Biol. Chem. Hoppe-Seyler
Vol. 370, pp. 503–510, May 1989

The Organization of the Human Myelin Basic Protein Gene

Comparison with the Mouse Gene

Rüdiger STREICHER and Wilhelm STOFFEL

Institut für Biochemie, Medizinische Fakultät, Universität zu Köln

(Received 13 February 1989)

Summary: The organization of the gene of the human myelin basic protein has been established. The gene contains seven exons distributed over 32–34 kb. The nucleotide sequences of the seven exons, extensive parts at the 5′ noncoding region and upstream and downstream regions adjacent to the exons were sequenced. A comparison of the nucleotide sequence of the human gene with mouse MBP cDNA and parts of the gene structure of mouse MBP has been carried out. There is a high degree of conservation in the seven exons and the 5′-regulatory and part of the 3′-noncoding region: the nucleotide sequence of the coding region of the human and mouse gene show a 93% homology. As far as the reported 5′-noncoding and upstream regulatory regions of 478 bp are comparable these regions are more than 80% homologous. Primer extension experiments positioned the transcription initiation sites at −55, −82 and −183. In the 5′-regulatory region of MBP gene three direct repeats were found, a nonameric at −401/−416 and two octameric at −192/−216 and −492/−508. A decameric sequence (−256 to −265 from the translation start) is completely homologous to a sequence of the human PLP regulatory 5′-noncoding region.

Die Organisation des menschlichen basischen Myelinprotein-Gens. Vergleich mit dem Mäusegen

Zusammenfassung: Die Organisation des menschlichen Gens für das basische Myelinprotein wurde ermittelt. Es enthält sieben Exons, die über 32–34 kb verteilt sind. Nucleotidsequenzen der sieben Exons, große Teile der 5′-nichtkodierenden Region sowie die der den Exon-/Introngrenzen benachbarten Regionen wurden bestimmt. Ein Vergleich der Nucleotidsequenz des menschlichen Gens mit der Maus-MBP-spezifischen cDNA sowie der Genstruktur des Maus-MBP wies einen hohen Grad an Homologie in den sieben Exons und der 5′-regulatorischen und 3′-nichtkodierenden Region auf: die Nucleotidsequenzen der MBP-kodierenden Exonsequenzen von Mensch und Maus sind zu 93% homolog. Soweit die berichteten 5′-nichtkodierenden und regulatorischen Regionen vergleichbar sind, besteht eine mehr als 80%ige Übereinstimmung der Nucleotidsequenz. Primer-Extension-Analysen legten potentielle Transcriptions-Startpunkte bei −55, −82 und −183 fest. Die 5′-nichtkodierenden Regionen der MBP-Gene von Mensch und Maus sind zu etwa 80% homolog. In der 5′-regulatorischen Region des MBP-Gens befinden sich drei direkte Repeats, eine nonamere Sequenz bei −401/−416 und zwei oktamere bei −192/−216 und −492/−508. Eine decamere Sequenz (−256 bis −265) ist völlig homolog zu einer Sequenz des menschlichen PLP-5′-regulatorischen Bereichs.

Abbreviations:
MBP, myelin basic protein; PLP, proteolipid protein; MAG, myelin associated glycoprotein; pMBP3, 603 bp cDNA clone encoding the complete myelin basic protein sequence; $NaDodSO_4$, sodium dodecyl sulfate; AMV, avian myeloblastosis virus; DDT, dithiothreitol; DM 20, PLP-derived isoform with apparent molecular mass 20 kDa.

Key words: Genes of myelin proteins, myelin basic protein gene, exon-intron structure, homology of man and mouse genes, transcription initiation sites.

Oligodendroglia cells contribute mostly to the process of myelination in the central nervous system. This glial cell type and complementary the Schwann cell in the peripheral nervous system produce the unique and highly specialized myelin sheath as extensions of the plasma membrane targeting several axons. The tightly compacted, spirally wrapped multilayer membrane system consists of approximately 80% complex lipids and cholesterol and 20% of protein. It is formed in the developmentally regulated process of myelinogenesis which in rodents – such as mouse and rat – peaks around days 10 to 30 after birth.

This cell-specific unusual process programmed in time and topology leads to the highly compacted myelin sheaths around a large number of axons. This insulation allows the rapid saltatory conduction of impulses with depolarization limited to the small areas at the nodes of Ranvier. This insulating membrane system saves considerable energy otherwise needed for the repolarization of an unmyelinated axonal membrane. Among the myelin-specific proteins proteolipid protein and myelin basic protein constitute approximately about 80–90%. In addition myelin associated glycoprotein (MAG) and 2′,3′-cyclic nucleotide 3′-phosphodiesterase are present as less abundant myelin proteins.

The protein structures of MBP and PLP of several species have been elucidated. It has become clear by immunochemical[1,2] and topochemical studies[3,4] that MBP is a peripheral membrane protein and located in the cytosolic cleft. It induces the apposition of the cytoplasmic surfaces of the plasma membrane processes whereas the PLP is an integral membrane protein oriented to both sides of the lipid bilayers, contributing to the apposition of the extracytoplasmic surfaces[4]. The cytosolic cleft appears as main dense line (MDL), the extracytosolic space between the layers as intermediate dense line (IDL).

