Neuronal Cell Thy-1 Glycoprotein: Homology with Immunoglobulin

Alan F. Williams and Jean Gagnon

Summary. The amino acid sequences of mouse brain Thy-1 glycoproteins are shown to be homologous to those of variable-region immunoglobulin domains. There is also good homology with constant domains and β_2-microglobulin; overall the results suggest that Thy-1 may be like the primordial immunoglobulin domain. Preliminary evidence for an invertebrate Thy-1 homolog supports this possibility.

Structural studies on proteins have suggested that many molecules with related functions belong to families that may have evolved from primitive genes by duplication and divergence. These families include the serine proteases, heme binding globins, cytochromes, and immunoglobulins (*1*). The sequences of genes of eukaryotes have further revealed that regions coding for amino acids (exons) are usually interrupted by noncoding regions (introns). It has been suggested that the exons may encode segments of protein sequence which correspond to particular functions and that rearrangement of exons may contribute to the rapid diversification in evolution of proteins with related functions (*2, 3*).

The heavy (H) and light (L) chains of antibodies are made up of domains of about 110 amino acids which show sequence and structural homology (*4–7*). This is illustrated for immunoglobulin M (IgM) in Fig. 1, which shows five domains in the heavy chain and two in the light chain. The homology includes variable (V) and constant (C) region domains, and it is widely accepted that the immunoglobulin (Ig) family was derived by gene duplication from a primordial molecule homologous to a single domain (*4*). This idea is supported by the sequences of Ig genes which show that each C domain corresponds to a single exon (*8*). For V domains the coding sequence from the amino terminus to the vicinity of the third hypervariable region appears to be in one exon with a small piece (V_L) or pieces (V_H) being spliced onto this to complete the carboxyl terminal region of the V domain during differentiation of B lymphocytes (*9*).

β_2-Microglobulin (β_2-m) was the first molecule, other than antibodies which showed homology with Ig, and it can be aligned with a single C domain (*10, 11*). Subsequently β_2-m was found to be the smaller of two noncovalently associated polypeptides which constitute the class 1 major histocompatibility antigens (HLA-A, -B, and -C antigens). The large polypeptide of the HLA-A, -B, -C antigens contains two disulfide-bonded loops of sequence and the COOH-terminal one of these is homologous with C domains of Ig (*12, 13*). The only other molecule known to be homologous to Ig is the Thy-1 antigen which consists of a single domain (*14, 15*). Studies on rat Thy-1 strongly suggested that it belongs to the Ig superfamily, but a particular relationship to V or C domains was not established (*15*). Immunoglobulins, HLA-A, -B, -C, and Thy-1 antigens are all found at cell surfaces, and their homology regions are shown in Fig. 1. Hitherto none of these molecules has been found in species more primitive than vertebrates.

In this article we report the sequences of the two allotypes of mouse Thy-1 glycoprotein and show that these sequences and those of rat Thy-1 fit best with Ig V domains. However, sequence homologies with C domains and β_2-m are also striking, and this suggests that Thy-1 may be like the primordial Ig domain. A putative Thy-1 homolog has been purified from squid brain, and the sequence of one tryptic peptide suggests that this may be the first invertebrate member of the Ig superfamily.

Dr. Williams is a member of the Medical Research Council Cellular Immunology Unit, Sir William Dunn School of Pathology, University of Oxford, Oxford, England OX1 3RE. Dr. Gagnon is a member of the Medical Research Council Immunochemistry Unit, Department of Biochemistry, University of Oxford, Oxford OX1 3QU.

Thy-1 Antigen

The Thy-1 antigen is a glycoprotein found at the cell surface of rodent thymocytes, neuronal cells, and some other cell types (*16*). There are 10^6 molecules per cell on rat thymocytes (*17*) and an equivalent amount on neuronal cells (*18*), and thus Thy-1 is probably the most abundant surface glycoprotein of both cell types. Thy-1 was first identified as the theta (θ) alloantigen of mouse thymocytes (*19*), which occurs in two allotypic forms now referred to as Thy-1.1 and Thy-1.2 (*20*). Rat Thy-1 also displays the Thy-1.1 determinant (*17, 21*), and other xenoantigenic determinants can be identified on rat and mouse Thy-1 with rabbit antibodies (*22*).

All the Thy-1 antigenic determinants are associated with rat brain and thymus glycoproteins that have molecular weights of 17,800 and 18,700, respectively (*23–25*). These glycoproteins have the properties of molecules that can find their way into the lipid bilayer, since they bind one micelle of detergent, and aggregate to form a regular complex if detergent is removed (*25*). Also Thy-1 is heavily labeled in thymocyte membranes by an affinity label that is subject to photoactivation and that partitions into the lipid bilayer (*26*). The Thy-1 glycoproteins have never been found associated with other membrane proteins, and the anchorage of Thy-1 to the membrane is thus quite different from that of β_2-m, which is not hydrophobic and binds to the large polypeptide of the HLA-A, -B, and -C antigens (Fig. 1). The protein moiety of the Thy-1 glycoproteins has a molecular weight of 12,500 (*25*) and is probably identical in the brain and thymus forms with differences in the carbohydrate accounting for the different molecular weights (*27*).

Glycoproteins that are structurally related to rat Thy-1 have been purified from brain tissue of mouse (*28, 29*), man (*30, 31*), dog (*32*), and chicken (*33*); and the expression of Thy-1 also seems to be conserved in evolution on fibroblasts (*30, 34*). In contrast, the expression of Thy-1 on lymphoid cells varies in different species (*16*). For example, whereas Thy-1 is the most abundant surface molecule on rodent thymocytes the human homolog is absent from human thymocytes (and the name Thy-1 is thus inap-

0036-8075/82/0514-0696$01.00/0

propriate) (*35*). These tissue distribution patterns are puzzling, but similarly patchy tissue distributions are being found for other cell surface antigens. In fact, surface antigens found on a few apparently unrelated cell types but not on all cells are much more common than molecules that are highly specific for one cell type.

The sequence of rat Thy-1 glycoprotein (*16*) consists of 111 amino acids with two disulfide bonds from Cys^{9} to Cys^{111} and Cys^{19} to Cys^{85} and three *N*-linked carbohydrate structures at residues 23, 74, and 98 (*36*). The molecule seems to have a nonprotein hydrophobic segment attached to the COOH-terminal Cys residue since small COOH-terminal peptides, which should have conventional properties, behave very anomalously, and can only be isolated by virtue of their binding to nonionic detergents (*16*). The postulated nonprotein segment would explain both the hydrophobic properties of the Thy-1 glycoprotein (*25, 26*) and its means of attachment to the lipid bilayer. Neither of these features can be explained by the amino acid sequence (*16*). A model summarizing these points and the new data presented below are shown in Fig. 2a.

