

The Oligodendrocyte–Myelin Glycoprotein Belongs to a Distinct Family of Proteins and Contains the HNK-1 Carbohydrate

Daniel D. Mikol, Jeffrey R. Gulcher, and Kari Stefansson

Departments of Neurology and Pathology (Neuropathology) and the Committees on Neurobiology and Immunology, University of Chicago, Chicago, Illinois 60637

Abstract. The complete primary structure of the human oligodendrocyte–myelin glycoprotein (OMgp), a glycosphospholipid-linked membrane protein of oligodendrocytes and central nervous system myelin, has been determined. The deduced amino acid sequence predicts a polypeptide of 433 amino acids which includes a 17–amino acid leader sequence. OMgp consists of four domains: (a) a short cysteine-rich motif at the NH₂ terminus; (b) a series of tandem leucine-rich repeats (LRs) present in several other proteins where they may play roles in adhesion; (c) a serine/threonine-rich region that contains probable attachment sites for O-linked carbohydrates; and (d) a

hydrophobic COOH-terminal segment that is likely to be cleaved concomitant with the attachment of lipid during biosynthesis of OMgp. OMgp shares the first three of its four domains with the platelet glycoprotein Ib, which is responsible for the initial adhesion of platelets to the exposed subendothelium during hemostasis. Together with glycoprotein Ib and several other proteins, OMgp belongs to a family of proteins that contain both an NH₂-terminal cysteine-rich motif and an adjacent series of LR. In addition, we report that a subpopulation of OMgp molecules contains the HNK-1 carbohydrate, which has been shown to mediate interactions among cells in the central nervous system.

WE have previously reported the identification of a peanut agglutinin-binding glycoprotein that appears in the human central nervous system (CNS)¹ at the time of myelination (Mikol and Stefansson, 1988). We subsequently isolated this protein and characterized some of its biochemical and immunohistochemical features. It was found to contain a substantial amount of both N- and O-linked oligosaccharides and was the predominant peanut agglutinin-binding protein in human CNS myelin. In the adult human CNS we were only able to immunohistochemically locate it in CNS myelin. However, it was also present on the surfaces of ovine oligodendrocytes in culture. It was not in Schwann cells, in peripheral nervous system myelin, or in any other tissue outside of the CNS. Therefore, we call this glycoprotein the oligodendrocyte–myelin glycoprotein (OMgp), as it appears to be specific for the oligodendrocyte–myelin unit.

OMgp is a membrane glycoprotein anchored in the outer leaflet of the membrane solely through a glycosylphospha-

tidylinositol (GPI) lipid intermediate. This mode of attachment places OMgp in a class of proteins with rather diverse functions, including the variant surface glycoproteins of *Trypanosoma brucei* (Ferguson et al., 1985a,b), the Thy-1 glycoprotein (Low and Kincade, 1985; Tse et al., 1985), acetylcholinesterase (Low and Finean, 1977; Futerman et al., 1983, 1985a,b,c), the smallest form of the neural cell adhesion molecule (He et al., 1986; Sadoul et al., 1986; Hemperly et al., 1986), and about 40 other proteins (Ferguson and Williams, 1988). The GPI linkage may grant a protein several unique features: (a) high lateral mobility in the membrane; (b) rapid cleavage from the membrane by the action of a phospholipase; and (c) generation of physiologic mediators such as diacylglycerol upon phospholipase cleavage (Ferguson and Williams, 1988). However, in light of the functional diversity of the GPI-linked proteins, it is unlikely that the lateral mobility, phospholipase-induced release, or transmembrane signaling are properties critical for the specific functions of many of the GPI-linked proteins.

Myelination in the CNS is a process that requires complicated interactions among the plasma membranes of oligodendrocytes and between the plasma membranes of oligodendrocytes and axons (Bunge et al., 1978). The molecular mechanisms of these interactions have yet to be determined although several candidates for participating molecules have been identified (Gulcher et al., 1986; Poltorak et al., 1987; Lemke, 1988). Because any component of myelin and the plasma membrane of oligodendrocytes could contribute to

Address correspondence and reprint requests to Dr. Kari Stefansson, Department of Neurology, P. O. Box 425, University of Chicago, 5841 S. Maryland, Chicago, Illinois 60637.

1. *Abbreviations used in this paper:* CNS, central nervous system; CR, cysteine-rich motif; GPI, glycosylphosphatidylinositol; LR, leucine-rich repeat; OMgp, oligodendrocyte–myelin glycoprotein; ST, serine/threonine-rich.

this process, the number of candidate molecules is quite large. It is well recognized the CNS myelin contains a complicated mixture of proteins, but with few exceptions only the major proteins have been isolated and characterized to a significant degree. OMgp is one of the minor proteins of CNS myelin. Here we report on the primary structure of OMgp and discuss structural features it shares with other proteins, including a series of tandem leucine-rich repeats (LRs), which in other proteins have been implicated in adhesive processes, and a short cysteine-rich motif (CR) that is fairly similar to an EGF motif. We propose that the coexistence of these two structural features in a protein defines a previously unrecognized family that we have termed the CR-LR family. We argue that one of the cardinal features of the members of this family is their ability to interact with other molecules through the LRs. We also report that the HNK-1 carbohydrate, which has been shown to mediate adhesion between cells in the CNS (Künemund et al., 1988), is present on a subpopulation of OMgp molecules from human brain. Hence, OMgp is well equipped to function as an adhesion molecule and could contribute to the interactive processes required for myelination.

Materials and Methods

General

Restriction enzymes and enzymes used in cDNA cloning were obtained from commercial sources. ^{32}P -labeled nucleotides were purchased from DuPont Co. (Wilmington, DE). Preparation of plasmid DNA, restriction digestions, agarose gel electrophoresis, RNA and DNA blotting, and hybridizations were from standard protocols described by Maniatis et al. (1982).

Screening of cDNA Libraries

$\lambda\text{gt}11$ cDNA libraries from the spinal cord (catalog No. 37434) and brainstem (catalog No. 37432) of a 1-d-old child were obtained from American Type Culture Collection (Rockville, MD). A human fetal (14–16-wk gestation) spinal cord $\lambda\text{gt}11$ cDNA library was kindly provided by Celia and Anthony Campagnoni (UCLA). Polyclonal rat anti-OMgp antiserum, obtained as previously described (Mikol and Stefansson, 1988), was first preabsorbed against $\lambda\text{gt}11$ lysates of *Escherichia coli* and then used to probe recombinant plaques from the neonatal spinal cord cDNA library on nitrocellulose plate replicas (Snyder and Davis, 1985). The single positive cDNA clone (S1) obtained from antibody screening was used to screen all three libraries on Colony/Plaque Screen nylon membrane (DuPont Co.) plate replicas. The S1 cDNA insert used to probe the libraries was prepared from agarose gels and labeled with ^{32}P [dCTP] by random priming (Feinberg and Vogelstein, 1983) using an oligolabeling kit (Pharmacia Fine Chemicals, Piscataway, NJ). After hybridization, the nylon disks were washed twice for 30 min with $0.1\times$ SSC + 1% (wt/vol) SDS at 65°C and exposed with intensifying screens at -80°C . All positive clones were plaque purified and subcloned into Bluescript (Stratagene, La Jolla, CA) by shotgun ligation for restriction mapping and sequencing.