Molecular biology has taken an enormous impact recently on the study of gene expression of these myelin proteins. We are now in the position to answer questions addressing the assembly of this specialized membrane, how the expression of the genes of these two main myelin proteins is regulated and how myelinogenesis is perturbed by alterations of these genes leading phenotypically to dysmyelinating diseases[5].

Cloned PLP from rat[6–8], mouse[9,10], bovine brain[11] and its DNA-derived amino-acid sequence confirmed the previously established polypeptide sequence of bovine and human PLP. Our amino-acid sequence analysis revealed almost complete homology between bovine and human PLP[3,12]. The complete homology of the coding region of PLP was extended by the cDNA-derived PLP polypeptide sequence of rat brain. As a consequence the human proteolipid gene organization and chromosomal X-assignment was established[13,14]. Remarkably the individual hydrophobic sequences and the adjacent hydrophilic loops are encoded by individual exons.

Among the dysmyelinating genetic diseases of the mouse the genetic trait of the jimpy mouse[15] could be traced to a point mutation at the splice acceptor site (A→G) of intron IV which leads to the 74 bp loss of exon V with a frame shift and thereby a missense amino-acid sequence at the C-terminus with a shortened 243-amino-acid polypeptide chain[16]. Also the DM 20 isoform of PLP results from an alternate splicing within exon III and thereby deleting amino-acid residues 116–150[17,18].

Myelin basic protein (MBP), the peripheral myelin membrane protein, has been sequenced quite early[19,20]. Again the screening of cDNA libraries with protein-derived oligonucleotide probes led to the isolation of mouse[21,22] and rat[23] MBP-specific cDNA clones. MBP isoforms were the translation products of different mRNAs, five isoforms of mouse MBP[24] with molecular masses of 21.5, 18.5, two of 17 and 14 kDa, and four isoforms of human MBP with 21.5, 20, 18.5 and 17 kDa[25]. All these isoforms result from alternative splicing of the primary transcript of the single mouse MBP gene which is larger than 30 kb. In mouse and man the MBP gene has been allocated to the distal end of chromosome 18[26,27]. The isoforms of MBP in the myelin of mouse and rat and again dysmyelinating mouse mutants, the shiverer[28] and myelin-deficient mouse[29,30] could be analysed on the molecular level after its gene structure had been elucidated.

So far most of the structural studies were confined to the mouse gene. Here we report on the structure and organization of the human myelin basic protein (MBP) gene. The 5′ part was derived from a clone isolated from the pcos2

EMBL[31] and the 3′ part of the MBP gene from the Charon 4A genomic library[32].

We compared the human and the mouse MBP gene. An extensive homology exists in the nucleotide sequence and similarly in the gene organization of the MBP gene of the two species.

The human MBP gene contains seven exons which are separated by six introns of divergent length distributed over approximately 32–34 kb. All exons and large parts of their adjacent upstream and downstream intron sequences have been sequenced, particularly extensively the 5′-noncoding region.

Materials and Methods

The pcos2 EMBL human genomic library was kindly provided by *H. Lehrach*, EMBL, Heidelberg.

Isolation of genomic myelin basic protein clones

Two genomic clones of the human MBP gene were isolated and further characterized: MBPNterm from 2×10^5 colonies of the pcos2 EMBL library[31] and MBPCterm from 7.5×10^5 plaques of the Charon 4A library[32]. ^{32}P-Nick-labelled rat cDNA fragments of clone pMBP3[8] were used as hybridization probes.

Colonies were plated on nitrocellulose filters BA85 (Schleicher & Schüll, D-3354 Dassel) and replicas taken as described before[33].

Mapping of the genomic MBP clones

DNA of both genomic MBP clones was digested with the respective restriction enzymes by single and double digestion. The fragments were size-fractionated by gel electrophoresis (0.8% agarose) and 5% (w/v) polyacrylamide respectively with TBE as running buffer (0.089M Tris/borate, 0.089M boric acid, 2mM EDTA, pH 8.4). The DNA in the agarose gels was alkali-denatured (0.5M NaOH, 1.5M NaCl for 45 min), neutralized (0.5M Tris/HCl, pH 7.5, 1.5M NaCl for 45 min) and blotted on a Gene Screen Plus membrane (Du Pont NEN, D-6380 Braunschweig) following the manufacturer's prescription. The membrane was hybridized with the respective radioactive probes of 5′-[^{32}P]labelled oligonucleotides with specific activities of approximately 5×10^6 cpm/pmol at 1 pmol/*ml*.

Hybridization probes

Rat MBP cDNA from clone pMBP3[8] and restriction fragments isolated from the genomic MBP clones were nick-labelled by standard methods[33]. MBP-specific synthetic oligonucleotides, derived from the human cDNA sequence and our own sequence data, were 5′ end-labelled with [γ-^{32}P]ATP (> TBq/mmol, Amersham) in the presence of T4 polynucleotide kinase[34] (Bethesda Research Laboratories, D-6078 Neu-Isenburg).

