Structure and Chromosomal Localization of the Gene for the Oligodendrocyte-Myelin Glycoprotein

Daniel D. Mikol,* Mark J. Alexakos,* Cynthia A. Bayley,* Richard S. Lemons,§ Michelle M. Le Beau,‡ and Kari Stefansson*

*Departments of Neurology and Pathology (Neuropathology) and the Committees on Neurobiology and Immunology; ‡Department of Medicine (Section of Hematology/Oncology), University of Chicago, Chicago, Illinois 60637; and §Department of Pediatrics, University of Utah, Salt Lake City, Utah 54132

Abstract. Utilizing a cDNA clone encoding the oligodendrocyte-myelin glycoprotein (OMgp) to screen a human genomic DNA library, we have obtained a clone that contains the OMgp gene. The genomic clone was restriction mapped and the OMgp gene and its 5′ and 3′ flanking regions were sequenced. A single intron is found in the 5′ untranslated region of the gene, while the coding region is uninterrupted by an intron. This placement of a single intron in the OMgp gene is identical to that of the gene for the α-chain of platelet glycoprotein Ib, which, along with OMgp, belongs to a family of proteins sharing two distinct structural domains: an NH_2-terminal cysteine-rich domain and an adjacent domain of tandem leucine-rich repeats. Hence, it is possible that this family of proteins is not only related in terms of primary structure, but also through similar gene structure. Sequence comparison of the 5′ and 3′ flanking regions did not reveal striking similarities to other DNA sequences, and no obvious promoter elements were noted. By hybridization of the genomic clone to metaphase cells, we have localized the human OMgp gene to chromosome 17 bands q11–12, a region to which the neurofibromatosis type 1 gene has been previously mapped.

THE oligodendrocyte-myelin glycoprotein (OMgp)[1] is a highly glycosylated protein of oligodendrocytes and central nervous system myelin which appears to be localized at the paranodal region of the myelin sheath (Mikol and Stefansson, 1988). OMgp is anchored in the plasma membrane as a 120-kD glycoslyphosphatidylinositol-linked form that can be released from the membrane upon incubation with phospholipase C to generate a soluble 105-kD polypeptide. Based on cDNA sequence, the predicted primary structure of OMgp consists of four domains (Mikol et al., 1990). At the NH_2-terminus there is a 32-amino acid cysteine-rich (CR) motif. This is followed by a domain consisting of 7 1/2 tandem leucine-rich repeats (LRs) of 24 amino acids each, and a domain of 4 1/2 repeats of 40 residues each that are rich in serines and threonines. A hydrophobic COOH-terminal segment is most likely cleaved concomitant with the attachment of a phosphatidylinositol-containing glycan (Cross, 1990).

Several proteins have been found to contain LRs (see Table I) since the first report of the leucine-rich α-2 glycoprotein of serum (Takahashi et al., 1985*b*). Some of these proteins, including OMgp, share both an NH_2-terminal CR motif and a contiguous series of LRs. On the basis of the sharing of two structurally unrelated domains, we believe that these proteins are best classified as belonging to a distinct family, the CR-LR family (Mikol et al., 1990). Among these proteins, the platelet glycoprotein Ib is most similar to OMgp, sharing not only the CR and LR domains, but also the serine/threonine-rich region which in both cases probably contains an abundance of O-linked oligosaccharides. Glycoprotein Ib is a transmembrane disulfide-linked αβ heterodimer that plays an important role in the initial adhesion of platelets to the exposed subendothelium of blood vessels during hemostasis, by virtue of the fact that it is a receptor for von Willebrand factor (Clemetson and Lüscher, 1988). Interestingly, both the α- and β-chains of glycoprotein Ib, as well as glycoprotein IX of platelets with which it associates in a noncovalent complex, contain LRs (Table I). The α-chain of platelet glycoprotein Ib (gpIbα) is the only other member of the CR-LR family whose gene has been characterized (Wenger et al., 1988).

We have provided evidence that OMgp is encoded by a single gene (Mikol et al., 1990). Here we describe the structure of the gene encoding OMgp, and localize the gene to human chromosome 17q bands 11–12. As is the case for the gpIbα gene, the OMgp gene has an uninterrupted open reading frame and an intron that is six nucleotides upstream of the transla-

Correspondence and reprint requests may be addressed to Dr. Kari Stefansson, Department of Neurology, P.O. Box 425, University of Chicago, 5841 S. Maryland Avenue, Chicago, IL 60637.

1. *Abbreviations used in this paper*: CR, cysteine-rich motif; GpIbα, α chain of platelet glycoprotein Ib; LR, leucine-rich repeat; NF1, neurofibromatosis type 1; OMgp, oligodendrocyte-myelin glycoprotein; TSS, transcriptional start site.

tional start codon. This similarity between the gpIbα and OMgp genes raises the possibility that members of the CR-LR family may be related not only by the sharing of two structurally unrelated domains, but also through gene structure.

Materials and Methods

General

Restriction enzymes and other enzymes used in cloning were obtained from commercial sources. [^{32}P]-labeled nucleotides were obtained from DuPont Co. (Wilmington, DE) or ICN Radiochemicals (Irvine, CA). Total RNA was prepared from human brain biopsy specimens, or from a cell line that does not express OMgp, by the guanidinium acid-phenol method as described by Chomczynski and Sacchi (1987). Isolation of genomic DNA and plasmid DNA, restriction digestions, preparation of restriction-deleted clones, end labeling of DNA with [γ-^{32}P]-ATP, agarose gel electrophoresis, and hybridizations were done according to standard protocols (Sambrook et al., 1989).

