Research Signpost
37/661 (2), Fort P.O., Trivandrum-695 023, Kerala, India

Recent Res. Devel. Human Genet., 2(2004): 111-136 ISBN: 81-7736-212-7

6

IMGT standardization for alleles and mutations of the V-LIKE-DOMAINs and C-LIKE-DOMAINs of the immunoglobulin superfamily

Elodie Duprat, Quentin Kaas, Valérie Garelle, Véronique Giudicelli Gérard Lefranc and Marie-Paule Lefranc[1]
IMGT, Laboratoire d'ImmunoGénétique Moléculaire, LIGM, Université Montpellier II, UPR CNRS 1142, Institut de Génétique Humaine, IGH Montpellier, France

Abstract

The immunoglobulin superfamily (IgSF) comprises the immunoglobulins (IG) and T cell receptors (TR) involved in antigen recognition, and also a great number of proteins other than IG and TR that are involved in many different functions (in ligand-receptor interactions in development, differentiation, activation, adhesion, regulation, etc.). The IgSF proteins

[1]Institut Universitaire de France, 103 Boulevard Saint-Michel, 75005 Paris, France
Correspondence/Reprint request: Dr. Marie-Paule Lefranc, IMGT, the international ImMunoGeneTics information system®, Laboratoire d'ImmunoGénétique Moléculaire, LIGM UPR CNRS 1142, IGH, 141 rue de la Cardonille 34396 Montpellier Cedex 5, France. E-mail: lefranc@ligm.igh.cnrs.fr

are defined by having at least one immunoglobulin-like (Ig-like) domain. Despite a large divergence in the amino acid sequences, the Ig-like domains share the IG structural fold which typically consists of about one hundred amino acids in antiparallel beta strands, linked by beta turns or loops, and located on two layers maintained by a disulfide bridge. The number of antiparallel beta strands defines two sets: 9 strands for the V-set (which comprises the V-DOMAINs of the IG and TR, and the V-LIKE-DOMAINs of the IgSF proteins other than the IG or TR) and 7 strands for the C-set (which comprises the C-DOMAINs of the IG and TR, and the C-LIKE-DOMAINs of the IgSF proteins other than the IG or TR). IMGT, the international ImMunoGeneTics information system® (http://imgt.cines.fr), has set up a unique numbering system which takes into account the structural features of the Ig-like domains. In this paper, we describe the IMGT Scientific chart rules for the description of the IgSF V-set and C-set domains, that are applicable for the sequence and structure analysis, whatever the species, the IgSF protein or the chain type. We present examples of 2D graphical representations (IMGT Colliers de Perles) based on the IMGT unique numbering, that are particularly useful for the visualization and comparison of the positions of mutations and polymorphisms in the Ig-like domains.

Introduction

IMGT, the international ImMunoGeneTics information system® (http://imgt.cines.fr) [1] is a high quality integrated knowledge resource specializing in immunoglobulins (IG), T cell receptors (TR), major histocompatibility complex (MHC), and related proteins of the immune system (RPI) of human and other vertebrates [2-12]. IMGT provides a common access to expertly annotated data on the genome, proteome, genetics and structure of the IG, TR, MHC and RPI, based on the IMGT Scientific chart and IMGT-ONTOLOGY [13]. The IMGT standardized description of mutations, allelic polymorphisms, 2D and 3D structure representations, is based on the IMGT unique numbering [14-17]. The IMGT unique numbering is used for the IG and TR variable (V-DOMAIN) and constant domain (C-DOMAIN) of all jawed vertebrates whatever the species, the antigen receptor, or the chain type [18-35]. The IMGT unique numbering led to the first complete description of the human IG and TR genes and alleles [18,19], the standardized 2D graphical representations or IMGT Colliers de Perles [16,17,34,36], the standardized definition of the framework (FR-IMGT) and complementarity-determining-regions (CDR-IMGT) of the V-DOMAINs [16] and the description of the strands and loops of the C-DOMAINs [17]. The many advantages of the IMGT unique numbering naturally led us to extend it to members of the immunoglobulin superfamily (IgSF) other than IG or TR [37]. Indeed, the IgSF not only comprises the IG and TR involved in antigen recognition, but

also a great number of proteins that are involved in many different functions (in ligand-receptor interactions in development, differentiation, activation, adhesion, regulation, etc.) [38]. The common feature of the IgSF proteins is to have at least one immunoglobulin-like (Ig-like) domain [37-41]. Despite a large divergence in the amino acid sequences, the Ig-like domains share the IG structural fold which typically consists of about one hundred amino acids in antiparallel beta strands, linked by beta turns or loops, and located on two layers maintained by a disulfide bridge. The number of antiparallel beta strands defines two sets: 9 strands for the V-set (which comprises the V-DOMAINs of the IG and TR, and the V-LIKE-DOMAINs of the IgSF proteins other than the IG or TR) [16], and 7 strands for the C-set (which comprises the C-DOMAINs of the IG and TR, and the C-LIKE-DOMAINs of the IgSF proteins other than the IG or TR) [17]. By taking into account the structural features of the Ig-like domains, the IMGT unique numbering [14-17] and its graphical representations, the IMGT Colliers de Perles [34,36,37], allow to fill in the gap between linear amino acid sequences and three-dimensional (3D) structures.

In this paper, we describe the IMGT Scientific chart rules for the description of the IgSF V-set and C-set domains, which are applicable for the sequence and structure analysis, whatever the species, the IgSF protein or the chain type. We present examples of IMGT Colliers de Perles based on the IMGT unique numbering. This standardization is particularly useful in the absence of 3D structural data, for the visualization and comparison of mutation and polymorphism positions in the Ig-like domains.

1. IMGT Colliers de Perles for V-LIKE-DOMAIN

The IMGT Colliers de Perles for V-LIKE-DOMAIN is based on the IMGT unique numbering for V-DOMAIN and V-LIKE-DOMAIN [16]. Indeed, the 3D structure of a V-LIKE-DOMAIN is very similar to that of an IG and TR V-DOMAIN (Fig. 1). Both domain types are made of 9 antiparallel beta strands (A, B, C, C', C", D, E, F and G) linked by beta turns (AB, CC', C"D, DE and EF) or loops (BC, C'C" and FG) forming a sandwich of two sheets (Figure 1). The sheets are closely packed against each other through hydrophobic interactions giving a hydrophobic core, and joined together by a disulfide bridge between strand B in the first sheet and strand F in the second sheet. In the IMGT unique numbering, the conserved amino acids always have the same position, for instance cysteine 23 (1st-CYS), tryptophan 41 (CONSERVED-TRP), conserved hydrophobic (leucine) 89, cysteine 104 (2nd-CYS). The hydrophobic amino acids of the framework regions are also found in conserved positions [14-17]. It is remarkable that the IG fold 3D structure has been conserved through evolution, despite the particularities of the IG and TR synthesis compared to the other proteins [18,19] and the sequence divergence of the IgSF domains. Indeed, the V-LIKE-DOMAIN is usually encoded by a

