Gene duplication in bovine brain myelin proteolipid and homology with related proteins

Richard A. Laursen, Mohammed Samiullah and Marjorie B. Lees

Department of Chemistry, Boston University, Boston, MA 02215 and *E.K. Shriver Center, Waltham, MA 02154, USA*

Received 27 June 1983

Analysis of the amino acid sequence of bovine brain myelin proteolipid reveals not only extensive internal homology, but also homology with portions of the myelin basic protein, the peripheral nerve myelin protein, P_O, and with the small proteolipid subunit of mitochondrial ATP synthase. These results suggest that the myelin proteolipid gene has been constructed from a small number of genetic elements, and that these elements are also found in non-myelin proteins. Furthermore, the proteolipid appears to have evolved by acquisition of elements from a 'gene pool' over a period of time, rather than by a simple duplication mechanism.

Gene duplication *Myelin proteolipid* *P_O glycoprotein* *Membrane protein* *ATP synthase* *Myelin basic protein*

1. INTRODUCTION

The major protein component of central nervous system myelin is the intrinsic membrane protein, myelin proteolipid. Recently, the amino acid sequence of bovine brain myelin proteolipid was determined through the efforts of our laboratory and those in [1–3]. Based on its sequence and other known properties, we have proposed [4] a model that describes the organization of the proteolipid in the lipid bilayer and incorporates 3 hydrophobic *trans*-membrane and 2 *cis*-membrane (defined here as entering and exiting the same side) domains which are inserted in the lipid bilayer, and which alternate with hydrophilic extra-membrane segments (fig.1). Repeating structural features in the model, namely a putative *cis*-membrane domain followed by a *trans*-membrane one, led us to search for internal homology within the proteolipid. As described here, there is indeed homology. Furthermore the homology extends to other proteins whose overall functions are related to, but different from that of the proteolipid.

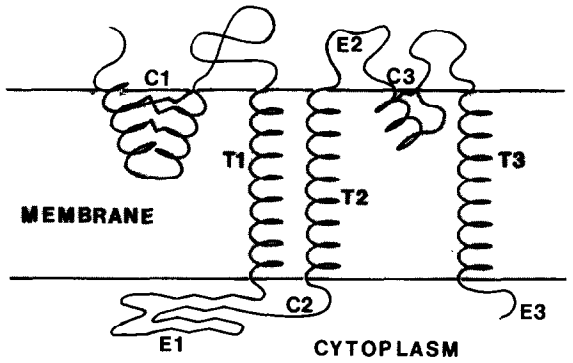

Fig.1. Hypothetical model of brain myelin proteolipid (PLP) and its organization in the membrane lipid bilayer [4]. T1–T3 are homologous, α-helical, hydrophobic *trans*-membrane segments that span the bilayer; C1 and C3 are homologous, partially hydrophobic *cis*-membrane structures that enter and exit the membrane on the same side (C2 is homologous with C1 and C3 but does not penetrate the membrane); and E1–E3 are homologous, charged extra-membrane domains. See fig.2 for alignment of domains. We have postulated [4] that in native myelin, the hydrophobic headpiece of C1, and possibly of C3, can dissociate from the bilayer as shown, and become imbedded in the apposite membrane lamella, thus spanning the extracellular space and helping to stabilize the lamellar myelin structure.

Published by Elsevier Science Publishers B.V.
00145793/83/$3.00 © 1983 Federation of European Biochemical Societies

2. RESULTS AND DISCUSSION

As indicated in fig.2, the proteolipid can be thought of as being composed of 3 types of domain: *cis*-membrane (C1–C3; C2 appears to be degenerate), *trans*-membrane (T1–T3), and the polar, predominantly basic extra-membrane domains (E1–E3). The carboxyl terminal segment (residues 197–276) can be aligned with the amino-terminal region (residues 1–134), with the insertion of only a few gaps. In cases where peptide alignments were not obvious, we used the model (fig.1) to guide us to a final choice. For example, C1 and C3 contain hydrophobic stretches that are proposed [4] to contain hydrophobic interior turns, so we aligned the helix-breaking prolines (residues 14 and 210) that are common to these turns (fig.2). Similarly we attempted to align analogous cysteine residues (Cys 24 and 32 with 219 and 227) that are postulated [4] to crosslink and stabilize the membrane-binding domains in C1 and C3. The putative α-helical *trans*-membrane segments (T1–T3) have several alignment possibilities. They have been aligned so as to minimize gaps and to maximize mutual overlap of hydrophobic zones, each of which contains about 30 residues.