Construction and Screening of Human Genomic DNA Library

Human genomic DNA was incompletely digested with Mbo I to create fragments of various sizes. Fragments sized at 15–20 kb were partially filled in and ligated into Xho I-digested λFIX II (Stratagene Cloning Systems, La Jolla, CA) that was also partially filled in according to the manufacturer's instructions. Plaques obtained upon infection of bacterial strain P2PLK17 with the genomic DNA recombinants were adsorbed onto Colony/Plaque Screen nylon membrane disks (DuPont Co.) for screening with a partial OMgp cDNA clone (S1) described previously (Mikol et al., 1990). The S1 cDNA insert was prepared from agarose gels and labeled with [α-^{32}P]-dCTP by random priming using an oligolabeling kit (Pharmacia Fine Chemicals, Piscataway, NJ) (Feinberg and Vogelstein, 1983). The hybridized nylon disks were washed twice for 30 min with 0.1× SSC, 1% (wt/vol) SDS at 65°C and exposed with intensifying screens at −80°C. Positive clones were plaque purified and subcloned into Bluescript KS+ plasmid (Strategene Cloning Systems, La Jolla, CA) for restriction mapping and sequencing.

Chromosomal Localization

The OMgp gene was localized by hybridizing a radiolabeled 12-kb genomic probe (M5E10) to normal human metaphase chromosomes prepared from phytohemagglutinin-stimulated peripheral blood lymphocytes (Le Beau et al., 1984), as well as by fluorescent in situ chromosomal hybridization using a biotin-labeled M5E10 probe (Lichter et al., 1988). In addition, a panel of rodent–human somatic cell hybrids was used to confirm the chromosomal localization.

DNA Sequencing

The M5E10 genomic clone or restriction deletion clones of the M5E10 clone were used as templates for DNA sequencing. Sequences were determined on both strands, using either double- or single-stranded DNA, by dideoxynucleotide chain termination (Sanger et al., 1977) with Sequenase or Taquence (United States Biochemical Corp., Cleveland, OH) (Tabor and Richardson, 1987). Oligonucleotide primers to be used for sequencing were synthesized by the phosphoramidite method with a DNA synthesizer (model 381A; Applied Biosystems; Inc., Foster City, CA). Sequences were analyzed by the sequence analysis software package of the University of Wisconsin Genetics Computer Group (Devereux et al., 1984), using the Dayhoff table and the GAP program in the above software package.

Primer Extension

Primer extension analysis of adult human brain RNA, which contains considerable OMgp mRNA (Mikol et al., 1990), or of RNA isolated from a cell line that does not express OMgp mRNA, was carried out by modifications of the methods of Boorstein and Craig (1989) and Sambrook et al. (1989). A 24-mer oligonucleotide primer complementary to the 5′ end of OMgp cDNA: OM13, CGTCTTGTTCTGTAGAGTCGGAAC, corresponding to nucleotides No. 647–670 (see Fig. 3), and a 24-mer oligonucleotide further upstream on the same strand: OM25, CCACATAGGAGGACTGCAGAGCTG (nucleotides No. 541–564), were used. Primers were end labeled with [γ-^{32}P]-ATP and T4 polynucleotide kinase. 40 μg of total RNA coprecipitated with 100 fmol of end-labeled primer was resuspended in 30 μl aqueous hybridization buffer (0.4 M NaCl, 1 mM EDTA, 40 mM PIPES [pH 6.4]), denatured at 80°C for 4 min, and annealed at 55°C for 3 h. The hybridized primer/transcripts were then diluted 10-fold in 10× reverse transcriptase buffer (10 mM dATP, 10 mM dCTP, 10 mM dGTP, 10 mM dTTP, 10 mM DTT, 600 mM KCl, 500 mM Tris-HCl [pH 7.6], 100 mM $MgCl_2$) and extended for 2 h at 37°C after addition of 25 U AMV reverse transcriptase plus 100 U RNase inhibitor. Extension was stopped with the addition of 5 μl 0.5 M EDTA, and the RNA was then digested with DNase-free RNase (5 μg/ml) for 30 min at room temperature. After phenol/chloroform extraction and ethanol precipitation the samples were resuspended in 4 μl of TE buffer (10 mM Tris-HCl [pH 7.5], 1 mM EDTA) and 6 μl of formamide loading buffer (95% formamide, 20 mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol FF). The primer extension samples, end-labeled molecular size standards (Hae III digest of PhiX174) (Bethesda Research Laboratories, Bethesda, MD), and sequenced products of the genomic clone using end-labeled oligonucleotide primers were analyzed on the same 6% polyacrylamide denaturing gel for comparison.

Results

Isolation of the Human OMgp Gene

Using the S1 cDNA probe, a 12-kb Eco RI genomic fragment (λM5E10) in λFIX II vector was isolated. The genomic clone was then subcloned into Bluescript KS+ plasmid for further analysis by restriction mapping (Fig. 1). Interestingly, cDNA restriction sites were found to map with the same distribution and spacing on M5E10, suggesting an intronless structure for the OMgp gene.

