The EMBO Journal vol.4 no.1 pp.113–118, 1985

MRC OX-2 antigen: a lymphoid/neuronal membrane glycoprotein with a structure like a single immunoglobulin light chain

Melanie J. Clark, Jean Gagnon[1], Alan F. Williams and A. Neil Barclay

MRC Cellular Immunology Unit, Sir William Dunn School of Pathology, University of Oxford, Oxford OX1 3RE, and [1]MRC Immunochemistry Unit, Department of Biochemistry, University of Oxford, Oxford OX1 3QU, UK

Communicated by A. F. Williams

The MRC OX-2 antigen is a rat cell surface glycoprotein of mol. wt. 41 000–47 000 found on neurones, thymocytes, B cells, follicular dendritic cells and endothelium. We now report the amino sequence for this antigen as deduced from the nucleotide sequence of cDNA clones detected by use of an oligonucleotide probe. The sequence contains 248 amino acid residues of which 202 residues are likely to be outside the cell with two domains that show homology with immunoglobulins. The N-terminal domain fits best with Ig V domains and Thy-1 antigen while the C-terminal part is like an Ig C domain. Thus the structure overall is similar to an Ig light chain or the T cell receptor β chain. Three glycosylation sites are identified on each of the MRC OX-2 antigen domains.

Key words: MRC OX-2 antigen/thymic antigen/neuronal antigen/membrane glycoprotein/immunoglobulin superfamily

Introduction

The MRC OX-2 mouse monoclonal antibody was raised against rat thymocyte membrane glycoproteins (McMaster and Williams, 1979) and binds to rat thymocytes, neurones, follicular dendritic cells of lymphoid organs, vascular endothelium, some smooth muscle and B lymphocytes (Barclay, 1981; Webb and Barclay, 1984). The thymocytes and brain forms of the antigen have been purified by monoclonal antibody affinity chromatography and are highly glycosylated proteins of apparent mol. wts. 47 000 (thymocyte OX-2) and 41 000 (brain OX-2). The antigenicity and amino acid compositions of these forms were indistinguishable and the size difference is probably due to known differences in the carbohydrate structures (Barclay and Ward, 1982).

The finding of OX-2 antigen on neurones and thymocytes was reminiscent of Thy-1 antigen which is also found on these cell types. Furthermore, both Thy-1 and OX-2 antigens are found on diverse cell types without an apparent functional correlation. The function of neither antigen is known but the structure of Thy-1 antigen is of particular interest because it is homologous to Ig V domains (Williams and Gagnon, 1982). Rat Thy-1 antigen has an apparent mol. wt. of 25 000 on SDS gels and contains a polypeptide sequence of 111 amino acids (Campbell *et al.*, 1981).

On the basis of the tissue distributions and molecular characteristics it seemed possible that OX-2 and Thy-1 antigens may be related in evolution and to evaluate this the OX-2 antigen has been sequenced. Tryptic peptides were prepared from the pure glycoprotein and their sequences used to identify a region suitable for the synthesis of an oligonucleotide probe which was used to screen a cDNA library prepared from rat thymocyte mRNA. From a number of cDNA clones the full sequence of the coding region for the OX-2 glycoprotein has been determined.

Results

Partial amino acid sequence of OX-2 antigen

OX-2 antigen was purified from rat brains and after reduction, alkylation and succinylation, was digested with trypsin. The peptides were fractionated and sequenced as described in Materials and methods. The sequences of five peptides are shown in Table I. They were all unambiguous except at the first position of T2 where small amounts of Lys, Tyr and Trp

Table I. Amino acid sequences of OX-2 tryptic peptides

Peptide	Sequence	Position in cDNA
T1	No sequence obtained. Thus is presumably blocked at N terminus. Composition $(Glx)_3(Val)_2(Arg)_1(Asx)_1(Thr)_1(Gly)_1(Ser)_1$	1–10?
T2	G L L H T T A S L R	11–20
T3	C S L K T T Q E P L I V T W Q K K K A V G P E N M V T Y	21–48
T4	P A P A I S W K G T G S G I E (–) S T E S ...	136–155
T5	V K D P K T Q V G K E V I C Q V L Y	171–188
T6	G E S S Q G M Q R	238–246

A dash (–) indicated no phenylthiohydantoin amino acid derivative was determined at this position. The residue is likely to be glycosylated asparagine as this peptide contained carbohydrate. The dotted line (...) indicates that T4 was not sequenced to the end of the peptide. The region of peptide T3 used to design the oligonucleotide probe is underlined. Peptides T3 and T5 ended at a tyrosine residue. This type of cleavage has been observed before (Campbell *et al.*, 1981).