The existence of homologous extra-membrane regions (E1–E3) became apparent when a comparison was made with the myelin basic protein. We have proposed [4] that the highly basic peptide segment comprising residues 91–150 is located on the cytoplasmic side of the membrane, partly on the basis of analogy with myelin basic protein [5], which is located cytoplasmically. As shown in fig.2, there is a remarkable similarity in sequence, if gaps are included, between E1 and basic protein residues 113–168. With this relationship established, the segments contained in E2 and E3 can also be aligned, even though they are much shorter.

The regular arrangement of elements, *cis*-membrane–*trans*-membrane–extra membrane, suggests that the myelin proteolipid gene has been constructed by duplication of the genetic material of smaller ancestral proteins; i.e., a small, basic protein and a small hydrophobic, membrane-penetrating protein containing around 80 amino acids. A contemporary protein with the latter characteristics is the proteolipid subunit of the mitochondrial ATP synthase F_o complex [6]; this protein also shows strong homology with myelin proteolipid (fig.2). It should be emphasized that the alignment in fig.2 does not represent the maximal homology between each pair of peptide segments, since a shift to maximize overlap of one pair can result in a mismatch of another pair. In fig.3 we present the results of maximal pairwise alignment of all segments. The percentage of iden-

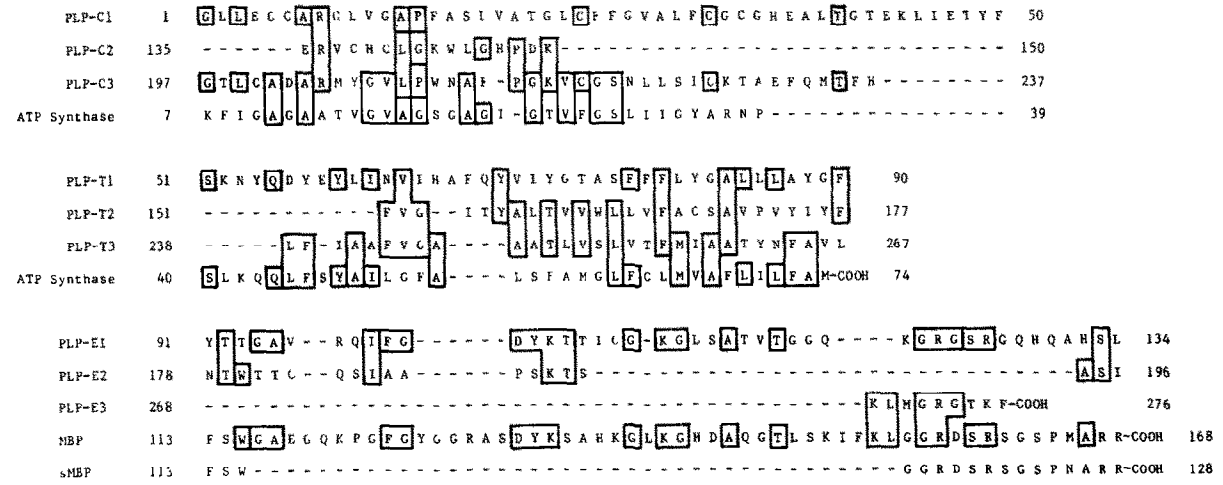

Fig.2. Alignment of homologous segments in bovine brain myelin proteolipid (PLP); C, T, and E refer to postulated *cis*-, *trans*- and extra-membrane domains, respectively. Comparisons are made with bovine ATP synthase proteolipid subunit [6] and bovine myelin basic protein (MBP) [5]; sMBP is a small form of basic protein found in the rat.