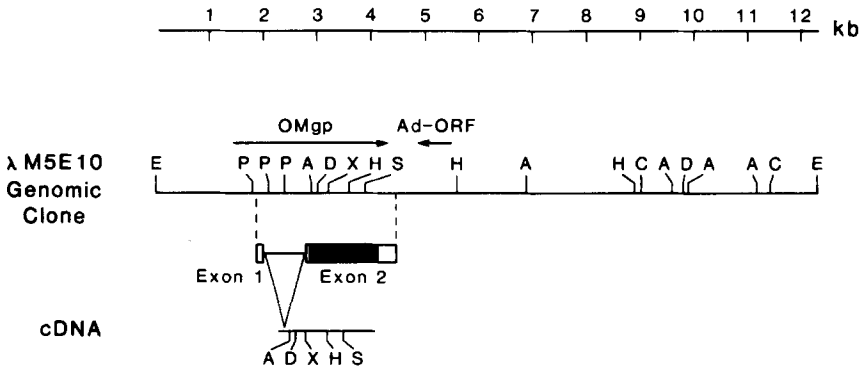

Figure 1. Restriction map of the 12-kb M5E10 genomic clone. Orientation of *OMgp* within this clone is shown by the arrow (5′ to 3′, left to right). The positions of exons 1 and 2 are indicated by boxes (*open boxes*, noncoding; *solid boxes*, coding) below M5E10, with the intron represented by a line between them. The cDNA restriction map is shown at the bottom, with the intron spliced out as indicated. An adjacent open reading frame (*Ad-ORF*) 3′ of *OMgp* on the antisense strand has also been found within M5E10, and may correspond to an exon of the NF1 gene. Restriction enzymes sites are *A*, Acc I; *C*, Sac I; *D*, Dra II; *E*, Eco RI; *H*, Hind III; *P*, Pst I; *S*, Spe I; *X*, Xho I.

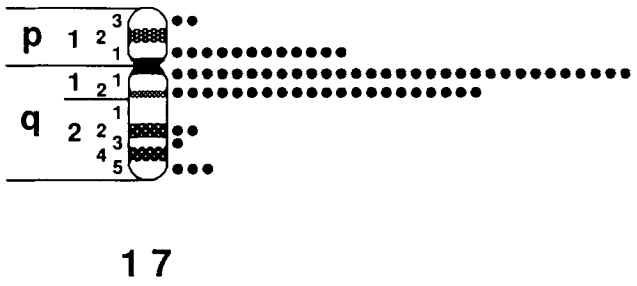

Figure 2. The figure shows the distribution of labeled sites on chromosome 17 in 100 normal human metaphase cells that were hybridized with an OMgp genomic probe (M5E10, see Fig. 1). Each dot indicates one labeled site observed in the corresponding band. Of 100 metaphase cells examined, 37 (37%) were labeled on region p1-q1 of one or both chromosome 17 homologues. Of 266 total labeled sites observed, 72 (27.1%) were located on chromosome 17. These sites were clustered at band p11-q12, and this cluster represented 24.1% (64/266) or all labeled sites (cumulative probability for the Poisson distribution is <0.0005). The largest number of grains was observed at 17q11. Similar results were obtained in a second hybridization experiment using this probe. Human metaphase cells were prepared from phytohemagglutinin-stimulated peripheral blood lymphocytes.

Chromosomal Localization

We localized the OMgp gene by hybridizing radiolabeled M5E10 genomic probe to normal human metaphase cells. Hybridization resulted in specific labeling of only chromosome 17, bands q11–12 (Fig. 2). To confirm and refine this localization, we performed fluorescence in situ chromosomal hybridization using a biotin-labeled M5E10 probe. Specific labeling of 17q11 was observed on one (3 cells) or both (17 cells) chromosome 17 homologues in each of 20 cells examined. Signal was not detected on other chromosomes. A panel of rodent–human hybrids was used to confirm the mapping of the OMgp gene to human chromosome 17 (data not shown).

Structure of the OMgp Gene

Comparison of the cDNA and genomic DNA restriction maps showed the presence of several shared restriction sites, spaced in a similar manner in each case, suggesting that the OMgp gene might not contain introns. Upon sequencing M5E10 (Fig. 3), however, an intron of 815 nucleotides was found at the 5′ end of the gene, interrupting the gene between the sixth and seventh nucleotides upstream of the ATG start codon (relative to cDNA). No introns were found within the coding region of OMgp. Hence, the first exon encodes 5′-untranslated mRNA sequences, while the second exon encodes the remainder of the 5′-untranslated sequence (six nucleotides), the entire coding sequence, and the entire 3′-untranslated sequence. Several proteins are known to be encoded by genes with uninterrupted coding regions (see Discussion). The most relevant example is the gene for gpIbα, which also contains a single intron in the 5′ noncoding end of its gene, in exactly the same position as the intron we have found for OMgp. In the case of OMgp, there are two possible translational start sites based on deduced sequence (see Fig. 3). The ATG codon for the first methionine is placed just six nucleotides downstream from the intron splice site, which is exactly the same distance between intron and the ATG in the gpIbα gene. The significance of this intron placement is presently unclear.

Sequence of the OMgp Gene

Where the cDNA and genomic sequences overlap they are identical. Furthermore, the OMgp gene does not appear to be very polymorphic, based on sequence identity of the genes from six individuals (data not shown). Overall the OMgp gene is extremely A-T rich. The coding region consists of 60% A + T, while the intron of OMgp consists of 64% A + T. There are three possible polyadenylation signals, the first two of which were noted in the cDNA; the third signal was discovered 3′ to the end of the cDNA sequence. Curiously, ~600 bp downstream of the 3′ end of the OMgp gene there is a long open reading frame of ~170 amino acids on the antisense strand (Fig. 1). This potential coding region may correspond to a hitherto undescribed exon of the recently described NF1 gene, which is a large gene spanning at least 100 kb (Cawthon et al., 1990; Wallace et al., 1990), previously mapped to the centromeric region of chromosome 17 (Barker et al., 1987; Seizinger et al., 1987; Goldgar et al., 1989).