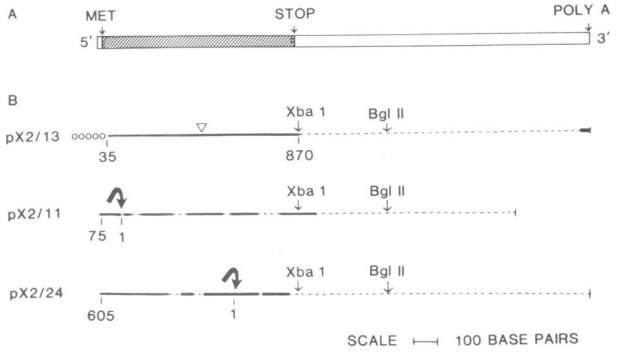

Fig. 1. Composite OX-2 mRNA structure and the three clones used to derive the sequence. **(A)** Structure of OX-2 mRNA. The hatched portion of the bar represents protein coding region, and the open portions represent non-coding regions. MET shows position of the initiator methionine and POLY A the beginning of the poly(A) region. **(B)** The three overlapping cDNA clones used to determine the sequence. The numbers refer to the nucleotide position in the composite sequence (Figure 2). Solid lines indicate those regions of the clones that have been sequenced, dashed lines those not sequenced. Cloning artifacts are marked 00000: double ligated insert, ▽ : position of deletion, ↷: position of inversion.

were identified together with the predominant Gly residue. The N terminus of OX-2 is blocked because intact OX-2 gave no sequence when analysed by automated Edman degradation. Peptide T1 also gave no sequence and is probably the corresponding N-terminal tryptic peptide. Its composition is given in Table I.

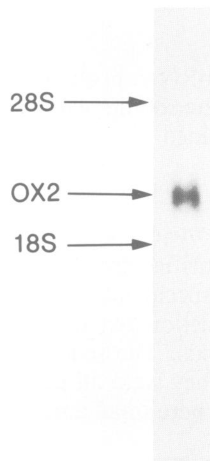

Fig. 2. Northern blot analysis of thymocyte RNA. 20 μg total rat thymocyte RNA was fractionated on an agarose gel in formaldehyde and transferred to Gene Screen membrane. The blot shown here was hybridised with the coding region probe of clone pX2/13 (5′ end of clone to *Xba*I site). This fragment contains ~360 bases of extraneous DNA other than OX-2 coding sequence (see text) but an identical result was obtained when the same filter was stripped and probed with a fragment from the coding region only of clone pX2/11 which does not contain this extra sequence. Size markers are rat ribosomal 18S and 28S RNA.

Isolation and characterization of OX-2 cDNA clones

The amino acid sequence of the C-terminal portion of peptide T3 was used to design a 17 residue long oligonucleotide probe consisting of all 64 possible sequences complementary to the OX-2 mRNA as shown below:

```
                                       A      A
               T      A                G      G
PROBE:  3'  C T    T T    T A C  C A       T G      A T   5'
               C      G                C      C
                                       T      T
```

A rat thymocyte cDNA library was constructed and screened with this probe and 30 positive cDNA clones isolated from a total of 150 000 clones. Twelve clones examined in detail all contained characteristic *Xba*I and *Bgl*II sites (see above, Figure 1). The nucleotide sequence was determined from three cDNA clones as shown in Figure 1 and all of these showed some cloning artifacts. Clone pX2/13 has apparently undergone a double ligation as 360 nucleotides of unidentified cDNA have become joined to the 5′ end of the OX-2 cDNA. This clone also has a 5-bp deletion in the coding region. The other two clones both have an inversion at their 5′ ends, involving the first 75 nucleotides of pX2/11, and the first 605 nucleotides of pX2/24. Similar inversions and deletions have been reported before (Anson *et al.*, 1984; Belt *et al.*, 1984). In view of these artifacts, the sequence of the coding region has been determined on both strands of the DNA and on more than one clone.