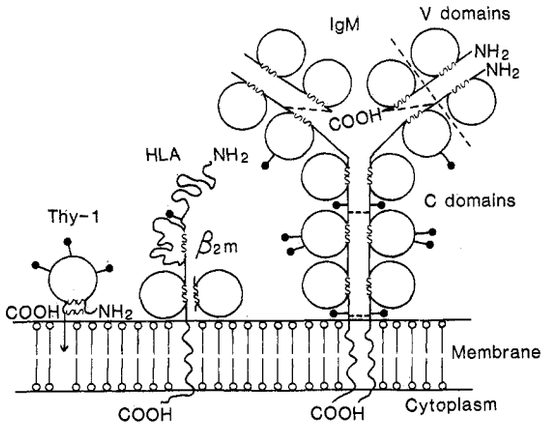

Fig. 1. Molecules in the Ig superfamily. IgM, HLA-A, -B, and -C and Thy-1 glycoproteins are shown schematically at a cell surface. Ig domains and their homologous regions in Thy-1 and HLA-A, -B, and -C are represented by circles. Intrachain disulfide bonds are shown by the S-S symbols and interchain bonds in IgM by dashed lines. *N*-Linked carbohydrate structures are shown by the club (🏌), and NH_2 and COOH identify the NH_2-terminus and COOH-terminus of the polypeptides, with the exception of β_2-m. It is believed that HLA-A, -B, and -C and IgM are integrated into the membrane by a hydrophobic protein sequence; but for Thy-1 the evidence suggests the existence of a nonprotein structure for membrane integration. [The diagrams are based on (*12–13, 15, 16, 54, 55*)]

Mouse Thy-1 Sequences

Thy-1 glycoproteins were purified from brains of mice of Thy-1.1 (AKR and A/Thy-1.1 strains) and Thy-1.2 (CBA strain) allotype (Fig. 3). The amino acid compositions of both glycoproteins were very similar, but the Thy-1.1 glycoprotein had one more Arg and one less Glx residue than Thy-1.2 glycoprotein (Fig. 3) in agreement with the results of Cotmore *et al.* (*30*). Different tryptic peptides were obtained from Thy-1.1 and Thy-1.2 glycoproteins (Fig. 4) and in the sequences (Fig. 5) the only difference was an Arg (Thy-1.1)/Gln (Thy-1.2) interchange at residue 89. This fits well with the amino acid composition (Fig. 3) and also with the finding that Thy-1.1 is more basic than Thy-1.2 in isoelectrofocusing (*37*).

The sequence for mouse Thy-1 is very similar to that of rat Thy-1, apart from two patches of sequence (see below), and the molecules of both species have all the features shown in Fig. 2a.

Thy-1 Alloantigenic Determinants and Immune Response to Thy-1

The finding of a sequence difference between Thy-1.1 and Thy-1.2 glycoproteins suggests a protein basis for the mouse alloantigenic determinants. The rat sequence is consistent with this since rat Thy-1 has the Thy-1.1 determinant and also has an Arg residue at the equivalent position to Arg^{89} of mouse Thy-1.1. Previous experiments suggested that the Thy-1.1 determinant was protein-based since the antigenicity was destroyed by proteolysis and protein denaturation (*27, 38*).

The single-sequence difference between mouse Thy-1.1 and Thy-1.2 also suggests that there is only one antigenic determinant recognized in immunization between mouse strains. Thus the Thy-1 antigen alone would not be expected to stimulate immune responses that require collaboration between B and T lymphocytes recognizing different determinants. This is consistent with the analysis of the alloimmune response to Thy-1, which

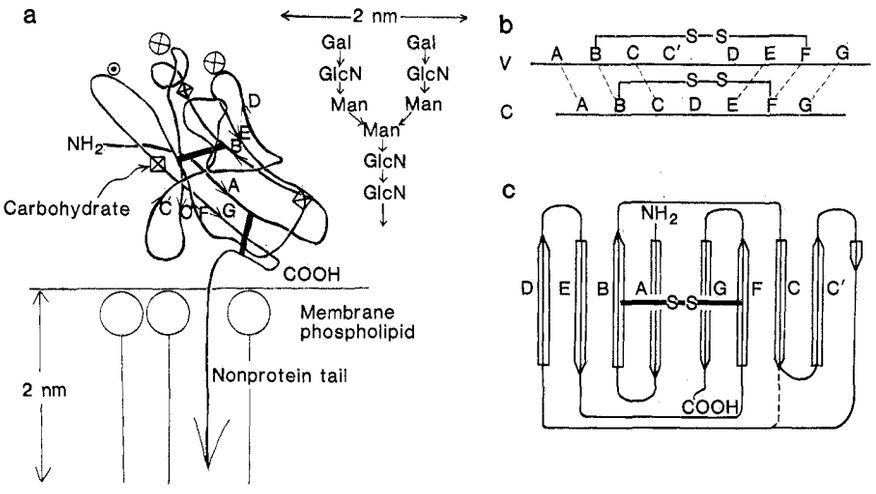

Fig. 2. (a) Model for Thy-1 at the cell surface drawn roughly to scale. The polypeptide is folded as for the V_λ Ig domain (*41*). ⊠ The positions of *N*-linked carbohydrate structures at residues 23, 74, and 98 (rat Thy-1) or 23, 75, and 99 (mouse Thy-1) (Fig. 5) and a typical asialo-*N*-linked structure is shown roughly to scale at top right; ⊙, the position of the allotype-related sequence difference at residue 89 in mouse Thy-1; ⊗, indicates the two blocks of four and five residues that differ in sequence between mouse and rat Thy-1 at residues 26 to 29 and 63 to 67 (Fig. 6). The bars (—) show disulfide bonds with the Thy-1 bond Cys^{19}-Cys^{85} (rat) or Cys^{19}-Cys^{86} (mouse) drawn as for the conserved bond between β-strands B and F in the V domain (Fig. 2c). The other Thy-1 disulfide bond, Cys^{9}-Cys^{111} (rat) or Cys^{9}-Cys^{112} (mouse) is drawn to fit with the Ig fold and to show how a nonprotein hydrophobic "tail" might be attached and inserted into the lipid bilayer. (b) Disposition of segments of β-strand along Ig V- and C-domain sequences. (c) Folding pattern in Ig domains. The broad arrows in (c) indicate β-strand segments and the single line sequences connecting β-strands. The lengths of the lines are not in proportion to the length of sequence [compare (a) and (c)]. The antiparallel β-strands form two β-sheets that are held together by hydrophobic interactions and the disulfide bond between β-strands B and F. The pattern for a C domain in (c) is shown by the dashed line directly connecting β-strands C and D. For V domains, there is extra sequence in the middle of the domain and this can vary in length. In (b) one extra β-strand (C') is shown, but often two extra β-strands are found as indicated in (c). [The diagrams in (b) and (c) are derived from (*7*)]

showed that antibody is only found if other antigenic determinants are present on the immunizing cells as well as the Thy-1 determinant (*39*). Alloimmunization with brain tissue does not elicit an immune response (*40*) perhaps because brain neuronal cells lack the viral or histocompatibility antigenic determinants that act as helper determinants when lymphocytes are used as antigens.

Homology Between Thy-1 and Immunoglobulin V Domains

All Ig domains consist of two β-sheets with antiparallel β-strands (*7, 41*). The segments forming the β-strands occur regularly along the sequences (Fig. 2b) and their folding pattern is illustrated schematically for V and C domains in Fig. 2c. Both domain types share a core of three β-strands in each β-sheet labeled A, B, E and G, F, C in Fig. 2. The sequences for these β-strands come from the ends of the domain while the central portion shows a different folding pattern in V and C domains. In V domains there is a longer central sequence which is folded in two main β-strands (C′ and D in Fig. 2, b and c), and C domains have only one central β-strand, which is labeled D. Also there are V- or C-specific residues in the other β-strands together with shared residues which characterize the whole Ig superfamily (*6*).