Nitrocellulose filters were prehybridized for 6 h at 42 °C in 50% (v/v) formamide, 5 x Denhardt's solution (0.02% polyvinylpyrrolidone, 0.02% Ficoll, 0.02% bovine serum albumin), 5 x SSPE (1 x SSPE = 0.15M NaCl, 10mM NaH_2PO_4, 1mM EDTA, pH 7.4), 100 μg/*ml* sonified salmon sperm DNA and subsequently hybridized with the probe at 10^5 cpm/*ml*. Filters were washed in 2 x SSC (1 x SSC = 0.15M NaCl, 0.015M sodium citrate, pH 7.0), 0.1% $NaDodSO_4$ at 65 °C for 30 min and autoradiographed at − 70 °C overnight with intensifying screens.

DNA sequence analysis

Restriction fragments were size-fractionated by agarose gel electrophoresis (0.7% in TBE buffer), transferred onto DEAE filter membrane (NA45, Schleicher & Schüll) by electrophoresis, and eluted with 1M NaCl in the presence of 1% arginine[35]. Nucleotide sequences were determined by plasmid sequencing[36,37] and the chain-termination method[38] of restriction fragments cloned into the polylinker sequence of pUC13 with the M13 sequencing and M13 reverse-sequencing primer (PL-Pharmacia). MBP-specific hexadecameric primers were synthesized by the solid-phase phosphoamidite method on an automated Applied Biosystems synthesizer (model 380A, Foster City, CA 94404).

Total RNA of fetal brain (2. trimenon) was isolated by the guanidine-thiocyanate method[39] and human leucocyte DNA as described before[40].

Transcription initiation sites of human MBP gene determined by primer extension analysis of human brain RNA[41]

10 μg total human brain RNA was annealed with 5′-labelled 20^{mer} oligonucleotide (CATGGTACTTGCTTG-GGCCA) at 65 °C for 1 h, cooled to room temperature, extended with 5 units AMV reverse transcriptase in the presence of 0.25mM dNTP, 80mM KCl, 0.3mM EDTA, 18mM Tris/HCl, pH 8.3, 10mM $MgCl_2$ and 5mM DDT and incubated for 1 h at 37 °C.

The primer extension products were extracted with phenol and ethanol, precipitated with ethanol, resuspended in 80% formamide, 10mM NaOH, EDTA and analysed by 5% acrylamide, gel electrophoresis, 7M urea.

Results and Discussion

Human MBP genomic clone

Two genomic MBP clones were isolated from genomic libraries: MBPNterm from the pcos2 EMBL library and MBPCterm from the Charon 4A library. Approximately 2×10^5 colonies and 7.5×10^5 plaques, respectively, were screened with nick-labelled DNA from MBP cDNA clone pMBP2 and/or synthetic oligonucleotides which were kinased with [γ-^{32}P]ATP.

Characterization of genomic clones MBPNterm and MBPCterm

Restriction endonuclease digestions of the two genomic clones were carried out. By single and double digestions the restriction fragments were cleaved, subsequently separated by agarose gel

electrophoresis and further processed for Southern blot analysis with the labelled probes indicated in the table. The restriction map resulting from these analyses is shown in Fig. 1.

The hybridization analysis with a 42^{mer} oligonucleotide homologous to the N-terminal sequence (exon I) of the cDNA of MBP and the complete 603 bp MBP-specific cDNA revealed that the clone from the pcos2 EMBL library encodes the N-terminal domain and the Charon 4A MBP clone probed with the MBP cDNA and a 30^{mer} C-terminal nucleotide probe encodes the C-terminal end of MBP. The two clones were named MBPNterm and MBPCterm, respectively (cf. Fig. 1a).

The human gene encoding MBP spans a region of 32–34 kb. The restriction analysis and hybridization with the aforementioned probes yielded seven exons the first two of which were located in the 36 kb insert of clone MBPNterm and exons III to VII in the 12 kb insert of clone MBPCterm. The gap between exons II and III was approximately 3 kb.

The size of the gap in intron II was determined by hybridization analysis of *Xba* I restricted human genomic leucocyte DNA: a 7.5 kb *Xba* I fragment hybridized with a 4.4 kb *Hin*d III fragment from the 3′-end of MBPNterm and a 1.5 kb *Eco* RI/*Nco* I fragment containing the 5′ end of MBPCterm. The *Xba* I site of MBPNterm is 2.6 kb distant from the 3′ end of the insert and the *Xba* I site of MBPCterm is 1.6 distant from the 5′ end. Therefore the size of the gap between the 3′ and the 5′ ends of the genomic clones is 3.3 kb.