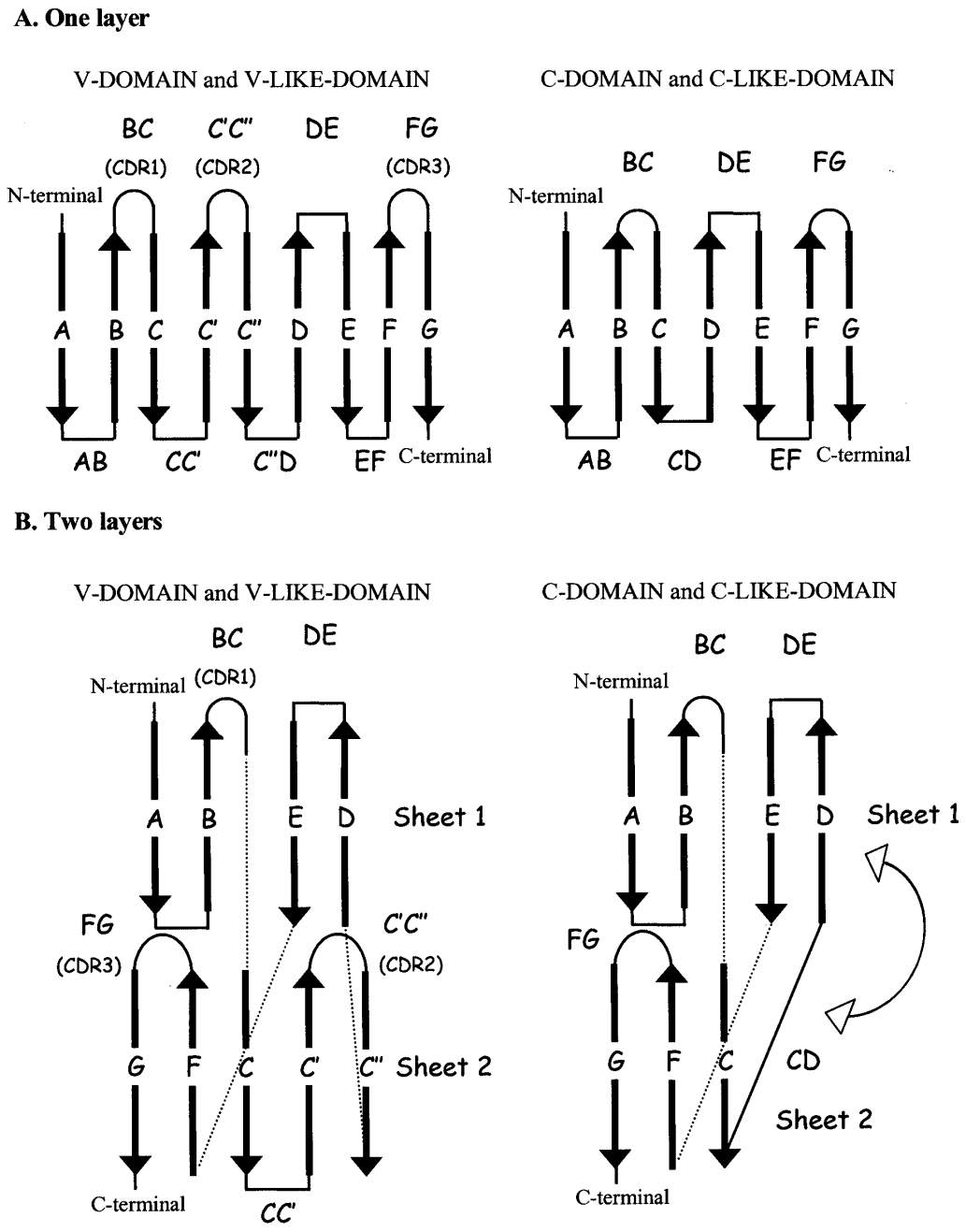
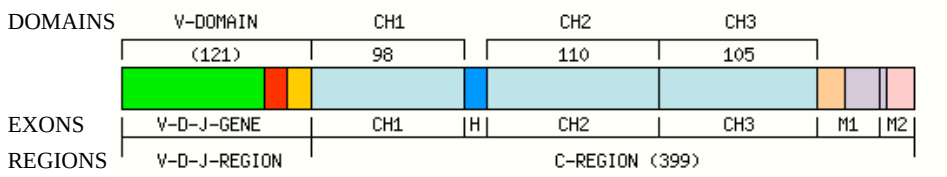


Figure 1. Schematic representation of the V-DOMAIN and V-LIKE-DOMAIN (V-set) and C-DOMAIN and C-LIKE-DOMAIN (C-set). (A) Representation on one layer. (B) Representation on two layers. A double-headed arrow shows that the D strand of the C-DOMAIN and C-LIKE-DOMAIN can be localized in sheet 1 (on the back) or in sheet 2 (on the front) depending from the length of the CD transversal strand.

unique exon, whereas the IG and TR V-DOMAIN results from the rearrangement of two (V, J) or three (V, D, J) genes [18,19] (Figure 2). The V-LIKE-DOMAIN is usually, as the IG and TR V-DOMAIN, the most N-terminal (and extracellular)

A. IG and TR

Homo sapiens membrane IG gamma 1



B. IgSF other than IG or TR

Homo sapiens CD4

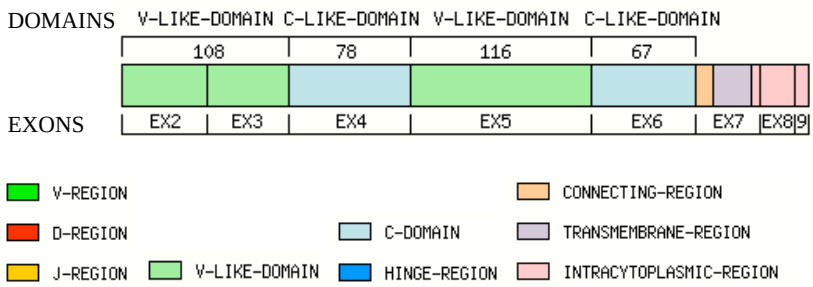


Figure 2. Correspondence between domains and exons. (A) IG and TR, (B) IgSF other than IG and TR. Lengths of the domains and exons are in number of amino acids or codons, respectively. IMGT standardized labels are in capital letters and are described in the IMGT Scientific chart (http://imgt.cines.fr). (A) *Homo sapiens* membrane IG gamma 1 heavy chain (IMGT/LIGM-DB M98324) as example of IG and TR. An IG or TR chain comprises two types of structural units: one V-DOMAIN and one (for the IG light chains and TR chains) or several (for the IG heavy chains) C-DOMAINs (CH1, CH2 and CH3). The unique V-DOMAIN (encoded by a rearranged V-J or V-D-J gene) of a IG or TR chain corresponds to the V-J-REGION or V-D-J-REGION, and is associated to a C-REGION encoded by the C-GENE [18, 19]. (B) *Homo sapiens* CD4 (EMBL/GenBank/DDBJ NT_009759) as example of an IgSF protein other than IG or TR. The general organization of the IgSF other than IG and TR is more diverse and follows the modular shuffling between domains ranging from a unique V-LIKE-DOMAIN or a unique C-LIKE-DOMAIN or to any combination of those domains [38].

