA clones*

The rat brain cDNA library was screened using the mixed 14-mer oligonucleotide probe. From 4 independent libraries of 5000 tetracycline-resistant clones only one colony (P-23 clone) gave a positive signal after ordered replication. Plasmid DNA was prepared from the P-23 clone and cleaved with *Pst*I. Two fragments were obtained, an 800 bp and a 560 bp. Southern hybridization of *Pst*I-generated fragments from the positive clone revealed that only the 800 bp band hybridized to the mixed oligonucleotide probe.

3.2. *Sequencing strategy*

The 2 fragments were directly subcloned in the phage vector M13mp8 which had been linearized with *Pst*I. A detailed DNA sequence analysis of the P-23 insert was undertaken using the strategy presented in fig.1. The cDNA clone isolated by the above described procedure contained the sequence of the rat PLP cDNA (fig.2).

Table 1

Differences observed between the bovine PLP sequence [2,3] and the deduced rat PLP sequence

Amino acid residue number	Bovine PLP sequence	Deduced rat PLP sequence
88	Tyr	Glu
188	Ala	Phe
198	Thr	Ser
254	Val	Leu

The complete 1334 bp cDNA sequence contains a single open reading frame, beginning with the first encountered ATG codon at nucleotide 52 and ending with the stop codon TGA at position 883. This open reading frame codes for a protein of 276 amino acids starting with Gly according to the PLP amino acid sequence [4]. In the 3′-noncoding region neither poly(A) tail nor polyadenylation signals were detected most probably due to cloning defaults.

4. DISCUSSION

Here we report the isolation and nucleotide sequence of a cDNA clone encoding for the entire rat PLP protein. Comparison of the deduced amino acid sequence with that of bovine PLP protein shows a complete sequence homology except for 4 amino acids located at positions 88, 188, 198 and 254 (table 1). According to the model of Laursen et al. [13], residues 88, 188 and 198 are located in the extracellular and cytoplasmic spaces while residue 254 is located in the C-terminal region. Among these 4 differences between rat and bovine PLP, 2 amino acids, 188 and 198, were implicated in the tertiary structure of an antigenic determinant [14]. The location of the initiator methionyl codon would support the conclusion that PLP is synthesized as a precursor protein of 277 amino acids with a processing limited to the removal of the first methionine residue. It is noteworthy that the codon ATG is contained in the sequence $^{A}_{G}$XX-ATGG which is frequently found in the functional initiation site [15]. Nevertheless the presence of a signal or leader peptide as described for the major structural protein of peripheral myelin [16] and the proteolipid subunit of the mitochondrial ATP synthase [17] cannot be excluded because of the short

length of the 5′-cloned noncoding region. This cDNA clone should prove a valuable tool for investigating the structure, organisation and chromosomal location of the gene coding for PLP protein.

ACKNOWLEDGEMENTS

This work was supported by the CNRS (Unité no.102 alliée à l'INSERM and LP 101) and the INSERM (SC 15). The authors thank Claude Mugnier and Christian Fondrat (CITI 2) for use of their computer facilities.

REFERENCES

[1] Lees, M.B. and Brostoff, S.W. (1984) in: Myelin (Morell, P. ed.) pp.197–224, Plenum, New York.
[2] Stoffel, W., Hillen, H., Schröder, W. and Deutzmann, R. (1983) Hoppe-Seyler's Z. Physiol. Chem. 364, 1455–1456.
[3] Lees, M.B., Chao, B., Lin, L.-F.H., Samiullah, M. and Laursen, R.A. (1983) Arch. Biochem. Biophys. 226, 643–656.
[4] Jollès, J., Nussbaum, J.-L. and Jollès, P. (1983) Biochim. Biophys. Acta 742, 33–38.
[5] Alvarado-Urbina, G., Sathe, G.M., Liu, W.C., Gillen, M.F., Duck, P.D., Bender, R. and Ogilvie, K.K. (1981) Science 214, 270–274.
[6] Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) in: Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
[7] Wickens, M.P., Buell, G.N. and Schimke, R.T. (1978) J. Biol. Chem. 253, 2483–2495.
[8] Hanahan, D. (1983) J. Mol. Biol. 166, 557–580.
[9] Grunstein, M. and Hogness, D.S. (1975) Proc. Natl. Acad. Sci. USA 72, 3961–3965.
[10] Wallace, R.B., Johnson, M.J., Hirose, T., Miyake, T., Kawashima, E.M. and Itakura, K. (1981) Nucleic Acids Res. 9, 879–894.
[11] Southern, E.M. (1975) J. Mol. Biol. 98, 503–517.
[12] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[13] Laursen, R.A., Samiullah, M. and Lees, M.B. (1984) Proc. Natl. Acad. Sci. USA 81, 2912–2916.
[14] Lin, L.-F.H. and Lees, M.H. (1985) J. Biol. Chem. 260, 4371–4377.
[15] Kozak, M. (1981) Nucleic Acids Res. 9, 5233–5252.
[16] Lemke, G. and Axel, R. (1985) Cell 40, 501–508.
[17] Viebrock, A., Perz, A. and Sebald, W. (1982) EMBO J. 1, 565–571.