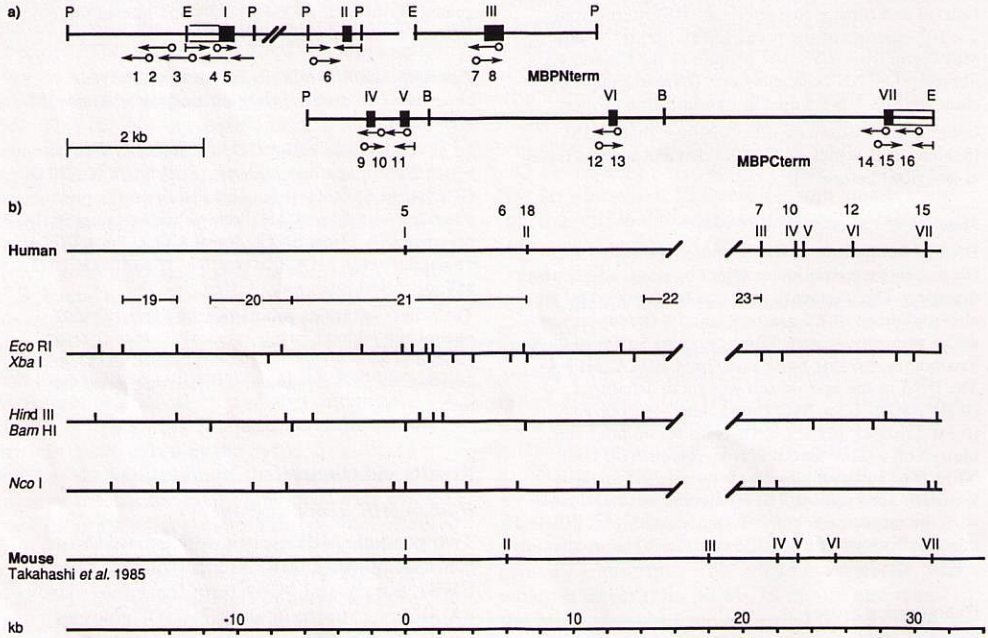

Fig. 1. a) DNA sequencing strategy of the seven exons.

Genomic fragments of MBPNterm and MBPCterm subcloned into the multicloning site pUC13 were analysed by double-strand sequencing. Arrows starting with a vertical line represent sequences derived from M13 and/or double-strand sequencing or with the regular and/or M13 reversed sequencing primer. Arrows with a circle at the beginning indicate those oligonucleotides which were synthesized according to the genomic DNA of MBP and used as sequencing primers. Roman numerals and solid bars represent coding sequences, open bars 5′- and 3′-untranslated sequences. Hybridization probes are numbered sequentially and are listed in detail in the table.

b) Restriction map and exon-intron organization of the human gene encoding myelin basic protein (MBP). MBPNterm, a pcos2 EMBL clone, and MBPCterm, a Charon 4A-genomic clone, were restricted with *Bam* HI (B), *Eco* RI (E), *Hin*d III (H), *Nco* I (N) and *Xba* I (X). Roman numerals and solid bars represent coding sequences.

Table. Hybridization probes used for the mapping and sequencing strategy.

Oligonucleotides

1)	TCA TTT AGT TAG TTT CC
2)	CAA AAC TTG CCA CAT CT
3)	AGC TAA CCT GGA TGG AGC
4)	CTT GGA TCC GTG CCT CT
5)	ATG GCG TCA CAG AAG AGA CCC TCC CAG AGG CAC GGA TCC AAG
6)	CCA CTT TAC TTG CAG AGT TG
7)	CCG GCC GTG TGA CTT CTG GGG CAG GGA GCC
8)	AGG AGG ACA AGC CGC
9)	TCT GGT AGC TCG GAG CC
10)	ACG CCT CGC ACA CCA CCC
11)	CAG CTA AAT CTG CTC
12)	TCC CTT GTG AGC CGA T
13)	TGG CAG CTC CCA GTG GCT
14)	GCG TCT AGC CAT GGG TGA TCC AGA
15)	CCC TGC GGC ACT CTC
16)	TCC TCT CTG CCG CCC ACG TC
17)	CAT GGT ACT TGC TGT GGC CA
18)	CCT CTG CCC TCT CAT GCC CGC AGC CGT CCT GGG C

Restriction fragments of the two genomic clones:

19) *Eco* RI 3.4 kb	MBPNterm
20) *Bam* HI 4.9 kb	MBPNterm
21) *Bam* HI 6.2 kb	MBPNterm
22) *Hin*d III 4.4 kb	MBPNterm
23) *Eco* RI/*Nco* I 1.4 kb	MBPCterm

Sequence determination of the seven exons of human MBP

The exon-harboring fragments were inserted into the multicloning site of pUC13 or M13 for double-strand sequencing or M13 mp8 and mp9 for dideoxy chain termination sequencing. The strategy is summarized in Fig. 1a.

Exon sequences with their 5′- and 3′-adjacent intron sequences and about 1300 bp of the 5′ and 580 bp of the 3′-untranslated regions are summarized in Fig. 2.