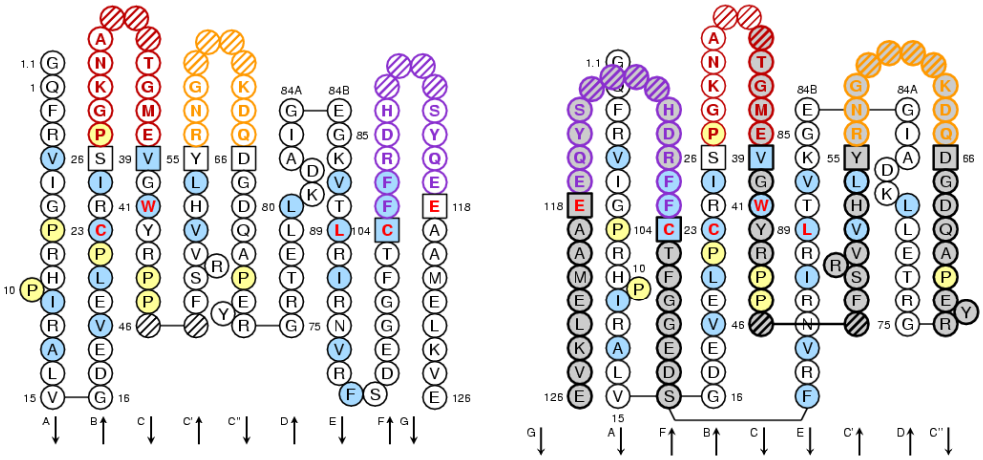
domain of the chain. However, in contrast to the IG and TR V-DOMAIN which is always unique, the V-LIKE-DOMAIN may be present in several copies in the same chain and interspersed with C-LIKE-DOMAINs (Figure 2) or with domains of other superfamilies [39].

1.1 Strands and loops of the V-LIKE-DOMAINs

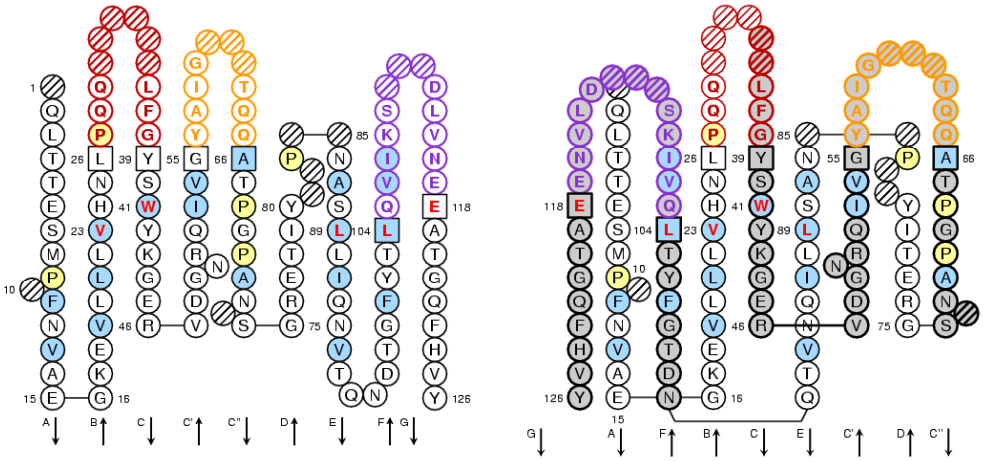
Three examples of V-LIKE-DOMAINs are shown in Figure 3: the myelin oligodendrocyte glycoprotein (MOG) [D], the carcinoembryonic antigen-related cell adhesion molecule 1 (CEACAM1) [D1] and the myelin protein zero (MPZ) [D].

