Vol. 178, No. 1, 1991 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
July 15, 1991 Pages 282-290

STRUCTURE OF MOUSE MYELIN-ASSOCIATED GLYCOPROTEIN GENE

Ryoichi Nakano[1], Nobuya Fujita[1], Shuzo Sato[1], Takashi Inuzuka[1], Kenji Sakimura[2], Hideaki Ishiguro[1], Masayoshi Mishina[2] and Tadashi Miyatake[1]

Departments of [1] Neurology and [2] Neuropharmacology, Brain Research Institute, Niigata University, Niigata 951, Japan

Received May 29, 1991

The mouse myelin-associated glycoprotein gene was isolated from a mouse gene library. This gene was split into 13 exons distributed about 15 kb in length. Each extracellular immunoglobulin-related domain was encoded by a single exon, and RNA splicing between those exons occurred between the first and second nucleotides of the junctional codon, the features of which are conserved in most of the genes of the immunoglobulin superfamily. The sequence of the 5'-flanking region appeared to have some regions homologous to other myelin proteins, which suggested that they were possible cis-elements for specific expression of oligodendrocytes.

Myelin-associated glycoprotein (MAG) is a heavily glycosylated glycoprotein with a molecular mass of approximately 100 kDa and is present in the central and peripheral nervous system myelin. This protein is mainly expressed at the periaxonal regions and is supposed to act as a glial-neuronal adhesion molecule during myelinogenesis (1). The recent isolation of MAG cDNA clones from several species revealed that MAG shares many features with immunoglobulins, as well as neural cell adhesion molecule (N-CAM), P0 protein and Thy-1 antigen, which are also found in the nervous system (2-7).

Abbreviations: MAG, myelin-associated glycoprotein; L-MAG, large MAG; S-MAG, small MAG; MBP, myelin basic protein; PLP, proteolipid protein; CNP, 2',3'-cyclic-nucleotide 3'-phosphodiesterase; N-CAM, neural cell adhesion molecule; nt, nucleotide; kbp, kilo base pairs; kDa, kilo dalton.

0006-291X/91 $1.50

At least two different polypeptide isoforms with apparent molecular masses of 72 kDa (large MAG; L-MAG) and 67 kDa (small MAG; S-MAG) exist in rodents (8). Sequence analysis of mouse MAG cDNA in our laboratory indicated that L-MAG is longer by 45 amino acid residues in the cytoplasmic domain at the C-terminus than S-MAG (5). The different MAG proteins are produced by alternative splicing of the 45-nucleotide (nt) exon portion containing a termination codon, and expression of two MAG isoforms is supposed to depend on myelinogenesis. That is, expression of L-MAG is apparently induced at the early stage of active myelination including remyelination, whereas S-MAG which contains the 45 nt exon portion is expressed at the later stage (9,10). Our previous works also demonstrated that in the quaking mouse brain, a mutant mouse with impaired myelination, L-MAG mRNA and protein are scarcely expressed (9,11). These findings suggest that expression of L-MAG is indispensable for the early stage of myelination.

The mouse MAG gene must be isolated to analyze the different functions of these two MAG isoforms, and the mechanism of regulation of oligodendrocyte specific expression. Furthermore, the MAG gene must constitute a remarkable system for studying the regulation of RNA processing. Nevertheless, the structure of MAG gene has not been reported in detail yet. In the rat, the structure of the MAG gene is partially reported to span 16 kilobases (kb) which contains 13 exons (2). Here we have isolated the full length of the mouse MAG gene for the first time, and have determined the exon-intron organization and the sequence of the 5'- flanking region.

METHODS

A full length mouse MAG cDNA (pMMAG 26) (5) was digested into 3 parts with EcoRI and used as the cDNA probes, and an EMBL-3 gene library of DBA mouse (Clontech Laboratories, Palo Alto, CA,

U.S.A.) was screened with these ^{32}P-labeled cDNA probes. The phage DNA, prepared from positive clones were then restriction mapped by Southern blot hybridization using regional cDNA probes. The DNA fragments obtained were subcloned into pUC118, and the sequence of the plasmid DNA was determined by the dideoxy method (12). Synthetic oligonucleotide primers were available to determine the sequence of the exon-intron junctions.