Primer Extension

In four separate experiments a single major primer extension product (Fig. 4) was obtained using the OM13 oligonucleotide primer that marks the transcriptional start site (TSS) (at the guanine at position 527 in Fig. 3). The sequence around the TSS, TCAG, is also found in the gene for the PO protein of peripheral myelin (Lemke et al., 1988), the myelin basic protein gene (Takahashi et al., 1985*a*), and the proteolipid protein gene (Diehl et al., 1986). No extension product was seen from OM13 primer hybridized to RNA from a cell line that does not express OMgp (data not shown). The OM25 primer did not generate any detectable extension product, perhaps because the TSS as defined by the OM13 primer extension is too close to the end of OM25.

There are several TATA and CAAT elements upstream of the proposed transcriptional start site (Fig. 3), although the positioning of these elements is not conventional. TATA and CAAT boxes are typically found at 20–30 and 40–100 nucleotides, respectively, upstream of the TSS of a gene (Breathnach and Chambon, 1981). Three TATA sites are found at positions −79, −81, and −95 relative to the proposed TSS. However, a segment 17 nucleotides in length containing only A/T is also found 42–58 bases upstream of the TSS, which may be analogous to the A-T rich segment in another myelin gene, the PO gene, where it is thought to serve as a TATA box (Lemke et al., 1988). A CCAATCT sequence is found at position −158, and three CAAAT sequences are found at positions −90, −148, and −179. A potential AP-1 binding site TGAGTCA spans the TSS and is of questionable significance. No GC boxes or binding sites for other transcriptional factors were identified.

Discussion

OMgp is a protein containing four domains and is encoded by a single exon. A number of genes have been discovered

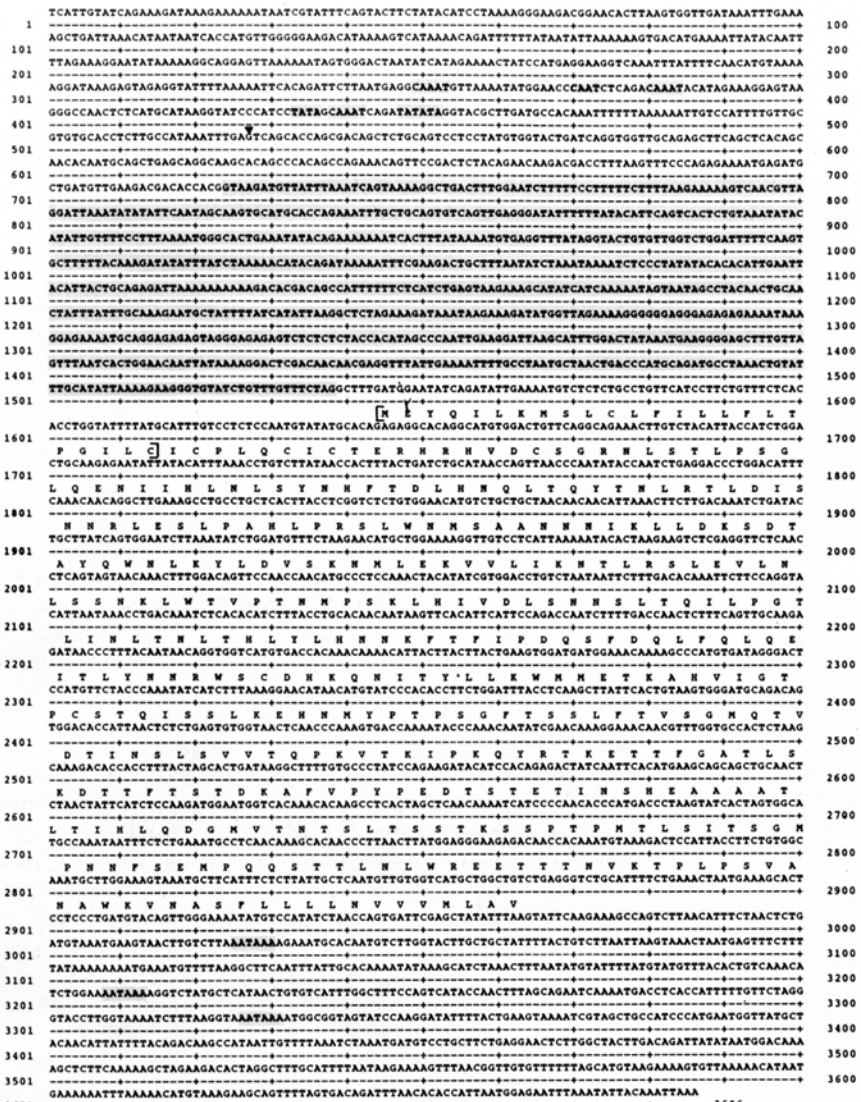

Figure 3. Nucleotide sequence of the OMgp gene and 5′ and 3′ flanking regions. A single intron consisting of 815 nucleotides (No. 724–1538) is shaded as are possible TATA and CAAT elements and potential polyadenylation signals. The guanine at position No. 527 (*arrowhead*) represents the proposed major TSS. The deduced amino acid sequence is also shown, beginning with the methionine whose ATG codon is at nucleotides No. 1545–1547. The theoretical NH_2-signal sequence is enclosed in brackets. It is also possible that the second methionine, whose ATG codon is at nucleotides No. 1566–1568, begins the signal sequence, although the former site is more appealing in light of the intron position of the gpIbα gene, which would place the intron 6 bp upstream of the ATG initiation codon in both the gpIbα and OMgp genes (see text). The isoleucine following the bracketed signal sequence begins the mature NH_2-terminus, based on amino acid sequencing of isolated protein (Mikol and Stefansson, 1988).