Figure 3

A. *Homo sapiens* MOG [D] V-LIKE-DOMAIN



B. *Homo sapiens* CEACAM1 [D1] V-LIKE-DOMAIN



C. *Homo sapiens* MPZ [D] V-LIKE-DOMAIN

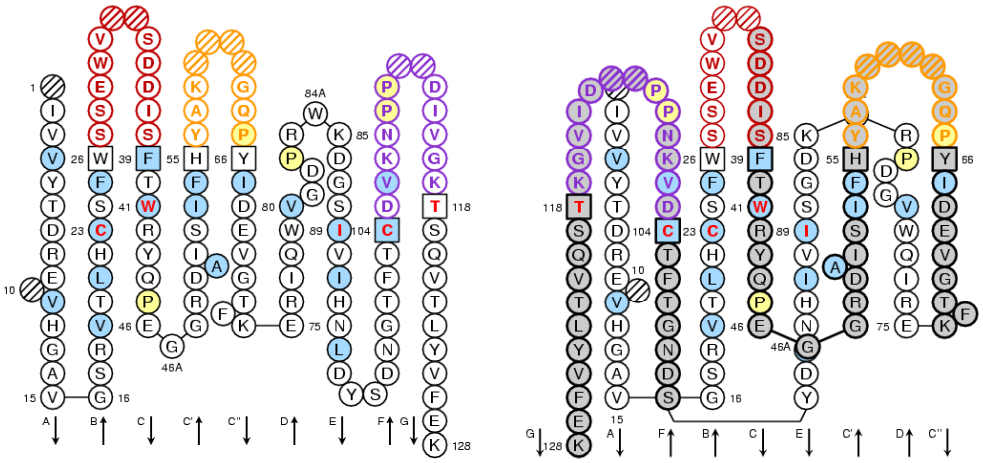


Figure 3. IMGT Colliers de Perles of V-LIKE-DOMAINs on one and two layers. (A) *Homo sapiens* MOG [D], (B) *Homo sapiens* CEACAM1 [D1], (C) *Homo sapiens* MPZ [D]. Amino acids are shown in the one-letter abbreviation. Position at which hydrophobic amino acids (hydropathy index with positive value: I, V, L, F, C, M, A) and tryptophan (W) are found in more than 50% of analysed sequences are shown in blue. All proline (P) are shown in yellow. The loops BC, C'C" and FG (corresponding to the CDR-IMGT) are limited by amino acids shown in squares (anchor positions), which belong to the neighbouring strands (FR-IMGT in V-DOMAINs). Arrows indicate the direction of the beta strands and their different designations in 3D structures (from IMGT Repertoire, http://imgt.cines.fr). BC loops are represented in red, C'C" loops in orange and FG loops in purple. The IMGT Colliers de Perles on two layers (on the right hand) show, on the forefront, the GFCC'C" strands and, on the back, the ABED strands. Hatched circles or squares correspond to missing positions according to the IMGT unique numbering. MPZ has two additional positions (46A and 84A) that interestingly are located at the apex of beta turns and do not disturb the general architecture of the domain.
Swiss-Prot accession numbers: Q16653 for the *Homo sapiens* MOG protein, P13688 for the *Homo sapiens* CEACAM1 protein, and P25189 for the *Homo sapiens* MPZ protein. IMGT Colliers de Perles were checked with the following PDB [42] entries: 1py9_A (*Mus musculus* MOG [D1]), 1pkq_E (*Rattus norvegicus* MOG [D1]), 1l6z_A (*Mus musculus* CEACAM1 [D1]), 1neu (*Rattus norvegicus* MPZ [D1]), as the human MOG [D], CEACAM1 [D1] and MPZ [D] have not yet been crystallized.

1.1.1 Strands

The antiparallel beta strands of the V-LIKE-DOMAIN correspond to the conserved regions or framework (FR-IMGT) described in the V-DOMAIN [16]. The A strand (A-STRAND, positions 1 to 15) and the B strand (B-STRAND, positions 16 to 26) with the conserved cysteine (1st-CYS) at position 23 correspond to the FR1-IMGT (Table 1). The C strand (C-STRAND, positions 39 to 46) with the tryptophan (CONSERVED-TRP) at position 41 and the C' strand (C'-STRAND, positions 47 to 55) correspond to FR2-IMGT. The C" strand (C"-STRAND, positions 66 to 74), the D strand (D-STRAND, positions 75 to 84), the E strand (E-STRAND, positions 85 to 96) with a conserved hydrophobic amino acid at position 89 and the F strand (F-STRAND, positions 97 to 104) with 2nd-CYS at position 104 correspond to the FR3-IMGT. The G strand (G-STRAND, positions 118 to 128) corresponds to FR4-IMGT (in the IG and TR V-DOMAINs, the G strand is the C-terminal part of the J-REGION, with J-PHE or J-TRP 118 and the canonical motif F/W-G-X-G at positions 118-121). Hatched circles or squares in Figure 3 correspond to missing positions according to the IMGT unique numbering. For example, unoccupied positions 46 and 47 in MOG [D], 10 or 73 in CEACAM1 [D1] and 10 in MPZ [D], are shown as hatched circles. In the IMGT Protein display (Figure 4), unoccupied positions according to the IMGT unique numbering are shown by dots.

1.1.2 Loops

The BC, C'C" and FG loops of the V-LIKE-DOMAIN correspond to the complementarity-determining regions (CDR-IMGT) described in the IG and TR V-DOMAINs [16]. The BC loop (BC-LOOP) comprises positions 27 to 38; the longest BC loops have 12 amino acids. For BC loops shorter than 12 amino acids, gaps are created at the apex (missing positions, hatched in IMGT Collier de Perles (Fig. 3), or not shown in structural data representations). The gaps are placed at the apex of the loop with an equal number of codons (or amino acids) on both sides if the loop length is an even number, or with one more codon (or amino acid) in the left part if it is an odd number. As an example, in a BC loop with 9 amino acids (MOG in Figure 3), positions 27 to 31 and 35 to 38 are present, whereas positions 32 to 34 are missing. The C'C" loop (C'C"-LOOP) comprises positions 56 to 65. The longest C'C" loops have 10 amino

Table 1. Gaps and additional positions for FG loop

A - Gaps for FG loops less than 13 amino acids

FG loop (CDR3-IMGT) lengths													
13	105	106	107	108	109	110	111	112	113	114	115	116	117
12	105	106	107	108	109	110	-	112	113	114	115	116	117
11	105	106	107	108	109	110	-	-	113	114	115	116	117
10	105	106	107	108	109	-	-	-	113	114	115	116	117
9	105	106	107	108	109	-	-	-	-	114	115	116	117
8	105	106	107	108	-	-	-	-	-	114	115	116	117
7	105	106	107	108	-	-	-	-	-	-	115	116	117
6	105	106	107	-	-	-	-	-	-	-	115	116	117
5	105	106	107	-	-	-	-	-	-	-	-	116	117

B - Additional positions for FG loops more than 13 amino acids

FG loop (CDR3-IMGT) lengths										

21	**111**	111.1	111.2	111.3	111.4	112.4	112.3	112.2	112.1	**112**
20	**111**	111.1	111.2	111.3	-	112.4	112.3	112.2	112.1	**112**
19	**111**	111.1	111.2	111.3	-	-	112.3	112.2	112.1	**112**
18	**111**	111.1	111.2	-	-	-	112.3	112.2	112.1	**112**
17	**111**	111.1	111.2	-	-	-	-	112.2	112.1	**112**
16	**111**	111.1	-	-	-	-	-	112.2	112.1	**112**
15	**111**	111.1	-	-	-	-	-	-	112.1	**112**
14	**111**	-	-	-	-	-	-	-	112.1	**112**

For FG loops (CDR3-IMGT) more than 13 amino acids, additional positions are created between positions 111 and 112 (in bold). In a given sequence set with FG loops more than 13 amino acids, gaps are created based on the FG loops in the set. As an example, gaps are shown by comparison to a 21 amino acid long FG loop.