	T1	T2	T3	Synthase trans	
C1		7/27(1),26	6/30(0),20	6/35(0),17	T1
C2	3/16(0),19		8/27(1),30	6/27(1),22	T2
C3	8/41(1),20	5/16(1),31		9/30(1),30	T3
Synthase cis	5/35(0),14	3/16(1),19	9/34(0),26		Synthase trans
	C1	C2	C3	Synthase cis	

Fig.3. Comparison of homologous proteolipid segments (cf. fig.2) aligned so as to maximize identities. *Trans* segments are compared on the right-hand side of the diagonal; *cis* on the left. For each pair, the fraction represents the number of identities/number of residues in the peptide segments compared; the values in parentheses are the number of gaps inserted; and the value on the right is the percent identity.

tities in all cases is greater than the average percentage (~10%) expected for purely random matches, although in some cases (e.g., the C2 comparisons) the number of matches is so small as to be of questionable significance. Nevertheless, one trend seems significant, namely that the percentage similarity between ATP synthase proteolipid and myelin proteolipid segments increases in the order T3 > T2 > T1 and C3 > C2 > C1. This suggests that myelin proteolipid may have evolved by subsequent borrowing of genetic elements of more recent ancestors of the ATP synthase proteolipid (fig.4), rather than by duplication of a primitive myelin proteolipid gene. Such a mechanism explains why T3 and C3 are more similar to the synthase proteolipid than they are to T1 and C1, respectively. Variations on the scheme in fig.4, for example, the precise time at which deletions were introduced, are also possible.

The evolutionary mechanism postulated in fig.4 also suggests a possible origin for the DM 20 form of myelin proteolipid. The DM 20 is still poorly characterized, but it appears to have a molecular mass at least 5 kDa lower than the proteolipid, and has the same amino- and carboxyl-terminal amino acids [7,8]. The DM 20 could be the immediate precursor (or a modified form of it) of the myelin proteolipid, lacking the segment C3 + T3, which has a molecular mass of about 75000 (fig.4). As a consequence, DM 20 would have a structure in the membrane similar to that of intact myelin proteolipid (fig.1), except that C3 and T3 would be absent, and E3 would be joined to E2 on the ex-

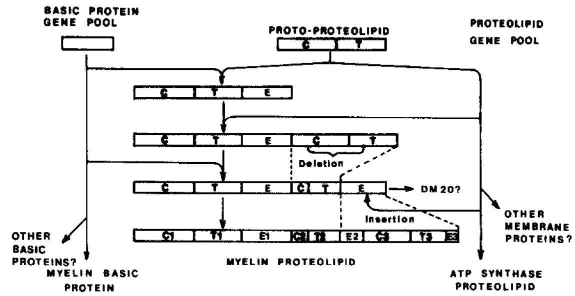

Fig.4. Hypothetical scheme for the evolution of myelin proteolipid. It is proposed that myelin proteolipid evolved by successive incorporation of genetic elements from 'proteolipid' and 'basic protein' pools, rather than by duplications of an early form of myelin proteolipid. Deletions leading to shortened polypeptide domains may have occurred at times other than those implied here. The DM 20 may represent the ancestral precursor of myelin proteolipid that existed before insertion of the third proteolipid segment C3 + T3.

tracellular side of the membrane. Sequence analysis of the DM 20 is needed to confirm or refute this hypothesis.

It is also interesting to note an analogy between the proteolipid segments E1 and E3 and the corresponding regions in myelin basic protein. In the rat there is a small form of basic protein that is identical to the larger form, except for a deletion of 40 residues between residue 115 and 156 (fig.2) [5]. This deletion corresponds very nearly to the gap found in E3, compared with E1.

Brain myelin proteolipid also appears to be homologous with the corresponding protein P_O from peripheral nerve myelin. Sequence data have been reported on 3 fragments of P_O [9]. One of these, a glycopeptide, aligns well with residues 182–210 of the brain proteolipid:

```
                                 190                 200                 210
                                  .                   .                   .
Proteolipid         C Q S I A A P S K T S A S I G T L C A D A R M Y G V L P
                        | |                     | |     |           | |
P0 Glycopeptide     D G S I V I N H L D Y S D N G T F T C D V E E A P V L K
                                      .       |         .
                                     10      CHO       20
```