Previous studies clearly showed that Thy-1 and Ig domains were related in evolution since assignments of β-strands A, B, C and E, F, G could be made in the Thy-1 sequence on the basis of structural rules and sequence homology (*15*). A categorization of Thy-1 as being more V- or C-like was not achieved since the molecule seemed to show some characteristics of both types of domain.

With the mouse sequence and further analysis, it is now clear that Thy-1 fits better along the whole sequence with V rather than C domains. This was particularly suggested by analysis of identities between blocks of four or more Thy-1 residues with those of all proteins listed by Dayhoff *et al.* (*42*) (see Tables 1 and 2). An alignment of the Thy-1 sequences with a V_H and V_L sequence is shown in Fig. 6. In this alignment only three breaks were needed to match the sequences after V_H and V_L had been aligned. There are 25 and 23 sequence identities between mouse Thy-1 and the V_L and V_H domains, respectively, and nine of these in each case are with residues that are conserved in more than 90 percent of all V_L or V_H sequences (*43*). Between the V_L and V_H sequences shown there are 29 identities.

In considering the homologies (Fig. 7) in detail it is convenient to begin with the COOH-terminal half at Cys^{86} (mouse Thy-1 numbering) which is found in the center of β-strand F and forms the conserved disulfide bond with the Cys of β-strand B (Fig. 2b). A sequence like $Tyr^{84}..Cys^{86}$ is conserved in all Ig domains, but a Phe at 85 as in mouse Thy-1 is V-like since all V domains but no C domains have Phe or Tyr at this position (*43*). Also an Asp at residue 80 and Gly at 82 are indicative of V domains since an Asp six residues from the Cys is highly conserved in all V domains, and at four residues on the NH_2-terminal to the Cys V domains almost always have Ala (89.6 percent) or Gly (9 percent) (*43*). The sequence Asp^{83} Tyr Phe Cys is found in three V_λ domains but in no other proteins in the protein dictionary (Table 1), and the mouse Thy-1 sequence from residue 80 to 88 has up to seven identities with V_λ sequences (see pig V_λ in Fig. 8). If Thy-1 was like a C domain, then a conserved sequence around Cys^{86} of Tyr (Phe) x Cys x Val x His would be expected (*6*) (x indicates a nonconserved amino acid).

	Thy-1.2	Thy-1.1	Thy-1.1 sequence	Squid brain glycoprotein
SCMC	3.6	3.7	4	8
ASX	13.8	13.8	14	15
THR	9.8	9.8	10	7
SER	10.7	10.8	11	6.8
GLX	12	10.5	10	12.4
PRO	3.2	3.1	3	2.7
GLY	5.5	5	4	8
ALA	4.3	3.7	3	4.7
VAL	8	7.9	8	7.7
MET	1.3	1.1	1	1.7
ILE	4	3.9	4	4.9
LEU	11.2	11.7	12	6.1
TYR	3.9	3.9	4	4.4
PHE	3.1	2.8	3	3.6
HIS	3.4	3.8	4	3.7
LYS	7.9	8.6	9	9
ARG	6.3	7.9	8	6.3

Fig. 3. Amino acid analyses of glycoproteins. Mouse Thy-1 glycoproteins were purified from the brains of AKR (Thy-1.1) and CBA (Thy-1.2) mice. Thy-1.1 was purified by means of an OX 7 (MRC) antibody affinity column (*16*) and Thy-1.2 by means of a lentil lectin affinity column (*24, 29*) after preparation of deoxycholate extract from brain (*16*). The amino acid compositions were derived from at least six analyses and the Ser and Thr values were corrected for degradation; 1/2 Cys values (SCMC) were determined on reduced and alkylated glycoprotein samples. The squid brain glycoprotein was prepared as for mouse Thy-1.2 and the analysis of a reduced and alkylated sample is shown with values normalized to 112 residues.

On the COOH-terminal side of Cys^{86} there are few identities between Thy-1 and V domains, and, in β-strand G, Thy-1 looks more like a C domain [(*15*) and K(100)SIS identity in Table 1]. When V_L and V_H domains form dimers, a number of contacts occur through the amino acids of β-strand G (*7*) and these have presumably been selected for such interactions. There is no evidence that the homologous regions of Thy-1 and C domains interact with other proteins.

In β-strand E, sequences which distinguish V and C domains are not found. A pattern like the Leu^{71} Thr Leu of Thy-1 is conserved in virtually all Ig domains in this position (*6*).

In β-strand D, all but one V_L and 90 percent of V_H sequences have Arg (the rest have Lys) at position 59 and a patch like Arg Phe Ser (less commonly Arg Val Ser or Thr) is highly conserved (*43*). Around Arg^{59} mouse Thy-1 has five residues that are identical to V_H (New M) in this position (Table 1), and the more extended sequence from Thy-1 residue 55 to 63 is highly homologous in terms of conservative substitutions with the predominant sequence in V_H domains (*43*). This suggests that Thy-1 runs from β-strand E directly into a β-strand D like that in V_H domains.

At the NH_2-terminus there is good homology in terms of conservative substitutions between Thy-1 and V domains through β-strand A, provided that a deletion of three Thy-1 residues is made between the postulated β-strands A and B. Cys^{19} defines β-strand B and a Leu at 17 is commonly found in both V and C domains. However, the sequence Leu^{15} Arg^{16} is characteristic of V_H domains (Table 1) and Arg^{20} of V_L domains (*43*).

For β-strand C, the homology of Thy-1 with Ig has been previously discussed (*15*) and Leu^{36} of Thy-1 can be confidently aligned with the invariant Trp of Ig domains. In this position β_2-m also has a Leu residue (Fig. 6) as does a pseudogene V_κ sequence (*44*). Toward the COOH-terminus from Leu^{36} there is excellent homology between Thy-1 and the V_L and V_H domains shown. In this region V domains move from β-strand C into the extra C′ strand not found in C domains (see Fig. 2, b and c). The homology between Thy-1 and V domains breaks down at a point (Leu^{49}) one residue from the beginning of the second hypervariable region of V domains. Immediately after this hypervariable region

V_L but not V_H domains have a highly conserved sequence of Gly Val Pro (*43*). If the hypervariable region of V_L domains is deleted this can be aligned with Gly^{50} Val Pro of Thy-1. Mouse Thy-1 has Ile not Val at position 51, but this reinforces the homology since, of 152 V_L sequences, 126 have Val and 23 have Ile at this position (*43*). The sequence Gly Ile Pro Glu is found in mouse Thy-1 and a number of human V_λ sequences, but in no other protein in the protein dictionary (Table 1).