The coding sequence (see below) contained 834 nucleotides and this was followed by a long 3′-non-coding region of ~1300 nucleotides as determined by sizing restriction enzyme fragments on agarose gels. After this a poly(A) tail was iden-

```
                        |--------------------- LEADER ---------------------------| |---
                        -30                                                      -1  1
                        M  G  S  P  V  F  R  R  P  F  C  H  L  S  T  Y  S  L  L  W  A  I  A  A  V  A  L  S  T  A  Q  V
GTCTTACTGCGGCCCAGAGCAAGGATGGGCAGTCCGGTATTCAGGAGACCTTTCTGCCATCTGTCCACCTACAGCCTGCTCTGGGCCATAGCAGCAGTAGCGCTGAGCACAGCTCAAGTG
                                                                                                                      120
--------------------------------------------- DOMAIN I ---------------------------------------------
                                                    20                                                    40
 E  V  V  T  Q  D  E  R  K  L  L  H  T  T  A  S  L  R (C) S  L  K  T  T  Q  E  P  L  I  V  T  W  Q  K  K  K  A  V  G  P
GAAGTGGTGACCCAGGATGAAAGAAAGCTGCTGCACACAACTGCATCCTTACGCTGTTCTCTAAAAACAACCCAGGAACCCTTGATTGTGACATGGCAGAAAAAGAAAGCCGTAGGCCCA
                                                                                                                      240
----------------------------------------------------------------------------------------------------
                                                    60                                                    80
 E  N  M  V  T  Y  S  K  A  H  G  V  V  I  Q  P  T  Y  K  D  R  I [N] I  T  E  L  G  L  L [N] T  S  I  T  F  W [N] T  T
GAAAACATGGTCACTTACAGCAAAGCCCATGGGGTTGTCATTCAGCCCACCTACAAAGACAGGATAAACATCACTGAGCTGGGACTCTTGAACACAAGCATCACCTTCTGGAACACAACC
                                                                                                                      360
---------------------------------------------------------------------------------------| |----------
                                                   100                                                   120
 L  D  D  E  G  C  Y  M (C) L  F  N  M  F  G  S  G  K  V  S  G  T  A  C  L  T  L  Y  V  Q  P  I  V  H  L  H  Y  N  Y  F
CTGGATGATGAGGGTTGCTACATGTGTCTCTTCAACATGTTTGGATCTGGGAAGGTCTCTGGGACAGCTTGCCTTACTCTCTATGTACAGCCCATAGTACACCTTCACTACAACTATTTT
                                                                                                                      480
--------------------------------------------- DOMAIN II --------------------------------------------
                                                   140                                                   160
 E  D  H  L [N] I  T (C) S  A  T  A  R  P  A  P  A  I  S  W  K  G  T  G  S  G  I  E [N] S  T  E  S  H  S  H  S [N] G  T
GAAGACCACCTAAACATCACGTGCTCTGCAACTGCCCGCCCAGCCCCTGCCATCTCCTGGAAGGGCACTGGGTCAGGAATTGAGAATAGTACTGAGAGTCACTCCCATTCAAATGGGACT
                                                                                                                      600
---------------------------------------------------------------------------------------------------|
                                                   180                                                   200
 T  S  V  T  S  I  L  R  V  K  D  P  K  T  Q  V  G  K  E  V  I (C) Q  V  L  Y  L  G  N  V  I  D  Y  K  Q  S  L  D  K  G
ACATCTGTCACCAGCATCCTCCGGGTCAAAGACCCCAAAACTCAGGTTGGAAAGGAAGTGATCTGCCAGGTTTTATACTTGGGGAATGTGATTGACTACAAGCAGAGTCTGGACAAAGGA
                                                                                                                      720
|-------------------- TRANSMEMBRANE -------------------------| |------ CYTOPLASMIC ----------------
                                                   220                                                   240
 F  W  F  S  V  P  L  L  L  S  I  V  S  L  V  I  L  L  V  L  I  S  I  L  L  Y  W  K  R  H  R  N  Q  E  R  G  E  S  S  Q
TTTTGGTTTTCAGTCCCACTGCTGCTGAGCATTGTTTCTCTGGTAATTCTTCTGGTCTTGATCTCCATCTTATTATACTGGAAACGGCACCGAAATCAGGAGCGGGGTGAGTCATCACAG
                                                                                                                      840
--------------|     XBA 1
             248      |
 G  M  Q  R  M  K  *
GGGATGCAAAGAATGAAATAAGAGCTCTAGA
```

Fig. 3. Composite nucleotide sequence of the coding region of a thymocyte cDNA clone with its corresponding amino acid sequence. ○ = Cys residues believed to participate in the conserved disulphide bond within each domain. □ = Asn residues in potential glycosylation sites.

tified by sequencing clone pX2/13 from the 3′ end. Thus the coding and non-coding regions total ~2200 nucleotides and this is consistent with a size of ~2500 nucleotides determined for the mRNA by Northern blot analysis (Figure 2).

Sequence of the coding region

The sequence shown in Figure 3 is a composite from the clones in Figure 1 and is strongly supported by the peptide data. One large open reading frame extends from a potential initiator methionine codon at nucleotide 25 contained within the consensus sequence A/G NNATG A/G found around most eukaryotic initiator methionine codons (Kozak, 1983). The first sequence established by peptide data is at amino residue 11 (peptide T2 Table I) and the other peptides occur throughout the sequence until residue 246 out of 248 where peptide T6 ends (Table I). The end of the protein is defined by a stop codon after residue 248. In the peptide data there is only one discrepancy, namely the first residue of peptide T2 which is a Gly in the peptide sequence compared with a predicted Lys in the cDNA sequence. Small amounts of Lys were found with the Gly in peptide T2 (see above) but the data were most consistent with the Gly designation shown in Table I. The other nine residues of T2 fitted with the nucleotide sequence and the inconsistency could thus be due to a polymorphism in the antigen; a difference between OX-2 of brain (peptides) and thymus (cDNA) or an artifact in the protein sequencing. Further confirmation of the coding region is shown by the good agreement in the amino composition determined for the glycoprotein compared with that predicted from the coding sequence minus leader sequence (Table II).

In Figure 3 a leader sequence extending to 30 amino acids is suggested with the protein sequence thus beginning at a Gln residue. If this Gln existed as a pyroglutamic acid residue then the blocked N terminus of OX-2 glycoprotein would be accounted for. Also the composition of peptide T1 which is believed to be the N-terminal peptide (Table I) fits reasonably well with a tryptic peptide predicted by residues 1 – 10. There is a discrepancy in the presence of Gly and Ser in T1 composition but this could be due to small amounts of contamination with peptide T6 and also the T1 composition cannot be accounted for elsewhere in the OX-2 sequence. The putative leader sequence is similar to leader sequences of other glycoproteins with the only unusual point that it is longer than the average (von Heijne, 1983; Watson, 1984).