DNA Sequencing

Either double- or single-stranded DNA was sequenced using dideoxynucleotide termination (Sanger et al., 1977) with Sequenase (United States Biochemical Corp., Cleveland, OH) (Tabor and Richardson, 1987). Nested deletions of selected subclones were made with exonuclease III (Henikoff, 1984) and all clones were sequenced in both directions. The sequence analysis was performed using the sequence analysis software package of the University of Wisconsin Genetics Computer Group (Devereux et al., 1984). Polypeptide sequences were compared using the Dayhoff table and the GAP program in the above software package.

Northern, Southern, and Western Blotting

Total RNA (10 $\mu\text{g}/\text{lane}$) isolated (Chomczynski and Sacchi, 1987) from bi-

opsied human cerebral white matter that had been quick frozen in liquid nitrogen was electrophoresed on a 1.2% (wt/vol) agarose formaldehyde gel and capillary blotted onto Gene Screen nylon membrane (DuPont Co.). For Southern blotting, human peripheral blood lymphocyte DNA was capillary blotted onto Gene Screen Plus nylon membrane (DuPont Co.). Both Northern and Southern blots were probed with random primed ^{32}P [dCTP] cDNA insert fragments. Blots were washed twice for 30 min with $0.1\times$ SSC + 1% (wt/vol) SDS at 65°C and exposed with intensifying screens at -80°C . OMgp isolation, SDS-PAGE on 7.5% Laemmli gels, and Western blotting were done as previously described (Mikol and Stefansson, 1988).

Antibodies

The spleens of rats that displayed significant titers against OMgp (Mikol and Stefansson, 1988) were fused with SP2/O-Ag-14 mouse myeloma cells (Galfré and Milstein, 1981). Hybridomas were screened for binding to OMgp using the Bio-Dot apparatus (Bio-Rad Laboratories, Richmond, CA) and the positive hybridomas were cloned by limiting dilution. While the majority of clones appeared to recognize carbohydrate moieties of OMgp, mAb 16 was found to be highly specific and stained OMgp that had been treated with N-glycanase (Genzyme Corp., Boston, MA) to remove N-linked carbohydrates.

New Zealand white female rabbits were immunized with 20 μg human OMgp in complete Freund's adjuvant (Difco Laboratories Inc., Detroit, MI), and boosted twice at 3-wk intervals using 20 μg OMgp in incomplete Freund's adjuvant (Difco Laboratories, Inc.). Rabbit antiserum was preabsorbed against nitrocellulose plate replicas of $\lambda\text{gt}11$ infected *E. coli* lysates, and then affinity purified on S1 lysate replicas (Snyder and Davis, 1985). The mouse anti-HNK-1 mAb (Becton Dickinson & Co., Mountain View, CA) was used as described previously (Mikol et al., 1988).

Anion Exchange Chromatography

The protein sample (1 mg total) was applied to a MonoQ anion exchange fast protein liquid chromatography column (Pharmacia Fine Chemicals) in 20 mM triethanolamine, pH 7.5. Proteins were eluted with a concentration gradient of NaCl at a flow rate of 0.2 ml/min.

Results

Cloning of Human OMgp cDNA

Screening of a cDNA library prepared from human neonatal spinal cord using polyclonal rat anti-OMgp antibodies yielded a single positive ~ 1 -kb clone (S1). To support our contention that the S1 clone encoded OMgp, rabbit polyclonal antibodies were affinity purified on nitrocellulose replicas of plates containing lysed *E. coli* expressing S1 fusion protein. The S1 affinity-purified antibodies bound to purified OMgp on Western blots (Fig. 1 A). When total RNA from human white matter was probed with S1 cDNA, a single ~ 1.8 -kb band was observed (Fig. 1 B). A Southern blot containing human genomic DNA cleaved with any of three restriction endonucleases was probed with S1 cDNA, and only one band was revealed in each case (Fig. 1 C), indicating that there is probably only one copy of OMgp in the haploid human genome. The S1 clone was fully sequenced in both directions, and was used to reprobe the neonatal spinal cord library, as well as to probe a neonatal brainstem library and a fetal spinal cord library. The neonatal spinal cord library yielded no new clones, while the neonatal brainstem library gave two clones, B1 and B2. Screening of the fetal spinal cord library gave several positive clones (C1, C4, C5, C6). All of the clones obtained were restriction mapped, and formed ~ 1.8 kb of overlapping cDNA (Fig. 2), which is in good agreement with the mRNA size seen on Northern blot analysis. When Northern blots containing RNA from other tissues were examined with the same probes, nothing hybridized (data not shown). Western blots containing electropho-

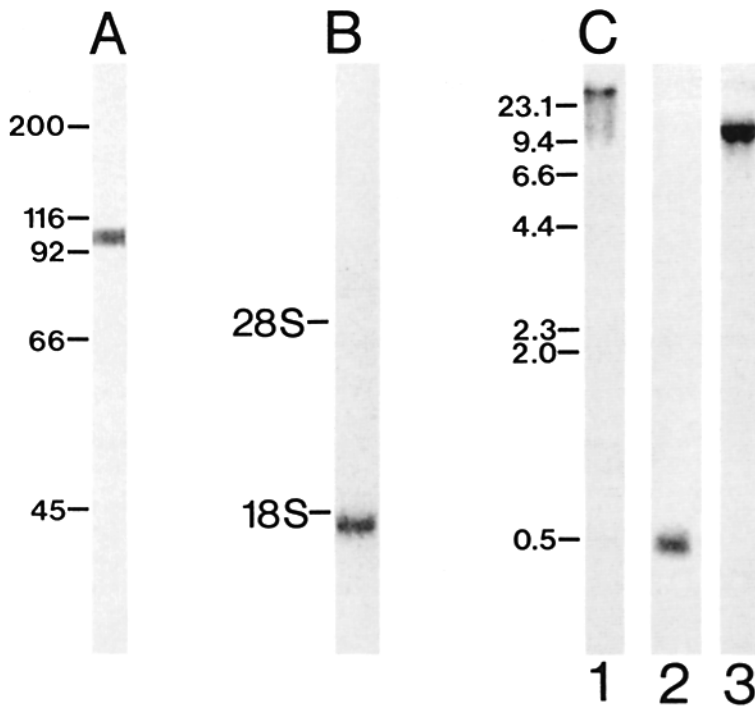


Figure 1. Evidence that the cDNA clones encode OMgp (A), that OMgp is translated from a single mRNA (B), and most likely transcribed from a single gene (C). (A) Western blot containing purified human OMgp after separation by SDS-PAGE. The blot was stained with rabbit anti-OMgp polyclonal antibodies affinity purified on S1 fusion protein, and shows staining of OMgp at 105 kD. Molecular mass standards (kD) are shown: myosin, 200; β -galactosidase, 116; phosphorylase b, 92; BSA, 66; ovalbumin, 45. (B) Northern blot containing 10 μ g total RNA from human white matter probed with S1 cDNA after electrophoresis in 1.2% agarose with formaldehyde and capillary blotting. Relative positions of the 28-S and 18-S ribosomal RNA bands are indicated. A distinct signal is apparent just below the level of the 18-S ribosomal RNA at \sim 1.8 kb. (C) Southern blot containing 5 μ g genomic DNA from human peripheral blood lymphocytes per lane, cut with the three enzymes Eco RI (lane 1), Taq I (lane 2), and Bam HI (lane 3), and probed with radiolabeled S1 cDNA. A single band is seen in each case. Molecular size standards (kb) are shown.

