Ann. Rev. Immunol. 1988. 6 : 381–405

THE IMMUNOGLOBULIN SUPERFAMILY—DOMAINS FOR CELL SURFACE RECOGNITION[1,2]

Alan F. Williams and A. Neil Barclay

MRC Cellular Immunology Unit, Sir William Dunn School of Pathology, University of Oxford, Oxford OX1 3RE, United Kingdom

ENCOUNTERING THE Ig SUPERFAMILY

When Ig chains were first sequenced, segments within the constant regions of H and L chains showed sequence similarities, and this led to the idea that the Ig chains had all evolved from a primordial gene coding for about 100 amino acids (1). The domains within the Ig chains all contained a characteristic intrachain disulfide bond, and the idea of the domain as an independent structural unit was proposed (2). The domain hypothesis was firmly established when the structures of V and C domains were determined to reveal a common fold forming a sandwich of two β-sheets that was stabilized by the conserved disulfide bond (3, 4).

Beta-2 microglobulin (β_2-m), identified in the urine of patients with kidney disease, was the first nonimmunoglobulin structure found to share sequence similarities with Ig-domains. The β_2-m sequence looked like an Ig C-domain (5, 6). β_2-m was subsequently discovered to be part of the major histocompatibility complex (MHC) class I structure, and sequencing of the class I heavy chain showed that a segment of sequence adjacent to the transmembrane region was also similar to Ig C-domains (7, 7a). MHC antigens were known to play some role in the specificity of T lymphocyte

[1] Abbreviations: NBRF data base is a protein sequence data base from Protein Identification Resource (1987) National Biomedical Research Foundation, Washington, DC. Terms are defined in Table 1.

[2] Because of limited space, full referencing for this article has not been possible. Recent key references that lead to the full literature on each topic are cited.

0732-0582/88/0410-0381$02.00

recognition, and thus the finding that the MHC class I was Ig-related was in accord with the concept that Ig domains were uniquely concerned with immune recognition. This view was clearly dominant in the 1970s.

The Thy-1 differentiation antigen was sequenced at the same time as MHC class I H chain, and the finding that this molecule was like a single Ig V domain was not consistent with the immune recognition concept (8). Thy-1 was expressed in large amounts in neural tissue and thymocytes in rodents, but expression in lymphoid cells was not conserved in all species. It thus seemed likely that Ig-related structures would have a general role in cell surface recognition (9).

In the 1980s, many new cell surface structures have been identified and sequenced, and the Ig-related family of molecules can now be argued to include the structures shown in Figure 1 and Table 1. The molecules have a diversity of functions (Table 2), but in most cases the common denominator is a recognition role at the cell surface. The genetic loci are widely spread throughout the chromosomes, but a number of interesting linkage groups are seen. This family of molecules is undoubtedly one of the key groups not only in immunity but also in the mediation of cell surface recognition to control the behavior of cells in various tissues.

Figure 1 Models for molecules in the Ig-superfamily. One model is shown for each main molecular type from one species (Table 1), and in some cases the same model suffices for the additional structures named in brackets. The circles show sequence segments that fold as for an Ig domain or are predicted to do so at least to the extent of two sheets with β-strands ABE:GFC (Figure 2). Segments labelled V, C1, and C2 are in the categories indicated in Figure 2 and in the text. Domain numbers as used in the text and figures are from the NH_2-terminus. In CD4 four domains are counted even though the second domain is not typically Ig-related. In the MHC class I H chain and related structures, the three obvious segments starting from the NH_2-terminus are called α_1, α_2, and α_3 while in MHC class II β and α chains, the segments are β_1, β_2 and α_1, α_2. CD1 is shown independently from class I to indicate that it has much lower sequence identity to class I than do Qa and T1 (full CD1a data by personal communication from L. H. Martin, F. Calabi, and C. Milstein). Intrachain disulfide bonds that are like the conserved Ig disulfide bond are shown by $\overset{S}{S}$ symbols within the circles, and cases where these are confirmed are given in Table 1. Other intrachain bonds also exist, but these are mostly not shown. Interchain bonds are indicated by SS between chains known to be disulfide linked. CTLA4 could well exist in a dimer since a free sulphydryl is predicted in a membrane proximal position. N-linked carbohydrate sites as indicated by the presence of an Asn × Thr or Ser sequence are shown by the symbol (ψ) unless absence of glycosylation is known. The presence of glycophospholipid anchors is indicated by an arrow for Thy-1, LFA-3, and one NCAM form, and the possibility of this is indicated for CEA by an arrow plus ?. The Qa2 antigen also has a lipid tail (see text). A second form of LFA3 with a protein anchor has also been identified (84). References are in Table 1.

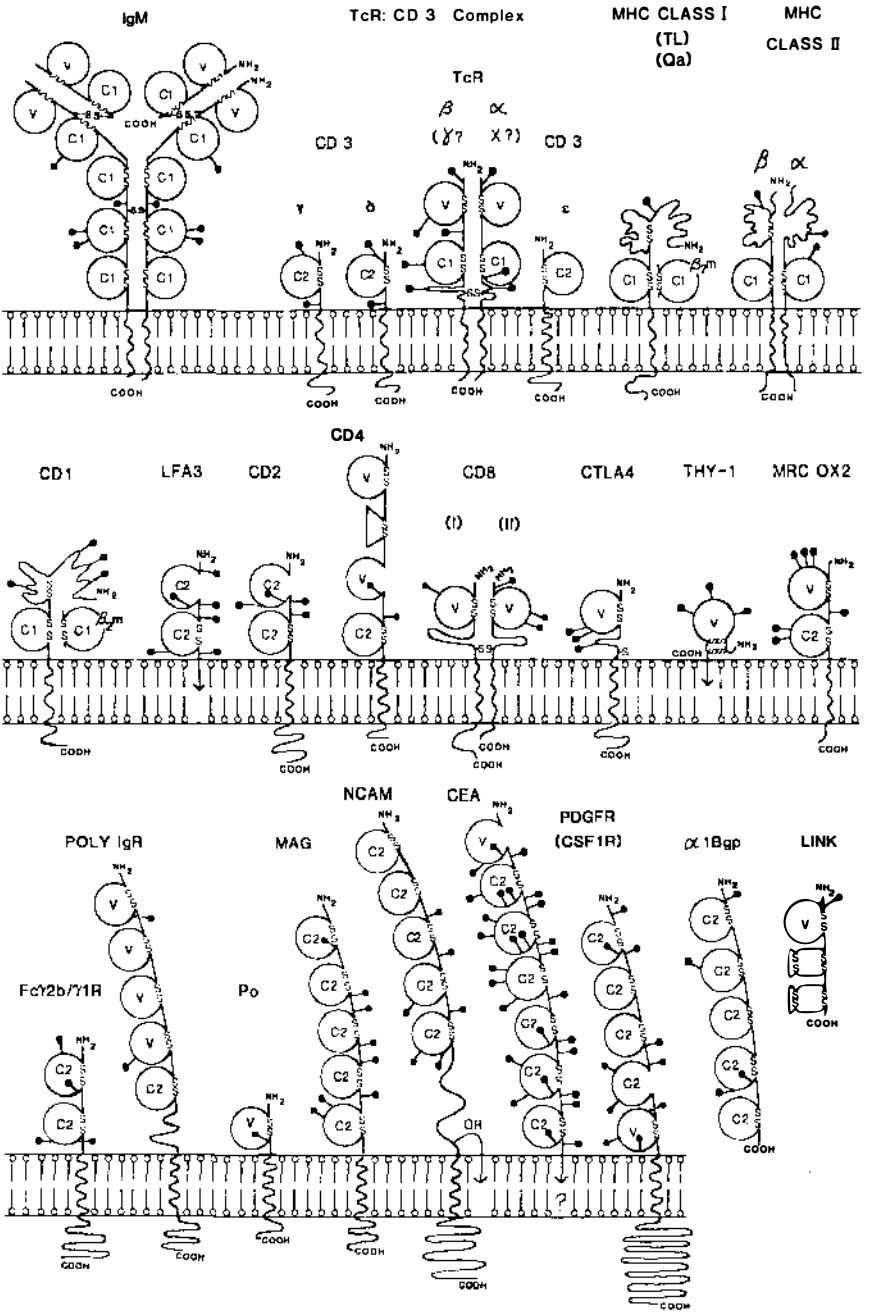


Table 1 Molecules of the Ig superfamily

Category	Sequence number	Human chromosome	Ig-like disulfide bonds	References
Immunoglobulins				
H chains (IgM)	572 (M)	14q32.33	Y	(2, 10)
L chain kappa	214 (M)	2p12	Y	
L chain lambda	213 (M)	22q11.12	Y	
T cell receptor (Tcr) complex				
Tcr α-chain	250 (M)	14q11.2	ND	(11)
β-chain	282 (M)	7q35	ND	
γ-chain	286 (M)	7p15	ND	
X-chain	272 (M)	ND	ND	(12)
CD3 γ-chain	160 (H)	11q23	ND	(13, 14, 15)
δ-chain	150 (H)	11q23	ND	
ε-chain	185 (H)	11q23	ND	
Major histocompatibility complex (MHC antigens)				
Class I H-chain	339 (H)	6p21.3	Y	(7, 16)
β_2-m	99 (H)	15q21-q22	Y	(6)
Class II α	229 (H)	6p21.3	Y	(17)
β	237 (H)	6p21.3	Y	
β_2-m associated antigens				
TL H chain	335 (M)	mouse only	ND	(18)
Qa H chain	313 (M)	mouse only	ND	(19, 20)
CD1a H chain	311 (H)	1	ND	(21)
T cell adhesion molecules				
CD2	322 (R)	1p13	ND	(22, 23, 24)
LFA-3	207 (H)	1	ND	(22, 25, 26)
T subset antigens				
CD4	435 (H)	12pter.p12	Y	(27, 28)
CD8 chain I	210 (R)	2p12	ND	(27)
chain II	187 (R)	2p12	ND	(27, 29)
CTLA4	188 (M)	ND	ND	(30)
Brain/lymphoid antigens				
Thy-1	111 (R)	11q23	Y	(8, 31, 32)
MRC OX-2	248 (R)	3	ND	(33)
Immunoglobulin receptors				
Poly Ig R	755 (RB)	ND	Y	(34, 35, 36)
Fcγ2b/γ1R	351 (M)	ND	Y	(37, 38, 39)
Neural molecules				
Neural adhesion molecule (NCAM)	1072 (CH)	11q23	ND	(40, 41)
Myelin associated gp (MAG)	610 (R)	ND	ND	(42, 43)
P_0 myelin protein	219 (R)	ND	ND	(42, 44)