Table II. Comparison of experimentally determined amino acid composition of thymocyte OX-2 with that deduced from the nucleotide sequence of the thymocyte cDNA

Residue	Determination	Prediction
Asx	21.3	19
Glx	26.0	24
His	8.0	8
Lys	15.0	16
Arg	10.0	9
Thr	22.8	23
Ser	20.8	22
Pro	8.7	8
Ala	10.9	8
Cys	5.2	6
Gly	18.4	16
Tyr	8.4	9
Val	18.6	21
Ile	13.1	16
Leu	25.0	27
Phe	6.7	6
Met	3.4	5
Trp	5.2	5

The figures for the experimentally determined composition are taken from Barclay and Ward (1982) and are expressed as residues per 248 amino acids.

The mature protein as shown in Figure 3 contains 248 amino acids and amongst these only one segment is seen that seems likely to span the lipid bilayer. This is found at residues 203 – 229 and contains predominantly hydrophobic amino acids with no charged residues or amides. The latter residues are virtually never seen in transmembrane segments (Komaromy *et al.*, 1983; Kaufman *et al.*, 1984). The moderately hydrophobic amino acids Thr and Ser are commonly found in membrane-spanning segments and the sequence 203 – 229 contains four Ser residues. To the C-terminal side of residue 229 is found a stretch of basic residues and this is a typical feature of sequences on the cytoplasmic side of a membrane spanning piece. Thus it seems likely that residues 230 – 248 are inside the cell and that the sequence 1 – 202 is outside.

Carbohydrate structures have only been found on the extracellular parts of membrane glycoproteins and OX-2 sequences suitable for N-glycosylation are found at residues 65, 73, 80, 127, 151 and 160. At position 151 the peptide data (Table I) shows that glycosylation has occurred since no amino residue was detected in the sequencer at this position in peptide T4 (the expected result for Asn residues that have carbohydrate attached).

Homology of OX-2 glycoprotein with immunoglobulins, Thy-1 glycoprotein and the poly Ig receptor

Inspection of the sequence for OX-2 glycoprotein showed obvious homologies with the immunoglobulin superfamily with the N-terminal and C-terminal parts looking like Ig V and C domains respectively. This is shown in Figure 4a for V domain homologies and Figure 4b for the Ig C domains. In both cases the segments that form β-strands in the Ig domains are identified with letters and can be seen in the folding pattern for Ig domains shown in Figure 7a. In both domains I and II of the OX-2 sequence are found Cys residues that could form the conserved disulphide bonds between β-strands B and F and the tryptophan residue that is highly characteristic of β-strand C in Ig domains is also present in both OX-2 domains. Most Ig sequences are also characterised by a pattern of alternating hydrophobic residues in β-strands E and the sequence Ser Ile Thr Phe in OX-2 domain I fits well with this as does the Ile Leu Arg Val of domain II.

With regard to the V-like domains, identities are particularly seen around the Cys of β-strand F (residue 91) and this is a region that shows strong homologies in Ig V domains, Thy-1 antigen and the poly Ig receptor. At positions 6 and 4 residues before the Cys in β-strand F almost all Ig V domains have Asp and Ala or Gly, respectively, and OX-2 fits with this. The Asp residue may be conserved because it forms a salt bridge with a conserved Arg (more rarely Lys) residue near the base of β-strand D (position 63) (R. Poljak, personal communication; Williams *et al.*, 1984) and in this position an Arg is seen in the OX-2 domain I sequence.