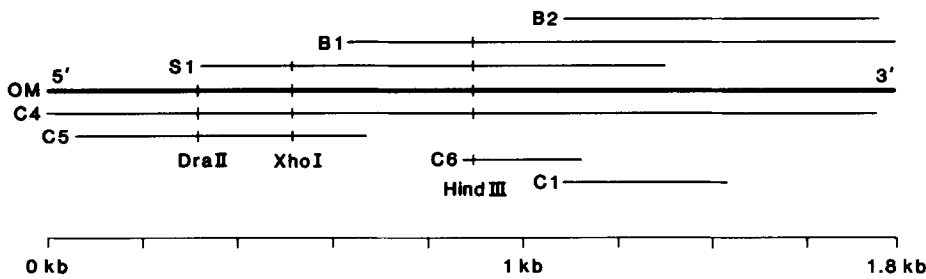


Figure 2. Restriction map of OMgp cDNA clones. The positive clones from the three libraries (B, neonatal brainstem; S, neonatal spinal cord; and C, fetal spinal cord) are shown. Vertical lines mark three restriction endonuclease cleavage sites. Note that C4 is \sim 40 nucleotides short of the full-length composite, which is represented by the solid line and is marked OM.

resed proteins from several tissues were stained with the affinity-purified antibodies. Only in the CNS was there a polypeptide that bound the antibodies, which is in keeping with our previous description of the distribution of OMgp (Mikol and Stefansson, 1988).

Nucleotide Sequence of OMgp cDNA

The full-length sequence of the overlapping clones (Fig. 3) consisted of 1,808 nucleotides, containing both the entire coding sequence and 3' and 5' noncoding sequences. One of the clones, C4, spanned the entire coding region. There is one nucleotide, No. 145, that remains ambiguous in the segment encoding the leader sequence. In one of the two clones spanning this segment it is G and in the other it is A, resulting in a glycine or aspartic acid in the deduced amino acid sequence. The coding sequence begins with a 17-amino acid leader sequence (Fig. 3), which is followed by a 32-amino acid NH_2 -terminal domain containing the 28-amino acid segment that had previously been revealed by protein sequencing (Mikol and Stefansson, 1988) (Fig. 4). This domain contains four cysteines spaced in a manner similar to what is found in the so-called EGF motifs that have been de-

scribed in several proteins including the EGF precursor, the EGF receptor (Stoscheck and King, 1986), the low density lipoprotein receptor (Russell et al., 1984; Yamamoto et al., 1984), the neurogenic Notch protein (Wharton et al., 1985), laminin (Pikkarainen et al., 1988), and the hexabrachions (Gulcher et al., 1989). However, the EGF motif contains at least six cysteines rather than four. Hence, we have elected to call this domain the cysteine-rich motif, or CR.

The NH_2 -terminal domain leads into a 172-residue domain that consists of $7\frac{1}{2}$ tandem LRs of \sim 24 amino acids each. Several proteins not only contain a CR at their NH_2 terminus, but also share with OMgp the LRs (Fig. 5 A). In addition to the proteins listed in Fig. 5 A, Toll (Hashimoto et al., 1988) and chaoptin (Reinke et al., 1988) from *Drosophila* contain both domains. LRs have been described in several other polypeptides including the α chain of human platelet glycoprotein Ib (Lopez et al., 1987), yeast adenylate cyclase (Kataoka et al., 1985), and the human leucine-rich glycoprotein of serum (Takahashi et al., 1985) (Fig. 5 B).

The LRs are followed by $4\frac{1}{2}$ serine/threonine-rich repeats (ST repeats). The α chain of platelet glycoprotein Ib also possesses similar ST repeats which are known to carry the