RESULTS AND DISCUSSION

Overlapping 4 clones covering the MAG gene were isolated, and the restriction maps were determined (Fig. 1). The Southern blot hybridization bands corresponded well to those of total mouse genomic Southern hybridization using regional cDNA fragments as probes (data not shown). The nucleotide sequence around exon-intron junctions was determined, and the exon-intron organization was revealed. The size of introns was confirmed by polymerase chain reactions. The mouse MAG gene was about 15 kb in length and split into 13 exons by 12 introns. The nucleotide sequence of the protein coding regions started in the middle of exon 4 (Fig. 1).

We previously reported that the regions consisted of 53 nt and 45 nt alternatively spliced in transcription (5). The former region were found to be corresponded to exon 2, and the latter region to exon 12.

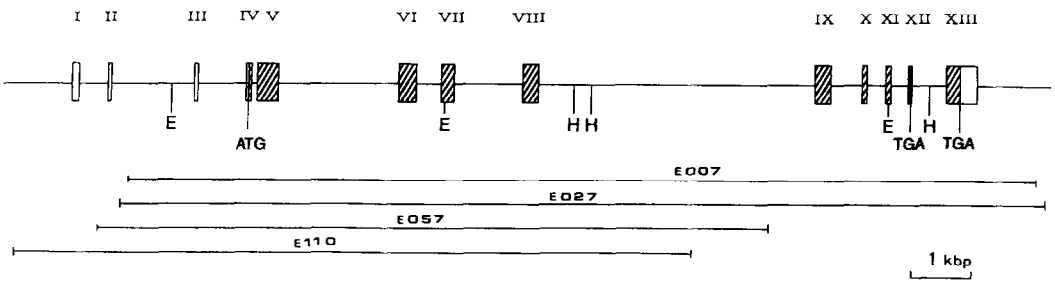

Fig. 1. The structure of the mouse MAG gene. The 4 overlapping phage clones covering the MAG gene are shown at the bottom. EcoRI (E) and HindIII (H) sites are indicated. The squares represent exons (I to XIII) and the bent line represents the coding regions. There are two termination codons (TGA) in exon XII and exon XIII. Exon II and exon XII are alternatively spliced in transcription.

	Intron	Exon	Intron	
		I -------ATCATCCAA	gtgagtac----	0.5kbp
	----ctgtgcag	CCTTCTGTGT-------AGAGACTTG (53bp)	gtaggtgt----	1.4kbp
	----tgttctag	III GTTGAAAGGC-------GAGACTAAG (56bp)	gtgagtgg----	0.8kbp
	----ctctgcag	IV CCCTAGCTCA-------ATT TCA G (73bp)	gtaatggc----	0.1kbp
D1	----atctacag	V CT TCT CGA-------ATC GTC A (369bp)	gtgagtcg----	2.0kbp
D2	----ctacacag	VI AC ACC CCC-------GTC AAG T (297bp)	gtgagcct----	0.4kbp
D3	----ggcctcag	VII AC CCC CCA-------GTC ATG T (258bp)	gtgagtct----	1.1kbp
D4	----ccccgcag	VIII AT GCA CCT-------GTG GAG T (261bp)	gtgagtgc----	4.6kbp
D5	----ttttgcag	IX TT GCC CCC-------GGA GCA C (288bp)	gtgagtga----	0.5kbp
D6	----acccttag	X AC CGA CTG-------AGA AGA AA (97bp)	gtgagtgc---	0.3kbp
D7	----ccctgcag	XI A AAG AAC -------G TAT GAG (100bp)	gtgagagg----	0.3kbp
	----ttcaatag	XII TCC AGA GA-------CCCCAGGAG (45bp)	gtaggtcc----	0.6kbp
D8	----ctccatag	XIII AGT AAG AA-------TAATGACAA (519bp)		

Fig. 2. Sequence around the exon-intron junctions of the MAG gene. Triplet codons of nucleotide sequence in exons are underlined. Intron sequence is given in lowercase letters. The size of each exon is indicated below the sequence, and the size of each intron confirmed by polymerase chain reactions (PCR) is shown in the right-hand column. The exons surrounded by dashed line encode the immunoglobulin-related extracellular domains of MAG. Each number of domain (D1 to D8) is denoted in the left-hand column of its encoding exon. The nucleotides which are in disagreement with the cDNA sequence are indicated by dots.