Table 1 (*continued*)

Category	Sequence number	Human chromosome	Ig-like disulfide bonds	References
Tumor antigen				
Carcinoembryonic antigen (CEA)	668 (H)	ND	ND	(45)
Growth factor receptors				
Platelet-derived growth factor (PDGF) receptor	1067 (M)	5q31-q32	ND	(42, 46)
Colony stimulating factor-1 (CSF1) receptor	953 (H)	5q33.2-q33.3	ND	(42, 47)
Non-cell surface molecules				
α_1B-glycoprotein	474 (H)	ND	Y	(48)
Basement membrane link protein	339 (R)	ND	Y	(49, 50)

Most sequences are predicted from cDNA, but all have been also characterized as proteins except Tcr X-chain and CTLA4. The sequence length is for the fully processed form of one sequence in each category, and most sequence data is from the NBRF data base. For LFA-3 and CEA, the residue number includes the hydrophobic COOH-terminal sequence that will be processed off if a lipid tail is present. Differential exon splicing of genes, giving alternative products, is seen for Igs (membrane and secreted forms), poly IgR (2 forms), class I MHC and related molecules, NCAM (3 forms), MAG (2 forms), and CD8 chain I (2 forms). The letters in brackets after the sequence number indicate species as follows: M, mouse; H, human; R, rat; RB, rabbit; CH, chicken. The human chromosome assignments are from Ref. (51) or other references given in the table. Under the heading *Ig-like disulfide bonds* a Y indicates the presence of conserved disulfide bonds as expected from sequence similarities, and ND denotes "not determined."

CHARACTERISTICS OF THE Ig-FOLD AND SEQUENCE PATTERNS

The two β-sheets of the Ig-fold consist of anti-parallel β-strands containing 5–10 amino acids. Between the sheets a hydrophobic interior is formed from in-pointing hydrophobic amino acids that alternate in the β-strands with out-pointing hydrophilic residues (3, 4). The interaction between the sheets is further stabilized by the conserved disulfide bond. The Ig fold for V- and C-domains is shown in Figure 2, and the core of the fold consists of β-strands A, B, E in one sheet and G, F, C in the other (52). These strands come from the first and last parts of the domain sequence, while in the middle considerable variation in sequence length occurs. V and V-related domains have about 65–75 amino acid residues between the conserved disulfide bond, and there are four β-strands in each β-sheet plus a short β-strand segment across the top of the domain. In C-domains the sequence between the disulfide bond is shorter at 55–60 residues, yielding sheets with 4 and 3 β-strands. In some non-Ig domains (see below) as few as 40 residues exist between the disulfide bond, and in some cases the fold

Table 2 Functions of the Ig-related molecules (References in Table 1)

Molecules and tissue expression	Functions	Recognition within the superfamily
Immunoglobulins: B lymphocytes only	B lymphocyte antigen receptors and in secreted form, antibodies	No, antibodies recognize antigen without involvement of other molecules
T cell receptors: T lymphocytes and thymocytes	T lymphocyte antigen receptors: no known soluble forms	Yes, heterophilic; Tcr binds MHC antigens plus peptide but recognition does not involve Ig-related MHC segments
CD3 chains: T lymphocytes and thymocytes	Part of the Tcr complex; role in signal transduction?	CD3 associates with Tcr but no known recognition of other molecules
MHC antigens: Many cell types, induced by interferon	Present peptides from foreign antigen to the Tcr; some soluble forms	Yes, heterophilic; Tcr interacts with class I and class II MHC antigen
β_2-m associated antigens: Subsets of lymphoid cells	Functions not known	No natural ligands known
T lymphocyte adhesion molecules: CD2, thymocytes and T cells (some macrophages in rat); LFA-3, widespread expression	CD2 of T cells interacts with LFA-3 on other cells in adhesion reactions, Anti-CD2 antibodies can trigger T cell division	Yes, heterophilic; CD2 binds LFA-3
T subset markers: CD4 and CD8 on thymocytes and T subsets, CD4 on macrophages, CD8 on NK cells; CTLA4, activated T cells	CD4 and CD8 appear to control the bias of T cells towards interaction with class I or class II MHC. CTLA4 function unknown	Perhaps heterophilic: CD4 and CD may bind class II and class I MH antigens respectively. CTLA4 unknown
Brain/lymphoid antigens: Thy-1, neurons, fibroblasts, various lymphoid; MRC OX-2, neurons, endothelium, various lymphoid	Anti-Thy-1 antibody triggers mouse T lymphocyte division; MRC OX-2 function unknown	No natural ligands known
Immunoglobulin receptors: PolyIgR, gut and liver epithelium; Fcγ2b/γ1R, macrophages	PolyIgR transports multimeric IgA or IgM across epithelium; macrophage Fcγ2b/γ1R binds aggregated IgG	Yes, heterophilic for both PolyIgr and Fcγ2b/γ1R. First domain of PolyIgR binds IgA
Neural-associated molecules: NCAM, neurons and glia, early embryo; MAG, peripheral and central myelin, some neurons; P_0 peripheral myelin	NCAM mediates adhesion of neural cells. MAG may function in myelination. P_0 constitutes 50% of peripheral myelin protein	Yes, homophilic for NCAM via Ig-related parts and perhaps for P_0. MAG not known
CEA: Epithelial cells and their tumors, early embryos	Tumor marker but function unknown	Natural ligand unknown
Growth factor receptors: PDGFR, widespread on mesenchymal cells; CSF1R, monocyte lineage	Interact with growth factors to trigger cell division and other activities	No, PDGFR and CSF1R not known to react with molecules other than growth factors
Link protein: Basement membrane	Acts as a binding molecule between proteoglycan and hyaluronate chain	No
α_1B-glycoprotein: Found in serum	Function unknown	Natural ligands unknown

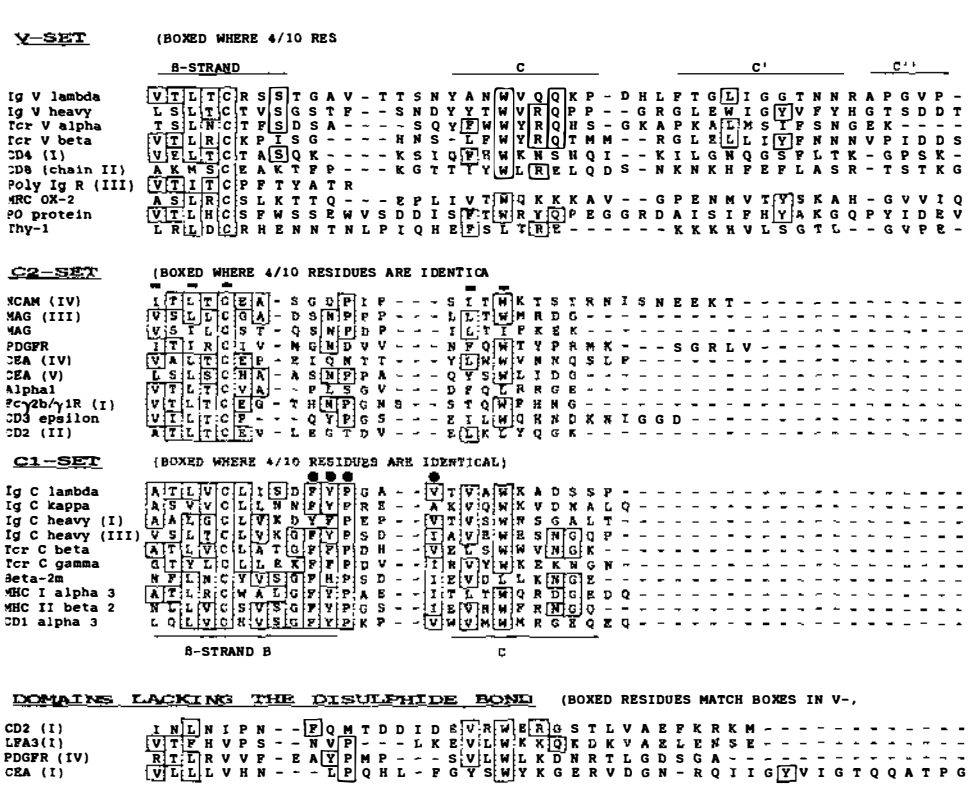
may consist of 3 β-strands in each sheet, plus a short connecting sequence across the top of the β-sheet sandwich.

Interactions between domains in Igs occur between the faces of the β-sheets, while in the interaction between antibody and antigen the contact residues consist of sequences in bends at the end of the V-domains (53). Thus, it seems that Ig-domains can interact with other molecules via any accessible part of their surface.