The overall conclusion from Table 1 and Fig. 6 is that Thy-1 has strong homology with V domains and that it is likely to have the extra β-strand homologous to C′ of V domains. It should be noted that Thy-1 has sequences that are characteristic of V_L (particularly λ) compared with V_H domains (residues 20, 50 to 53, 81, 83, and 88) and vice versa (residues 15, 16, and 55 to 63), as well as having sequences commonly found in V_L

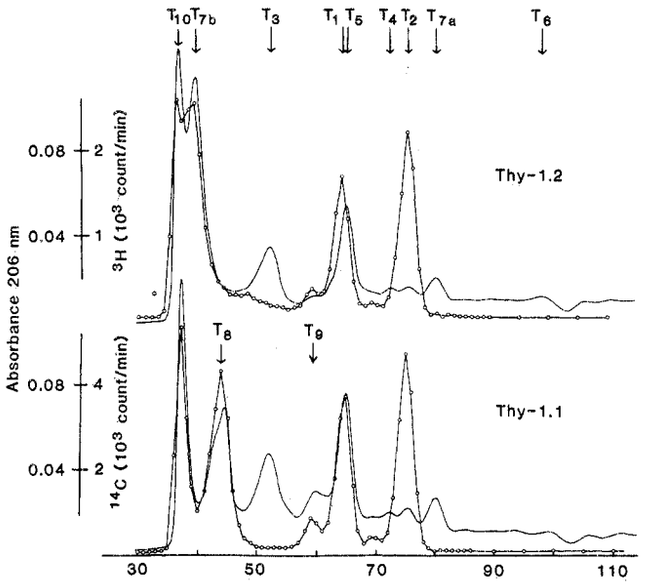

Fig. 4. Separation of tryptic peptides from mouse Thy-1.2 and Thy-1.1 glycoproteins. For tryptic digestion the glycoproteins were completely reduced and alkylated (with [^{3}H]- or [^{14}C]iodoacetic acid) and succinylated. The digestion was carried out in Brij 96 detergent, which allows the recovery of the COOH-terminal tryptic peptide, which binds to the detergent micelle (*16*). The fractionation of tryptic digests on Biogel P-10 in 0.1*M* ammonium bicarbonate is shown for Thy-1.1 (10 nanomoles) and Thy-1.2 (8 nanomoles). The absorbance at 206 nanometers (—) was read on an LKB Uvicord S monitor and the radioactivity (—○—) in *S*-carboxymethylcysteine was counted. The arrows indicate the elution positions of the tryptic peptides (Fig. 5).

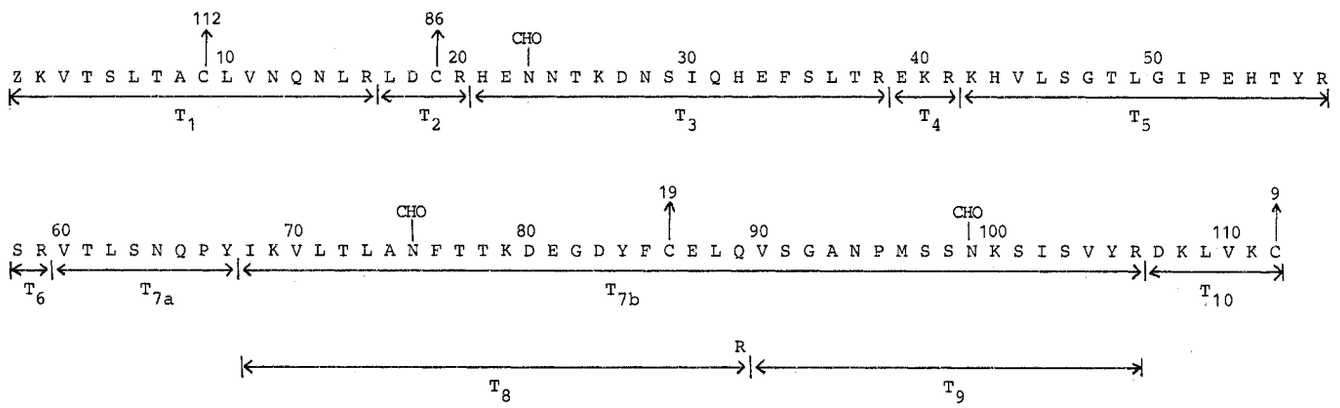

Fig. 5. Sequences of mouse Thy-1.1 and Thy-1.2 glycoproteins. Peptides were prepared for sequencing by digesting 100 nanomoles of Thy-1.1 and Thy-1.2 with trypsin and fractioning on Biogel P-10. They were further purified by high-performance liquid chromatography (HPLC) (*56*) on a Waters μBondapak C_{18} column in 0.1 percent ammonium bicarbonate or 0.2 percent formic acid buffers with a gradient of acetonitrile. The efferent was monitored on an LKB Uvicord S monitor with HPLC flow cell at 206 nanometers. The peptides were identical for Thy-1.1 and Thy-1.2, except for T_{7b}, T_8, and T_9, which differ because of the presence of Arg at position 89 in Thy-1.1 compared with glutamine for Thy-1.2. The tryptic peptides from both Thy-1.1 and Thy-1.2 were sequenced in a Beckman automatic sequencer (with Polybrene), and the phenylthiohydantoin derivatives were identified by HPLC. In all cases the sequences satisfied the analyses. The NH_2-terminal peptides T_1 were blocked, but were sequenced after enzymatic removal of the NH_2-terminal pyroglutamic acid (*57*). The peptide analysis was started with 25 to 80 nanomoles, and the initial yield was in the range of 30 to 80 percent, except for T_6 (12 percent), with repetitive yields of 83 to 97 percent for peptides with more than three residues. There was no indication of more than one residue at any position and no ambiguity about any of the residues, except for residues 23, 75, and 99. At these positions no amino acid was determined, but the residues are likely to be Asn with *N*-linked carbohydrate attached as for rat Thy-1. The peptides contained carbohydrate, and in each case one Asx residue in the composition was not accounted for by the peptide sequence. There was one unexpected cleavage, which occurred after Tyr 67 in both Thy-1.1 and Thy-1.2. The peptides have been aligned by homology with rat Thy-1 (Fig. 6). The alignment of T_5, T_6, and T_{7a} was confirmed by isolation and sequencing of a pepsin peptide (see below) YRSRVTL. The alignment of T_{7a} and T_{7b} was confirmed by the identification of a minor sequence in T_{7b} and T_8 which was for T_{7a} plus T_{7b} and T_{7a} plus T_8, respectively, and was presumably due to failure of the unscheduled cleavage to go to completion. A pepsin peptide of composition SNQPYIKV was also isolated. The alignment of T_{7b} or T_9 with T_{10} was confirmed by isolation and sequencing of a pepsin peptide of sequence YRDKL. To identify the disulfide bonds, 100 nanomoles of unreduced mouse Thy-1 was digested with pepsin in the presence of Brij 96 and 0.5*M* acetic acid as for rat Thy-1 (*16*). The peptides linked by Cys^9 and Cys^{112} bound to the detergent micelle and were eluted at the front of a Biogel P-10 column in 0.1*M* acetic acid while the peptides linked by Cys^{19} and Cys^{86} were retarded on the column. The dipeptides were oxidized, and the Cys peptides were purified by gel filtration and HPLC. The amino acid compositions were in complete accord with the sequences: TAC(9)L; RLDC(19)RHENNTKDNSIQHE; FC(86)E or FCEL; VKC(112).

and V_H domains but rarely in C domains (for example, residues 59, 80, 82, 85). The mean minimum base change per codon needed to convert the aligned sequences in Fig. 6 from one to the other is not much greater for Thy-1 compared to V_L (1.12) or V_H (1.13) than it is for V_L versus V_H (1.02) (Table 3). The probability that these scores are due to chance is much less than 0.01 percent according to the methods of Moore and Goodman (*45*).

Comparisons Between Mouse and Rat Thy-1 and the V-Domain Fold

In rat and mouse Thy-1, there are a number of long stretches of identical sequences, especially residues 5 to 25, 30 to 40, 42 to 50, 52 to 60, and 68 to 84 (Fig. 6). In two places these are interrupted by blocks of residues that differ between the species. At residues 26 to 29 there are four differences including a deletion, while at residues 63 to 67 there is a block of five different amino acids.