Fig. 2 shows the exon-intron organization. Each donor and acceptor site was conventional. The nucleotide sequence in the exons appeared to be different in several points from that of the

mouse cDNA we previously reported . The discrepancy between the genomic and cDNA sequence is supposed to be due to the difference in mouse strain. The segregation of exons was consistent with the functional segregation of MAG into extracellular, membrane-spanning, and cytoplasmic domains, respectively. The membrane-spanning domain was encoded by exon 10. The cytoplasmic domain was encoded by exon 11 and exon 12 in S-MAG, and by exon 11 and exon 13 in L-MAG. The 5 extracellular domains of MAG, which were similar in structure to the immunoglobulin domains (13), were encoded by exon 5 to exon 9.

Each immunoglobulin-related domain was encoded by a single exon, and RNA splicing between those exons occurred between the first and second nucleotides of the junctional codon, designated as the type 1 junction (14). These rules were broken in exon 10 and the following exons. These features of the exons encoding immunoglobulin-related domains were highly conserved in the immunoglobulin gene (15) which suggested that MAG and immunoglobulin genes had evolved from the same ancestral gene. Each immunoglobulin-related domain is also encoded by a single exon in most other immunoglobulin supergene families (16), for example T-cell receptors, major histocompatibility antigens and Thy-1 antigen, and these exons have type 1 junctions (15). Therefore, the formation of the genes of immunoglobulin superfamily is hypothesized to have occurred by duplication of an ancestral gene encoding a complete domain (13). On the other hand, in N-CAM (14), P0 protein (17) and lymphocyte CD4 protein (18), each immunoglobulin-related domain is encoded by two exons. A progenitor gene may properly encode a half-domain (17). Together with each two exons encoding the complete domain, interestingly, the rule of the type 1 junction is conserved in the N-CAM, P0 and lymphocyte CD4 protein genes.

Primer extension analysis was carried out to determine the 5'-terminus of the gene, and gave several products distributed from -243 to -433 (data not shown). There was a TATA box-like sequence around nt -422, and the GC box was absent (Fig. 3). NF-Y binding sequences were found around nt -433 and -520. NF-Y was first identified as a sequence-specific DNA binding protein that interacts with the conserved Y motif of the major histocompatibility complex class II genes (19). Since the Y motif ($T^{T}/_{G}CTG\underline{ATTGG}^{T}/_{C}T^{A}/_{C}{}^{A}/_{C}$) actually contains a CCAAT sequence in reverse, NF-Y is in fact a CCAAT box-binding protein and has a ubiquitous tissue distribution (20).

In the central nervous system, only several proteins are known to constitute the myelin; myelin basic protein (MBP), proteolipid protein (PLP), 2',3'-cyclic-nucleotide 3'-phospho-diesterase (CNP) and MAG. MBP, PLP and MAG are thought to be oligodendrocyte specific proteins, whearas CNP is highly enriched in myelin, but is also expressed elsewhere, particularly in the lymphoid tissue. The 5'-flanking regions of the genes of MBP (21,22), PLP (23) and CNP (24) were compared to determine cis-elements for oligodendrocyte specific expression.