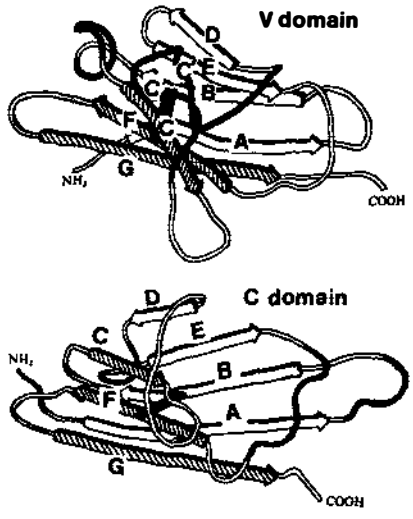
The Ig-related domains of non-Ig molecules are described as being V- or C-like according to whether they are likely to have a pattern of β-strands approximating to a V- or C-domain. A designation of a sequence as being V-like does not indicate sequence variation in the molecule concerned.

Conserved patterns of sequence are seen among the Ig-superfamily domains, and some alignments are shown in Figure 2. The first and last β-strands have been omitted to simplify the data, but this does not imply that conserved patterns are absent in these regions. Also, the whole of the domain is included in statistical analyses. The sequences are grouped into three categories called the V-SET, C1-SET, and C2-SET (54). The V-SET includes antigen receptor V-domains and other sequences likely to have a V-type fold. The extra sequences that form or are thought to form the C′ and C″ β-strands are obvious in Figure 2. The C1-SET includes mostly receptor C-domains and MHC antigen domains, and these are distinguished from the C2-SET in some conserved sequence patterns. However, both C1-SET and C2-SET sequences are likely to be folded as for Ig C-domains.

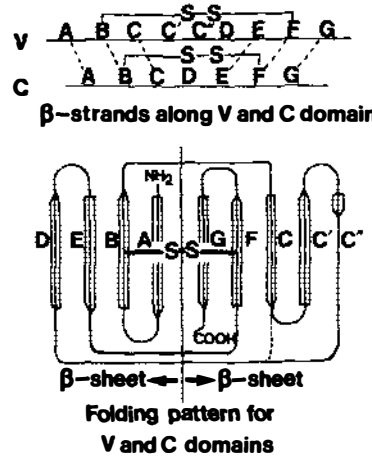
Across all these sequences, identities or conservative amino acid substitutions are seen in β-strands B, C, E, and F; in particular, the alternating hydrophobic residues are evident. In regions outside the β-strands, conserved patterns characterize V-SET and C1-SET sequences (marked in Figure 2). The C2-SET sequences seem somehow in between the V-SET and C1-SET since they are likely to have a C-type fold, but their sequence patterns in the region of β-strands E and F are like those of the V-SET. In Figure 1 all the domains are labelled as V, C1, or C2, and these assignments are mostly clear-cut. However, in some cases the designation is somewhat arbitrary, and among the C-like domains, sequences are placed in the C2-SET unless they clearly show the conserved C1-SET residues. The one exception to this is the C domain of TCR alpha chain which is not typical of sequences in the C1 or C2-SET. In this case, assignment to the C1-SET was made on the basis that the alpha chain is part of the heterodimer that dictates antigen specificity in one category of Tcrs. In general, statistical analysis of sequence similarities (see below) supports the domain assignments shown in Figure 1.



(a)



(c)



(d)

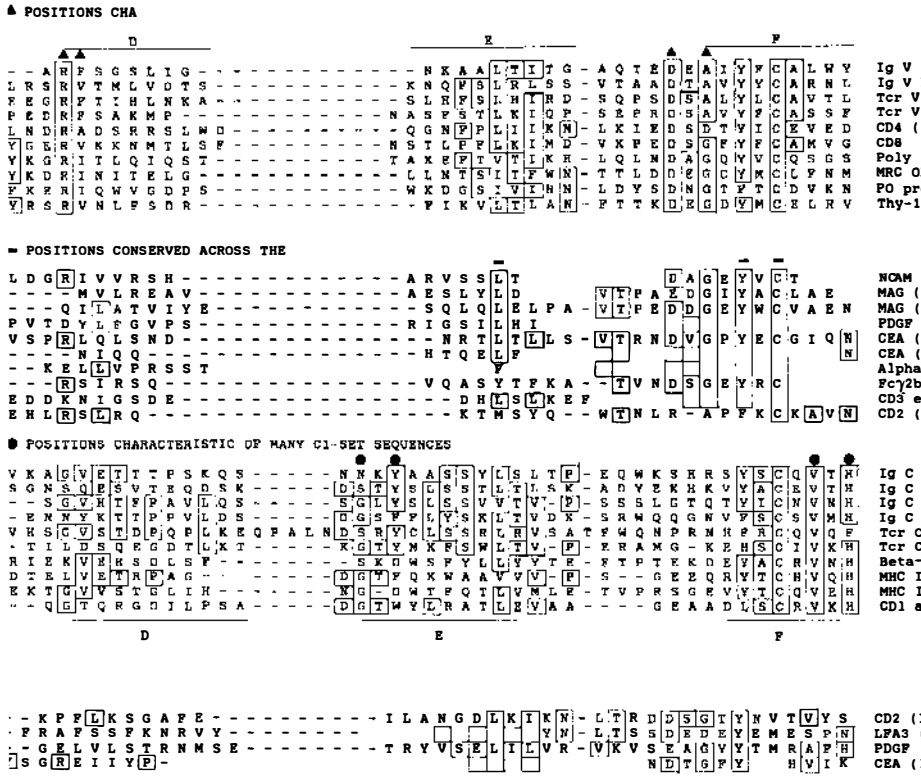


Figure 2 Alignments of Ig-related sequences and Ig-folding patterns. Sequences are aligned by eye and with the use of the ALIGN program (55), and sequence categories are defined in the text. The positions of the β-strands known for Ig V and C domains are indicated above and below the V-SET and C1-SET sequences, respectively. Sequences in regions corresponding to β-strands A and G have been omitted to simplify the data. Some conserved sequence positions are indicated by symbols, and these often involve conservation of amino acids of similar type rather than identities. The sequences are referenced by NBRF protein data base code in square brackets or literature references in (round brackets. Ig V lambda, Mouse [L1MS4E]; Ig VH, human [G1HUNM]; Tcr V alpha, mouse [RWMSAV]; Tcr V beta, human [RWHUVY]; CD4, human [RWHUTA]; CD8 chain II, rat (56); PolyIgR, rabbit [QRRBG]; MRC OX-2, rat [TDRTOX]; P_0, rat (44); Thy-1, rat [TDRT]; NCAM, chicken (40); MAG, rat (42); PDGFR, mouse (46); CEA, human (45); Alpha1V-gp, human [OMHU1B]; Fcγ2b/γ1R, mouse (37); CD3 epsilon, human (14); CD2, rat (24); Ig C lambda, human [L2HU]; Ig C kappa, human [K3HU]; Ig C heavy, human [GHHU]; TcR C beta, human [RWHUCY]; TcR C gamma, mouse [RWMSC1]; β_2-M, human [MGHUB2]; MHC I alpha 3, human [HLHUB2]; MHC II beta 2, human [HLHU3D]; CD1a alpha 3, human (21); LFA-3, human (25).

Below the sequence alignments diagrams of the folding patterns of V and C domains are shown. The folds as determined for an Ig V_L and C_L domain are from Ref. (3), and the labelling of the β-strands along the domain is illustrated in the schematic diagram. The schematic view of the fold at bottom right is adapted from Ref. (4).

Although conserved sequence patterns are seen in Figure 2, no residue is invariant in all Ig-related domains. The conserved disulfide bond was once considered the hallmark of Ig domains, but recently a functional antibody has been described that has a Tyr residue instead of Cys in β-strand F of the V_H domain (57). Also, it can be argued that domains in CD2, LFA3, CD4, CEA, PDGFR, and CSF1R may be Ig-related even though they have no Cys residues in putative β-strands B and F (54, 58). Four of these sequences are shown below the C1-SET in Figure 2. In these and the other cases the Cys residues are replaced by hydrophobic amino acids that would presumably be suitable as inpointing residues that stabilize an Ig-like fold, just as do the other hydrophobic residues in the β-strands.

There is great diversity of sequence in the Ig-related molecules, and the question arises—why is a conserved pattern seen at all? The biological functions (see below) require unique recognition specificities, and these cannot directly be responsible for conserved sequence patterns. The conserved sequences are mostly seen in the β-strands that make up the core of the fold. It can be argued both that the fold itself is selected for, because it is stable to proteolysis, and that this is an essential feature for molecules operating in the extracellular environment (54).

CRITERIA FOR INCLUSION OF MOLECULES IN THE Ig SUPERFAMILY

The initial argument for related domains within the Ig chains was based on sequence similarities. However, an evolutionary relationship between V and C domains was only accepted when similarities in tertiary structures were established. It remains a consensus that the first criterion for an Ig relationship should be the presence of a domain-sized sequence with significant similarity to Ig or Ig-related domains, but in addition there should also be the probability that the sequence shares key structural features of the Ig-fold. If evolutionary selection acts on the basis of domain stability, then a rationale for requiring sequence and structural similarity is evident.

To evaluate sequence similarities, a statistical test must be used, and the ALIGN program of Dayhoff and colleagues (55) is now widely available as part of the NBRF data base package [Protein Identification Resource (1987) Protein Sequence Database, National Biomedical Research Foundation, Washington, DC]. The ALIGN program scores the best match between two sequences on the basis of a scoring matrix derived by determining the frequency of amino acid replacements in equivalent molecules

between widely divergent species. Then the sequences are scrambled and rescored a number of times (e.g. 150) to yield a mean best random score and standard deviation (SD). The score for the real sequences is expressed as the number of SD units away from the random mean score. Assuming a normal distribution and no effect of sequence selection, scores of 3.1, 4.3, and 5.2 SD units indicate chance probabilities of 10^{-3}, 10^{-5}, and 10^{-7}, respectively.