The locations of these blocks of differing sequence are shown on the V fold in Fig. 2a, and they strongly support this model for Thy-1. Residues 26 to 29 would come in the bend region connecting β-strands B and C and residues 63 to 67 in the bend connecting β-strands D and E. This is to be expected, since in Ig domains the sequences connecting β-strands vary much more than those in the β-strands (*6*), and the hypervariable regions consist of interconnecting sequences. Thy-1 residues 26 to 29 correspond closely to the first hypervariable region of V domains and residues 63 to 67 correspond to a stretch that can show variability in V domains (*43*) even though it is not considered to be part of the combining site. The amino acid difference between mouse Thy-1.1 and Thy-1.2 is also found in a sequence that is postulated to connect β-strands (residue 89 between strands F and G) in a position corresponding to the third hypervariable region of V domains. This bend and the other two mentioned above are all adjacent in the model shown in Fig. 2a, and this corner of the molecule might account for all the antigenic determinants of Thy-1. Other parts of the surface of Thy-1 may be covered by the carbohydrate structures interacting with the polypeptide as is seen in IgG C_H2 domains (*46*) and in the influenza hemagglutinin molecule (*47*). The attachment points for *N*-linked carbohydrate are shown in Fig. 2a, which also shows a typical *N*-linked carbohydrate structure drawn roughly to scale. Three such structures could cover most of the outer protein surface of Thy-1.

Homologies Between Thy-1 and C Domains

In previous studies homologies between Thy-1 and C domains were noted (*14, 15*). The extra disulfide bond of Thy-1 is like one that can occur in C_H1 domains, and homologies with C domains were much more significant than with V domains in sequence comparisons restricted to 35 residues assigned to the β-strands A, B, C, E, F, and G (*15*). In these comparisons the blocks of sequences in Tables 1 and 2 which strongly suggest an exact fit with a V fold were mostly not considered.

This likeness to both V and C domains would be explained if Thy-1 was closer to the primordial domain than any other contemporary Ig domain. McLachlan (*48*) has argued on structural grounds that the primordial domain should have had a V domain fold with the C domain pattern being derived by sequence deletion in the middle of the molecule (Fig. 2b). A similarity between Thy-1 and the primordial domain would be supported if it could be shown that β-strand D of C domains had sequence homologies with the Thy-1 sequences found between β-strands C and E. In Fig. 7 it is seen that this is the case, and the homology with rabbit C_H1 includes one of the blocks of four identities listed in Tables 1 and 2.

If C domains are aligned along the full sequence of Thy-1 with alignments at each end as before (*15*) and those in the middle as in Fig. 7, then good scores for identities and mean minimum base changes are obtained (Table 3). The best fit is seen with the C_H1 domain whose scores are as good as those for V domains. However with C domains, unlike V domains, patches of four and five identities in highly conserved regions are not seen and more gaps are needed for maximum homology. An excellent alignment between Thy-1 and β_2-m is also possible and this is shown in Fig. 6. To achieve maximum homology the deletion in the middle of the molecule must be

Tables 1 and 2. Identities between segments of rat or mouse Thy-1 sequences (four or more residues) and Ig sequences. In mouse and rat Thy-1 there are 129 short sequences of at least four residues which can be compared with sequences in the protein segment dictionary (*42*), which contains about 120,000 segments and was compiled in 1978 from 1081 sequences or major fragments of sequence containing 119,000 residues (about 11,500 residues are from immunoglobulin V-domain sequences and 4200 from C-domain sequences). On a random basis it would be expected that identities to a minimum of 40 of the Thy-1 tetrapeptide sequences should be found in the dictionary since 30 percent of the possible tetrapeptides are present (presumably > 30 percent of those beginning with common amino acids). The number of Thy-1 sequences found with identities in the dictionary was 72, of which the 19 shown in Tables 1 and 2 were with Ig domains [only sequences which cross-checked as correct with (*43*) were used]. The sequence LRLXC, where X is a nonidentity, is also included since it involves homology with the conserved NH_2-terminal Cys of Ig domains. There were also seven segments identical to hemoglobin sequences (8500 residues in the dictionary), but there is no extended relation between Thy-1 and hemoglobin sequences. The identities with Ig in nonhomologous positions (Table 2) are presumably due to chance association biased by the presence of patches of sequence suitable for β-strand formation in both Thy-1 and Ig sequences. Table 1. Identities between segments of Thy-1 sequence (four or more residues) and Ig sequences in homologous positions within domain (see Figs. 6 and 7 for alignments). The numbers in parentheses refer to the amino acid position in the Thy-1 sequences.

Thy-1 sequences*	Ig domains with identity	In other proteins
L(15)RLXC	V_H (19 sequences)	One
S(46)GTL	Rabbit C_H1 IgG	Three
G(50)IPE	Three human V_λ	None
R(57)SRVT	Human V_H (New M)	None
D(83)YFC	Three V_λ (two human, pig)	None
K(100)SIS	Rabbit C_H3 IgG	None; nonhomologous V_κ

*The one-letter abbreviations for amino acids are used here (*36*).

Table 2. Identities between segments of Thy-1 sequences (four or more residues) and Ig sequences (*36*) in nonhomologous positions. Other details as in the legend for Table 1.

K(2)VTS[C_H2 IgM 320]*; I(4)SLT [C_H2 IgG 266]; T(4)SLTA[V_H 81]; S(5)LTA[V_H 81]; A(8)CLV[C_H1 IgA 141]; L(27)PIQH[C_H2 IgG 325]; K(42)HVL[V_λ 96]; R(59)VTL[V_κ 18]; V(60)NLF[$C_L\lambda$ 115]; S(64)DRF[V_λ 59]; T(78)KDE[$C_L\kappa$ 182]; V(90)SGA[V_λ 11; C_H1 IgE 193]; T(96)SSN[V_λ 92]; S(98)NKSIS[V_κ 26].

*In brackets the Ig sequence is identified as domain, class, and sequence position with numbering as in (*43*).