Transcription initiation sites and upstream elements

The 20^{mer} oligonucleotide CATGGTACTTGC-TGTGGCCA corresponding to nucleotides 47 to 66 of the coding region of exon I was 5′ end labelled in the kinase reaction and used in the primer extension reaction following the standard procedure[36]. Three transcription initiation points at −55, −82 and −183 were observed. Fig. 3 shows the analysis of the −183 position. Multiple transcription starts have been observed with rat (−47, −51 and −175) and mouse (−25 and −42). No well-defined sequences of a TATA and/or CAAT box can be recognized. A potential TATA box can be located only at −81 (26 bp upstream of the −55 transcription initiation site), whereas no obvious TATA box is apparent further upstream of the other transcription initiation sites. There are four remarkable nucleotide sequences in the 5′-noncoding region: a GC box is located 139 bp upstream of the translation initiation site. Three direct repeats are present, a nonameric repeat at −409/−424 bp and two octameric at −192/−216 and −492/−508. Furthermore the decameric nucleotide sequence at −256 to −265 shows absolute homology to a decameric sequence in the regulatory 5′ region of proteolipid protein. Whether these elements are of relevance for the transcription regulation awaits further clarification.

After this work had been completed we got notice of the publication of Kamholz et al.[42]. The seven exons of the MBP gene were identified in six overlapping MBP clones by oligonucleotide hybridization. Their data differ considerably from our findings: a length of 45 kb was proposed for the human MBP gene as compared to 32–34 kb derived from our two genomic clones. Besides those in the 5′-noncoding sequence other differences consist of the location of the transcription initiation point which the authors localize −47 upstream of the translational start site, whereas our primer extension studies revealed multiple transcription start sites at −55, −82 and −183 bp upstream of the translational start. Sequence motifs such as homologies and repeats of the upstream regulatory regions were not described.

Exon I

```
tgggtaggtgggtgtgtgtatggatggatggatagatggatggatgggtaaatggactgttatgtggatggatggatggatggatagagagatagatgga -1086
tgactggtattacagggatatgtgagtgaatcctgttttctgtagataagtaatagagtttggagaggaaactaactaaatgatatttatttaaacctaa  -986
cactctaacttgaaagcaaaatggattcattgcccttcgtgacagaaatgtggtatttttttggagaaagctatgagatgctggtatacaacatgaaatat  -886
ctcaatcccacttcagatttctaattgtttctgcttccagaggagaagccaagtcaaaatgtcctgaataagcagttctcttattgtgagaggcctcttgt  -786
ggaatctgggattgaaacaattctaaatgccccacttctttcatgcatgaattgcaaaaagatgtggcaagttttgtttctaccaagaaaactaaaaaca  -686
ccttttgtcaaataaatgctccttgcatatttaacttatgcaccagtggccttttaaacagtcaatgtcccatcaaggtgcctgcacatctgggctctcc  -586
gggagcagccatggcagcacccgggaagaaacgctgatgtggctgctctgcatgctcagatgacttcatcgggaagcctgggtgcattttacgctgggtg  -486
ccaaatctgagtaactgaggaattcccagagccttctgaaacacagagctgcaataaggctgctccatccaggttagctccatc.........ctaggcc  -395
    >>>   a c    a  t gg  cct agcca  gt t  ca   ..........aa    a                  c  tggttgggg a  cta
aagggctttatgaggactgcacatattctgtgggttttataggagacagctaggtcaagacccctcagagaaagctgctttgtccggtgctcagctttgc  -295
 c t   cct  g cc c  g g  .a     c                  t          agga       g          ccct  ag  c c t
acaggcccgtattcatatctcattgttgttgcaggagaggcagatgcgaaccagaacaatgggacctcctctcaggacacagcggtgactgactccaag  -195
        ac                  .          g                  t                  gg  g        g        a
cgcacagcggacccgaagaatgcctggc.aggatgccacccagctgacccagggagccgcccccacttgatccgcctcttttcccgagatgccccgggga  -96
 a       a  ↑ a      aa     a   cgc                      a                                          a
gggaggacaacaccttcaaagacaggccctctgagtccgacgagctccagaccatccaagaagacagtgcagccacctccgagagcctggatgtgATGGC   5
              ↑                    a        ↑     t                    ccca     ag t    gag
GTCACAGAAGAGACCCTCCCAGAGGCACGGATCCAAGTACCTGGCCACAGCAAGTACCATGGACCATGCCAGGCATGGCTTCCTCCCAAGGCACAGAGAC  105
A                  A   ......C
ACGGGCATCCTTGACTCCATCGGGCGCTTCTTTGGCGGTGACAGGGGTGCGCCCAAGCGGGGCTCTGGCAAGgtgagctctgaggagtagaggagtttta  177
                                  A                                                  <<<
gtttaaatggaaaaagcaaaggagaaatcagtaggtgaactcagccattagaggaagaactggcacgtagcctcttgctgtctaaggtctcgttccgtgc
tggagaatgcatatgagcccaagagtgtgggcctgagaggctgcttaggacgttttcgtttaactcacccccctctttcctcacaagggatggtggccgg
ggtgtggctcaggaatgtaaggacatgctgaattc
```