In choosing sequences for ALIGN analyses, a putative domain is defined by taking sequence within positions that are 20 residues before and after Cys residues that might approximate to the conserved Ig disulfide bond. If Cys residues are absent, then possible replacements are identified and used to define the domain segment. The selection of Cys residues to define the domain carries the possible problem that this will bias ALIGN scores since matches between Cys residues carry a high value in the Dayhoff scoring matrix. To check this effect, 11 segments from membrane molecules that are not Ig-related were chosen on the basis of the presence of a suitable pair of Cys residues. These were scored against Ig-related domains (54). From 682 scores, a mean and standard deviation of 0.6 ± 1 SD units was obtained. In the controls there were three scores of >4 (4.1, 4.2, 4.2) and 13 of $>3<4$ SD units.

In testing a new sequence, ALIGN scores should be determined with as many distinct sequences as possible from the V-SET, C1-SET, and C2-SET. This overcomes the problem that the ALIGN program scores throughout the sequence and takes no account of the conserved sequence patterns that have great significance when assessing sequence similarities by eye (Figure 2). By chance a reasonable score might result in one comparison, but repeated good scores should indicate that a test sequence contains a conserved Ig-related pattern of sequence, since this is the only common denominator between the sequences in Figure 2 (54, 58).

In Table 3 some ALIGN scores are shown for β_2-m, Thy-1, NCAM, and LCA (control) against sequences from the V-SET and C1-SET. β_2-m scores well with the sequences from the C1-SET, but a significant relationship is not seen with V-SET sequences, and THY-1 shows good scores in the opposite direction. NCAM gives good scores with both C1-SET and V-SET sequences because, although its length matches with C-domains, it has some of the conserved sequence patterns of the V-SET. The control sequence from LCA gives no good scores even though the sequence has two Cys residues that can match with those in the Ig-domains and it also has a Trp that can match the conserved Trp of β-strand C in Figure 2.

Some molecules that have been claimed to be Ig-related fail the ALIGN test against the sets of domains (54, 58). These include the adenoviral E3 glycoprotein (59) and the CD5 antigen (60). Also, the enzyme superoxide

Table 3 ALIGN scores for comparisons of Ig related sequences

V or V-related	β_2-m	Thy-1	NCAM (III)	LCA	C or C-related	β_2-m	Thy-1	NCAM (III)	LCA
Ig lambda	−0.9	7.4	3.3	−0.6	Ig lambda	5.6	1.4	4.7	−1.0
Ig kappa	1.7	3.7	5.4	−0.1	Ig kappa	6.0	1.3	4.0	−1.5
Ig heavy	1.1	3.9	3.9	1.5	Ig CH1	4.0	3.0	4.1	0.7
Tcr beta	1.8	3.3	4.6	−0.2	Ig CH2	2.4	2.9	3.8	0
Tcr alpha	2.1	2.3	4.4	−1.5	Ig CH3	6.3	3.1	3.7	0.7
Tcr gamma	−0.2	1.6	3.9	1.4	Tcr beta	4.4	2.3	3.0	2.2
CD8 (chain I)	2.6	4.5	4.7	−0.3	Tcr alpha	2.1	−0.3	1.7	−0.7
CD4 (I)	2.4	2.5	5.5	−0.1	Tcr gamma	1.9	0.8	3.6	−0.6
PolyIgR (I)	1.5	5.7	2.7	0	MHC I α3	8.2	2.2	2.9	1.3
PolyIgR (III)	1.7	5.8	4.3	0.9	MHC II α2	11.2	3.7	4.9	1.2
MRC OX-2 (I)	−1.1	5.0	5.3	−0.6	MHC II β2	11.3	2.4	4.3	1.6
P_0 Protein	0.7	3.5	6.0	−0.4	CD1 α3	9.1	1.3	5.4	1.2

Domains were defined from a position 20 residues before the first Cys to 20 residues after the second Cys of a putati Ig-like disulfide bond. The leucocyte common antigen (LCA) sequence is a control and includes residues 88-189 fro the partial rat LCA sequence (54). All other sequences are referenced in Figure 2 except the following with NBR data base code given in square brackets or reference in parentheses: Ig kappa [K1HURY], Tcr gamma V [RWMSV CD8 chain I (56), Tcr alpha C [RWHUAC], MHC class I [HLHU12], MHC class II beta [HLHU3D]. The ALIG program (55) was run with a bias of 6 and a break penalty of 6, and 150 random runs were performed.

dismutase which has a fold like an Ig-domain (61) shows no sequence similarity and thus is not regarded as being in the Ig-superfamily (58). Other domains with no statistically significant relationship are the α_1 and α_2 domains of MHC class I antigen and the α_1 and β_1 domains of MHC class II antigens. A possible Ig-related segment in the sequence of HIV glycoprotein gp110 does not include sequence that can be matched with a full domain (62). However, there are matches with Ig C-domains over 44 residues that give ALIGN scores of 7–9 SD. This is sufficiently high to raise the possibility that the viral sequence has been captured from Igs without the maintenance of the full domain. A sequence similarity might remain if it were selected for unknown reasons or if the capture of a piece of Ig-related sequence by virus was a recent event.

Structural proof for an Ig-related domain can only convincingly be established by tertiary structure determined by X-ray crystallography. Thus far, this is only reported for Ig V and C domains and for the β_2-m and α3 domains of MHC class I antigens which have structures exactly like Ig C_H3 domains (6, 7a). Unambiguous but limited evidence comes from the determination of disulfide bonds. Thus far in all cases where this has been determined the bonding pattern is in accord with the Ig-fold (Table 1). Circular dichroism can convincingly establish the presence of pure β-structure without α-helix, and this has been shown for Thy-1 (63).

Secondary structure prediction can be used, but this remains imprecise (64) and in our view should be used as a test to see whether a domain assignment indicated by sequence similarities might be improbable on structural grounds. Finally, the exon pattern in the genes can support domain assignments, and this is discussed below.

OVERALL MOLECULAR CHARACTERISTICS

The Ig superfamily is notable because no extracellular sequence includes an enzymatic activity or segments from more than one protein superfamily. This generalization does not apply to intracellular parts since the PDGFR and CSF1R molecules have cytoplasmic domains that have tyrosine kinase enzymatic activities and are related in sequence to other tyrosine kinase domains (46). Mixing of segments from different superfamilies is commonly seen in other cell or immune-related surface molecules. For example, in the complement proteins, repeats of a disulfide-linked domain of about 60 amino acids can be found together with serine protease domains, plus in some cases epidermal growth factor-like segments (65).

The Ig-related molecules commonly form dimers (Figure 1), and binding between domains can be homophilic, as seen between the C_H3 domains of Igs, or heterophilic, as in interactions between V domains and Ig C_H1 and C_L domains. Stable dimers are often disulfide-linked, but this is not the case for the MHC antigens and β_2-m associated antigens. Metastable interactions can also occur, as is seen in the Tcr complex where the α and β chains in the disulfide-linked dimer associate with the CD3 ε, γ, δ, and ξ chains (13). (The ξ chain is not shown in Figure 1 since the sequence is unknown.) This precedent raises the possibility that even weaker interactions that cannot be detected by conventional techniques might occur between Ig-related molecules at the cell surface during functional responses.

Carbohydrate structures can be dominant features in some Ig-superfamily molecules, and in CEA up to six possible N-linked sites have been observed on one Ig-related domain (Figure 1). Between tissues the same protein can be differentially glycosylated, and all the complex N-linked structures of Thy-1 differ between brain and thymus (66). Also on NCAM differences occur in glycosylation between fetal and adult forms (40). The fetal forms show extensive polysialation, and this is thought to modulate the adhesive potential of NCAM molecules.

The transmembrane sequences and cytoplasmic domains of Ig-related molecules show great diversity. Most molecules have a hydrophobic transmembrane protein sequence, but for Thy-1 (31), one form of NCAM (40, 41), and of LFA-3 (25, 84) and Qa-2 antigen (20, 20a) membrane

attachment via a glycophospholipid anchor has been established. Such an attachment may also be the case for CEA, since this molecule has hydrophobic residues at its COOH-terminus, but there are no basic residues (as is to be expected) on the cytoplasmic side of an authentic transmembrane sequence (45). The COOH-terminal sequence of CEA is like that of Thy-1, which is absent from the mature molecule and is presumed to be cleaved off when the glycophospholipid tail is attached (31, 67). A similar sequence is also predicted from NCAM cDNA clones that are thought to encode the form of NCAM that has a glycophospholipid anchor (40, 41; reviewed in 85).

Ig-superfamily molecules with predicted transmembrane sequences show only one such sequence per chain, and in most cases these sequences show no amino acids with amide or charged residues. However, in all the CD3 chains there is one acidic residue in the midst of the hydrophobic domain (13, 14), and in Tcr chains basic residues are found in similar positions (11). In Tcr γ and X chains two basic residues and an Asn are found in the 22 residues most likely to cross the bilayer (12). The charged residues in the hydrophobic domains of the Tcr complex may stabilize interactions between CD3 and Tcr chains, or alternatively, they may be involved in signal transduction. Cytoplasmic domains of Ig-superfamily sequences are mostly completely unrelated, and they vary in length from 3 amino acids for IgM (10) to 543 residues for PDGFR (46). Their roles in general are a mystery, although a function in intracellular traffic is known for the cytoplasmic domain of PolyIgR (68). The unknown aspects of signal transduction are illustrated by considering Thy-1, IgM, and PDGFR, all of which can act as targets in the triggering of mitogenesis in various circumstances (Table 2). Thy-1 has no transmembrane protein sequence; IgM has such a sequence but has almost no cytoplasmic domain; and PDGFR has both a transmembrane sequence and a very large cytoplasmic domain with a tyrosine kinase activity.

GENETIC LINKAGE AND EXON STRUCTURE

It can be seen from Table 1 that loci for Ig-superfamily molecules are spread across the chromosomes. Genetic separation is common for loci that are coregulated and whose products are found in molecular complexes, but four notable cases of linkage are seen.

Firstly, in all cases the loci for V, J, and D segments of antigen receptors are linked to the C domain genes to which they can be rearranged (11, 69). Presumably a relatively close *cis* orientation is essential for the gene rearrangement mechanism.