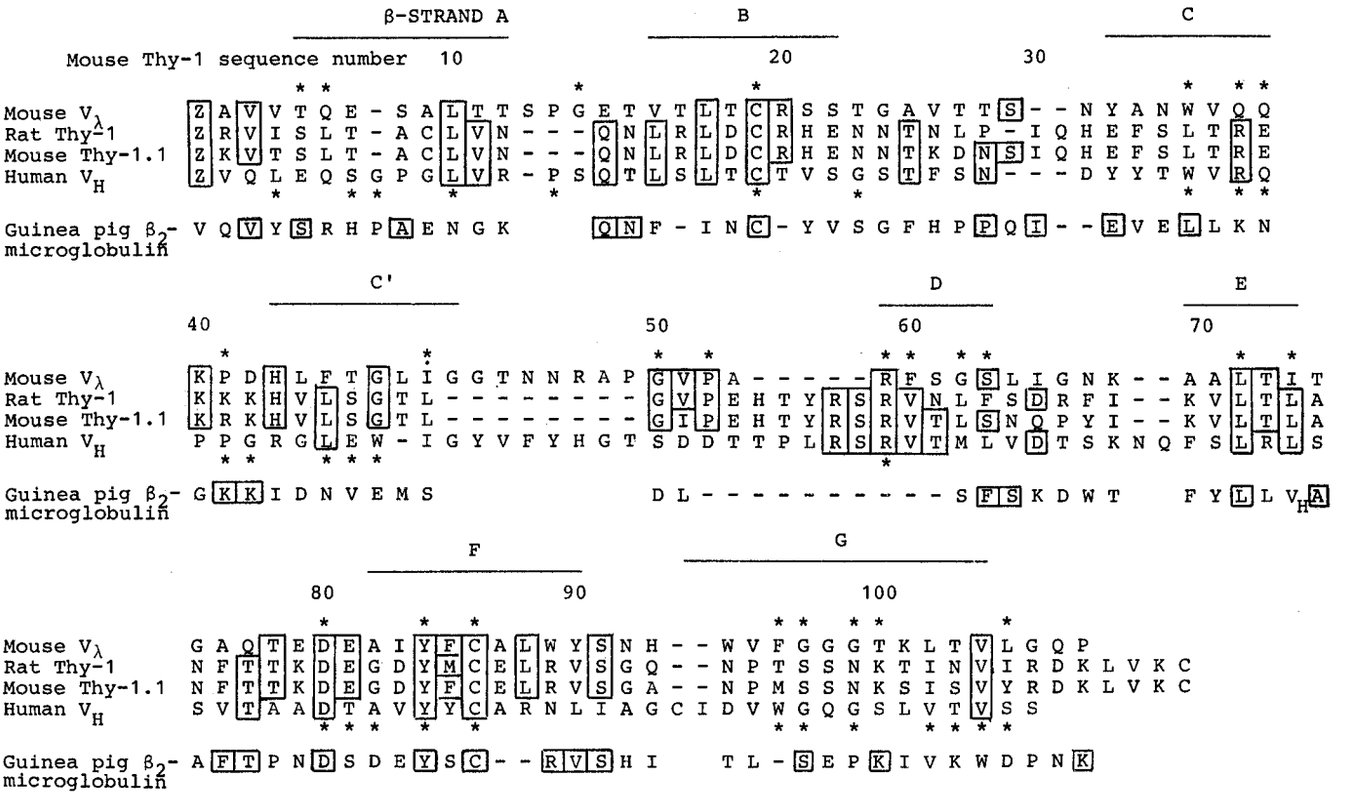

Fig. 6. Alignment of rat and mouse Thy-1 sequences with a mouse V_λ sequence (MOPC 104E) (*41*), a human V_H sequence (New M) (*43*) and guinea pig β_2-m sequence (*58*). In the comparison with V domains, identical residues are boxed and dashes show gaps in the sequence inserted to maximize the alignment. Residues with an asterisk (*) above or below are those present at a frequency of > 0.9 in all V_L or V_H sequences. The bars with letters A, B, C, C', D, E, F, and G above the sequences indicate residues involved in forming β-strands as shown in Fig. 2. Some variations occur between V_H and V_L in the beginning and end of these segments, but the sequences shown are a reasonable approximation to those involved in the strands for both V_L and V_H domains (*15*). In the alignment with guinea pig β_2-m the boxed residues in the β_2-m sequence are those that are identical to rat Thy-1.

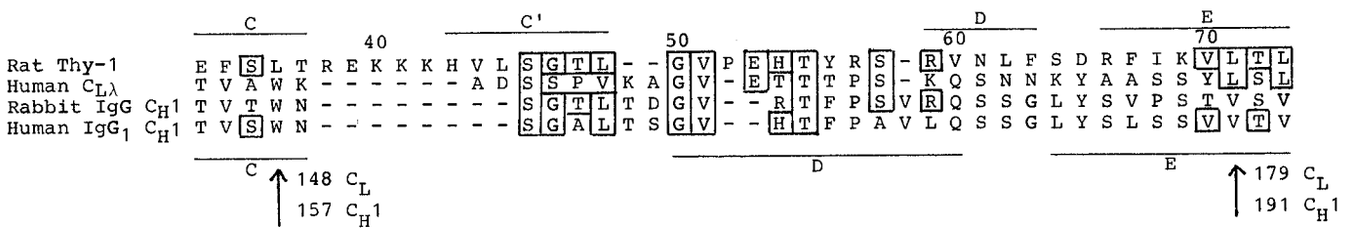

Fig. 7. Homologies between rat Thy-1 and C_H1 and $C_L\lambda$ sequences in the middle of the domain. The β-strands C and E were identified in the Ig sequences and in Thy-1 as in (*15*), and the sequences in between then aligned to give maximum homology with a minimum of gaps inserted. The proposed β-strands C' and D for Thy-1 are as in Fig. 6 and are marked above the sequence while the D β-strand for C domains is marked below and is as designated for human $C_L\lambda$ in (*15*) from the structure of Saul *et al.* (*59*). The sequence numbers above the sequences are for mouse Thy-1 as in Fig. 6, and the numbers below are for the arrowed residues of the C_L and C_H1 sequences with numbering as in (*43*).

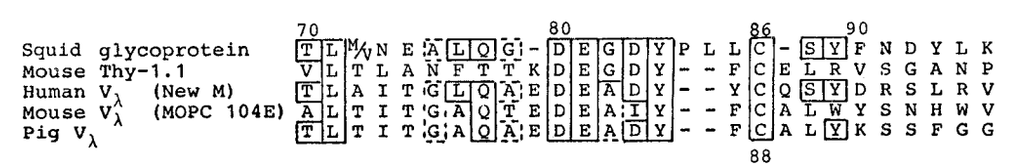

Fig. 8. Homology between a squid brain "Thy-1" homolog, mouse Thy-1.1, and immunoglobulin V_λ domains. A partial sequence from a squid brain glycoprotein is compared with mouse Thy-1.1 and V_λ sequences (*43*) around the COOH-terminal Cys of the conserved disulfide bond of Ig domains. The numbering above the sequences is as for mouse Thy-1 (Fig. 6) and that below for V_λ sequences as in (*43*). The boxes with unbroken lines show identities and those with broken lines the conservative substitution Ala or Gly. The sequence of the squid peptide is from automatic sequencer runs on two tryptic peptides. One was a peptide from reduced, alkylated, and succinylated brain glycoprotein isolated by gel filtration on Biogel P-10, as in Fig. 4 in a similar position to T_{7b} of mouse Thy-1.2. This peptide gave sequence to residue 36 and sequence up to 26 as shown was unambiguous with the exception of the Met or Val assignment at residue 3. The other tryptic peptide was from reduced and alkylated glycoprotein and was isolated by gel filtration on Biogel P-10 and was further purified by HPLC on a Waters μBondapak C_{18} column in 0.1 percent ammonium bicarbonate with elution by a gradient of 5 to 35 percent of a solvent mixture containing equal volumes of acetonitrile, methanol, and isopropanol. This peptide gave unambiguous sequence (with the exception of residue 3 which is probably Met) to residue 24.

Table 3. Evaluation of alignments between Thy-1 and other sequences in the immunoglobulin superfamily. The V domain and β_2-m sequences were aligned as in Fig. 6. Immunoglobulin C domain sequences were aligned as in (*15*) except that sequences were aligned in the middle of the molecule as shown in Fig. 7; the conserved Val of C domains two residues from the COOH-terminal Cys of the conserved disulfide bond was aligned with Val^{90} of Thy-1 as shown for β_2-m in Fig. 6; residues at the beginning of the molecule and in stretches connecting β-strands were aligned to maximize homologies without excessive insertion of gaps. The probabilities of the mean base change scores being a random occurrence was read from the statistical tables of Moore and Goodman (*45*). According to their test the probability of obtaining a mean score of 1.17 for 100 residues aligned is 0.01 percent. Ig sequences are from (*43*) and guinea pig and mouse β_2-m from (*58*, *60*).