The numbers of identities and conservative substitutions

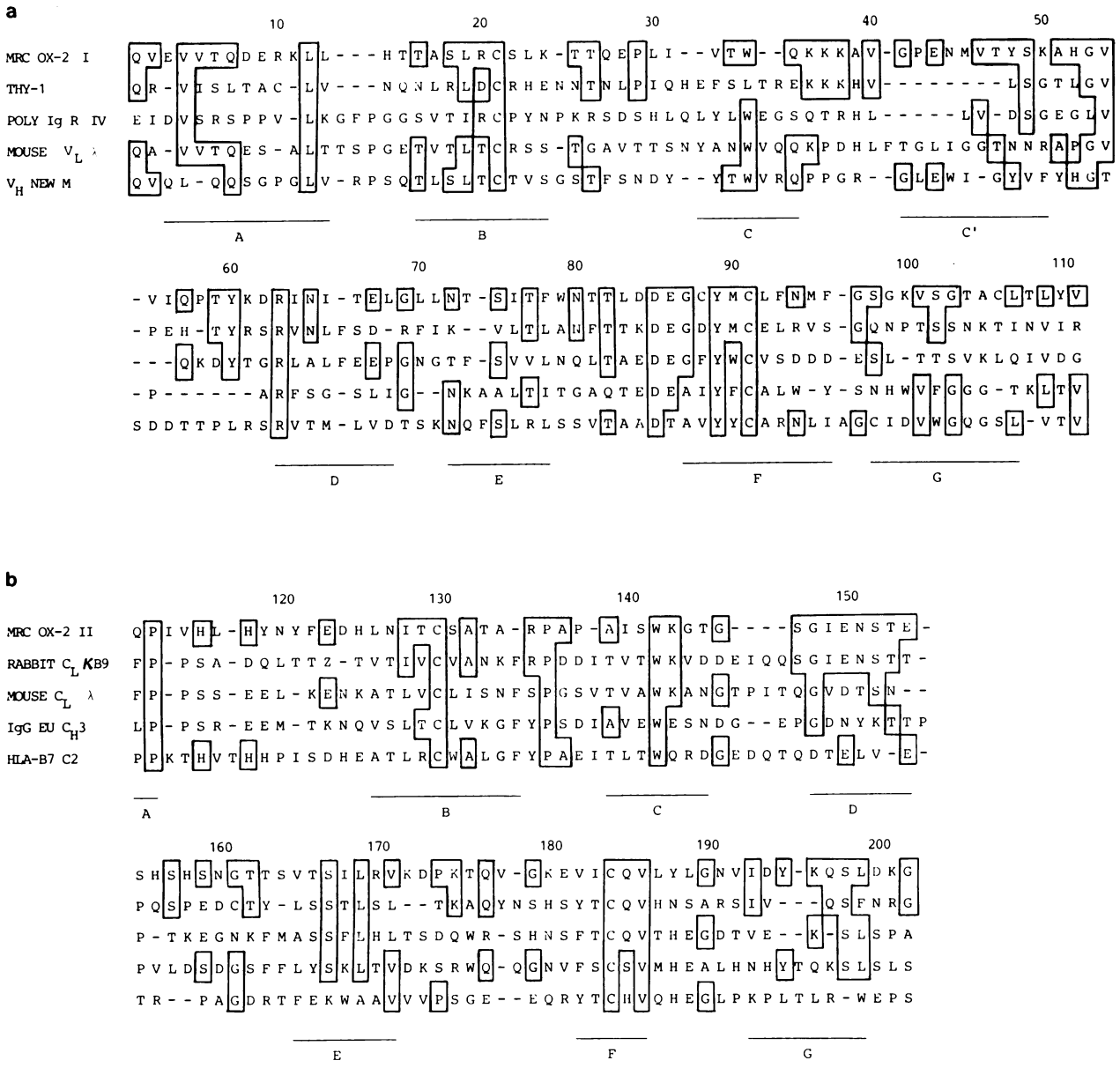

Fig. 4. Alignment of OX-2 sequence with immunoglobulin domains. **(A)** Alignment of OX-2 domain I sequence with rat Thy-1 (Campbell *et al.*, 1981), poly Ig receptor domain IV (Mostov *et al.*, 1984) mouse MOPC 104E V_λ, V_H NEW M (Kabat *et al.*, 1983). **(B)** Alignment of OX-2 domain II sequence with rabbit κ B9, mouse MOPC 315$_\lambda$, IgG EU C_H3 and HLA-B7 C2 (Kabat *et al.*, 1983). In the comparisons with V-type domains **(A)** or C-type domains **(B)**, residues identical with OX-2 are boxed and dashes show gaps in the sequence inserted to maximize the alignment. The bars with letters A, B, C, C′, D, E, F and G below the sequences indicate residues involved in forming β strands as shown in Figure 7a.

between all the sequences shown in Figure 4a are given in Figure 5 where it can be seen that the scores for OX-2 with the other sequences are at least as convincing as are the scores within the Thy-1, poly Ig R and V domain sequences.

The OX-2 domain II sequence clearly aligns well along the length of the Ig light chain C domain and heavy chain C_H3 domain sequences shown in Figure 4b. Fewer identities are seen with the Ig-like domain of HLA-B7. In Figure 5 the scores amongst all the sequences shown in Figure 4b are given and it can be seen that OX-2 domain II fits well with these Ig C domain sequences as does the HLA-B7 sequence.

One other homology of interest is that between mouse Thy-1 residues 2–10 and OX-2 residues, 100–107, namely:

rat OX-2	K V	–	S	G	T A C L
mouse Thy-1	K V	I	S	L	T A C L

The full Thy-1 sequence fits best with the OX-2 domain I sequence but the above identities are between the beginning of the Thy-1 sequence and the junction of the OX-2 domain I and II sequences. In the book of short sequences (Dayhoff *et al.*, 1978b) there are 3708 sequences containing Cys residues but only one TACL sequence (in bacteriophage φX174 gene D protein). Thus the above patches of identity may not have occurred by chance but might indicate a gene duplication event of a Thy-1 like sequence occurring long ago.