Secondly, all the polymorphic MHC antigens are found in one large

chromosomal region, and this has been found in all species that have been investigated (16). Also in the mouse, the Qa and T1 products that are very closely related to MHC class I chains are coded by loci linked to the MHC. In contrast the CD1 antigens of humans are not MHC linked (21). The H chains of CD1 antigens are associated with β_2-m, but their sequences are much less closely related to class I MHC than are the Qa or T1 antigen sequences. The linkage of the polymorphic MHC antigens may be due to selective advantages that result if a set of polymorphic variants favorable for antigen presentation are inherited on one chromosome.

Thirdly, the two chains of the CD8 antigens are closely linked to each other and to the V_κ loci in human and mouse (27). No functional reason for this linkage is obvious. The CD8 antigens seem particularly closely related to the V-domains of antigen receptors (56). It may be that the V_κ and CD8 genes have remained together in a region of chromosome where extensive duplication occurred of genes for heterodimer structures related to immunity.

Finally, the loci for the CD3 chains, NCAM, and Thy-1 are all linked on the q23 band of chromosome 11, and these loci are also linked in the mouse (15). The CD3 δ and γ chain genes are arranged in reverse orientation and are separated by only 1.4 kb of sequence (70, 71). These genes are within 400 kb of the CD3 ε gene. The distance between the CD3 genes and NCAM and Thy-1 genes is not known.

There is no obvious functional reason for the chromosome 11 linkage group. Thy-1 and NCAM have in common their expression in neural tissues and also the fact that both molecules can be attached to the membrane by a glycophospholipid tail. The CD3 chains are found only on cells in the T lymphocyte lineage, but among Ig-related sequences the CD3 domains seem to match best with NCAM domains (14). Extensive gene duplication and divergence of Ig-related sequences may have occurred in the region now constituted by band q23 of chromosome 11, and Ig-related genes now found in this position may be those that have not moved to other chromosomes in evolution.

The most characteristic aspect of gene structure for Ig-superfamily molecules is that the majority of the domain sequence is often encoded within one exon. This is true for all domains in Igs and Tcrs and for 20 out of 30 domains in nonreceptor genes whose structures are shown in Figure 3A. However, a number of examples now exist of introns found between sequences coding for the Cys residues of the conserved disulfide bonds. The positions of these introns in relation to the postulated folding pattern of the domains are shown in Figure 3B. One other exception to the one domain: one exon rule is seen in the PolyIgR where the domains (II) and (III) are both encoded in one exon (76). In one PolyIgR mRNA form, this

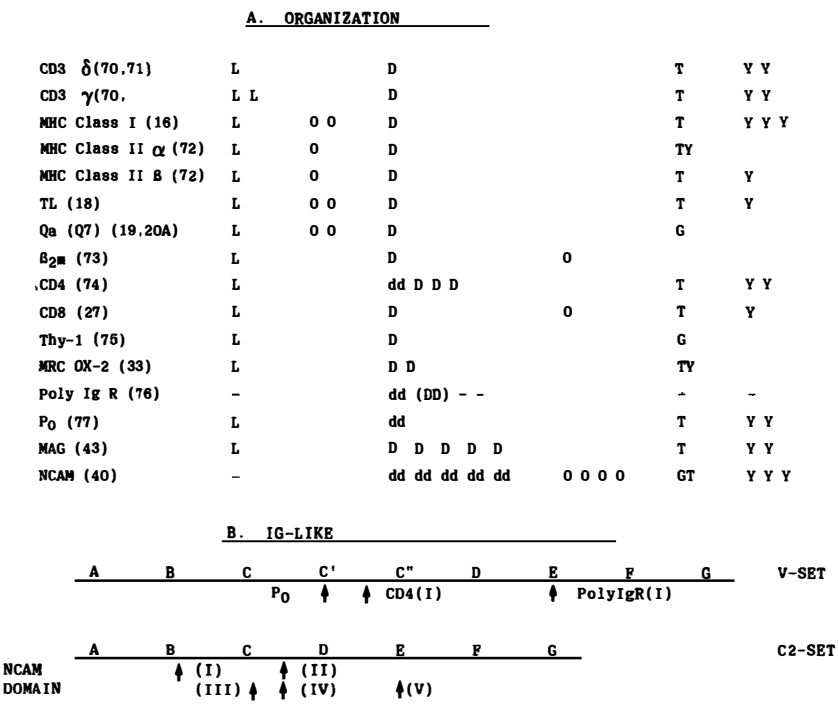


Figure 3 Organization of coding exons from genes for Ig-superfamily molecules excepting antigen receptor genes. *A: Organization of exons.* References to each gene structure are given after the molecule name. The letters showing exon organization are coded: L, exon for leader sequence; O, an exon encoding extracellular sequence that is not Ig-related; D, an exon encoding an Ig-related domain with no introns between codons for the conserved cys residues; dd, two exons encoding an Ig-related domain with an intron between codons for the conserved cys residues; (DD), an exon encoding two Ig-related domains; G, an exon for a hydrophobic sequence that is or may be cleaved from the protein when lipid is attached at the COOH-terminus; T, an exon for a transmembrane sequence; TY, an exon for a transmembrane plus cytoplasmic sequence; Y, an exon for cytoplasmic region sequence; — indicates exon structure was not established. *Comments:* the second D in CD4 applies to the second domain in Figure 1 which will not form a standard Ig-fold but has some sequence similarities to Ig-domains (28). The exon marked G in Thy-1 is established to encode a sequence that is cleaved in processing, and this is likely to be so for NCAM in which the shortest mRNA form has the exon marked G but not the TYYY exons. The longer NCAM mRNA forms splice out the G exon and include two possible combinations of the other exons. No attempt is made in this figure to show the alternative splicing events that can be seen, and noncoding exons are not shown. *B. Ig-like domains encoded by two exons.* The arrows indicate the positions of introns within the Ig-like domains. The letters A, B, C, C′, C″, D, E, F, G indicate the positions of putative β-strands determined by sequence similarity as indicated in Figure 2. (Data for PolyIgR from J. Harris and K. Mostov, personal communication.)

exon is spliced out and the mRNA is translated to give the small form of rabbit PolyIgR (76).

FUNCTIONAL ASPECTS

Known functions of Ig-superfamily molecules are given in Table 2; in all cases they include adhesion functions or binding functions that trigger a subsequent event at the cell surface. A key functional feature is that homophilic or heterophilic binding occurs between Ig-related molecules; this is often between molecules on opposed membrane surfaces. Most of the Ig-superfamily molecules function at cell surfaces; the exceptions are antibody, the link protein, and α_1B-glycoprotein. The functions of antibody involve interactions with antigen and then with effector molecules via the Fc regions to trigger subsequent events. The link protein could be considered an adhesion molecule for binding together hyaluronic acid and proteoglycan; the function of α_1B-gp is unknown. It remains possible that α_1B-gp is a cleaved product from a cell surface receptor in much the same way that secretory component is a product derived by proteolysis from PolyIgR.

In all functions in which Ig-related domains are known to be involved, the domain can be considered as providing a stable platform for the display of specific determinants for recognition reactions on the faces of β-sheets or at the bends between β-strands (9, 52). The determinants involved are likely to be mostly protein in nature, but there is always the possibility that the chemical entities recognized are carbohydrate structures.

EVOLUTION

It is commonly accepted that all the Ig-related sequences shown in Figure 1 will have been derived by gene duplication and divergence from one primordial domain. Preceding this there may have been a half domain structure (78) that formed a homodimer with a structure like the V-domain (79). The half domain fold is postulated to be like that of β-strands ABCC′ or GFED in Figure 2, and to associate to form a homodimer in the same way that the ABCC′ and GFED β-strand loops associate in the V-domains (79). The half-domain idea seems to be supported by the existence of genes that have introns in the midst of sequence coding for residues between the conserved disulfide bonds, and it is notable that in a number of cases these come in a position that would roughly demarcate a half domain as

proposed by McLachlan (79) (Figure 3B). Genes with the intra-domain introns may be thought more directly derived from ancient genes than those lacking this feature, but this is not a reliable conclusion since the probability of intron loss is unknown. Equally unknown is the probability of intron acquisition.

The possibility that the V-like fold is the most ancient is supported by the fact that Thy-1 and P_0 are the only single domain molecules thought to exist without association with other chains, and both of these are in the V-SET. P_0 also has an intron in the middle of the domain. If the V-fold were the most ancient type, then the C-like domain would be derived by loss of sequence from the middle of a V-type fold. A possible lineage would be V-type to C2-type to C1-type. However, one could argue alternatively that the C2-type is the most primitive and that both V-type and C1-type sequences were derived from this. A start from a C-type fold would not fit with the idea that the primordial domain is derived from a homodimer of half-domain structures. The C1-type seems an unlikely candidate for the most primitive domain since thus far C1-set sequences have been seen only in structures associated with the immune system and mostly with immune recognition. Also no intradomain introns have been seen in the C1-set sequences.

It is difficult to suggest detailed evolutionary trees for the Ig-related molecules. Some molecules can be grouped because they show greater than average similarity within the superfamily. Some such groupings might be: (IgV, TcrV, CD8); (IgC, TcrC, MHC, CD1); (CD3 ε, δ, γ); (CD2, LFA-3); (MAG, NCAM); (PDGFR, CSF1R). The difficulty in trying to connect up these groups and the other molecules is that the Ig-related molecules appear to be diverging very rapidly, as assessed by the percentage of identity for equivalent chains between species. This is as low as 42% for the V-like domain of CD8 chain I between rodents and humans (27) which suggests that when new genes are created they might rapidly diverge to a level where only the basic conserved patterns remain. Intermediate stages in evolution may be hard to detect in the contemporary sequences. In addition, multichain sequences have probably arisen repeatedly, and perhaps sometimes single domain structures have been rederived from multidomain forms. Examples of differences in evolution of multidomain structures can be seen with PDGFR and CSF1R in one case and CEA in another. PDGFR and CSF1R have similar 5 domain patterns (46); these were probably derived from an immediate common precursor with the same domain pattern. In contrast, CEA has seven domains among which the last 6 appear to have been derived by a recent double duplication of a two domain segment (45). This is likely because within domains II, IV, VI and III, V, VII there is about 70% identity, while the level of identity

between these two sets is about 25%. The nature of the immediate precursor of CEA is thus unpredictable. In the case of rabbit PolyIgR, reduction in size can be seen in the alternative splicing that produces a variant lacking domains II and III of the structure as shown in Figure 1 (76).