Exon II

```
ctgcagaaaatcggaaaaggtccgtcctcggttcactgaaccttcacagagcagaaaaagctttaacctgctaaatccaatgcagaagataacacacttg
ctataaaagaaaaaatacaactctgcaagtaaagtggaatgtaaaatttatccattccatggtagacgtcaaaaaacccttttagtgcactttgctccag
actgggaaaggccttctgcactggatcctagaaatctttcacagctggtgcctgctgccatattaagaactcttggtcccattgtttgatacagggcctc
agaatagactttggaggagagaattcttccctaaactttccttttgctctcgatttcctgagtccttcagggcgcatgctgcccctctgacctcccatcac
ctcttgctcttcttccatcatccatcatctgcacatgccttcttctctccctgtcgtccttcatcctccacccccccgctcactcctcctcactctgggc
tcttgccaagccagctctagaggagatttgctggaggactttggggcattgccggcggcgcccacccggactcacgcagcccactctgtgtcccccggca
                                                                                          >>>a   c cctt  tc
gGTACCCTGGCTAAAGCCGGGCCGGAGCCCTCTGCCCTCTCATGCCCGCAGCCAGCCTGGGCTGTGCAACATGTACAAGgtaagacgccggcgggtcctc  255
                  A A                                  GT            C                tg  <<<

acccatcggggccaggggtgacctgccgtttcctgagctctcagccgactgtccctcggggcaggtagtgtcactgccaggggcccaccccccagcct
```

Exon III

```
ttccctgaggaggacaagccgcagggactgtggacttgtcctgaggtcaccgcgcctctgtgtttcagGACTCACACCACCCGGCAAGAACTGCTCACTA  287
                                                  >>>   tt    c t  ca            ......A G      A C  T
CGGCTCCCTGCCCCAGAAGTC...ACACGGCCGGACCCAAGATGAAAACCCCGTAGTCCACTTCTTCAAGAACATTgtaagtgacgatcgatgggaagag  360
T                    GCAG                          A        T                          <<<
gtagcaactgtgagggggaggagggg
```

Exon IV + V

```
gccagggttctctgtgcctttcagGTGACGCCTCGCACACCACCCCCGTCGCAGGGAAAGgtaagaccttggaatgttttgattgatcatcacttttctg  396
   >>>cac c  c ctt c         A     A        T  A  C  A  G        g  <<<
atagaccttctctaaaatcccataatgtaccaaagagagagttaggctccgagctaccagaatccatcccaaaacgtgttgccaggcagctcccaagtag
aacaggtcggagatccatgcacccctcctgtccctcccgcacctgcacagccgctgtggccctagctgcggcccccctcggagctccgggtggaacctgt
ttttaccacctcagctccactgtgctttgactgtgtttcctgttgattgaaaggactttcccttcactgaccaccatgtcattatttctctgtcttcctc
                                                                                          >>>c       t  c  ...
atgcagGGGAGAGGACTGTCCCTGAGCAGATTTAGCTGGgtaggtgacgaacgcacttccatcggcttcctcttccgtcccagtcctcacagccccgcaa  429
..a            C        C                    <<<
cttttgtgttctgctctgtttcggttgcttcctggcctcctttctctcctctcga
```

Exon VI

```
cctcagcgtggtgctggcccgtggctcctgaaccactcaccagtccagtccgggcctgggcccttccccggggccctggtggcagctcccagtggctcaa
gcagcgtgcccagcaccgcgggtggaggttgagctccgtggtcttctcttgcagGGGGCCGAAGGCCAGAGACCAGGATTTGGCTACGGAGGCAGAGCGT  475
                                  >>> t t gtc t     g          G  G    AG                              T
CCGACTATAAATCGGCTCACAAGGGATTCAAGGGAG...TCGATGCCCAGGGCACGCTTTCCAAAATTTTAAGCTGgtaaggtaccctctgctcagctc  549
                                  G CCTA   C                          C             t  <<<
cactga
```

```
Exon VII
agaaaatgctcatcaaagcaaagcaatggagcggctgcacttcacccacagaatgcacttgtttgcagagcatggttatgccaggcaccactccctgggg
cctgggaatccgattccagttctctctcacagaactggaaattcattagcatttgctgagggaggtcgggagaaatggaatgaaaagaccagctctccgg
ggtgccatttctattagcagatgagagagacacccaatggctcagcatggacccgggaccacagaagaggggatgctggtgtgggaggaccctgcggcac
tctcgtcctaactcctctctcttcctttcagGGAGGAAGAGATAGTCGCTCTGGATCACCCATGGCTAGACGCtgaaaacccacctgg.......ttcc   591
                              >>>          C  C          T       G         g g   t  cc ctcagcc
ggaatcctgtcctcagcttcttaatataactgccttaaaactttaatcccacttgcccctgttacctaattagagcagatgacccctcccctaatgcctg
c        c    g                          ct      t t      a  ga   g   g          .....  ct t  t at  c
cggagttgtgcacgtagtagggtcaggccacggcagcctaccggcaatttccggccaacagttaaatgagaacatgaaaacagaaaacggttaaaactgt
t    cc   at  cg  g   c   c        c  cga t ....................g t a   tattcgt cctttt  t  g  ga tg
ccctttctgtgtgaagatcacgttccttcccccgcaatgtgcccccagacgcacgtgggtcttcagggggccaggtgcacagacgtccctccacgttcac
agt     cg g .......tc   tcag   t act   t    gt c a tt  tc a  c g a  tt         g   ac  t t  a
cc...ctccacccttggactttcttttcgccgtggctcggcacccttgcgcttttgctggtcactgccatggaggcacacagctgcagagacagagagga
  cta     t   ca            a g   agg acac    cca    cc tg  cactgcgg  a  cagg    c gc  ag  gcc    ag
cgtgggcggcagagaggactgttgacatccaagcttcctttgtttttttttcctgtccttctctcacctcctaaagtagacttcatttttcctaacagga
 a ....   gc  g      .  tgg    g              c ................  c      ..ag    tg
ttagacagtcaaggagtggcttactacatgtgggagctttttggtatgtgacatgcgggctgggcagctgttagagtccaacgtggggcagcacagagag
 c        tt           taca c g   ...c  g          .gca a              g        ga   t  tg  c ......
ggggccacctccccaggccgtggctgcccacacaccccaattagctgaattc
.t  gg  actt       ..a   at  c tgca   c cc      t
```