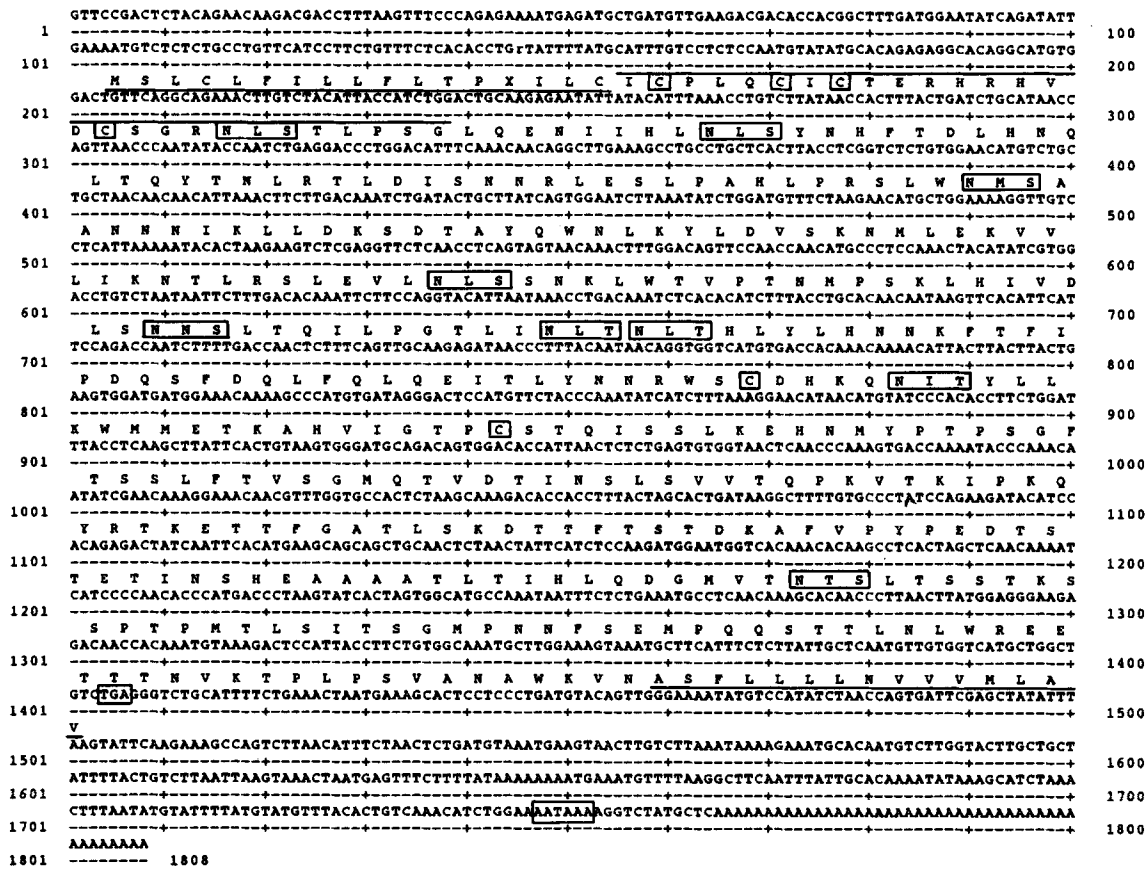


Figure 3. Nucleotide sequence of the overlapping clones and the deduced amino acid sequence of OMgp. The nucleotide sequence begins with 104 nucleotides of 5' untranslated sequence. The NH₂-terminal and potential COOH-terminal signal peptides are underlined. In the sequence of the mature polypeptide, the cysteines and the potential N-glycosylation sites are boxed. In the 3' untranslated sequence, the stop codon and polyadenylation signal are also boxed. Nucleotide No. 145 is still ambiguous; in one clone it is G and in another it is A, leading to either glycine or aspartic acid at position 14 in the NH₂-terminal signal peptide. The segment which had previously been revealed by NH₂-terminal amino acid sequencing is overlined.

N - TERMINAL EGF MOTIF

H₂N-I C P L Q C I C T E R H R H V D C S G R N L S T L P S G L Q E N

LEUCINE-RICH REPEATS

33 I I H L N L S Y W H F F D L E N Q L T Q Y T N
 L R T L D I S N N R L E S L P A H L P R S
 L W N M S A A N N N I K L L D K S D T A Y Q W N
 L K Y L D V S R N M L E K V V L I K N T L R S
 L E V L N L S S N K L W T V P T N M P S K
 L H I V D L S N N S L F T Q I L P G T L I N L T N
 L T H L Y L H N N R F T F I P D Q S F D Q L F Q
 L Q E I T L Y M N R W S * 204

SERINE-THREONINE REPEATS

205 C D H . K Q N I T Y L L K W M M E . . T R A H V I G . T . P C S T Q I S S L . K E H N M Y P T P
 S G F T . S S L F T V . S G M Q T V D T I N S L * . T . P C S T Q I S S L . K E H N M Y P T P
 S V T T Q P K V T K I P K . . Q Y R . T R E T T F G A T L S K D T T F T S T D K . A F V . P Y P
 E D T S T E T I N S H E A A A A T L . T I H L Q D G . M V T N . T S L T S S T K . . . S S P F P
 . . M T . L S I T S G . . . M P N N F S E M P Q Q S T T L N L W R E E T T T . . . N V K T E L P

C-TERMINUS

393 S V A N A W K V N - COOH

Figure 4. Domains of OMgp. The figure shows the four major domains of what we believe is the mature OMgp. It begins with the CR that is directly followed by 7½ LRs, then 4½ ST repeats. The predicted amino acid sequence ends in a short COOH-terminal hydrophobic region which probably serves as a signal for GPI attachment that is cleaved during biosynthesis of OMgp, leaving some unknown residue as the COOH-terminal amino acid, possibly asparagine (residue 401). The NH₂-terminal cysteines, as well as some of the sequence similarities within the LRs and ST repeats are highlighted. The two serine residues in the CR that may have been derived from ancestral cysteine residues are overlined (see text). Note the high content of proline residues (underlined) within the LRs.

A

OMgp	1 -	I C P L Q C I C T E R H R H V D C S G R N L S T . . L P S G L Q E N	- 32
Ib β	1 -	. C P A P C S C A G T L . . V D C G R R G L T W A S L P T A F P D	- 30
Decorin	25 -	V C P F R C Q C . . H L R V V Q C S D L G L D K . . V P K D L P D	- 54
Biglycan	25 -	M C P F G C H C . . H L R V V E C S D L G L . K . S V P K T S P D	- 54

B

L X X L d L S n N X L t X L p X X X l X X X X n	OMgp
L X X L X L X n N k I X X V P k X X f s X X X X	Biglycan
L X X L X L X n N k I s x V X X g X f X X l k X	Decorin
L t X L X L s X n X L X X L P s X L F X H X t N	dToll
L r X L D L s h N X L X X I p X d l F X X L X X	dChaoptin
L X X L X L S X N k l X X i P X e l X X L X X	yAdCyclase
L r X L d L s g N X L e s L P p g L L q g l p q	LRG
L t X L X L s X N X L t t L P P g l l d a l p a	Ib β
L S X L X L s x N X L t t L P X G l l X X l p X	Ib α

Figure 5. The CR (A) and LRs (B) of OMgp. (A) This figure shows portions of four polypeptides which are composed of a CR at the NH₂-terminus with a contiguous series of LRs. Numbers refer to amino acid positions in the mature polypeptides. Sequence similarities are highlighted, accounting for conservative substitutions. *Ib β* , the β chain of platelet glycoprotein Ib (Lopez et al., 1988); *Decorin* and *biglycan*, two proteoglycans from bone (Fisher et al., 1989). (B) This shows the consensus sequences of the LRs in OMgp and eight other proteins. *dToll*, the product of the *Toll* gene in *Drosophila*; *dChaoptin*, the product of the *chaoptic* gene in *Drosophila*; *yAdCyclase*, adenylate cyclase from yeast; *LRG*, the serum leucine-rich glycoprotein; *Ib α* , the α chain of platelet glycoprotein Ib.

bulk of its O-linked carbohydrate and directly follow its LRs (Lopez et al., 1987). The second ST repeat of OMgp is truncated with only 23 amino acids as compared with ~40 amino acids in the other four. The predicted COOH terminus consists of a hydrophobic stretch of ~15 amino acids that is in keeping with the weak sequence requirements for attachment of proteins to GPI (Fig. 6). We do not know where the mature protein ends and the COOH-terminal signal peptide begins, but based on comparison to other GPI-linked proteins, we believe it could be after residue number 401 in the mature protein. The 3' untranslated part contains a stop codon, a polyadenylation signal, and a poly(A) tail (Fig. 3).

There are nine potential asparagine-linked oligosaccharide attachment sites, most of which are probably used, because OMgp contains at least 25 kD of N-linked carbohydrates (Mikol and Stefansson, 1988). The deduced size of the mature polypeptide backbone after NH₂-terminal and COOH-terminal peptide cleavage is ~46 kD, which is at odds with our previous evaluation of ~75 kD after enzymatic deglycosylations. However, it has proved difficult for us to enzymatically remove all of the O-linked oligosaccharide

chains, based on residual peanut agglutinin binding. Clearly, the presence of residual O-linked chains can alter electrophoretic mobility dramatically (Johnson and Spear, 1983; Cummings et al., 1983; Lublin et al., 1986). Recent attempts to enzymatically remove both O-linked and N-linked glycans from OMgp have resulted in a polypeptide that binds the polyclonal antibodies but not peanut agglutinin and measures ~52 kD (our unpublished results), which is in fair agreement with the deduced size. A factor contributing to the overestimation of the size of the polypeptide backbone of OMgp is the presence of the COOH-terminal inositol-containing glycan, which by virtue of its amide linkage through an ethanolamine group is likely to be resistant to the commonly used glycosidases.