In terms of functional evolution the phenomenon of interactions within the superfamily (Table 2) is strongly suggestive that the primordial function involved a single domain interacting with itself probably between opposed membranes. Such a function has been suggested for the P_0 myelin protein (44). If correct, this may be an interesting model for the function of the primordial domain. Heterophilic receptor pairs presumably evolved from a homophilic interaction system, and highly specific interactions become possible with heterophilic recognition between different cell types. It seems likely that the first functions were related to adhesion or triggering at cell surfaces to control the behavior of cells within a multicellular organism. The Ig-related molecules that function in neural tissues may be mediating functions of the primitive type. A scheme for the evolution of heterophilic interactions from a homophilic adhesion system between cells is illustrated in Figure 4A.

Another possibility is that heterophilic pairs giving recognition between cells may have originated from chains that interacted on one cell to form a heterodimer (80). This is illustrated in Figure 4B starting from a homodimer diverging to a heterodimer and then to modified forms of the heterodimer chains interacting between cells. The LFA-3: CD2 adhesion molecules may have evolved in this way from a chain that formed a homodimer and contained one Ig-related domain with a disulfide bond and one without (NH_2 terminal). From this type of origin the result could be a heterophilic pair, each member of which is more closely related to the other than to other members of the superfamily (as is the case for LFA-3 and CD2) (25).

If it is accepted that involvement in cell recognition was the primary role of the Ig-superfamily, the question then arises—how might the vertebrate immune system be derived from this? One possible functional antecedent is the phenomenon of programmed cell death. In the invertebrate *Caenorhabditis elegans*, 25% of developing neural cells die in a predictable way, and this commonly involves an apparent differentiation to cell death followed by phagocytosis (81, 82). In some cases, however, one cell appears to kill another, and this function is of the type that might be turned outwards to produce an immune system. Figure 4C suggests that Ig-related molecules control the specificity of a primitive natural killer cell and that modification of this specificity to include determinants of a common pathogen resulted in a killer system to eliminate pathogen-infected cells.

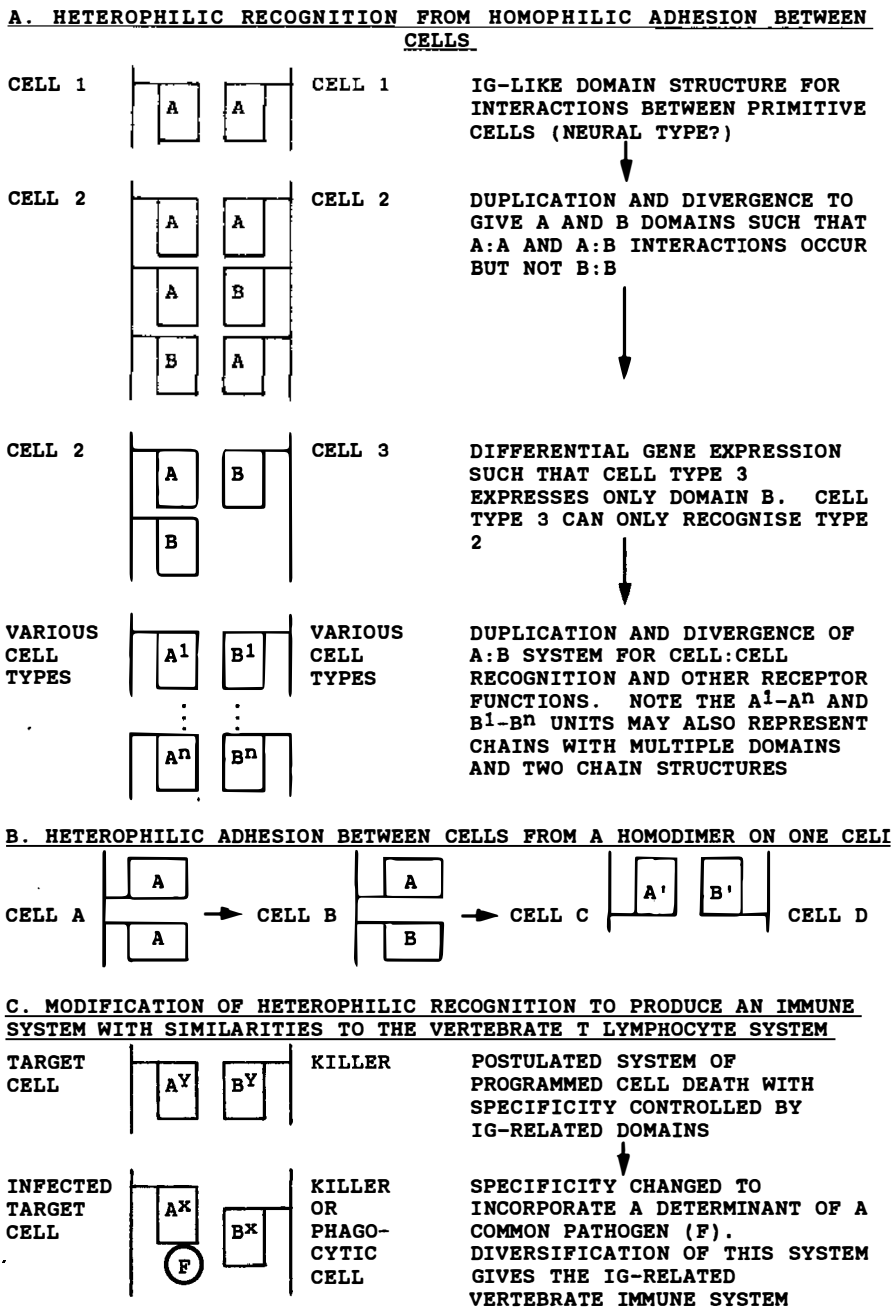


Figure 4 A possible scheme for structural and functional evolution in the Ig superfamily.

Duplication and divergence of this system could lead to an immune system with the properties of the vertebrate T lymphocyte system. The B lymphocyte arm of the immune response may have evolved from this as a recognition system freed from the constraint of interaction with MHC antigens. The possibility that T cell immunity is the more primitive is supported by the finding that the T cell CD8 antigen chain II has a sequence that is very like receptor J pieces without an intron or other intervening genomic sequence between the main V-like exon and the J-related piece (56). The CD8 heterodimer may be similar in its V-like domains to the structure that gave rise to the antigen receptor heterodimers.

Structural and functional evolution would be greatly illuminated if invertebrate Ig-related molecules were identified. Thus far only one invertebrate sequence that might be Ig-related is known, and this is a glycoprotein of 84 amino acids with a glycophospholipid tail that was identified in a search for Thy-1-like molecules from squid neural tissue (83). The squid glycoprotein has some interesting sequence similarities to Thy-1 and Ig V-domains but does not have a standard domain pattern. Thus, it cannot at this stage be added to the Ig-superfamily with the level of confidence that was applied for molecules in Figure 1. Are Ig-related structures common in invertebrate neural cells and do invertebrate immune systems use Ig-related molecules at all? Answers to questions of this type are needed to further elucidate the structural and functional evolution of the Ig-superfamily.

ACKNOWLEDGMENTS

We thank Denise Roby for help with the manuscript and Catherine Lee for photography.

Literature Cited

1. Hill, R. L., Delaney, R., Fellows, R. E., Lebovitz, H. E. 1966. The evolutionary origins of the immunoglobulins. *Proc. Natl. Acad. Sci. USA* 56: 1762–69
2. Edelman, G. M. 1970. The covalent structure of a human IgG. XI. Functional implications. *Biochemistry* 9: 3197–3205
3. Edmundson, A. B., Ely, K. R., Abola, E. E., Schiffer, M., Panagiotopoulos, N. 1975. Rotational allomerism and divergent evolution of domains in Ig light chains. *Biochemistry* 14: 3953–61
4. Amzel, L. M., Poljak, R. J. 1979. Three-dimensional structure of Igs. *Ann. Rev. Biochem.* 48: 961–97
5. Peterson, P. A., Cunningham, B. A., Berggard, I., Edelman, G. M. 1972. β_2 microglobulin—a free Ig domain. *Proc. Natl. Acad. Sci. USA* 69: 1697–1701
6. Becker, J. W., Reeke, G. N. Jr. 1985. Three-dimensional structure of β_2-microglobulin. *Proc. Natl. Acad. Sci. USA* 82: 4225–29
7. Orr, H. T., Lancet, D., Robb, R. J., Lopez de Castro, J. A., Strominger, J. L. 1979. The heavy chain of human histocompatibility antigen HLA-B7 contains an Ig-like region. *Nature* 282: 266–70