Figure 4


A
B
V-DOMAIN
V-LIKE-DOMAIN
C-LIKE-DOMAIN
FR1-IMGT
CDR1-IMGT
FR2-IMGT
CDR2-IMGT
FR3-IMGT
CDR3-IMGT
FR4-IMGT

Figure 4. IMGT Protein display. (A) Examples of V-DOMAINs and V-LIKE-DOMAINs (V-set). (B) Examples of C-DOMAINs and C-LIKE-DOMAINs (C-set). #c: rearranged cDNA, g: genomic DNA. Amino acids resulting from a splicing with a preceding exon are shown between parentheses. (A) IG and TR V-DOMAINs: VH (AB027433, #c IGHV3-30-IGHD4-17-IGHJ3), V-KAPPA (AB022654, #c IGKV1-39-IGKJ2), V-ALPHA (AK026255, #c TRAV26-1-TRAJ39), V-BETA (AF043183, #c TRBV28-TRBD1-TRBJ2-3). V-LIKE-DOMAINs: MOG [D] (Z48051, g), BTN1A1 [D1] (U39576, #c, delimitated by homology with MOG [D]), CEACAM1 [D1] (AC004785, g, leader delimited according to [43]), MPZ [D] (D14720, g, MPZ [D] encoded by EX2 and EX3 with (I)53 resulting from the splicing), CD4 [D1] (NT_009759, g, CD4 [D1] encoded by EX2 and EX3 with (G)68 splicing site), CD8A [D] (M27161, g), CD8B1 [D] (M17514, partial g, [44]), CEACAM5 [D1] (M59255, g, leader delimited according to [45]), CTLA4 [D] (AF411058, g), PIGR [D1] (S43441, partial g, limited to EX2; (A) is deduced from S43437), VPREB1 [D] (D86992, g). (B) IG and TR C-DOMAINs: CH1 (J00228, g), C-LAMBDA1 (X51755, g), C-ALPHA (X02883, g), C-BETA2 (M12888, g). C-LIKE-DOMAINs: KIR2DL2 [D1] (AL133414, g), KIR2DL1 [D1] (L41267, #c, delimitated by homology with KIR2DL2 [D1]), CEACAM1 [D3] (AC004785, g), VCAM1 [D1] (M73255, g), CD1A [D3] (M22165, partial g, limited to EX4; (V) is deduced from M22164), CD3E [D] (M23319, M23320 and M23321, partial g, limited to EX4, EX5 and EX6, respectively; (G) is deduced from M23318), CD4 [D2] (NT_009759, g), CEACAM5 (M59257, partial g, limited to EX4; (Y) is deduced from M59256), FCER1A [D1] (L14075, g), FCGR1A [D1] (M63832, partial g, limited to EX3; (D) is deduced from M63831), FCGR2A [D1] (M90723, partial g, limited to EX3; (A) is deduced from M90722). The accession numbers are from IMGT/LIGM-DB (http://imgt.cines.fr) [11,46] for IG and TR, and from EMBL/GenBank/DDBJ [47-49] for IgSF other than IG and TR. Beta strands are shown by arrows. Dots indicate missing amino acids according to the IMGT unique numbering. Putative N-glycosylation sites (N-X-S/T) are underlined.

(1) Gene names (symbols) for IG and TR are according to the IMGT Nomenclature committee (IMGT-NC) [18,19] and the HUGO Nomenclature Committee (HGNC) [50]. Full gene designations are the following: MOG: myelin oligodendrocyte glycoprotein; BTN1A1: butyrophilin, subfamily 1, member A1; CEACAM1: Carcinoembryonic antigen-related cell adhesion molecule 1; MPZ: myelin protein zero (Charcot-Marie-Tooth neuropathy 1B); CD4: CD4 antigen (p55); CD8A: CD8 antigen, alpha polypeptide (p32); CD8B1: CD8 antigen, beta polypeptide 1 (p37); CEACAM5: Carcinoembryonic antigen-related cell adhesion molecule 5; CTLA4: cytotoxic T-lymphocyte-associated protein 4; PIGR: polymeric immunoglobulin receptor; VPREB1: pre-B lymphocyte gene 1; IGHG1: Immunoglobulin heavy constant gamma 1; IGLC1: immunoglobulin lambda constant 1; TRAC: T cell receptor alpha constant; TRBC2: T cell receptor beta constant 2; KIR2DL2: killer cell immunoglobulin-like receptor, two domains, long cytoplasmic tail, 2; VCAM1: vascular cell adhesion molecule 1; CD1A: CD1A antigen, a polypeptide; CD3E: CD3E antigen, epsilon polypeptide (TiT3 complex); CD4: CD4 antigen (p55); FCER1A: Fc fragment of IgE, high affinity I, receptor for alpha polypeptide; FCGR1A: Fc fragment of IgG, high affinity Ia, receptor for (CD64); FCGR2A: Fc fragment of IgG, low affinity IIa, receptor for (CD32).

(2) Domain name. The C-DOMAINs are designated with the IMGT labels (IMGT Scientific chart, http://imgt.cines.fr) The C-LIKE-DOMAINs are designated by the

Figure legend continued

letter D between brackets with a number, corresponding to the position of the domain from the N-terminal end of the protein, and relative to the other domains. There is no number if there is a unique C-LIKE-DOMAIN in the chain.
Amino acid one-letter abbreviation: A (Ala), alanine; C (Cys), cysteine; D (Asp), aspartic acid; E (Glu), glutamic acid; F (Phe), phenylalanine; G (Gly), glycine; H (His), histidine; I (Ileu), 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; Y (Tyr), tyrosine.

acids. For C'C" loops shorter than 10 amino acids, gaps are created (missing positions, hatched in IMGT Collier de Perles, or not shown in structural data representations). As an example, in a C'C" loop with 6 amino acids (MOG and MPZ in Figure 3), positions 56 to 58, 63 to 65 are present, whereas positions 59 to 62 are missing. The FG loop (FG-LOOP) comprises position 105 to 117. These positions correspond to a FG loop of 13 amino acids. For FG loops shorter than 13 amino acids, gaps are created from the apex of the loop, in the following order 111, 112, 110, 113, 109, 114, etc (Table 1). For FG loops longer than 13 amino acids (which is rare), additional positions are created, between positions 111 and 112 at the top of the FG loop (Table 1).

1.1.3 Loop length

The loop length (number of codons or amino acids, that is number of occupied positions) is a crucial and original concept of IMGT-ONTOLOGY [13]. The lengths of the BC, C'C" and FG loops characterize the V-LIKE-DOMAINs, as the lengths of the CDR1-IMGT, CDR2-IMGT and CDR3-IMGT characterize the IG and TR V-DOMAINs. Thus, the length of the three loops BC, C'C" and FG is shown, in number of codons (or amino acids), into brackets and separated by dots (Table 2). For examples: *Homo sapiens* MOG [D] [9.6.9] means that in the human MOG [D] domain, the BC, C'C" and FG loops have a length of 9, 6 and 9 codons, respectively; *Homo sapiens* CEACAM1 [D1] [6.7.10] means that in the human CEACAM1 [D1] domain, the BC, C'C" and FG loops have a length of 6, 7 and 10 codons, respectively; *Homo sapiens* MPZ [D] [10.6.11] means that in the human MPZ [D] domain, the BC, C'C" and FG loops have a length of 10, 6 and 11 amino acids, respectively (Figure 3, Table 2).