OMgp Contains the HNK-1 Carbohydrate Epitope

Contrary to what we previously reported (Mikol and Stefansson, 1988), OMgp contains the HNK-1 carbohydrate epitope. Most of OMgp elutes from an anion exchange column (MonoQ) between 0.10 and 0.60 M NaCl, and between 0.80

OMgp	ANAWKVN	ASFLLLLVVVMLAV
<i>T. brucei</i> V.S.G.	DKEDCRN	GSFLTQKQFALMVSAAFVTLIF
<i>H. Thy-1</i>	RDKLVKC	EGISLLAQNTSWLLLLLSLLQATDFMSL
<i>P. falciparum</i> gp 190	DGIFCSS	SNFLGISFLLIIMLILYSFI
N-CAM	QPTVIPA	TLGSPSTSSSFVSLLSAVTLLLLC
<i>D. discoideum</i> CsA	EETEAPS	SATTLISPLSLIVIFISFVLLI
Scrapie prion	YYDGRRS	SAVLFSSPPVILLISFLIFLNVG

Figure 6. COOH-terminal amino acid sequences of several proteins attached to membranes via a glycosylphosphatidylinositol moiety. The list is adapted from Ferguson and Williams (1988). The point of cleavage during attachment of the GPI anchor is shown by spacing where it is known (*Trypanosoma brucei* variant surface glycoprotein [*T. brucei* V.S.G.], human Thy-1 [*H. Thy-1*]) or presumed to occur (all others). The sequences shown correspond to *T. brucei* variant surface glycoprotein Group 3 MIT at 1.5b; human Thy-1; *Plasmodium falciparum* gp190 (*P. falciparum* gp 190), the neural cell adhesion molecule (N-CAM), *Dictyostelium discoideum* contact site A (*D. discoideum* CsA), and the scrapie prion protein.

and 1.00 M NaCl, with a peak at 0.15–0.35 M NaCl, as detected by staining with a mAb against OMgp (Fig. 7 A). However, the HNK-1 epitope is only on OMgp that elutes between 0.15 and 0.40 M NaCl, with a peak at 0.25–0.35 M NaCl (Fig. 7 B). Therefore, it appears that a subpopulation of OMgp from human brain contains the HNK-1 epitope. It is likely that one of the reasons that the HNK-1 carbohydrate on OMgp eluded us initially is the fact that it is only found on a subpopulation of the molecules, which has also been observed for other HNK-1-containing molecules (Kruse et al., 1984; Poltorak et al., 1987; Faissner, 1987). Furthermore, the expression of HNK-1 on OMgp may show regional and/or developmental variation. No evidence of HNK-1 was observed in our early isolates of OMgp, which were done using cerebral white matter as starting material, whereas recent isolates using whole brain have unequivocally contained HNK-1 carbohydrate on a subpopulation of OMgp molecules.

Discussion

In this paper we describe the cloning and sequencing of OMgp, which is a protein we have previously identified, isolated, and characterized somewhat biochemically (Mikol and Stefansson, 1988). We propose that, based on the sharing of two structurally unrelated domains, OMgp together with several other proteins form a hitherto unrecognized family. In addition we report that OMgp contains the HNK-1 carbohydrate.

The NH₂-terminus of the deduced amino acid sequence contains four cysteines spaced in a manner similar to what is found in the EGF motifs (Russell et al., 1984; Wharton et al., 1985; Pikkariainen et al., 1988; Gulcher et al., 1989), although the full EGF motif typically contains six or eight cysteines. However, there are two serines that follow the last cysteine in this group with spacing that would fit for the missing two cysteines (Fig. 4), and both serines could have been derived from cysteines through a one-nucleotide substitution (Brenner, 1988). We are tempted to postulate that originally there was a full EGF motif in the NH₂ terminus of OMgp.

It is not clear what role the CR plays in OMgp. However, what we know about the molecules that contain the authentic EGF motifs suggests that the role of OMgp may be either in adhesion or as a mitogen.

The NH₂-terminal CR of OMgp is followed by 7½ LRs that are similar to repeats that have been described in several other proteins, including bone small proteoglycans I (biglycan) and II (decorin) (Fisher et al., 1989), the α and β chains of platelet glycoprotein Ib (Lopez et al., 1988), the leucine-rich glycoprotein of serum (Takahashi et al., 1985), yeast adenylate cyclase (Kataoka et al., 1985), and *Drosophila Toll* (Hashimoto et al., 1988) and chaptin (Reinke et al., 1988). It has been postulated that the LRs mediate adhesion between the proteins containing the repeats and other molecules. There is little experimental evidence in support of this theory except in the case of the α chain of platelet glycoprotein Ib, which has been shown to bind to von Willebrand factor through a polypeptide fragment that includes the LRs (Vicente et al., 1988). In any case, the leucine residues of the LRs must impose on a protein a hydrophobic quality such that the protein may either bury these residues within its interior, or it may keep them outside in order to interact with hydrophobic molecules. Takahashi et al. (1985) have observed that the LR structure may be functionally similar to 11–22-amino acid repeats of apolipoproteins that have amphipathic helical potential (Kaiser and Kézdy, 1983). The LRs do not have the appropriate amino acid spacing for either α -helical or β -sheet conformation but are likely to form amphipathic structures (Eisenberg, 1984). In the case of chaptin, it has been postulated that the LRs may mediate adhesion between adjacent photoreceptor membranes through interactions with residues on the surface of the lipid bilayer rather than by penetration into it (Reinke et al., 1988). This hypothesis is supported by the fact that chaptin behaves in many ways as an integral membrane protein, although it is lacking both a likely transmembrane segment and evidence for a covalent linkage to molecules intrinsic to the plasma membrane. It is also supported by the observation that loss of function mutations in the *chaptin* gene cause separation of otherwise closely apposed membranes of adjacent photoreceptor elements (Van Vactor et al., 1988).