7a. Bjorkman, P. J., Saper, M. A., Samraoui, B., Bennett, W. S., Strominger, J. L., Wiley, D. C. 1987. Structure of the

human class I histocompatibility antigen HLA-A2. *Nature*. In press
8. Williams, A. F., Gagnon, J. 1982. Neuronal cell Thy-1 glycoprotein: homology with Ig. *Science* 216: 696–703
9. Williams, A. F. 1982. Surface molecules and cell interactions. *J. Theoret. Biol.* 98: 221–34
10. Kehry, M., Ewald, S., Douglas, R., Sibley, C., Raschke, W., Fambrough, D., Hood, L. 1980. The Ig μ chains of membrane-bound and secreted IgM molecules differ in their C-terminal segments. *Cell* 21: 393–406
11. Kronenberg, M., Siu, G., Hood, L. E., Shastri, N. 1986. The molecular genetics of the T-cell antigen receptor and T-cell antigen recognition. *Ann. Rev. Immunol.* 4: 529–91
12. Chien, Y.-H., Iwashima, M., Kaplan, K. B., Elliott, J. F., Davis, M. M. 1987. A new T-cell receptor gene located within the alpha locus and expressed early in T-cell differentiation. *Nature* 327: 677–82
13. Weiss, A., Imboden, J., Hardy, K., Manger, B., Terhorst, C., Stobo, J. 1986. The role of the T3/antigen receptor complex in T-cell activation. *Ann. Rev. Immunol.* 4: 593–619
14. Gold, D. P., Clevers, H., Alarcon, B., Dunlap, S., Novotny, J., Williams, A. F., Terhorst, C. 1987. Evolutionary relationship between the T3 chains of the T cell receptor complex and the Ig supergene family. *Proc. Natl. Acad. Sci. USA* 84: 7649–53
15. Gold, D. P., van Dongen, J. J. M., Morton, C. C., Bruns, G. A. P., van den Elsen, P., Guerts van Kessel, A. H. M., Terhorst, C. 1987. The gene encoding the ε subunit of the T3/T-cell receptor complex maps to chromosome 11 in humans and to chromosome 9 in mice. *Proc. Natl. Acad. Sci. USA* 84: 1664–68
16. Lew, A. M., Lillehoj, E. P., Cowan, E. P., Maloy, W. L., Van Schravendijk, M. R., Coligan, J. E. 1986. Class I genes and molecules: an update. *Immunology* 57: 3–18
17. Kaufman, J. F., Auffray, C., Korman, A. J., Shackelford, D. A., Strominger, J. 1984. The class II molecules of the human and murine major histocompatibility complex. *Cell* 36: 1–13
18. Fisher, D. A., Hunt, S. W. III, Hood, L. 1985. Structure of a gene encoding a murine thymus leukaemia antigen, and organization of T1a genes in the Balb/c mouse. *J. Exp. Med.* 162: 528–45
19. Devlin, J. J., Weiss, E. H., Paulson, M., Flavell, R. A. 1985. Duplicated gene pairs and alleles of Class I genes in the Qa 2 region of the murine major histocompatibility complex: a comparison. *EMBO J.* 4: 3203–7
20. Stiernberg, J., Low, M. G., Flaherty, L., Kincade, P. W. 1987. Removal of lymphocyte surface molecules with phosphatidylinositol-specific phospholipase C: Effects on mitogen responses and evidence that ThB and certain Qa antigens are membrane-anchored via phosphatidylinositol. *J. Immunol.* 138: 3877–84
20a. Stroynowski, I., Soloski, M., Low, M. G., Hood, L. 1987. A single gene encodes soluble and membrane-bound forms of the Qa-2 antigen: anchoring of the product by a phospholipid tail. *Cell* 50: 759–768
21. Martin, L. H., Calabi, F., Milstein, C. 1986. Isolation of CD1 genes: a family of major histocompatibility complex-related differentiation antigens. *Proc. Natl. Acad. Sci. USA* 83: 9154–58
22. Selvaraj, P., Plunkett, M. L., Dustin, M., Sanders, M. E., Shaw, S., Springer, T. A. 1987. The T lymphocyte glycoprotein CD2 binds the cell surface ligand LFA-3. *Nature* 326: 400–403
23. Brown, M. H., Gorman, P. A., Sewell, W. A., Spur, N. K., Sheer, D., Crumpton, M. J. 1987. The gene coding for the human T lymphocyte CD2 antigen is located on chromosome 1p. *Hum. Genet.* 76: 191–95
24. Williams, A. F., Barclay, A. N., Clark, S. J., Paterson, D. J., Willis, A. C. 1987. Similarities in sequences and cellular expression between rat CD2 and CD4 antigens. *J. Exp. Med.* 165: 368–80
25. Seed, B. 1987. An LFA-3 cDNA encodes a phospholipid-linked membrane protein homologous to its receptor, CD2. *Nature* 329: 840–42
26. Barbosa, J. A., Mentzer, S. J., Kamark, M. E., Hart, J., Strominger, J. L., Biro, P. A., Burakoff, S. J. 1985. Somatic cell hybrid analysis of human lymphocyte function associated antigen (LFA-3): Gene mapping and role in CTL-target cell interactions. *I.C.S.U. Short Rep.* 2: 107–108
27. Littman, D. R. 1987. The structure of the CD4 and CD8 genes. *Ann. Rev. Immunol.* 5: 561–84
28. Clark, S. J., Jefferies, W. A., Barclay, A. N., Gagnon, J., William, A. F. 1987. Peptide and nucleotide sequences of rat CD4 (W3/25) antigen: evidence for derivation from a structure with four immunoglobulin-related domains. *Proc. Natl. Acad. Sci. USA* 84: 1649–53
29. Johnson, P. 1987. A human CD8 (LYT-3) homologue exists: Genomic sequence

and expression. *Immunogenetics* 26: 174–77
30. Brunet, J.-F., Denizot, F., Luciani, M.-F., Roux-Dosseto, M., Suzan, M., Mattei, M.-G., Goldstein, P. 1987. CTLA-4, a new member of the immunoglobulin superfamily. *Nature* 328: 267–70
31. Tse, A. G. D., Barclay, A. N., Watts, A., Williams, A. F. 1985. A glycophospholipid tail at the carboxyl-terminus of the Thy-1 glycoprotein of neurons and thymocytes. *Science* 230: 1003–8
32. Gunter, K. C., Germain, R. N., Kroczek, R. A., Saito, T., Yokoyama, W. M., Chan, C., Weiss, A., Shevach, E. M. 1987. Thy-1-mediated T-cell activation requires co-expression of CD3/Ti complex. *Nature* 326: 505–7
33. McCaughan, G. W., Clark, M. J., Barclay, A. N. 1987. Characterization of the human homologue of the rat MRC OX-2 membrane glycoprotein. *Immunogenetics* 25: 329–35
34. Mostov, K. E., Friedlander, M., Blobel, G. 1984. The receptor for transepithelial transport of IgA and IgM contains multiple immunoglobulin-like domains. *Nature* 308: 37–43
35. Eiffert, H., Quentin, E., Decker, J., Hillemeir, S., Hufschmidt, M., Klingmuller, D., Weber, M. H., Hilschmann, N. 1984. Die primärstruktur der menschlichen freien sekretkomponente und die anordnung der disulfidbrücken. *Hoppe-Seyler's Zeitschrift für Physiologische Chemie* 365: 1489–95
36. Frutiger, S., Hughes, G. J., Hanly, W. C., Kingzette, M., Jaton, J.-C. 1986. The amino-terminal domain of rabbit secretory component is responsible for noncovalent binding to Immunoglobulin A dimers. *J. Biol. Chem.* 261: 16673–81
37. Lewis, V. A., Koch, T., Plutner, H., Mellman, I. 1986. A complementary DNA clone for a macrophage-lymphocyte Fc receptor. *Nature* 324: 372–75
38. Ravetch, J. V., Luster, A. D., Weinschank, R., Kochan, J., Pavlovec, A., Portnoy, D. A., Hulmes, J., Pan, Y.-C. E., Unkeless, J. C. 1986. Structural heterogeneity and functional domains of murine immunoglobulin G Fc receptors. *Science* 234: 718–25
39. Hibbs, M. L., Classon, B. J., Walker, I. D., McKenzie, I. F. C., Hogarth, P. M. 1987. The structure of the murine Fc receptor for IgG. *J. Immunol.* In press
40. Cunningham, B. A., Hemperley, J. J., Murray, B. A., Prediger, E. A., Brackenbury, R., Edelman, G. M. 1987. Neural cell adhesion molecule: structure, Ig-like domains, cell surface modulation and alternative RNA splicing. *Science* 236: 799–806
41. Barthels, D., Santoni, M.-J., Wille, W., Ruppert, C., Chaix, J.-C., Hirsch, M.-R., Fontecilla-Camps, J. C., Goridis, C. 1987. Isolation and nucleotide sequence of mouse NCAM cDNA that codes for a M_r 79,000 polypeptide without a membrane-spanning region. *EMBO J.* 6: 907–14
42. Lai, C., Brow, M. A., Nave, K.-A., Noronha, A. B., Quarles, R. H., Bloom, F. E., Milner, R. J., Sutcliffe, J. G. 1987. Two forms of 1B235/myelin associated glycoprotein, a cell adhesion molecule for postnatal neural development, are produced by alternative splicing. *Proc. Natl. Acad. Sci. USA* 84: 4337–41
43. Salzer, J. L., Holmes, W. P., Colman, D. R. 1987. The amino acid sequences of the Myelin-associated glycoproteins: homology to the Ig-superfamily. *J. Cell. Biol.* 104: 957–65
44. Lemke, G., Axel, R. 1985. Isolation and sequence of a cDNA encoding the major structural protein of peripheral myelin. *Cell* 40: 501–508
45. Oikawa, S., Imajo, S., Noguchi, T., Kosaki, G., Nakazato, H. 1987. The carcinoembryonic antigen (CEA) contains multiple immunoglobulin-like domains. *Biochem. Biophys. Res. Commun.* 144: 634–42
46. Yarden, Y., Escobedo, J. A., Kuang, W.-J., Yang-Feng, T. L., Daniel, T. O., Tremble, P. M., Chen, E. Y., Ando, M. E., Harkins, R. N., Francke, U., Fried, V. A., Ullrich, A., Williams, L. T. 1986. Structure of the receptor for platelet-derived growth factor helps define a family of closely related growth factor receptors. *Nature* 323: 226–32
47. Sherr, C. J., Rettenmier, C. W., Sacca, R., Roussel, M. F., Look, A. T., Stanley, E. R. 1985. The c-fms Proto-oncogene product is related to the receptor for the mononuclear phagocyte growth factor CSF-1. *Cell* 41: 665–76
48. Ishioka, N., Takahashi, N., Putnam, F. W. 1986. Amino acid sequence of human plasma alpha$_1$B-glycoprotein: homology to the immunoglobulin supergene family. *Proc. Natl. Acad. Sci. USA* 83: 2363–67
49. Bonnet, F., Perin, J.-P., Lorenzo, F., Jolles, J., Jolles, P. 1986. An unexpected sequence homology between link proteins of the proteoglycan complex and Ig-like proteins. *Biochem. Biophys. Acta.* 873: 152–55