1.2 Characteristics of the MOG, CEACAM1 and MPZ V-LIKE-DOMAINs

1.2.1 Myelin oligodendrocyte glycoprotein

The myelin oligodendrocyte glycoprotein (MOG) is a specific component of the central nervous system (CNS) localized on the outermost lamellae of

Table 2. Delimitation of the strands and loops for V-LIKE-DOMAINs

		V-LIKE-DOMAIN strands and loops		MOG	BTN1A1	CEACAM1	MPZ	CD4	CD8A	CD8B1	CEACAM5	CTLA4	PIGR	VPREB1
		Numbering	Length	[D] [9.6.9]	[D1] [9.6.9]	[D1] [6.7.10]	[D] [10.6.11]	[D1] [6.6.6]	[D] [7.2.9]	[D] [7.5.9]	[D1] [6.7.10]	[D] [8.9.12]	[D1] [9.2.11]	[D] [9.7.14]
A-STRAND	FR1-IMGT	1.4-1.1	+4	+1	+1			+2	+4	+4			+3	+4
		1-15	15	15	15	13	13	15	15	15	13	14	15	14
B-STRAND		16-26	11	11	11	11	11	11	11	11	11	11	11	11
BC-LOOP	CDR1-IMGT	27-38	12	9	9	6	10	6	7	7	6	8	9	9
C-STRAND		39-46	8	7	7	8	8	6	8	8	8	8	8	8
C'C"-TURN	FR2-IMGT	46A,46B,46C 47B,47A	+5				+1		+4	+4				
C"-STRAND		47-55	9	8	8	9	9	7	9	9	9	9	9	9
C'C"-LOOP	CDR2-IMGT	56-65	10	6	6	7	6	6	2	5	7	9	2	7
C"-STRAND		66-74	9	9	9	8	9	9	9	9	8	6	9	8
D-STRAND		75-84	10	10	10	7	10	10	10	10	7	8	10	10
DE-TURN	FR3-IMGT	84A,84B,84C	+3	+2	+2		+1	+1						
E-STRAND		85-96	12	12	12	11	12	12	12	12	11	12	12	12
F-STRAND		97-104	8	8	8	8	8	8	8	8	8	8	8	8
FG-LOOP	CDR3-IMGT	105-117	13	9	9	10	11	6	9	9	10	12	11	13
		111.1-111.6, 112.6-112.1	+12											+1
G-STRAND	FR4-IMGT	118-128	11	9	9	9	11	9	10	9	9	11	8	16
Total length			128 (+24)	116	116	107	120	108	118	120	107	116	115	126

The delimitations of the strands and loops for the V-LIKE-DOMAINs are identical to those of the IG and TR V-DOMAINs. For more details, see [16]. Lengths of the BC, C'C" and FG loops are shown within brackets. Blank cells indicate no amino acids. A plus sign indicates additional amino acids. Amino acid sequences are shown in Figure 4.

mature myelin [51], and may contribute to myelin maturation and maintenance [52] or as a signal to arrest further myelination [53]. MOG may have an unforeseen immunological status within the central nervous system, providing for instance a rudimentary molecular framework for presentation of pathogens to the immune sytem [54]; MOG is a candidate target antigen for autoimmune-mediated demyelination, and seems to be implicated in pathogenesis of encephalomyelitis and multiple sclerosis, an inflammatory disease of the central nervous system [55,56]. The human MOG gene contains 8 exons [57]. The N-terminal extracellular V-LIKE-DOMAIN of MOG encoded by the exon 2 (EX2) has significant sequence homologies with three non-myelin proteins: *Homo sapiens* butyrophilin (BTN1A1; subfamily 1, member A1) expressed in the mammary gland during lactation and facilitating the interaction between cytoplasmic lipid droplets and the apical membrane [58], *Gallus gallus* B-G antigen encoded by a gene mapping to the major histocompatibility complex [59], and *Gallus gallus* BEN adhesion glycoprotein expressed on the epithelial cells of the bursa of Fabricius and on various neuronal subsets during chicken embryonic development [60]. The IMGT unique numbering allows us to compare *H. sapiens* MOG [D] and *H. sapiens* BTN1A1 [D1] and to describe divergent positions (Figure 3). The MOG and BTN1A1 genes are colocalized near

the human MHC on chromosome 6p21.3-p22 [61]. Two motifs highly homologous to consensus sequences found in glial promoters of proteolipid protein (PLP), protein zero (MPZ), myelin basic protein (MBP), and mouse MOG were also found in human MOG [62]. The *H. sapiens* MOG V-LIKE-DOMAIN has not yet been crystallized, but interestingly PDB has an entry for the *Rattus norvegicus* MOG V-LIKE-DOMAIN in complex with Fab fragment of *Mus musculus* antibody anti-MOG 8-18C5 (1pkq), that allows to identify the MOG FG loop as an important epitope [63,64], and to confirm the N-glycosylated site in N31 (BC-LOOP) (conserved in *H. sapiens* MOG sequence, Figure 3); the MOG FG loop may be implicated in autoimmune recognition [63].

1.2.2 Carcinoembryonic antigen-related cell adhesion molecule 1

Members of the CEA family consist of a single N-terminal V-LIKE-DOMAIN, followed by a variable number of C-LIKE-DOMAINs. Based on sequence similarity and functional characteristics, the CEA family has been subdivided into the CEA subfamily and the pregnancy-specific glycoprotein (PSG) subfamily [65]. Members of the CEA subfamily are anchored in the cell membrane, whereas all of the PSGs appear to be secreted; however, the genes in the CEA and PSG subfamilies have a similar gene structure and organization. The carcinoembryonic antigen-related cell adhesion molecule 1 (CEACAM1, or biliary glycoprotein BGP) consists of a single N-terminal V-LIKE-DOMAIN followed by 3 C-LIKE-DOMAINs, and is expressed in cells of epithelial and myeloid origin [43]. In granulocytes, CEACAM1 is a main antigen of the CD66 cluster of differentiation antigens that mediate the binding to endothelial E-selectin. The loss or reduced expression of the CEACAM1 adhesion molecule is a major event in colorectal carcinogenesis [66].

1.2.3 Myelin protein zero

The myelin protein zero (MPZ or P0) gene is localized at 1q21.3-q23, about 130 kb of the FCGR2A gene [67]. MPZ is the major structural protein of peripheral myelin, accounting for more than 50% of the protein present in the sheath of peripheral nerves. Expression of the MPZ gene is restricted to Schwann cells; MPZ is not found in the CNS. MPZ corresponds to an integral membrane glycoprotein of 28 kD, and is thought to link adjacent lamellae and thereby stabilize the myelin assembly. The V-LIKE-DOMAIN [D] that is encoded by EX2 and EX3, plays a significant role in myelin membrane adhesion. Several mutations in the MPZ V-LIKE-DOMAIN are associated with the autosomal dominant form of Charcot-Marie-Tooth disease type 1, which is characterized by progressive slowing of nerve conduction and hypertrophy of Schwann cells: the amino acid changes D68>E (C’’-STRAND)

and K74>E (C''-STRAND, near the C''D-TURN) are independently implicated in Charcot-Marie-Tooth disease type 1B (CMT1B) [68], whereas all affected members of another CTM1B family had a 3-bp deletion in EX2 causing loss of the S38 (S38>del; BC-LOOP) [69]. S38>C was found in a 7-year-old boy (heterozygous for the mutation, which was absent in the parents and in 100 unrelated healthy controls) with delayed motor development, hypotonia, muscle weakness, and sensory disturbance thought to be typical of Dejerine-Sottas syndrome, or hereditary motor and sensory neuropathy type III (HMSN3) [68,70]. Partial symptom relief with corticosteroid treatment was reported [71] in a patient with demyelinating CMT1B and a heterozygous R76>H mutation (D-STRAND, near the C''D-TURN). Although this response is rare in such patients, poor myelin compaction by the MPZ protein, caused by the mutation, may have allowed circulating immune elements access to normally sequestered endoneurial components, thus accounting for the response to corticosteroid treatment [71] (OMIM: 159440).