Rat Thy-1 sequences compared with	Residues aligned*	Identities	Minimum base change per residue (mean)	Probability due to random event (%)
Mouse V_λ (MOPC 104E)	100	23	1.11	<<.01
Human V_λ (New M)	101	15	1.18	>.01 <.1
Human V_H (New M)	103	22	1.13	<<.01
Human C_L (lambda)	95	19	1.28	>1 <5
Human C_H1 (IgG_1)	96	23	1.03	<<.01
Human C_H2 (IgG_1)	94	20	1.15	<.01
Human C_H3 (IgG_1)	94	15	1.17	>.01 <.1
Mouse β_2-m	90	22	1.14	<.01
Guinea pig β_2-m	90	27	1.05	<<.01

*Alignments are available on request.

made in a different position for β_2-m than for C_H1 and C_L domains, and this could indicate an independent divergence from the primordial domain for β_2-m compared with C_H1 and C_L domains. The homologies between Thy-1 and β_2-m are at least as good as those between β_2-m and C domains (*11*). It is clear that the case for Thy-1 being like the primordial domain is as sound as could be expected on the basis of sequence homologies.

Identification of an Invertebrate Thy-1 Homolog

If Thy-1 were shown to be primitive, the argument for its likeness to the primordial domain would be greatly strengthened. A homolog of Thy-1 was thus sought in squid brain with the use of a biochemical approach. In all species so far examined the Thy-1 glycoprotein stands out because of its low molecular weight and high content in neuronal tissue. Glycoproteins were prepared and fractionated as for mouse Thy-1.2 (Fig. 3), and a possible homolog was purified. It has an apparent molecular weight on sodium dodecyl sulfate (SDS)–polyacrylamide gel electrophoresis (PAGE) of 18,000 (rodent Thy-1 is 23,000 but the protein molecular weight is only 12,500) and an amino acid composition similar to rodent Thy-1 (Fig. 3). Tryptic peptides were prepared and a partial sequence is shown in Fig. 8. A block of five identities can be aligned with rodent Thy-1 at the NH_2-terminal side of Cys^{86}. This homology seems highly significant since the five identical residues are not seen in any protein listed in the protein segment dictionary (*42*). With Ig V_λ domain sequences the homology is even more striking than with rodent Thy-1 (Fig. 8).

Implications of the Homologies

There is no doubt that Thy-1 is the best candidate for a molecule which may approximate to the primordial Ig domain, and the full sequence for the squid glycoprotein should be informative in further testing this possibility. Analysis of the variation of Thy-1 in evolution and of the disposition of exons and introns in Thy-1 genes are of interest in relation to the evolution of immunoglobulins. The sequence divergence between rat and mouse Thy-1 may be of relevance to the mechanism of generation of diversity in the hypervariable regions of V domains. Of the three hypervariable regions, a mechanism is known only for the third region of V_H where diversity is produced by variation in the splicing of small segments of sequence to the main exon (*9*, *49*). The patch of four residues which differ between rat and mouse Thy-1 at sequences 26 to 29 is almost exactly homologous in position to the first hypervariable region of V_H domains (*42*). This suggests that genetic drift at the rate seen in Thy-1 would be sufficient to produce hypervariability in this region in a multiple set of V genes.

The data on Thy-1 alters concepts about Ig-related structures on cell surfaces. In immunology there has been an unstated assumption that Ig-related structures, and particularly V regions, will be found only on receptors involved in immune functions. The possibility of such a relationship may have survived the finding of homologies between histocompatibility antigens and C-region domains since histocompatibility antigens appear to control T cell recognition (*50*). However, the Thy-1 data show that V-like domains must have a more basic function, since expression of this molecule is not conserved on lymphoid cells (*35*). In fact, Thy-1 seems very much a neuronal molecule rather than a lymphoid one, and if a molecule like it did give rise to the Ig superfamily there seems no reason why there may not be a related set of molecules involved in neuronal recognition. The finding that homology with V domains does not define an antigen receptor raises problems in the interpretation of the ever-popular experiments in which T lymphocyte receptors are sought on the basis of cross-reactions with antibodies to immunoglobulins. These problems are strikingly illustrated by the isolation of monoclonal antibody to Thy-1, which is specific for idiotypic determinants of the TEPC-15 mouse myeloma protein (*51*).

The sequence homologies between rat and mouse Thy-1 prove that the molecule must have an essential function, since it is not obvious how long stretches of sequence can remain identical on either side of a patch of variation other than by selection. Two possible functions are suggested by the roles of Ig domains in immunoglobulins where V-region domains mediate antigen recognition and C-domains effector functions. By analogy with V domains one could argue that Thy-1 functions as a receptor for another molecule (for example, a hormone) with the combining site being the region homologous to the antigen recognition site. In contradiction is the fact that the sequence in this part of Thy-1 shows great variation between the rat and mouse. Alternatively, one could argue that Thy-1 functions in a manner analogous to C domains of Ig and is in effect a ligand that is recognized by receptors on other cells. We have previously proposed a hypothesis on these lines, suggesting that Thy-1 is one of a set of Ig-related structures that mediate cell recognition in morphogenesis (*15*). In this hypothesis, the Ig domain can be considered as a stable "platform" for the display of determinants that could be protein as in Ig Fc regions and potential-

ly protein or carbohydrate in the case of Thy-1 antigen. The idea of the Ig domain as a basic recognition unit for cell interactions would then be the functional common denominator for the Ig superfamily rather than a function that is linked to immunity.

Evidence as to the role of Thy-1 is clearly needed. It has been shown that antibodies to Thy-1, when applied to the brain in vivo, can selectively affect the behavioral responses of rats (*52*), and that such antibodies can inhibit the virus-dependent proliferation of leukemia cells (*53*). Further experiments on the effects of antibodies against Thy-1 in functional systems may be rewarding.