50. Neame, P. J., Christner, J. E., Baker, J. R. 1986. The primary structure of link protein from rat chondrosarcoma proteoglycan aggregate. *J. Biol. Chem.* 261: 3519–35
51. McKusick, V. A. 1987. The human gene map. *Genomics.* In press
52. Lesk, A. M., Chothia, C. 1982. Evolution of proteins formed by β-sheets. II. The core of the Ig domains. *J. Mol. Biol.* 160: 325–42
53. Amit, A. G., Mariuzza, R. A., Phillips, S. E. V., Poljak, R. J. 1986. Three-dimensional structure of an antigen-antibody complex at 2.8 A resolution. *Science* 233: 747–53
54. Williams, A. F. 1987. A year in the life of the Immunoglobulin Superfamily. *Immunol. Today* 8: 298–303
55. Dayhoff, M. O., Barker, W. C., Hunt, L. T. 1983. Establishing homologies in protein sequences. *Meth. Enzymol.* 91: 524–45
56. Johnson, P., Williams, A. F. 1986. Striking similarities between antigen receptor J pieces and sequence in the second chain of the murine CD8 antigen. *Nature* 323: 74–76
57. Rudikoff, S., Pumphrey, J. G. 1986. Functional antibody lacking a variable-region disulphide bridge. *Proc. Natl. Acad. Sci. USA* 83: 7875–78
58. Barclay, A. N., Johnson, P., McCaughan, G. W., Williams, A. F. 1987. Immunoglobulin-related structures associated with vertebrate cell surfaces. In *The T Cell Receptors.* New York: Plenum. In press
59. Chatterjee, D., Maizel, J. V. Jr. 1984. Homology of adenoviral E3 glycoprotein with HLA-DR heavy chain. *Proc. Natl. Acad. Sci. USA* 81: 6039–41
60. Huang, H.-J. S., Jones, N. H., Strominger, J. L., Herzenberg, L. A. 1987. Molecular cloning of Ly-1, a membrane glycoprotein of mouse T lymphocytes and a subset of B cells. *Proc. Natl. Acad. Sci. USA* 84: 204–8
61. Richardson, J. S., Richardson, D. C., Thomas, K. A., Silverton, E. W., Davies, D. R. 1976. Similarity of three-dimensional structure between the immunoglobulin domain and the copper, zinc super-oxide dismutase subunit. *J. Mol. Biol.* 102: 221–35
62. Maddon, P. J., Dalgleish, A. G., McDougal, J. S., Clapham, P. R., Weiss, R. A., Axel, R. 1986. The T4 gene encodes the AIDS virus receptor and is expressed in the immune system and the brain. *Cell* 47: 333–48
63. Campbell, D. G., Williams, A. F., Bayley, P. M., Reid, K. B. M. 1979. Structural similarities between Thy-1 antigen from rat brain and immunoglobulin. *Nature* 282: 341–42
64. Zvelebil, M. J., Barton, G. J., Taylor, W. R., Sternberg, M. J. E. 1987. Prediction of protein secondary structure and active sites using the alignment of homologous sequences. *J. Mol. Biol.* 195: 957–61
65. Campbell, R. D., Law, S. K. A., Reid, K. B. M., Sim, R. B. 1988. Structure, organization and regulation of the complement genes. *Ann. Rev. Immunol.* 6: 000–000
66. Parekh, R. B., Tse, A. G. D., Dwek, R. A., Williams, A. F., Rademacher, T. W. 1987. Tissue-specific N-glycosylation, site-specific oligosaccharide patterns and lentil lectin recognition of rat Thy-1. *EMBO J.* 6: 1233–44
67. Seki, T., Chang, H.-C., Moriuchi, T., Denome, R., Ploegh, H., Silver, J. 1985. A hydrophobic transmembrane segment at the carboxyl terminus of Thy-1. *Science* 227: 649–51
68. Mostov, K. E., Kops, A. de B., Deitcher, D. L. 1986. Deletion of the cytoplasmic domain of the polymeric Immunoglobulin receptor prevents basolateral localization and endocytosis. *Cell* 47: 359–64
69. Wall, R., Kuehl, M. 1983. Biosynthesis and regulation of Igs. *Ann. Rev. Immunol.* 1: 393–422
70. Tunnacliffe, A., Buluwela, L., Rabbitts, T. H. 1987. Physical linkage of three CD3 genes on human chromosome 11. *EMBO J.* 6: 2953–57
71. Saito, H., Koyama, T., Georgopoulos, K., Clevers, H., Haser, W. G., Le Bien, T., Terhorst, C. 1987. Close linkage of the mouse and human T3γ and δ genes suggests a control of their transcription by common regulatory elements. In press
72. Trowsdale, J., Young, J. A. T., Kelly, A. P., Austin, P. J., Carson, S., Meunier, H., So, A., Erlich, H. A., Spielman, R. S., Bodmer, J., Bodmer, W. F. 1985. Structure, sequence and polymorphism in the HLA-D region. *Immunol. Rev.* 85: 5–43
73. Parnes, J. R., Seidman, J. G. 1982. Structure of wild-type and mutant β_2-microglobulin genes. *Cell* 29: 661–69
74. Littman, D. R., Gettner, S. N. 1987. Unusual intron in the immunoglobulin domain of the newly isolated murine CD4 (L3T4) gene. *Nature* 325: 435–55
75. Giguère, V., Isobe, K.-I., Grosveld, F. 1985. Structure of the murine Thy-1 gene. *EMBO J.* 4: 2017–24
76. Deitcher, D. L., Mostov, K. E. 1986.

Alternate splicing of rabbit polymeric Immunoglobulin receptor. *Mol. Cell. Biol.* 6: 2712–15
77. Lemke, G., Lamar, E., Patterson, J. 1987. Isolation and analysis of the gene encoding peripheral myelin protein zero. Submitted
78. Bourgois, A. 1975. Evidence for an ancestral immunoglobulin gene coding for half a domain. *Immunochemistry* 12: 873–76
79. McLachlan, A. D. 1980. Early evolution of the antibody domain. In *Proteins and Related Subjects, Protides of the Biological Fluids*, ed. H. Peeters, Pergamon 28: 29–32
80. Matsunaga, T. 1985. Evolution of the Ig-superfamily by duplication of complementarity. *Immunol. Today* 6: 260–63
81. Horvitz, H. R., Ellis, H. M., Sternberg, P. W. 1982. Programmed cell death in nematode development. *Neurosci. Commentaries* 1: 56–65
82. Hedgecock, E. M., Sulston, J. E., Thomson, J. N. 1983. Mutations affecting programmed cell deaths in the nematode *Caenorhabditis elegans*. *Science* 220: 1277–79
83. Williams, A. F., Tse, A. G.-D., Gagnon, J. 1987. Squid glycoproteins with structural similarities to Thy-1 and Ly-6 antigens. *Immunogenetics*. In press
84. Dustin, M. L., Selvaraj, P., Mattaliano, R. J., Springer, T. A. 1987. Anchoring mechanisms for LFA-3 cell adhesion glycoprotein at membrane surface. *Nature* 329: 846–48
85. Ferguson, M. A. J., Williams, A. F. 1988. Cell surface anchoring of proteins via glycosyl-phosphatidylinositol structures. *Ann. Rev. Biochem.* In press

Note Added in Proof

Additional Ig-related structures include the CD7 and CD28 T lymphocyte antigens, one of the chains of the mast cell Fc receptor for IgE, BLAST-1, a leucocyte antigen and the proto-oncogene c-*kit*. CD7 and CD28 antigens have a single V-like domain with transmembrane sequence and cytoplasmic domain. The CD28 chain exists as a disulfide linked homodimer. In structural architecture the FcRε chain is the same as the FcRγ chain, BLAST-1 is like LFA-3 and CD2. The proto-oncogene c-*kit* is the same as CSF1R and PDGFR.

Literature Cited

Aruffo, A., Seed, B. 1987a. Molecular cloning of two CD7 (T-cell leukemia antigen) cDNAs by a COS cell expression system. *EMBO J.* 6: 3313–16
Aruffo, A., Seed, B. 1987b. Molecular cloning of a CD28 cDNA by a high efficiency COS cell expression system. *Proc. Natl. Acad. Sci. USA*. In press
Kinet, J.-P., Metzger, H., Hakimi, J., Kochan, J. 1987. A cDNA presumptively coding for the α subunit of the receptor with high affinity for immunoglobulin E. *Biochemistry* 26: 4605–10
Staunton, D. E., Thorley-Lawson, D. A. 1987. Molecular cloning of the lymphocyte activation marker Blast-1. *EMBO J.* 6: 3695–3701
Yarden, Y., Kuang, W.-J., Yang-Feng, T., Coussens, T., Munemitsu, S., Dull, T. J., Chen, E., Schlessinger, J., Francke, U., Ulrich, A. 1987. Human proto-oncogene c-*kit*: a new cell surface receptor tyrosine kinase for an unidentified ligand. *EMBO J.* 6: 3341–51