2. IMGT Colliers de Perles for C-LIKE-DOMAIN

The IMGT Colliers de Perles for C-LIKE-DOMAIN is based on the IMGT unique numbering for C-DOMAIN [17]. This numbering is itself derived from the IMGT unique numbering for V-DOMAIN [14-16]. Indeed, the sandwich beta sheet of the C-set (C-DOMAIN and C-LIKE-DOMAIN) has the same topology and 3D structure than the V-set (V-DOMAIN and V-LIKE-DOMAIN), but they differ by the number of strands (Figure 1). The C-LIKE-DOMAIN, as the IG and TR C-DOMAIN, is made of seven beta strands linked by beta turns or loops, and arranged so that four strands form one sheet and three strands form a second sheet. A characteristic CD transversal strand links the two sheets; depending from the CD length, the D strand is in the first or second sheet (shown by an arrow in Figure 1). As shown in Table 3, the IMGT unique numbering for the C-LIKE-DOMAIN follows the same rules as those of the C-DOMAIN [17].

2.1. Strands, loops and turns of the C-LIKE-DOMAINs

Three examples of C-LIKE-DOMAIN are shown in Figure 5: the killer cell immunoglobulin-like receptor KIR2DL2 (two domains, long cytoplasmic tail, 2) [D1], the carcinoembryonic antigen-related cell adhesion molecule 1 (CEACAM1) [D3] and the vascular cell adhesion molecule 1 (VCAM1) [D1].

2.1.1 Strands

The A strand (A-STRAND, positions 1 to 15) and the B strand (B-STRAND, positions 16 to 26, with the 1st-CYS at position 23) are similar to those of the V-DOMAIN and V-LIKE-DOMAIN [16,17]. The C strand (C-STRAND, positions 39 to 45, with the CONSERVED-TRP at position 41) and the D strand (D-STRAND, positions 77 to 84) are shorter of one position and

Table 3. Delimitation of the strands and loops for C-LIKE-DOMAINs

	C-LIKE-DOMAIN strands and loops		KIR2DL2	KIR2DL1	CEACAM1	VCAM1	CD1A	CD3E	CD4	CEACAM5	FCER1A	FCGR1A	FCGR2A
	Numbering	Length	[D1]	[D1]	[D3]	[D1]	[D3]	[D]	[D2]	[D3]	[D1]	[D1]	[D1]
A-STRAND	1.4-1.1	+4	+2	+2	+2	+2		+3	+5	+2	+3	+1	+2
	1-15	15	15	15	15	15	14	15	15	15	15	15	15
AB-TURN	15.1-15.3	+3	+1	+1		+1			+3			+2	+2
B-STRAND	16-26	11	11	11	11	11	11	11	11	11	11	11	11
BC-LOOP	27-36	10	5	5	6	6	8	4	1	6	7	7	7
C-STRAND	39-45	7	7	7	7	7	7	7	7	7	7	7	7
CD-STRAND	45.1-45.7	+7	+5	+5		+4	+5	+4			+2	+2	+2
D-STRAND	77-84	8	8	8	6	6	8	8		6	5	5	6
DE-TURN	84.1-84.7 85.7-85.1	+14	+3	+3		+4	+7	+4					
E-STRAND	85-96	12	12	12	10	11	10	11	10	10	10	10	9
EF-TURN	96.1, 96.2	+2											
F-STRAND	97-104	8	8	8	8	7	7	8	8	8	8	8	8
FG-LOOP	105-117	13	13	13	13	9	11	13	11	13	7	7	7
	111.1-111.6	+12	+1	+1									
G-STRAND	118-128	11	9	9	7	9	8	6	7	6	10	10	10
Total length		95 (+42)	100	100	85	92	96	94	78	84	85	85	86

Delimitations of the strands and loops for C-LIKE-DOMAINs are identical to those of the IG and TR C-DOMAINs. For more details, see [17]. A plus sign indicates additional amino acids. KIR2DL2 [D1] has one additional position in 15.1, whereas 15.2 and 15.3 are unoccupied. VCAM1 [D1] has four additional positions 84.1, 84.2, 85.2 and 85.1. Amino acid sequences are shown in Figure 4.

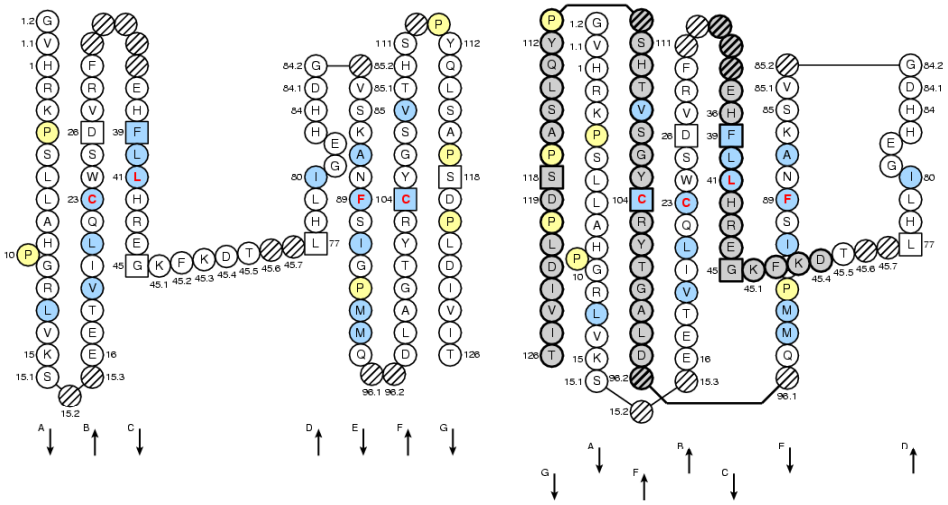
two positions, respectively, compared to the V-DOMAIN and V-LIKE-DOMAIN. As previously described [17], the C' and C" strands are missing and are replaced by the characteristic transversal CD strand (CD-STRAND, positions 45.1 to 45.7). The E strand (E-STRAND, positions 85 to 96, with a conserved hydrophobic amino acid at position 89), the F strand (F-STRAND, positions 97 to 104, with the 2nd-CYS at position 104) and the G strand (G-STRAND, positions 118 to 128, with a conserved hydrophobic amino acid at position 121) are similar to those of the V-DOMAIN and V-LIKE-DOMAIN.