a region which our model [4] predicts to be on the extracellular face of the membrane bilayer. Furthermore authors in [10] have reported that the tryptic peptide comprising residues 192–204 in rat brain proteolipid contains a fatty acid, which is presumably esterified to a Ser or Thr residue and

is therefore at a site close to the analogous carbohydrate site in P_O. Thus there appears to be both sequence homology and functional analogy. The N-terminal sequence of a 23 kDa form of P_0, can be aligned with proteolipid residues 51–77:

```
          51                  60
          .                   .
Proteolipid  S K N Y Q D Y E Y L I N V I H A F
                 |   |   |       |
23kDa P0     I V V Y T D Q E V S G A V G T L V.
          .                   .
          1                   10
```

A third P_0 peptide segment, MLLYLGIIVL, probably occurs in a *trans*-membrane segment; but there are not yet sufficient data for unambiguous alignment.

In conclusion, we provide evidence that myelin brain proteolipid consists of internally homologous segments, which are also homologous with proteins of different overall function: ATP synthase proteolipid, and a portion of myelin basic protein. The ATP synthase proteolipid, which is one component of a multiprotein complex, is also an intrinsic membrane protein and may have functions in common with the membrane-associated domains of myelin proteolipid. Myelin basic protein is almost identical in sequence to a hypothalamic protein and may be related to other basic proteins, such as histones [11]. Thus myelin proteolipid appears to have been constructed of genetic elements coding for specific polypeptide domains, as has been noted recently for β-crystallin [12] and certain sugar-binding proteins [13]. If this is so, then sequencing of the proteolipid gene may reveal introns, which could provide splicing sites [14], at positions corresponding to interdomain transitions [12]. Our model [4] for myelin proteolipid suggests that there are two types of domain that penetrate the lipid bilayer: *cis* as well as the generally recognized *trans* [15]. These domains may have a common genetic origin in other intrinsic membrane proteins as well.

ACKNOWLEDGEMENTS

This work was supported by NSF grant no. PCM 82-03004 and NIH grant no. NS 13649.

REFERENCES

[1] Lees, M.B., Chao, B., Lin, L.-F.H., Samiullah, M. and Laursen, R.A., submitted.
[2] Jolles, J., Nussbaum, J.L. and Jolles, P. (1983) Biochim. Biophys. Acta 742, 33–38.
[3] Stoffel, W., Hillen, H., Schroder, W. and Deutzmann, R. (1982) Hoppe-Seyler's Z. Physiol. Chem. 363, 1397–1407.
[4] Laursen, R.A., Samiullah, M. and Lees, M.B., submitted.
[5] Carnegie, P.R. and Moore, W.J. (1980) in: Proteins of the Nervous System (Bradshaw, R.A. and Schneider, D.M. eds) 2nd edn, 119–143, Raven Press, New York.
[6] Sebald, W. and Hoppe, J. (1981) Curr. Top. Bioenerg. 12, 1–64.
[7] Agrawal, H.C., Burton, R.M., Fishman, M.A., Mitchell, R.F. and Prensky, A.L. (1972) J. Neurochem. 19, 2082–2089.
[8] Vacher-Lepretre, M., Nicot, C., Alfsen, A., Jolles, J. and Jolles, P. (1976) Biochim. Biophys. Acta 420, 323–331.
[9] Ishaque, A., Roomi, M.W., Szymanska, I., Kowalski, S. and Eylar, E.H. (1980) Can. J. Biochem. 58, 913–921.
[10] Jolles, J., Nussbaum, J.-L., Schoentgen, F., Mandel, P. and Jolles, P. (1977) FEBS Lett. 74, 190–194.
[11] Martenson, R.E., Deibler, G.E. and Kies, M.W. (1971) Nature New Biol. 234, 87–89.
[12] Inana, G., Piatigorsky, J., Norman, B., Slingsby, C. and Blundell, T. (1983) Nature 302, 310–315.
[13] Müller-Hill, B. (1983) Nature 302, 163–164.
[14] Gilbert, W. (1978) Nature 271, 501.
[15] Engelman, D.M., Henderson, R., McLachlan, A.D. and Wallace, B.A. (1980) Proc. Natl. Acad. Sci. USA 77, 2023–2027.