to have their coding regions uninterrupted by introns. The proteins encoded by these genes are functionally and structurally diverse and include the histones (Kedes, 1979), the β_2-adrenergic receptor (Kobilka et al., 1987), gpIbα (Wenger et al., 1988), interferons (Lawn et al., 1981; Nagata et al., 1981), the JUN protooncogene (Hattori et al., 1988), thrombomodulin (Jackmann et al., 1987), and a gene within an intron of the Factor VIII gene (Levinson et al., 1990). Therefore, it is unlikely that a coding region uninterrupted by introns indicates a close evolutionary relationship between genes or a functional relationship between their products.

OMgp and gpIb are polypeptides consisting of several distinct structural elements, some of which they share. Each of the structural elements found in these polypeptides, for example the CR and LR domains, are found separately in other proteins, where they are likely to have specific functions. Remarkably, however, both OMgp and gpIbα are each encoded by a single exon. Therefore, it is conceiveable that these domains might have originally been encoded by separate exons and were assembled independently and the introns between them lost during evolution, though it is unclear what advantage the intron loss might grant.

We localized the OMgp gene to chromosome 17q11–12, a region which by linkage studies had previously been shown to contain the gene for NF1 (Barker et al., 1987; Seizinger et al., 1987;Goldgar et al., 1989). The NF1 gene has recently been discovered and its structure partially characterized (Cawthon et al., 1990; Wallace et al., 1990). It is a large (>100 kb) gene whose mRNA transcript measures 11–13 kb. Astonishingly, it appears that the NF1 gene may contain at least three genes, including *OMgp*, within its introns but in an antisense orientation (Cawthon et al., 1990). In mammals, the presence of a gene within the intron of another gene had previously been described only for a gene embedded in an antisense orientation within a Factor VIII gene intron (Levinson et al., 1990). Upon sequencing downstream of the OMgp gene, we have identified a long open reading frame on the antisense strand ~600 bp away which may correspond to an exon of the NF1 gene not yet characterized (our unpublished results). Thus, the A-T rich character of the OMgp gene may reflect the fact that it is contained within an intron,

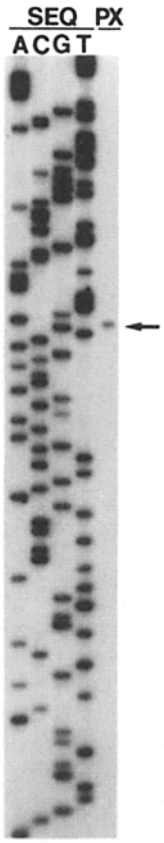

Figure 4. Primer extension analysis of total RNA from adult human brain using the OM13 oligonucleotide primer. The extension product (*PX*) and the sequenced products (*SEQ*) of the M5E10 clone using the same OM13 oligonucleotide (end-labeled) are shown. A single major product was obtained (*arrow*). Comigration of this product with one of the sequencing products establishes the transcriptional start site at guanine, nucleotide No. 527 in Fig. 3.

although this is not the case for the gene within the Factor VIII intron, which is G-C rich (Levinson et al., 1990). It is presently not known whether the gpIbα gene is contained within an intron of a larger gene.

We have previously described the primary structure of OMgp in the context of the CR-LR family of proteins (Table I) that share two structurally unrelated domains: an NH_2-terminal CR that is fairly similar to an EGF motif (Gray et al., 1983; Scott et al., 1983), and an adjacent series of LRs, which by virtue of their amphipathic character (Takahashi et al., 1985*b*) have been implicated in adhesive processes (Kataoka et al., 1985; Vicente et al., 1988; Reinke et al., 1988). We have now shown that two members of the CR-LR family, namely gpIbα and OMgp, are encoded by genes that are highly similar. Thus, proteins within the CR-LR family, which share features of their primary structure and perhaps some aspects of function, may also share gene structure, in particular an uninterrupted coding sequence and an intron in the 5′ noncoding region. This intron, just upstream of the translation start site, may serve some regulatory role (Smith et al., 1989).

In addition to *OMgp*, the genes of several myelin proteins have been reported, including the proteolipid protein (Diehl et al., 1986), which is expressed in the central nervous system, the PO protein of the peripheral nervous system (Lemke et al., 1988), and the myelin basic proteins (Takahashi et al., 1985*a*) which are found in both the central and peripheral nervous systems. All of the above genes, including *OMgp*, have been found to have the sequence TCAG around the TSS, but no other sequence similarities between the OMgp gene and the other myelin genes have been identified. In contrast to OMgp, these other proteins are abundant, constituting the major proteins of myelin. It is unresolved how these myelin proteins are coordinately regulated during the process of myelination.