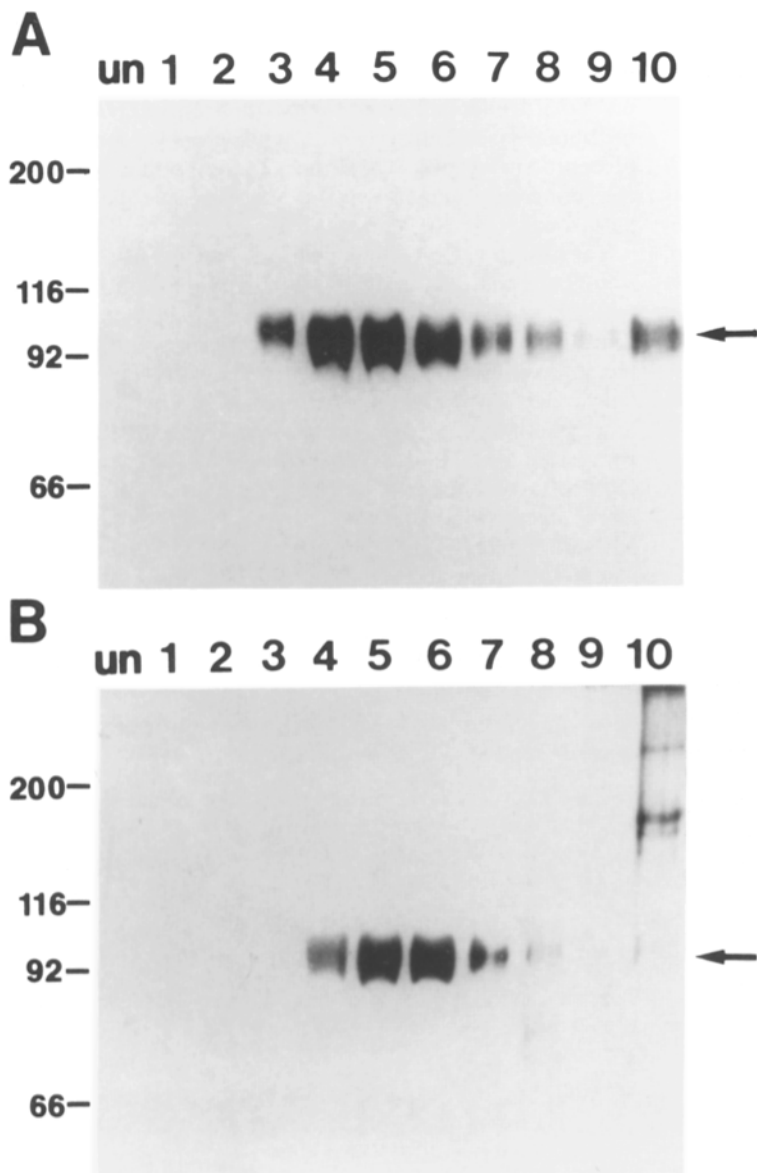


Figure 7. Western blots containing OMgp eluted from an anion exchange column. An enriched sample of OMgp which had been purified by peanut agglutinin affinity chromatography was applied to an anion exchange column (MonoQ) and eluted with a NaCl gradient (see Materials and Methods). Fractions were electrophoretically separated and the blots stained with either (A) mAb16, an anti-OMgp mAb, or (B) anti-HNK-1 mAb. The unbound fraction is shown in the far left lane, while fractions eluted at increasing NaCl concentration are shown in lanes 1–10: lane 1, 0.00–0.05 M NaCl; lane 2, 0.05–0.10 M; lane 3, 0.10–0.15 M; lane 4, 0.15–0.25 M; lane 5, 0.25–0.30 M; lane 6, 0.30–0.35 M; lane 7, 0.35–0.40 M; lane 8, 0.40–0.60 M; lane 9, 0.60–0.80 M; lane 10, 0.80–1.00 M. Staining of OMgp is indicated by arrows. Note that mAb16 stains OMgp mostly in lanes 3–8 and 10, whereas anti-HNK-1 mAb stains OMgp mostly in lanes 4–7, thus comprising a subset of OMgp molecules. Additional HNK-1 containing polypeptides of uncertain identity are found to elute in the 1.00 M NaCl wash (B, lane 10).

Of the nine polypeptides known to contain LRs, the yeast adenylate cyclase and the leucine-rich glycoproteins of serum are the only ones that do not contain a CR at the NH₂ terminus. The α chain of glycoprotein Ib is rich in cysteines at the NH₂ terminus but the cysteines do not have the same distribution as in the case of OMgp. The five remaining molecules contain four cysteines with distributions similar to that described here for OMgp (Fig. 5 A). In addition to the proteins listed in Fig. 5 A, chaoptin and the *Toll* gene product, both from *Drosophila melanogaster*, belong to this group. The β chain of platelet glycoprotein Ib contains both a CR and a single LR, and is similar in sequence to a part of the α chain of the same protein (Lopez et al., 1988). It is distinctly possible that the glycoprotein Ib of platelets provides us with an example of duplication of a gene encoding both the CR and the LRs, where in one of the resultant genes the CR has been lost. However, since the α and β chains are linked together in the same molecule it could be argued that the function of the CR lost from the α chain was taken over by the β chain. Decorin and biglycan have 55% overall

amino acid identity and consist almost solely of the NH₂-terminal CR and the LRs (Fisher et al., 1989). However, their NH₂-terminal domains have 72% amino acid identity. Thus in these two related molecules there appears to be more evolutionary pressure to maintain the sequence of the CR than the LRs. This may not reflect more evolutionary pressure to maintain the function of the CR, but merely a more rigid correlation between primary structure and function.

We propose that the coexistence in the same protein of the CR at the NH₂ terminus and the contiguous LRs defines a family of proteins that can be called the CR–LR family. The proteins in this family may therefore either share two functions or a function that requires the coordinated contribution of these two structurally unrelated domains. The fact that the CR and LRs occur separately could be regarded as an argument in favor of the first possibility. We believe that the members of this family are all likely to play their major roles by binding to other molecules through the LRs and in some instances also through other interactive domains. The membrane-bound proteins in the family may mediate interactions

between cells and thus serve as adhesion molecules. The leucine-rich glycoprotein of serum, which is of unknown origin and function, may be membrane derived (Takahashi et al., 1985), or it could serve as a transport vehicle for hydrophobic molecules.

The LRs of OMgp are followed by a domain of ST repeats and can be divided into 4½ repeats (Fig. 4). This ST domain of OMgp is shared by the α chain of glycoprotein Ib (Lopez et al., 1987) and a number of other proteins, including the low density lipoprotein receptor (Cummings et al., 1983), decay-accelerating factor (Lublin et al., 1986), and several viral proteins (Johnson and Spear, 1983). In these proteins, the serine and threonine residues are concentrated on the portion of the protein that juxtaposes the extracellular face of the plasma membrane and are used for attachment of O-linked carbohydrates. The function of such clustered O-linked carbohydrates is unknown, but it is conceivable that by a hydrating effect such proteins are allowed to exist in an extended conformation away from the plasma membrane.

OMgp is the only protein containing LRs that has been reported to be linked to membranes via a GPI moiety, although the platelet glycoprotein Ib β chain contains covalently bound fatty acid through a thioester linkage (Muszbek and Laposata, 1989). The predicted COOH terminus of OMgp consists of a hydrophobic region that conforms to the weak sequence requirements signaling protein attachment to GPI (Ferguson and Williams, 1988) (Fig. 6). This supports our previous work using enzymes to determine the nature of the membrane linkage (Mikol and Stefansson, 1988). We do not know where the COOH terminus of the mature protein ends and the COOH-terminal signal peptide begins but our guess, based on comparison to other GPI-linked proteins, is that the cleavage occurs after residue number 401 in the mature polypeptide. While a COOH-terminal hydrophobic domain can result in peptide cleavage and addition of a GPI tail, the very nature of the COOH-terminal signal does not appear to be critical. For example, when cDNA encoding the COOH-terminal hydrophobic sequence of decay-accelerating factor, a GPI-linked protein, was replaced with cDNA encoding NH₂-terminal signal peptides or random hydrophobic sequences, the protein appeared on the surfaces of transfected cells with a GPI anchor (Caras and Weddell, 1989).