Discussion

The sequence data on the OX-2 glycoprotein shows that this molecule is a member of the Ig superfamily along with the other structures shown in Figure 6. The OX-2 molecule looks like a T receptor β chain or an Ig light chain with a membrane

C-DOMAINS - ALIGNMENTS WITH 91 RESIDUES OF MRC OX-2 DOMAIN II

	C_K	C_λ	C_H3	HLA B7	
		28,46	18,43	20,47	17,38	MRC OX-2
THY-1	29,54		32,60	27,56	20,47	C_K
POLY IgR (IV)	21,48	22,45		30,49	21,45	C_λ
V_λ	29,49	24,44	21,44		22,47	C_H3
V_H	30,48	23,47	22,50	32,62		
	MRC OX-2	THY-1	POLY IgR	V_λ		

V-DOMAINS - ALIGNMENTS WITH 111 RESIDUES OF MRC OX-2 DOMAIN I

Fig. 5. Identities and conservative substitutions between MRC OX-2 sequence and Ig V and C domain sequences. The two numbers in each square give the number of amino acid identities and identities plus conservative substitutions respectively between the sequences indicated in Figure 4. Conservative substitutions are those that occur more frequently than random chance would predict and are those with a score of 1 or more in Figure 84 in Dayhoff *et al.* (1978a). For comparisons other than those involving the OX-2 sequence the alignments were altered from those in Figure 4 if adjustments, without excessive insertion of gaps, improved the scores. Thus in all cases the scores are for optimal adjustments as far as could be judged.

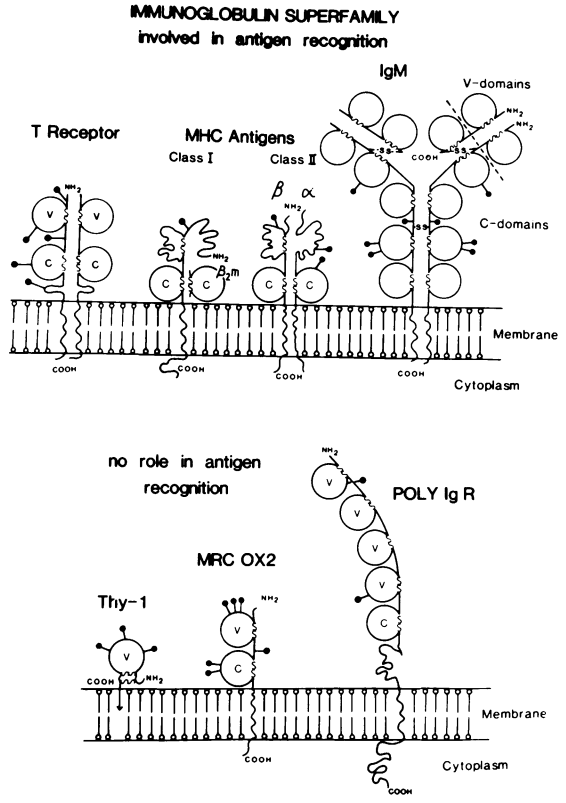

Fig. 6. Shows models for molecules in the Ig superfamily. Circles marked V are like an Ig V domain while those marked C are more like Ig C domains. Intrachain disulphide bonds are shown by symbols (S) interchain bonds by (s-s) and N-linked carbohydrate structures by ( ). The figure is modified from Williams *et al.* (1984). Data are from: mouse T cell receptor (Hedrick *et al.*, 1984; Saito *et al.*, 1984). Class I and Class II MHC (Ploegh *et al.*, 1981; Kaufman *et al.*, 1984). IgM (Kehry *et al.*, 1980). Thy-1 (Williams and Gagnon, 1982). Poly Ig R (Mostov *et al.*, 1984).

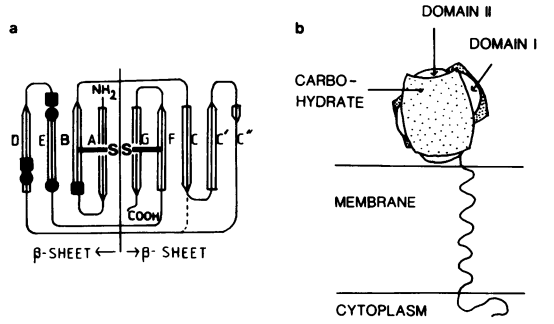

Fig. 7. (A) β strand folding pattern of Ig domains, modified from Amzel and Poljak (1979). The pattern for a C domain is shown by a dashed line directly connecting strands C and D. V domains have an extra loop of sequence in the middle of the domain, indicated here as strand C'. Positions of asparagine-linked carbohydrate are indicated as circles (●) for OX-2 domain I and squares (■) for OX-2 domain II. **(B)** A model for OX-2 at the cell surface. The stippled areas show how carbohydrate might cover most of the exterior surfaces of the molecule. The layer of carbohydrate would be thicker than drawn here.

integration segment. In Figure 6 we divide the Ig-related molecules into two categories based on whether or not they are involved in antigen recognition. The T receptors and immunoglobulins directly mediate this function and the MHC antigens are somehow co-recognised with foreign antigen by the T cell receptor. The poly Ig receptor clearly has no role in antigen recognition and the functions of Thy-1 and OX-2 glycoproteins are unknown. However it is very difficult to imagine that these latter molecules are in any way directly involved in the specificity of antigen recognition. A more likely possibility is that they are involved in mediating cell-cell interactions on various cell types (Cohen *et al.*, 1981; Williams, 1982; Williams *et al.*, 1984).