Four homologous regions had been pointed out between the MBP and PLP genes (23), and two of them, AATGTTTT and AAGGGAGGA, were also found in the MAG gene (Fig. 3,4A). The location of the former homologous region, AATGTTTT, was approximately the same to those of the MBP and PLP genes. The location of the latter homologous region, AAGGGAGGA, was also similar to those of the MBP and PLP genes, but the order between the sequence and the TATA box was reversed between the MAG and MBP genes. There were, furthermore, 3 additional homologous regions which were relatively short in length in the MBP, PLP and MAG genes (Fig. 4B). Each homologous region was located at approximately the same

```
CATACACATA GCACAAATAA ATCTAAAAGA AGTAAATCAA TAAATGTTTT    -963

TTTTTTTAAA GAGGGGGAAC CGCTGGCTCC AGTGTTAAGT GCTATGCTTC    -913

TGGAAGAGCA GGCTGGGGTC GGCAGGATGG TGTGAGGACA GCTTTCCAGG    -863

AATGAGAGTT AAGAATGACA TCTCTGAGTG CCAAGGTGCC ACAGGGCATT    -813

AGAAGAATAG ACATGAGCAC TTATTTTCTG TTTAACAGGT GGAGACAGAC    -763

TCAGACAATA GACAGCATCC CTGTCCTTAC CAAGGATCTC AAGGAGACAG    -713

CAGCTAATGT CATTTCAGAG GACAAGAAGT GTCAAGGAGA AAACAGACCA    -663

GGAAGGAAAG CATGTTGGGG GGGGGCAGTG ATTCTCAGCA AGGTGGAGGT    -613

GACTTTGGAG CAGAACTAGG AGAGGGTGTG TTTCCTCTTC CTGACAACCA    -563

                                          NF-Y
GTCTCTACCT ATCTCTTGGA CCAAGGTCAC AGTGTTGCCA CCAATTAATT    -513

CCCCGAAGAC AACAGGTTCC ACCTTTTAGC CTGGAGCTTT CAGAAAGATG    -463

                    NF-Y          TATA like
AGGCCACTCC TCCCTCCTCA ACTCCCAAAT TGGGGTGTAT ATTAATGGAG    -413

GAAGGGAGGA GCAAGGCAGA CCCCCATCTT TTCTAGGAAG GAGCTAGTAG    -363

GACCCAGGAA TCAAGCTGCC CAGTCCCAGC CCTAGCCCGA TGGTGACAAG    -313

GGCCCCTTTG TGCCCCCCTC CCCCGGGGGC AGGGAGGAGC TGGGCCCTGG    -263

AGGCAGGCGG CCCCTGGCAC CCAGGGGGCA GAGGAGGGGC TGGCAAGTGG    -213

GGGCCTAGAC CCTGGGAGCC CAAGGGACTG TAAGCCCGGC CAGCAGAGCA    -163

                          -137
GAAGGTGCAG AAGCCAGATC ATCCAA
                          Exon 1
```

Fig. 3. Exon 1 and 5'-flanking region of the mouse MAG gene. The nucleotide numbers are indicated in the right-hand column. For convenience, the first, second and third introns are omitted from the nucleotide numbering. The nucleotides in the upstream region are numbered in the negative direction with A of the methionine start codon at position +1. The boxes show the TATA box like sequence and NF-Y binding sequences. Each region underlined is homologous region with the MBP and PLP genes.

in 5'-flanking regions in all those genes. In the CNP gene, these consensus homologous regions were not found, but the core sequence of the last homologous region, AGGGAG, existed

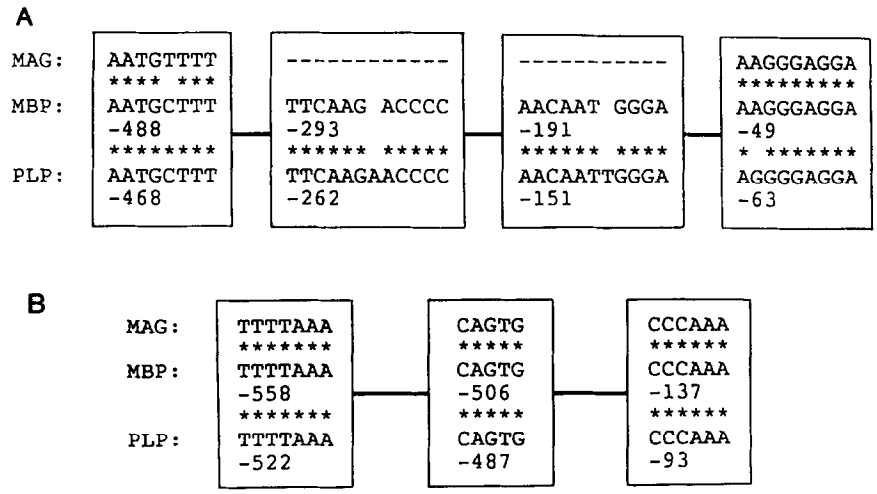

Fig. 4. A.Homologous nucleotide sequences between the 5'-flanking regions of the MAG, MBP and PLP genes. Two homologous regions between MAG gene and MBP or PLP gene are indicated. The numbers denote the positions from the transcription initiation sites in the MBP and PLP genes. The most upstream initiation site was designated as +1 in the PLP gene.
B.Three additional homologous sequences in the 5'-flanking regions of the MAG, MBP and PLP genes. The nucleotide numbers are indicated in the same manner as described above.

just under the TATA box. These homologous regions possibly play some role for the oligodendrocyte specific expression. The CNP gene had little region homologous with the MBP, PLP and MAG genes, because CNP is not expressed specifically in oligodendrocytes.