Fig. 2. Nucleotide sequence of exons of the human myelin basic protein gene and comparison with the mouse MBP gene sequence.

The noncoding 5′ region (475 bp) of the mouse gene was taken from Popko et al.[30] and the rest from de Ferra et al.[23]. Deviations from the human sequence are given in lower-case letters, deletions as dots. Arrows (>>>/<<<) embrace mouse sequences available for comparison. The three repeats in the 5′-noncoding region are given as underlined italic lowercase letters. The nucleotide sequence of the coding regions of human MBP was derived from the genomic clones MBPNterm and MBPCterm and determined as described under Material and Methods. Intron sequences are written in lower-case letters, coding sequences in upper case letters. Nucleotide 1 is the A of Met. The transcription initiation sites suggested by the primer extension experiments are indicated by arrows at −55, −82 and −183, two direct repeats at −409 and −424, one GC box and a potential TATA box are underlined. The decameric sequence homologous to a PLP 5′-noncoding decameric sequence (−140 to −149 from ATG) is boxed.

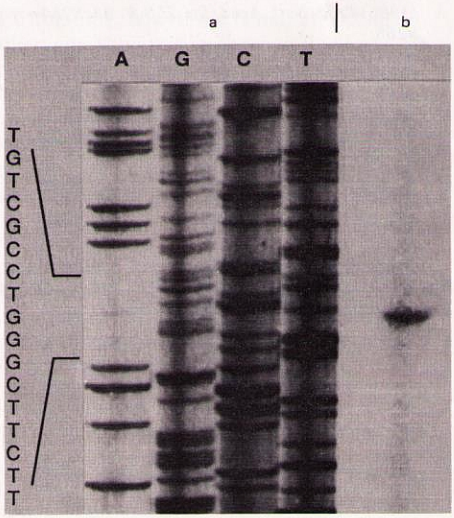

Fig. 3. Autoradiographic analysis of human MBP mRNA.

Oligonucleotide 17 complementary to nucleotides 47 to 66 was used as primer in the extension reaction and for sequencing a *Pst* I fragment encoding exon I and the 5′ upstream sequences. The oligonucleotide was annealed to human total brain mRNA and extended with AMV reverse transcriptase. The reaction products of the two reactions were analysed on a 5% polyacrylamide gel (7M urea). The dried gel was autoradiographed using a Kodak XAR film. a) Sequence ladder; b) primer extension.

We are grateful to the *Fritz-Thyssen-Stiftung* and the *Deutsche Forschungsgemeinschaft* (DFG) for the support of our work. We thank *J. Teufel* for his skillful assistance in the preparation of the manuscript.