The Thy-1 and OX-2 glycoproteins are of considerable interest with regard to the evolution of the Ig superfamily since these are the two simplest structures shown in Figure 6. All of the structures shown are ultimately believed to have come from a primordial single domain structure and Thy-1 is the only known contemporary molecule that exists in this way (β_2M is a single domain structure but is found associated with the large chain of Class I MHC antigens). Furthermore all of the molecules involved in antigen recognition were probably derived from a primordial precursor like an Ig L-chain and OX-2 glycoprotein is a contemporary molecule with these properties. It is interesting that both Thy-1 and OX-2 are found on neurones and in the case of Thy-1 this is the tissue on which this molecule is conserved between various species. It is thus possible that the Ig-related molecules first evolved to function in cell-cell interactions on sensory cells and this idea is discussed in detail elsewhere (Williams, 1982; Williams *et al.*, 1984).

In Figure 6 we show the OX-2 glycoprotein as an extended two domain structure at a cell surface but in reality interactions may occur between the domains. A relevant point here is the position of the N-linked carbohydrate structures in relation to the predicted Ig-like folding pattern for the OX-2 domains. This is shown in Figure 7a where it can be seen that in both domains the three glycosylation sites are in very similar positions. Furthermore they are found on the same face of the Ig-fold in both domains I and II. This can be contrasted with Thy-1 glycoprotein which has three N-linked glycosylation sites predicted to be on β-strands B, E and G and thus on the faces of both β-sheets of the domain. The proposed

restriction of carbohydrates to one side of each OX-2 domain suggests that the domains may interact *via* the non-glycosylated β-sheets. The faces that would be exposed might then in large part be covered by carbohydrate. A possible model for OX-2 glycoprotein taking these ideas into account is shown in Figure 7b.

Materials and methods

Purification of OX-2 antigen and sequencing of tryptic peptides

5 mg OX-2 antigen was purified from 750 g wet weight of frozen Sprague-Dawley rat brains by solubilisation with sodium deoxycholate, affinity chromatography with MRC OX-2 antibody and gel filtration as described previously (Barclay and Ward, 1982). After removal of sodium deoxycholate by dialysis against 0.1 M NH_4HCO_3, OX-2 antigen (2 mg protein; ~70 nmol) was reduced, alkylated, succinylated and digested with trypsin as described by Campbell *et al.* (1981). Peptides were purified by gel filtration and h.p.l.c. and sequenced on a Beckman spinning cup sequencer as in Campbell *et al.* (1981) and Williams and Gagnon (1982).

Screening the rat thymocyte cDNA library

The oligonucleotide mixture (see Results) was synthesised using the solid phase phosphotriester method developed by Gait *et al.* (1980) and purified by polyacrylamide gel electrophoresis. This probe was radiolabelled by phosphorylating the 5′ ends using T4 polynucleotide kinase (Maniatis *et al.*, 1982) and [γ-^{32}P]ATP and used to screen a rat thymocyte cDNA library. This library which will be described in details elsewhere (M. Thomas and A. N. Barclay, in preparation) was prepared from cDNA made by the loopback procedure and fractionated on sucrose gradients to give a >1 kb fraction which was blunt end-ligated into the pAT 153/*Pvu*II/8 vector (Anson *et al.*, 1984). A complexity of ~160 000 clones was achieved.

The cDNA library was transferred to Whatman 541 filter papers and chloramphenicol amplified as described by Gergen *et al.* (1979). The filters were screened with the oligonucleotide probe essentially as described by Wallace *et al.* (1981) employing a hybridisation temperature of 38°C and a washing temperature of 42°C.

DNA sequencing

Large mol. wt. fragments spanning the coding region of the cDNA clones were prepared by restriction enzyme digestion and purification by agarose gel electrophoresis and electro-elution (Girvitz *et al.*, 1980). The majority of these fragments were further digested with restriction enzymes and labelled at their 3′ termini by the Klenow fragment of *Escherichia coli* DNA polymerase I and the appropriate ^{32}P-labelled deoxynucleotide triphosphate and then fractionated on polyacrylamide gels or strand separated (Maniatis *et al.*, 1982). The sequence was determined by the chemical degradation method of Maxam and Gilbert (1980). A small proportion of the sequence was established by dideoxy sequencing of cDNA fragments subcloned into M13mp8 (Sanger *et al.*, 1977; Messing, 1983).

Northern blot analysis

Total thymocyte RNA was fractionated on a 1.5% agarose gel containing formaldehyde (Lehrach *et al.*, 1977). Transfer to Gene Screen membrane, hybridisation and washing followed the procedure recommended by the manufacturer (New England Nuclear, Boston, USA). Probes from the pX2/11 and pX2/13 cDNA clones were prepared by purification of the fragments extending from the 5′ end of the cDNA to the *Xba*I site (Figure 1). These were labelled with ^{32}P using a nick translation kit from Amersham International.

Acknowledgements

We are indebted to members of Professor G. G. Brownlee's group and to Dr Duncan Campbell for helping us establish the recombinant DNA techniques, to Keith Gould for assistance with oligonucleotide synthesis and Dr Matt Thomas for invaluable discussions. Expert technical assistance was provided by Ed Davey, Tony Gascoyne and Tony Willis, and Mrs Caroline Griffin typed the manuscript.