Our isolated clones of the MAG gene together with other myelin specific protein genes will be useful to analyze the mechanism of regulation of oligodendrocyte specific expression.

ACKNOWLEDGMENTS

This study was supported in part by grants-in-aid from the Ministry of Education, Science and Culture of Japan, the Ministry of Health and Welfare of Japan and from Uehara memorial fundation.

REFERENCES

1. Quarles,R.H. (1984) Dev. Neurosci. 6, 285-303.
2. Lai,C., Brow,M.A., Nave,K.-A., Noronha,A.B., Quarles,R.H., Bloom,F.E., Milner,R.J. and Sutcliff,J.G. (1987) Proc. Natl. Acad. Sci. USA 84, 4337-4341.

3. Salzer,J.L., Holmes,W.P. and Colman,D.R. (1987) J. Cell Biol. 104, 957-965.
4. Arquint,M., Roder,J., Chia,L.-S., Down,J., Wilkinson,D., Bayley,H., Braun,P. and Dunn,R. (1987) Proc. Natl. Acad. Sci. USA 84, 600-604.
5. Fujita,N., Sato,S., Kurihara,T., Kuwano,R., Sakimura,K., Inuzuka,T., Takahashi,Y. and Miyatake,T. (1989) Biochem. Biophys. Res. Commun. 165, 1162-1169.
6. Sato,S., Fujita,N., Kurihara,T., Kuwano,R., Sakimura,K., Takahashi,Y. and Miyatake,T. (1989) Biochem. Biophys. Res. Commun. 163, 1473-1480.
7. Spagnol,G., Williams,M., Srinivasan,J., Golier,J., Bauer,D., Lebo,R.V. and Latov,N. (1989) J. Neurosci. Res. 24, 137-142.
8. Frail,D.E. and Braun,P.E. (1984) J. Biol. Chem. 259, 14857-14862.
9. Fujita,N., Sato,S., Kurihara,T., Inuzuka,T., Takahashi,Y., and Miyatake,T. (1988) FEBS Lett. 232, 323-327.
10. Fujita,N., Ishiguro,H., Sato,S., Kurihara,T., Kuwano,R., Sakimura,K., Takahashi,Y. and Miyatake,T. (1990) Brain Res. 513, 152-155.
11. Fujita,N., Sato,S., Ishiguro,H., Inuzuka,T., Baba,H., Kurihara,T., Takahashi,Y. and Miyatake,T. (1990) J. Neurochem. 55, 1056-1059.
12. Sanger,F., Nicklen,S., and Coulson,A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
13. Hunkapiller,T. and Hood,L. (1986) Nature 323, 15-16.
14. Owens,G.C., Edelman,G.M. and Cunningham,B.A. (1987) Proc. Natl. Acad. Sci. USA 84, 294-298.
15. Evans,G. (1987) In Molecular Neurobiology, Heinemann,S. and Patrick,J. eds. (New york: Plenum Publishing Corp.), pp 225-257.
16. Williams,A.F. (1987) Immunol. Today 8, 298-303.
17. Lemke,G., Lamar,E. and Patterson,J. (1988) Neuron 1, 73-83.
18. Littman,D.R. and Gettner,S.N. (1987) Nature 325, 453-455.
19. Dorn,A., Durand,B., Marfing,C., Le Meur,M., Benoist,C. and Mathis, D. (1987) Proc. Natl. Acad. Sci. USA 84, 6249-6253.
20. Hooft van Huijsduijnen,R.A.M., Bollekens,J., Dorn,A., Benoist,C. and Mathis,D. (1987) Nucleic Acids Res. 15, 7265-7282.
21. Takahashi,N., Roach,A., Teplow,D.B., Prusiner,S.B. and Hood,L. (1985) Cell 42, 139-148.
22. Miura,M., Tamura,T., Aoyama,A. and Mikoshiba,K. (1989) Gene 75, 31-38.
23. Ikenaka,K., Furuichi,T., Iwasaki,Y., Moriguchi,A., Okano,H. and Mikoshiba,K. (1988) J. Mol. Biol. 199, 587-596.
24. Monoh,K., Kurihara,T., Sakimura,K. and Takahashi,Y. (1989) Biochem. Biophys. Res. Commun. 165, 1213-1220.