References and Notes

1. M. O. Dayhoff, L. T. Hunt, P. J. McLaughlin, W. C. Barker, in *Atlas of Protein Sequence and Structure*, M. O. Dayhoff, Ed. (National Biomedical Research Foundation, Georgetown University, Washington, D.C., 1972), vol. 5.
2. W. Gilbert, *Nature (London)* **271**, 501 (1978).
3. C. C. F. Blake, *ibid.* **291**, 616 (1981).
4. R. L. Hill, R. Delaney, R. E. Fellows, H. E. Lebowitz, *Proc. Natl. Acad. Sci. U.S.A.* **56**, 1762 (1966).
5. G. M. Edelman, *Biochemistry* **9**, 3197 (1970).
6. D. Beale and A. Feinstein, *Q. Rev. Biophys.* **9**, 135 (1976).
7. L. M. Amzel and R. J. Poljak, *Annu. Rev. Biochem.* **48**, 961 (1979).
8. H. Sakano *et al.*, *Nature (London)* **277**, 627 (1979).
9. J. M. Adams, *Immunol. Today* **1**, 10 (1980).
10. O. Smithies and M. D. Poulik, *Science* **175**, 187 (1972).
11. P. A. Peterson, B. A. Cunningham, I. Berggård, G. M. Edelman, *Proc. Natl. Acad. Sci. U.S.A.* **69**, 1697 (1972).
12. H. L. Ploegh, H. T. Orr, J. L. Strominger, *Cell* **24**, 287 (1981).
13. H. T. Orr, D. Lancet, R. J. Robb, J. A. Lopez de Castro, J. L. Strominger, *Nature (London)* **282**, 266 (1979).
14. D. G. Campbell, A. F. Williams, P. M. Bayley, K. B. M. Reid, *ibid.*, p. 341.
15. F. E. Cohen, J. Novotný, M. J. E. Sternberg, D. G. Campbell, A. F. Williams, *Biochem. J.* **195**, 31 (1981).
16. D. G. Campbell, J. Gagnon, K. B. M. Reid, A. F. Williams, *ibid.*, p. 15.
17. D. W. Mason and A. F. Williams, *ibid.* **187**, 1 (1980).
18. A. N. Barclay and H. Hydén, *J. Neurochem.* **31**, 1375 (1978).
19. A. E. Reif and J. M. V. Allen, *J. Exp. Med.* **120**, 413 (1964).
20. G. D. Snell and M. Cherry, in *RNA Viruses and Host Genome in Oncogenesis*, P. Emmelot and P. Bentvelzen, Eds. (North-Holland, Amsterdam, 1972), p. 221.
21. T. C. Douglas, *J. Exp. Med.* **136**, 1054 (1972).
22. R. J. Morris and A. F. Williams, *Eur. J. Immunol.* **5**, 274 (1975).
23. M. Letarte-Muirhead, A. N. Barclay, A. F. Williams, *Biochem. J.* **151**, 685 (1975).
24. A. N. Barclay, M. Letarte-Muirhead, A. F. Williams, *ibid.*, p. 699.
25. P. W. Kuchel, D. G. Campbell, A. N. Barclay, A. F. Williams, *ibid.* **169**, 411 (1978).
26. M. J. Owen, J. C. A. Knott, M. J. Crumpton, *Biochemistry* **19**, 3092 (1980).
27. A. N. Barclay, M. Letarte-Muirhead, A. F. Williams, R. Faulkes, *Nature (London)* **263**, 563 (1976).
28. L. D. McClain, M. Tomana, R. T. Acton, *Brain Res.* **159**, 161 (1978).
29. M. Letarte and G. Meghji, *J. Immunol.* **121**, 1718 (1978).
30. S. F. Cotmore, S. A. Crowhurst, M. D. Waterfield, *Eur. J. Immunol.* **11**, 597 (1981).
31. J. L. McKenzie, A. K. Allen, J. W. Fabre, *Biochem. J.* **197**, 629 (1981).
32. J. L. McKenzie and J. W. Fabre, *Transplantation* **31**, 275 (1981).
33. J. A. P. Rostas, T. A. Shevenan, C. M. Sinclair, P. L. Jeffrey, in *New Approaches to Nerve and Muscle Disorders*, A. D. Kidman, J. Tomkins, R. A. Westerman, Eds. (Excerpta Medica, Amsterdam, in press); also personal communications from J. A. P. Rostas, University of Newcastle, Australia.
34. P. L. Stern, *Nature (London) New Biol.* **246**, 76 (1973).
35. R. Dalchau and J. W. Fabre, *J. Exp. Med.* **149**, 576 (1979).
36. Abbreviations: A, Ala, alanine; B, Asx, aspartic acid or asparagine; C, Cys, cysteine; D, Asp, aspartic acid; E, Glu, glutamic acid; F, Phe, phenylalanine; G, Gly, glycine; H, His, histidine; I, Ile, isoleucine; K, Lys, lysine; L, Leu, leucine; M, Met, methionine; N, Asn, asparagine; P, Pro, proline; Q, Gln, glutamine; R, Arg, arginine; S, Ser, serine; T, Thr, threonine; V, Val, valine; W, Trp, tryptophan; X, —, unknown or "other"; Y, Tyr, tyrosine; Z, Glx, glutamic acid or glutamine.
37. D. Hoessli, C. Bron, J. R. L. Pink, *Nature (London)* **283**, 576 (1980).
38. A. F. Williams, A. N. Barclay, M. Letarte-Muirhead, R. J. Morris, *Cold Spring Harbor Symp. Quant. Biol.* **41**, 51 (1976).
39. P. Lake and T. C. Douglas, *Nature (London)* **275**, 220 (1978).
40. R. Asakuma and A. E. Reif, *Cancer Res.* **28**, 707 (1968).
41. A. B. Edmundson, K. R. Ely, E. E. Abola, M. Schiffer, N. Panagiotopoulos, *Biochemistry* **14**, 3953 (1975).
42. M. O. Dayhoff, L. T. Hunt, W. C. Barker, R. M. Schwartz, B. C. Orcutt, *Protein Segment Dictionary* (National Biomedical Research Foundation, Georgetown University, Washington, D.C., 1978).
43. E. A. Kabat, T. T. Wu, H. Bilofsky, *Sequences of Immunoglobulin Chains* (U.S. Department of Health Education and Welfare, Public Health Service, National Institutes of Health Publication No. 80-2008, Bethesda, Md., 1979).
44. E. E. Max, J. G. Seidman, H. Miller, P. Leder, *Cell* **21**, 793 (1980).
45. G. W. Moore and M. Goodman, *J. Mol. Evol.* **9**, 121 (1977).
46. R. Huber, J. Diesenhofer, P. M. Colman, M. Matsushima, W. Palm, *Nature (London)* **264**, 415 (1976).
47. I. A. Wilson, J. J. Skehel, D. C. Wiley, *ibid.* **289**, 366 (1981).
48. A. D. McLachlan, *Protides Biol. Fluids Proc. Colloq.* **28**, 29 (1980).
49. E. A. Kabat, *J. Immunol.* **125**, 961 (1980).
50. J. Klein, A. Juretič, C. N. Baxevanis, A. Z. Nagy, *Nature (London)* **291**, 455 (1981).
51. E. Pillemer and I. L. Weissman, *J. Exp. Med.* **153**, 1068 (1981).
52. C. A. Williams, J. Barna, N. Schupf, *Nature (London)* **283**, 82 (1980).
53. M. S. McGrath, E. Pillemer, D. Kooistra, I. L. Weissman, *Contemp. Top. Immunobiol.* **11**, 157 (1980).
54. F. W. Putnam, G. Florent, C. Paul, T. Shinoda, A. Shimizu, *Science* **182**, 287 (1973).
55. M. Kehry, S. Ewald, R. Douglas, C. Sibley, W. Raschka, D. Fambrough, L. Hood, *Cell* **21**, 393 (1981).
56. M. Waterfield, G. Scrace, J. Skehel, *Nature (London)* **289**, 422 (1981).
57. D. N. Podel and G. N. Abraham, *Biochem. Biophys. Res. Commun.* **81**, 176 (1978).
58. P. B. Wolfe and J. J. Cebra, *Mol. Immunol.* **17**, 1493 (1980).
59. F. A. Saul, L. M. Amzel, R. J. Poljak, *J. Biol. Chem.* **253**, 585 (1978).
60. F. T. Gates, J. E. Coligan, T. J. Kindt, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 554 (1981).

61. We thank Mrs. Gaynor Newton, Tony Gascoyne, and Tony Willis for expert technical assistance, Mrs. Christine Scott for typing the manuscript, and Dr. M. Waterfield for advice on purification of peptides by high-performance liquid chromatography.