The HNK-1 epitope contains a sulfated glucuronosyl residue (Shashoua et al., 1986) which is probably attached to a complex N-linked triantennary oligosaccharide (Mikol et al., 1988). In the CNS the HNK-1 carbohydrate may be a marker for adhesion molecules (Kruse et al., 1985), and the HNK-1 carbohydrate itself has been shown to mediate adhesion between cells (Künemund et al., 1988). It is of interest that a subpopulation of OMgp contains the HNK-1 carbohydrate. Heterogeneous expression of the HNK-1 carbohydrate has also been observed for other molecules and may allow for modulation of cell adhesion (Kruse et al., 1984; Poltorak et al., 1987; Faissner, 1987).

The OMgp is present on oligodendrocytes and in CNS myelin. Because oligodendrocytes are the cells that synthesize CNS myelin, it is probably not incorrect to say that OMgp is specific for oligodendrocytes. CNS myelin can be viewed as a complicated set of closely apposed oligodendrocytic membranes, and our hypothesis in its most general form is that OMgp plays a role in establishing and maintaining some aspects of the membrane interactions that go into making

myelin. Several other proteins have been implicated in determining the structural integrity of CNS myelin. Proteolipid protein and the myelin basic proteins are thought to contribute toward maintaining the apposition of compact myelin membranes (see Lemke, 1988), while the myelin-associated glycoprotein is a minor component believed to play a role in mediating the initial oligodendrocyte–neuron adhesion in periaxonal regions (Poltorak et al., 1987).

The OMgp is well equipped to mediate adhesion because it contains the LRs and the HNK-1 carbohydrate, both of which have been shown to participate in adhesive processes. Since OMgp contains domains that are not very hydrophobic, in addition to having a considerable complement of carbohydrates, some of which are charged, we believe that it is more likely to play a role in the noncompacted than in the compacted myelin. This is supported by our finding that OMgp fractionates into loose rather than compact myelin membranes (Mikol and Stefansson, 1988), as well as by our preliminary immunoelectron microscopic work which places OMgp at the nodes of Ranvier (our unpublished results). Hence, our hypothesis is that OMgp mediates adhesion between membranes at the nodes of Ranvier, either between two oligodendrocytic membranes (between two loops), between oligodendrocytes and axons, or even between oligodendrocytes and a third cell type (type-2 astrocytes) (French-Constant and Raff, 1986).

We are grateful to Celia and Anthony Campagnoni for generously providing us with a cDNA library human fetal spinal cord.

This work was supported in part by a grant from the Multiple Sclerosis Society to Kari Stefansson and by a grant from the National Institutes of Health (NIH) (5 P01 HS-21442-03) to Kari Stefansson. Daniel D. Mikol was supported through the Medical Scientist Training Program (NIH, Public Health Service grant 5 T32GM07281).

Received for publication 4 August 1989 and in revised form 3 October 1989.

References

- Brenner, S. 1988. The molecular evolution of genes and proteins: a tale of two serines. *Nature (Lond.)* 334:528–530.
- Bunge, R. P., M. B. Bunge, and M. Cochran. 1978. Some factors influencing the proliferation and differentiation of myelin-forming cells. *Neurology* 28(Suppl.):59–67.
- Caras, I. W., and G. N. Weddell. 1989. Signal peptide for protein secretion directing glycosylphospholipid membrane anchor attachment. *Science (Wash. DC)* 243:1196–1198.
- Chomczynski, P., and N. Sacchi. 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* 162:156–159.
- Cummings, R. D., S. Kornfeld, W. J. Schneider, K. K. Hobgood, H. Tolleshaug, M. S. Brown, and J. L. Goldstein. 1983. Biosynthesis of N- and O-linked oligosaccharides of the low density lipoprotein receptor. *J. Biol. Chem.* 258:15261–15273.
- Devereux, J., P. Haerberli, and O. Smithies. 1984. A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* 12:387–395.
- Eisenberg, D. 1984. Three-dimensional structure of membrane and surface proteins. *Annu. Rev. Biochem.* 53:595–623.
- Faissner, A. 1987. Monoclonal antibody detects carbohydrate heterogeneity on the murine cell adhesion molecule L1. *Neurosci. Lett.* 83:327–332.
- Feinberg, A. P., and B. Vogelstein. 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* 132:6–13.
- Ferguson, M. A. J., and A. F. Williams. 1988. Cell-surface anchoring of proteins via glycosylphosphatidylinositol structures. *Annu. Rev. Biochem.* 57:285–320.
- Ferguson, M. A. J., K. Haldar, and G. A. M. Cross. 1985a. *Trypanosoma brucei* variant surface glycoprotein has a sn-1,2-dimyristyl glycerol membrane anchor at its COOH terminus. *J. Biol. Chem.* 260:4963–4968.
- Ferguson, M. A. J., M. G. Low, and G. A. M. Cross. 1985b. Glycosyl-sn-1,2-dimyristyl-phosphatidylinositol is covalently linked to *Trypanosoma brucei* variant surface glycoprotein. *J. Biol. Chem.* 260:14547–14555.