References

Amzel,L.M. and Poljak,R.J. (1979) *Annu. Rev. Biochem.*, **48**, 961-997.
Anson,D.S., Choo,K.H., Rees,D.J.G., Gianelli,F., Gould,K., Huddleston, J.A. and Brownlee,G.G. (1984) *EMBO J.*, **3**, 1053-1060.
Barclay,A.N. (1981) *Immunology*, **44**, 727-736.
Barclay,A.N. and Ward,H.A. (1982) *Eur. J. Biochem.*, **129**, 447-458.
Belt,K.T., Carroll,M.C. and Porter,R.R. (1984) *Cell*, **36**, 907-914.
Campbell,D.G., Gagnon,J., Reid,K.B.M. and Williams,A.F. (1981) *Biochem. J.*, **195**, 15-30.
Cohen,F.E., Novotny,J., Sternberg,M.J.E., Campbell,D.G. and Williams, A.F. (1981) *Biochem. J.*, **195**, 31-40.
Dayhoff,M.O., Schwartz,R.M. and Orcutt,B.C. (1978a) in Dayhoff,M.O. (ed.), *Atlas of Protein Sequence and Structure*, Vol. **5**, National Biomedical Research Foundation, Baltimore, pp. 345-352.
Dayhoff,M.O., Hunt,L.T., Barker,W.C., Schwartz,R.M. and Orcutt,B.C., eds. (1978b) *Protein Segment Dictionary*, published by The National Biomedical Research Foundation, MD 20901, USA.
Gait,M.J., Singh,M., Sheppard,R.C., Edge,M.D., Greene,A.R., Heathcliffe,G.R., Atkinson,T.C., Newton,C.R. and Markham,A.F. (1980) *Nucleic Acids Res.*, **8**, 1081-1096.
Gergen,J.P., Stern,R.H. and Wensink,P.C. (1979) *Nucleic Acids Res.*, **7**, 2115-2136.
Girvitz,S.C., Bacchetti,S., Rainbow,A.J. and Graham,F.L. (1980) *Anal. Biochem.*, **106**, 492-496.
Hedrick,S.M., Nielsen,E.A., Kavaler,J., Cohen,D.I. and Davis,M.M. (1984) *Nature*, **308**, 153-158.
Kabat,E.A., Wu,T.T., Bilofsky,H., Reid-Miller,M. and Perry,H., eds. (1983) *Sequences of Proteins of Immunological Interest*, published by U.S. Department of Health, Bethesda, USA.
Kaufman,J.F., Auffray,C., Korman,A.J., Shackelford,D.A. and Strominger,J. (1984) *Cell*, **36**, 1-13.
Kehry,M., Ewald,S., Douglas,R., Sibley,C., Raschke,W., Fambrough,D. and Hood,L. (1980) *Cell*, **21**, 393-406.
Komaromy,M., Clayton,L., Rogers,J., Robertson,S., Kettman,J. and Wall, R. (1983) *Nucleic Acids Res.*, **11**, 6775-6785.
Kozak,M. (1983) *Microbiol. Rev.*, **47**, 1-45.
Lehrach,H., Diamond,D., Wozney,J.M. and Boedtker,H. (1977) *Biochemistry (Wash.)*, **16**, 4743-4751.
Maniatis,T., Fritsch,E.F. and Sambrook,J., eds. (1982) *Molecular Cloning, A Laboratory Manual*, published by Cold Spring Harbor Laboratory Press, NY.
Maxam,A.M. and Gilbert,W. (1980) *Methods Enzymol.*, **65**, 499-560.
McMaster,W.R. and Williams,A.F. (1979) *Eur. J. Immunol.*, **9**, 426-433.
Messing,J. (1983) *Methods Enzymol.*, **101**, 20-78.
Mostov,K.E., Friedlander,M. and Blobel,G. (1984) *Nature*, **308**, 37-43.
Ploegh,H.L., Orr,H.T. and Strominger,J.L. (1981) *Cell*, **24**, 287-299.
Saito,H., Kranz,D.M., Takagaki,Y., Hayday,A.C., Eisen,H.N. and Tonegawa,S. (1984) *Nature*, **309**, 757-762.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463-5467.
von Heijne,G. (1983) *Eur. J. Biochem.*, **133**, 17-21.
Wallace,R.B., Johnson,M.J., Hirose,T., Miyake,T., Kawashima,E.H. and Itakura,K. (1981) *Nucleic Acids Res.*, **9**, 879-894.
Watson,M.E.E. (1984) *Nucleic Acids Res.*, **12**, 5145-5164.
Webb,M. and Barclay,A.N. (1984) *J. Neurochem.*, **43**, 1061-1067.
Williams,A.F. (1982) *J. Theor. Biol.*, **98**, 221-234.
Williams,A.F. and Gagnon,J. (1982) *Science (Wash.)*, **216**, 696-703.
Williams,A.F., Barclay,A.N., Clark,M.J. and Gagnon,J. (1984) in Anderson,L.C., Gahmberg,C.G. and Ekblom,P. (eds.), *Proceedings of the Sigrid Juselius Symposium; Gene Expression during Normal and Malignant Differentiation*, Academic Press, in press.

Received on 22 October 1984