While we do not yet understand the function of OMgp, it is a glycoprotein of central nervous system myelin and oligodendrocytes which may play a role as an adhesion mole-

Table I. Proteins That Contain LRs and CR-LRs

Protein	Species	No. of repeats	Adjacent NH_2-terminal Cys-rich region	Reference
OMgp	Human	7 1/2	yes	(Mikol et al., 1990)
Platelet Ibα	Human	7	no	(Lopez et al., 1988)
Platelet Ibβ	Human	1	yes	(Lopez et al., 1988)
Platelet IX	Human	1	yes	(Hickey et al., 1989)
Biglycan	Human, bovine	12	yes	(Fisher et al., 1989)
Decorin	Human, bovine	10	yes	(Fisher et al., 1989)
Fibromodulin	Bovine	10	yes	(Oldberg et al., 1989)
α3 (collagen VI)	Avian	1	no	(Bonaldo and Colombatti, 1989)
Toll	*Drosophila*	15	yes	(Hashimoto et al., 1988)
Chaoptin	*Drosophila*	41	yes	(Reinke et al., 1988)
Leucine-rich α-2 glycoprotein	Human	8	no	(Takahashi et al., 1985*b*)
Adenylate cyclase	Yeast	26	no	(Kataoka et al., 1985)
Lutropin-choriogonadotropin receptor	Rat	14	yes	(McFarland et al., 1989)
Carboxypeptidase N (noncatalytic subunit)	Human	12	yes	(Tan et al., 1990)
RNase inhibitor	Human, porcine	15*	no	(Schneider et al., 1988)
yopM	*Yersinia pestis*	6‡	no	(Leung and Straley, 1989)

* Contains repeats of 28 amino acids.
‡ Contains repeats of 14 amino acids.
The table lists all of the proteins known to date that contain LRs; proteins that contain an adjacent NH_2-terminal CR domain (and hence belong to the CR-LR family) are indicated.

cule in the developing nervous system. It contains LRs, which in other proteins are believed to function as amphipathic adhesion modules. Also, a subpopulation of OMgp molecules contain the HNK-1 carbohydrate (Mikol et al., 1990), which has been shown by itself to mediate cell-cell adhesion (Künemund et al., 1988).

NF1 is one of the most common human autosomal dominant diseases, with an incidence between 1 in 4,000 and 1 in 3,000. The clinical features of NF1 are diverse and include café au lait spots, neurofibromas that increase in size and are the main cause of morbidity in the disease, iridial hamartomas (Lisch nodules), bone abnormalities, learning disabilities, and optic gliomas. The role of OMgp if any in the causation or phenotypic expression of NF1, such as learning disabilities or gliomas, is unclear. The OMgp gene is one of three genes situated in an antisense orientation within the NF1 gene (Cawthon et al., 1990). Because the finding of expressed genes within introns of large genes is such a recent one, we can only speculate about the control of their expression. The restricted tissue expression of OMgp, in contrast to the more general distribution of the NF1 gene (Wallace et al., 1990), may be controlled at the level of antisense RNA, whereby NF1 primary transcripts inhibit expression of OMgp. In contrast, activation of the OMgp gene might inhibit transcription of the NF1 gene. The OMgp and NF1 genes may be closely regulated, and altered expression of the OMgp gene may account for some of the phenotypes of NF1.

We should like to acknowledge the technical assistance of Rafael Espinaosa III, Anthony Fernald, Yogesh Patel, Beth A. Miller, Ji-hyon Kim, and Gladys Magnoni. We are grateful to Dr. Janet D. Rowley for her advice and support.

This work was supported in part by grants from the National Institutes of Health (K. Stefansson and M. M. Le Beau), the National Multiple Sclerosis Society, (K. Stefansson), and the Lucille P. Markey Charitable Trust (K. Stefansson). D. D. Mikol was supported by a fellowship from the Illinois Chapter of the Neurofibromatosis Society. M. M. Le Beau is a Scholar of the Leukemia Society of America. R. S. Lemons is a Pew Scholar in the Biomedical Sciences.

Received for publication 20 July 1990 and in revised form 16 August 1990.

References

Barker, D., E. Wright, K. Nguyen, L. Cannon, P. Fain, D. Goldgar, D. T. Bishop, J. Carey, B. Baty, J. Kivlin, H. Willard, J. S. Waye, G. Greig, L. Leinwand, Y. Nakamura, P. O'Connell, M. Leppert, J.-M. Laiouel, R. White, and M. Skolnick. 1987. Gene for von Recklinghausen neurofibromatosis is in the pericentromeric region of chromosome 17. *Science (Wash. DC).* 236:1100–1102.

Bonaldo, P., and A. Colombatti. 1989. The carboxyl terminus of the chicken α3 chain of collagen VI is a unique mosaic structure with glycoprotein Ib-like, fibronectin type III, and Kunitz modules. *J. Biol. Chem.* 264:20235–20239.

Boorstein, W. R., and E. A. Craig. 1989. Primer extension analysis of RNA. *Methods Enzymol.* 180:347–369.

Breathnach, R., and P. Chambon. 1981. Organization and expression of eucaryotic split genes coding for proteins. *Annu. Rev. Biochem.* 50:349–383.

Cawthon, R. M., R. Weiss, G. Xu, D. Viskochil, M. Culver, J. Stevens, M. Robertson, D. Dunn, R. Gesteland, P. O'Connell, and R. White. 1990. A major segment of the neurofibromatosis type 1 gene: cDNA sequence genomic structure, and point mutations. *Cell.* 62:193–201.