- french-Constant, C., and M. C. Raff. 1986. The oligodendrocyte-type-2 astrocyte cell lineage is specialized for myelination. *Nature (Lond.)*. 323:335-338.
- 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.
- Futerman, A. H., M. G. Low, and I. Silman. 1983. A hydrophobic dimer of acetylcholinesterase from *Torpedo californica* electric organ is solubilized by phosphatidylinositol-specific phospholipase C. *Neurosci. Lett.* 40:85-89.
- Futerman, A. H., R.-M. Fiorini, E. Roth, M. G. Low, and I. Silman. 1985a. Physicochemical behavior and structural characteristics of membrane-bound acetylcholinesterase from *Torpedo* electric organ. *Biochem. J.* 226:369-377.
- Futerman, A. H., M. G. Low, K. E. Ackerman, W. L. Sherman, and I. Silman. 1985b. Identification of covalently bound inositol in the hydrophobic membrane-anchoring domain of *Torpedo* acetylcholinesterase. *Biochem. Biophys. Res. Commun.* 129:312-317.
- Futerman, A. H., M. G. Low, D. M. Michaelson, and I. Silman. 1985c. Solubilization of membrane-bound acetylcholinesterase by a phosphatidylinositol-specific phospholipase C. *J. Neurochem.* 45:1487-1494.
- Galfré, G., and C. Milstein. 1981. Preparation of monoclonal antibodies: strategies and procedures. *Methods Enzymol.* 73:1-46.
- Gulcher, J. R., L. S. Marton, and K. Stefansson. 1986. Two large glycosylated polypeptides found in myelinating oligodendrocytes but not in myelin. *Proc. Natl. Acad. Sci. USA.* 83:2118-2122.
- Gulcher, J. R., D. E. Nies, L. S. Marton, and K. Stefansson. 1989. An alternatively spliced region of the hexabrachion contains a repeat of potential N-glycosylation sites. *Proc. Natl. Acad. Sci. USA.* 86:1588-1592.
- 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.
- He, H.-T., J. Barbet, J.-C. Chaix, and C. Goriadis. 1986. Phosphatidylinositol is involved in the membrane attachment of NCAM-120, the smallest component of the neural cell adhesion molecule. *EMBO (Eur. Mol. Biol. Organ.) J.* 5:2489-2494.
- Hemperly, J. J., G. M. Edelman, and B. A. Cunningham. 1986. cDNA clones of the neural cell adhesion molecule (N-CAM) lacking a membrane-spanning region consistent with evidence for membrane attachment via a phosphatidylinositol intermediate. *Proc. Natl. Acad. Sci. USA.* 83:9822-9826.
- Henikoff, S. 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. *Gene (Amst.)*. 28:351-359.
- Johnson, D. C., and P. G. Spear. 1983. O-linked oligosaccharides are acquired by herpes simplex virus glycoproteins in the Golgi apparatus. *Cell.* 32:987-997.
- Kaiser, E. T., and F. J. Kézdy. 1983. Secondary structures of proteins and peptides in amphiphilic environments (a review). *Proc. Natl. Acad. Sci. USA.* 80:1137-1143.
- Kataoka, T., D. Broek, and M. Wigler. 1985. DNA sequence and characterization of the *S. cerevisiae* gene coding adenylate cyclase. *Cell.* 43:493-505.
- Kruse, J., R. Mailhammer, H. Wernecke, A. Faissner, I. Sommer, C. Goriadis, and M. Schachner. 1984. Neural cell adhesion molecules and myelin-associated glycoprotein share a common carbohydrate moiety recognized by monoclonal antibodies L2 and HNK-1. *Nature (Lond.)*. 311:153-155.
- Kruse, J., G. Keilhauer, A. Faissner, R. Timpl, and M. Schachner. 1985. The J1 glycoprotein: a novel nervous system cell adhesion molecule of the L2/HNK-1 family. *Nature (Lond.)*. 316:146-148.
- 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.
- Lemke, G. 1988. Unwrapping the genes of myelin. *Neuron.* 1:535-543.
- Lopez, J. A., D. W. Chung, K. Fujikawa, F. S. Hagen, T. Papayannopoulou, and G. J. Roth. 1987. Cloning of the α chain of human platelet glycoprotein Ib: a transmembrane protein with homology to leucine-rich α 2-glycoprotein. *Proc. Natl. Acad. Sci. USA.* 84:5615-5619.
- 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.
- Low, M. G., and J. B. Finean. 1977. Non-lytic release of acetylcholinesterase from erythrocytes by a phosphatidylinositol-specific phospholipase C. *FEBS (Fed. Eur. Biochem. Soc.) Lett.* 82:143-146.
- Low, M. G., and P. W. Kincade. 1985. Phosphatidylinositol is the membrane-anchoring domain of the Thy-1 glycoprotein. *Nature (Lond.)*. 318:62-64.
- Lublin, D. M., J. Krsek-Staples, M. K. Pangburn, and J. P. Atkinson. 1986. Biosynthesis and glycosylation of the human complement regulatory protein decay-accelerating factor. *J. Immunol.* 137:1629-1635.
- Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY. 545 pp.
- 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., L. Wrabetz, L. S. Marton, and K. Stefansson. 1988. Developmental changes in the molecular weights of polypeptides in the human CNS that carry the HNK-1 epitope and bind *Phaseolus vulgaris* lectins. *J. Neurochem.* 50:1924-1928.
- Muszbek, L., and M. Laposata. 1989. Glycoprotein Ib and glycoprotein IX in human platelets are acylated with palmitic acid through thioester linkages. *J. Biol. Chem.* 264:9716-9719.
- Pikkarainen, T., T. Kallunki, and K. Tryggvason. 1988. Human laminin B2 chain. Comparison of the complete amino acid sequence with the B1 chain reveals variability in sequence homology between different structural domains. *J. Biol. Chem.* 263:6751-6758.
- Poltorak, M., R. Sadoul, G. Keilhauer, C. Landa, T. Fahrig, and M. Schachner. 1987. Myelin-associated glycoprotein, a member of the L2/HNK-1 family of neural cell adhesion molecules, is involved in neuron-oligodendrocyte and oligodendrocyte-oligodendrocyte interaction. *J. Cell Biol.* 105:1893-1899.
- 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.
- Russell, D. W., W. J. Schneider, T. Yamamoto, K. L. Luskey, M. S. Brown, and J. L. Goldstein. 1984. Domain map of the LDL receptor: sequence homology with the epidermal growth factor precursor. *Cell.* 37:577-585.
- Sadoul, K., A. Meyer, M. G. Low, and M. Schachner. 1986. Release of the 120 kDa component of the mouse neural cell adhesion molecule N-CAM from cell surfaces by phosphatidylinositol-specific phospholipase C. *Neurosci. Lett.* 72:341-346.
- Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA.* 74:5463-5467.
- Shashoua, V. E., P. F. Daniel, M. E. Moore, and F. B. Jungalwala. 1986. Demonstration of glucuronic acid on brain glycoproteins which react with HNK-1 antibody. *Biochem. Biophys. Res. Commun.* 138:902-909.
- Snyder, M., and R. W. Davis. 1985. Screening λ gt11 expression libraries with antibody probes. In *Hybridoma Technology in the Biosciences and Medicine*. T. A. Springer, editor. Plenum Publishing Corp., New York. 397-406.
- Stoscheck, C. M., and L. E. King, Jr. 1986. Functional and structural characteristics of EGF and its receptor and their relationship to transforming proteins. *J. Cell. Biochem.* 31:135-152.
- 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., Y. Takahashi, and F. W. Putnam. 1985. 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.
- Tse, A. G. D., A. N. Barclay, A. Watts, and A. F. Williams. 1985. A glycosphospholipid tail at a carboxyl terminus of the Thy-1 glycoprotein of neurons and thymocytes. *Science (Wash. DC)*. 230:1003-1008.
- Van Vactor, D., D. E. Krantz, R. Reinke, and S. L. Zipursky. 1988. Analysis of mutants in chaoptin, a photoreceptor cell-specific glycoprotein in *Drosophila*, reveals its role in cellular morphogenesis. *Cell.* 52:281-290.
- Vicente, V., P. J. Kostel, and Z. M. Ruggeri. 1988. Isolation and function of the von Willebrand factor-binding domain located between residues His¹-Arg²⁹³ of the α -chain of glycoprotein Ib. *J. Biol. Chem.* 263:18473-18479.
- Wharton, K. A., K. M. Johansen, T. Xu, and S. Artavanis-Tsakonas. 1985. Nucleotide sequence from the neurogenic locus Notch implies a gene product that shares homology with proteins containing EGF-like repeats. *Cell.* 43:567-581.
- Yamamoto, T., C. G. Davis, M. S. Brown, W. J. Schneider, M. L. Casey, J. L. Goldstein, and D. W. Russell. 1984. The human LDL receptor: a cysteine-rich protein with multiple Alu sequences in its mRNA. *Cell.* 39:27-38.