Literature

1 Poduslo, J.F. & Braun, P.E. (1975) *J. Biol. Chem.* **250**, 1099–1105.
2 Schwob, V.S., Clark, H.B., Agrawal, D. & Agrawal, H.C. (1985) *J. Neurochem.* **45**, 559–571.
3 Stoffel, W., Hillen, H., Giersiefen, H. (1984) *Proc. Natl. Acad. Sci. U.S.A.* **81**, 5012–5016.
4 Stoffel, W., Subkowski, T. & Jander, S. (1989) *Biol. Chem. Hoppe-Seyler* **370**, 165–176.
5 Raine, C.S. (1984) in *Myelin*, 2nd edition (P. Morell, ed.) pp. 259–310, Plenum Press, New York.
6 Dautigny, A., Alliel, P.M., d'Auriol, L., Pham Dinh, D., Nussbaum, J.L., Galibert, F. & Jollès, P. (1985) *FEBS Letters* **188**, 33–36.
7 Milner, R.J., Lai, C., Nave, K.A., Lenoir, D., Ogata, J. & Sutcliffe, J.G. (1985) *Cell* **42**, 931–939.
8 Schaich, M., Budzinski, R.M. & Stoffel, W. (1986) *Biol. Chem. Hoppe-Seyler* **367**, 825–834.
9 Macklin, W.B., Campagnoni, C.W., Deininger, P.L. & Gardinier, M.W. (1987) *J. Neurosci. Res.* **18**, 383–394.
10 Gardinier, M.V., Macklin, W.B., Diniak, J. & Deininger, P.L. (1986) *Mol. Cell. Biol.* **6**, 3755–3762.
11 Naismith, A.L., Hoffman-Chudzik, E., Tsui, L.C. & Riordan, J.R. (1985) *Nucleic Acids Res.* **13**, 7413–7425.
12 Stoffel, W., Giersiefen, H., Hillen, H., Schröder, W. & Tunggal, B. (1985) *Biol. Chem. Hoppe-Seyler* **366**, 627–635.
13 Willard, H.F. & Riordan, J.R. (1985) *Science* **230**, 940–942.
14 Diehl, H.J., Schaich, M., Budzinski, R.-M. & Stoffel, W. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 9807–9811.
15 Sidman, R.L., Dickie, M.M. & Appel, S.H. (1964) *Science* **144**, 309–311.
16 Nave, K.A., Lai, C., Bloom, F.E. & Milner, R.J. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 9264–9268.
17 Nave, K.A., Lai, C., Bloom, F.E. & Milner, R.J. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84**, 5685–5669.
18 Macklin, W.B., Campagnoni, C.W., Deininger, P.L. & Gardinier, M.V. (1987) *J. Neurosci. Res.* **18**, 383–394.
19 Eylar, E.H., Brostoff, S., Hashim, G., Caccam, J. & Burnett, P. (1971) *J. Biol. Chem.* **246**, 5770–5784.
20 Carnegie, P.R. (1971) *Biochem. J.* **123**, 57–67.
21 Roach, A., Boylen, K., Horvath, S., Prusiner, S.B. & Hood, L.E. (1983) *Cell* **34**, 799–806.
22 Takahashi, N., Roach, A., Taplow, D.B., Prusiner, S.B. & Hood, E. (1985) *Cell* **42**, 139–148.
23 de Ferra, F., Engh, H., Hudson, L., Kamholz, J., Puckett, C., Molineaux, S. & Lazzarini, R.A. (1985) *Cell* **43**, 721–727.
24 Kamholz, J., de Ferra, F., Puckett, C. & Lazzarini, R.A. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 4962–4966.
25 Roth, H.J., Kronquist, K.E., Karlero de Rosbo, N., Crandall, B.F. & Campagnoni, A.T. (1987) *J. Neurosci. Res.* **17**, 321–328.
26 Sidman, R.K., Conover, C.S. & Carson, J.H. (1985) *Cytogenet. Cell. Genet.* **39**, 241–245.
27 Roach, A., Takahashi, N., Pravtcheva, D., Ruddle, F. & Hood, L. (1983) *Cell* **42**, 149–155.
28 Readhead, C., Popko, B., Takahashi, N., Shine, H.D., Saavedra, R.A., Sidman, R.L. & Hood, L. (1987) *Cell* **48**, 703–712.
29 Popko, B., Puckett, C., Lai, C., Shine, H.D., Readhead, C., Takahashi, N., Hunt, S.W. III, Sidman, R.L. & Hood, L. (1987) *Cell* **48**, 713–721.
30 Popko, B., Puckett, C. & Hood, L. (1988) *Neuron* **1**, 221–225.
31 Poustka, A., Rackwitz, H.-R., Frischauf, A.-M., Hohn, B. & Lehrach, H. (1984) *Proc. Natl. Acad. Sci. U.S.A.* **81**, 4129–4133.
32 Maniatis, T., Hardison, R.C., Lacy, E., Laner, J., Connell, C.O. & Quon, D. (1978) *Cell* **15**, 687–761.
33 Maniatis, T., Fritsch, E.F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
34 Rigby, P.W.J., Dieckmann, M., Rhodes, C. & Berg, P. (1977) *J. Mol. Biol.* **113**, 237–251.
35 Dretzen, G., Bellard, M., Sassone-Corsi, P. & Chambon, P. (1981) *Anal. Biochem.* **112**, 295–298.
36 Chen, T.J. & Seeburg, P.H. (1985) *DNA* **4**, 165–170.
37 Heinrich, P. (1986) *Guidelines for Quick and Simple Plasmid Sequencing*, Boehringer Mannheim GmbH, Biochemica.
38 Sanger, F., Nicklen, S. & Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463–5467.
39 Choczynski, P., Sacchi, N. (1987) *Anal. Biochem.* **162**, 156–159.
40 Blin, N. & Stafford, O.W. (1976) *Nucleic Acids Res.* **3**, 2303–2308.
41 Epstein, L.M., Mahon, K.A. & Gall, J.G. (1986) *Cell Biol.* **103**, 1037–1044.
42 Kamholz, J., Toffenetti, J. & Lazzarini, R.A. (1988) *J. Neurosci. Res.* **21**, 62–70.

Rüdiger Streicher and Wilhelm Stoffel*, Institut für Biochemie, Med. Fakultät der Universität zu Köln, Joseph-Stelzmann-Str. 52, D-5000 Köln 41.

* To whom correspondence should be sent.