Chomczynski, P., and N. Sacchi. 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* 162:156–159.

Clemetson, K. J., and E. F. Lüscher. 1988. Membrane glycoprotein abnormalities in pathological platelets. *Biochim. Biophys. Acta.* 947:53–73.

Cross, G. A. M. 1990. Glycolipid anchoring of plasma membrane proteins. *Annu. Rev. Cell Biol.* 6:1–39.

Devereux, J., P. Haeberli, and O. Smithies. 1984. A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* 12:387–395.

Diehl, H.-J., M. Schaich, R.-M. Budzinski, and W. Stoffel. 1986. Individual exons encode the integral membrane domains of human myelin proteolipid protein. *Proc. Natl. Acad. Sci. USA.* 83:9807–9811.

Feinberg, A. P., and B. Vogelstein. 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* 132:6–13.

Fisher, L. W., J. D. Termine, and M. F. Young. 1989. Deduced protein sequence of bone small proteoglycan I (biglycan) shows homology with proteoglycan II (decorin) and several nonconnective tissue proteins in a variety of species. *J. Biol. Chem.* 264:4571–4576.

Goldgar, D. E., P. Green, D. M. Parry, and J. J. Mulvihill. 1989. Multipoint linkage analysis in neurofibromatosis type 1: an international collaboration. *Am. J. Hum. Genet.* 44:6–12.

Gray, A., T. J. Dull, and A. Ullrich. 1983. Nucleotide sequence of epidermal growth factor cDNA predicts a 128,000 molecular weight protein precursor. *Nature (Lond.).* 303:722–725.

Hashimoto, C., K. L. Hudson, and K. V. Anderson. 1988. The *Toll* gene of Drosophila, required for dorsal-ventral embryonic polarity, appears to encode a transmembrane protein. *Cell.* 52:269–279.

Hattori, K., P. Angel, M. M. Le Beau, and M. Karin. 1988. Structure and chromosomal localization of the functional intronless human JUN protooncogene. *Proc. Natl. Acad. Sci. USA.* 85:9148–9152.

Hickey, M. J., S. A. Williams, and G. J. Roth. 1989. Human platelet glycoprotein IX: an adhesive prototype of leucine-rich glycoproteins with flank-center-flank structures. *Proc. Natl. Acad. Sci. USA.* 86:6773–6777.

Jackman, R. W., D. L. Beeler, L. Fritze, G. Soff, and R. D. Rosenberg. 1987. Human thrombomodulin gene is intron depleted: nucleic acid sequences of the cDNA and gene predict protein structure and suggest sites of regulatory control. *Proc. Natl. Acad. Sci. USA.* 84:6425–6429.

Kataoka, T., D. Broek, and M. Wigler. 1985. DNA sequence and characterization of the *S. cerevisiae* gene encoding adenylate cyclase. *Cell.* 43:493–505.

Kedes, J. 1979. Histone genes and messengers. *Annu. Rev. Biochem.* 48:837–70.

Kobilka, B. K., T. Frielle, S. Collins, T. Yang-Feng, T. S. Kobilka, U. Francke, R. J. Lefkowitz, and M. G. Caron. 1987. An intronless gene encoding a potential member of the family of receptors coupled to guanine nucleotide regulatory proteins. *Nature (Lond.).* 329:75–79.

Künemund, V., F. B. Jungalwala, G. Fischer, D. K. H. Chou, G. Keilhauer, and M. Schachner. 1988. The L2/HNK-1 carbohydrate of neural cell adhesion molecules is involved in cell interactions. *J. Cell Biol.* 106:213–223.

Lawn, R., J. Adelman, A. E. Franke, C. M. Houck, M. Gross, R. Najarian, and D. V. Goeddel. 1981. Human fibroblast interferon gene lacks introns. *Nucleic Acids Res.* 9:1045–1052.

Le Beau, M. M., C. A. Westbrook, M. O. Diaz, and J. D. Rowley. 1984. Evidence for two distinct c-*src* loci on human chromosome 1 and 20. *Nature (Lond.).* 312:70–71.

Lemke, G., E. Lamar, and J. Patterson. 1988. Isolation and analysis of the gene encoding peripheral myelin protein zero. *Neuron.* 1:73–83.

Leung, K. Y., and S. C. Straley. 1989. The *yopM* gene of *Yersinia pestis* encodes a released protein having homology with the human platelet surface protein GPIbα. *J. Bacteriol.* 171:4623–4632.

Levinson, B., S. Kenwrick, D. Lakich, G. Hammonds, Jr., and J. Gitschier. 1990. A transcribed gene in an intron of the human Factor VIII gene. *Genomics.* 7:1–11.

Lichter, P., T. Cremer, J. Borden, L. Manuelides, and D. C. Ward. 1988. Delineation of individual human chromosomes in metaphase and interphase cells by *in situ* suppression hybridization using recombinant DNA libraries. *Hum. Genet.* 80:224–234.

Lopez, J. A., D. W. Chung, K. Fujikawa, F. S. Hagen, E. W. Davie, and G. J. Roth. 1988. The α and β chains of human platelet glycoprotein Ib are both transmembrane proteins containing a leucine-rich amino acid sequence. *Proc. Natl. Acad. Sci. USA.* 85:2135–2139.

McFarland, K. C., R. Sprengel, H. S. Phillips, M. Köhler, N. Rosemblit, K. Nikolics, D. L. Segaloff, and P. H. Seeburg. 1989. Lutropin-choriogonadotropin receptor: an unusual member of the G protein-coupled receptor family. *Science (Wash. DC).* 245:494–499.

Mikol, D. D., and K. Stefansson. 1988. A phosphatidylinositol-linked peanut agglutinin-binding glycoprotein in central nervous system myelin and on oligodendrocytes. *J. Cell Biol.* 106:1273–1279.

Mikol, D. D., J. R. Gulcher, and K. Stefanson. 1990. The oligodendrocyte-myelin glycoprotein belongs to a distinct family of proteins and contains the HNK-1 carbohydrate. *J. Cell Biol.* 110:471–479.

Nagata, S., C. Brack, K. Henco, A. Schambock, and C. Weissmann. 1981. Partial mapping of ten genes of the human interferon-α family. *J. Interferon Res.* 1:333–336.

Oldberg, Å., P. Antonsson, K. Lindblom, and D. Heinegård. 1989. A collagen-binding 59-kd protein (fibromodulin) is structurally related to the small interstitial proteoglycans PG-S1 and PG-S2 (decorin). *EMBO (Eur. Mol. Biol. Organ.) J.* 8:2601–2604.

Reinke, R., D. E. Krantz, D. Yen, and S. L. Zipursky. 1988. Chaoptin, a cell surface glycoprotein required for Drosophila photoreceptor cell morphogenesis, contains a repeat motif found in yeast and human. *Cell.* 52:291–301.

Sambrook, J., E. F. Fritsch, and T. Maniatis. Molecular Cloning: A Laboratory Manual. 2nd edition. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York. 545 pp.

Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA.* 74:5463–5467.

Schneider, R., E. Schneider-Scherzer, M. Thurnher, B. Auer, and M. Schweiger. 1988. The primary structure of human ribonuclease/angiogenin inhibitor (RAI) discloses a novel highly diversified protein superfamily with a common repetitive module. *EMBO (Eur. Mol. Biol. Organ.) J.* 7:4151–4156.

Scott, J., M. Urdea, M. Quiroga, R. Sanchez-Pescador, N. Fong, M. Selby, W. J. Rutter, and G. I. Bell. 1983. Structure of a mouse submaxillary messenger RNA encoding epidermal growth factor and seven related proteins. *Science (Wash. DC).* 221:236–240.

Seizinger, B. R., G. A. Rouleau, L. J. Ozelius, A. H. Lane, A. G. Faryniarz, M. V. Chao, S. Huson, B. R. Korf, D. M. Parry, M. A. Pericak-Vance, F. S. Collins, W. J. Hobbs, B. G. Falcone, J. A. Iannazzi, J. C. Roy, P. H. St. George-Hyslop, R. E. Tanzi, M. A. Bothwell, M. Upadhyaya, P. Harper, A. E. Goldstein, D. L. Hoover, J. L. Bader, M. A. Spence, J. J. Mulvihill, A. S. Aylsworth, J. M. Vance, G. O. D. Rossenwasser, P. C. Gaskell, A. D. Roses, R. L. Martuza, X. O. Breakefield, and J. F. Gusella. 1987. Genetic linkage of von Recklinghausen neurofibromatosis to the nerve growth factor receptor gene. *Cell.* 49:589–594.

Smith, C. W. J., J. G. Patton, and B. Nadal-Ginard. 1989. Alternative splicing in the control of gene expression. *Annu. Rev. Genet.* 23:527–577.

Tabor, S., and C. C. Richardson. 1987. DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. *Proc. Natl. Acad. Sci. USA.* 84:4767–4771.

Takahashi, N., A. Roach, D. B. Teplow, S. B. Prusiner, and L. Hood. 1985*a*. Cloning and characterization of the myelin basic protein gene from mouse: one gene can encode both 14 kd and 18.5 kd MPBs by alternate use of exons. *Cell.* 42:139–148.

Takahashi, N., Y. Takahashi, and F. W. Putnam. 1985*b*. Periodicity of leucine and tandem repetition of a 24-amino acid segment in the primary structure of leucine-rich α_2-glycoprotein of human serum. *Proc. Natl. Acad. Sci. USA.* 82:1906–1910.

Tan, F., D. K. Weerasinghe, R. A. Skidgel, H. Tamei, R. K. Kaul, I. B. Roninson, J. W. Schilling, and E. G. Erdös. 1990. The deduced protein sequence of the human carboxypeptidase N high molecular weight subunit reveals the presence of leucine-rich tandem repeats. *J. Biol. Chem.* 265:13–19.

Vicente, V., P. J. Kostel, and Z. M. Ruggeri. 1988. Isolation and function of the von Willebrand factor-binding domain located between residues His^{1}-Arg^{293} of the α-chain of glycoprotein Ib. *J. Biol. Chem.* 263:18473–18479.

Wallace, M. R., D. A. Marchuk, L. B. Andersen, R. Letcher, H. M. Odeh, A. M. Saulino, J. W. Fountain, A. Brereton, J. Nicholson, A. L. Mitchell, B. H. Brownstein, and F. S. Collins. 1990. Type 1 neurofibromatosis gene: identification of a large transcript disrupted in three NF1 patients. *Science (Wash. DC).* 249:181–186.

Wenger, R. H., N. Kieffer, A. N. Wicki, and K. J. Clemetson. 1988. Structure of the human blood platelet membrane glycoprotein Ibα gene. *Biochem. Biophys. Res. Commun.* 156:389–395.