2.1.2 Loops

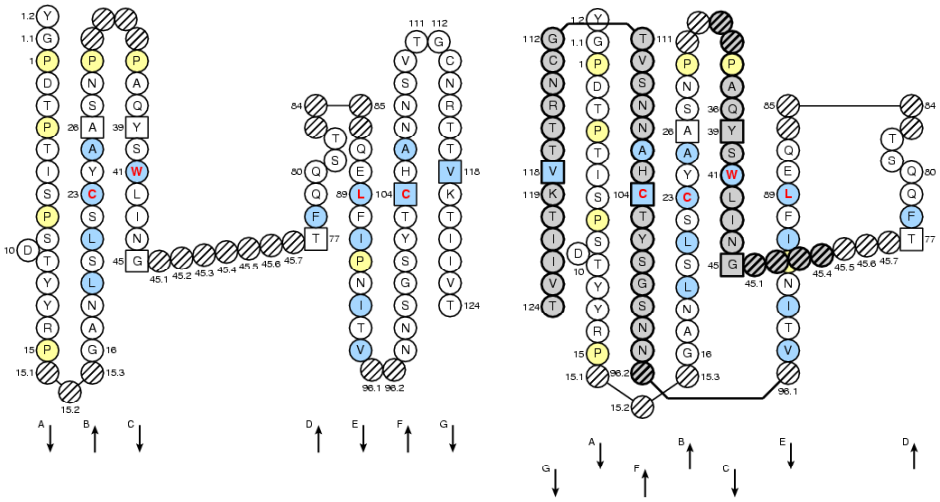
The BC loop (BC-LOOP) comprises positions 27 to 36; the longest BC loops have 10 amino acids, instead of 12 amino acids in the V-DOMAIN and V-LIKE-DOMAIN. For BC loops shorter than 10 amino acids, gaps are created from the apex in the following order 32, 31, 33, 30, 34, etc. As an example, in a BC loop with 5 amino acids (KIR2DL2 [D1] in Figure 5), positions 27 to 29 and 35 and 36 are present, whereas positions 30 to 34 are missing. The FG loop (FG-LOOP) comprises positions 105 to 117 and is similar to that of the V-DOMAIN and V-LIKE-DOMAIN. These positions correspond to a FG loop of 13 amino acids. Gaps for FG loops shorter than 13 amino acids and additional positions for FG loops longer than 13 amino acids, are created following the same rules as those of the V-DOMAIN and V-LIKE-DOMAIN (Table 1). As examples, CEACAM1 [D3] has a FG loop of 13 amino acids, VCAM1 [D] has a FG loop of

Figure 5

A. *Homo sapiens* KIR2DL2 [D1] C-LIKE-DOMAIN



B. *Homo sapiens* CEACAM1 [D3] C-LIKE-DOMAIN



C. *Homo sapiens* VCAM1 [D1] C-LIKE-DOMAIN

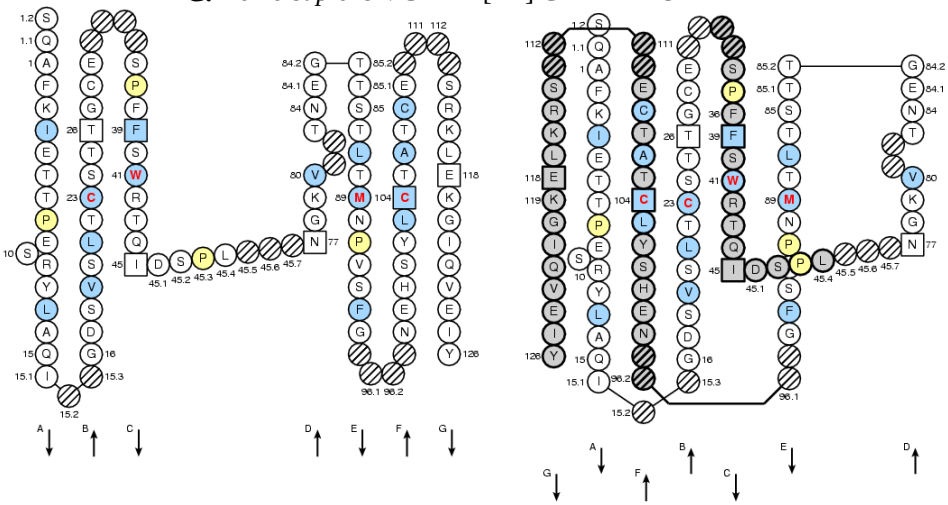


Figure 5. IMGT Colliers de Perles of C-LIKE-DOMAINs on one and two layers. (A) *Homo sapiens* KIR2DL2 [D1], (B) *Homo sapiens* CEACAM1 [D3], (C) *Homo sapiens* VCAM1 [D1]. Amino acids are shown in the one-letter abbreviation. Position at which hydrophobic amino acids (hydropathy index with positive value: I, V, L, F, C, M, A) and tryptophan (W) are found in more than 50% of analysed sequences are shown in blue. All proline (P) are shown in yellow. The positions 26, 39 and 104 are shown in squares by homology with the corresponding positions in the V-set (V-DOMAINs and V-LIKE-DOMAINs). Positions 45 and 77 which delimit the characteristic CD strand of the C-set (C-DOMAINs and C-LIKE-DOMAINs), and position 118 which corresponds structurally to J-PHE or J-TRP of the IG and TR J-REGION [16,17], are also shown in squares. Hatched circles correspond to missing positions according to the IMGT unique numbering for C-DOMAINs and C-LIKE-DOMAINs. Arrows indicate the direction of the beta strands and their different designations in 3D structures (from IMGT Repertoire, http://imgt.cines.fr). The IMGT Colliers de Perles on two layers (on the right hand) show, on the forefront, the GFC strands and, on the back, the ABE strands. Swiss-Prot accession numbers: P43627 for the *H. sapiens* KIR2DL2 protein, P13688 for the *H. sapiens* CEACAM1 protein, and P19320 for the *H. sapiens* VCAM1 protein. The IMGT Colliers de Perles were checked with the following PDB [42] entries: 1efx_D (*H. sapiens* KIR2DL2 [D1]), 1l6z_A (*Mus musculus* CEACAM1 [D4], sequence similar to the *H. sapiens* CEACAM1 [D3]; no available 3D structure of the *H. sapiens* CEACAM1 [D3]), 1vsc_A (*H. sapiens* VCAM1 [D1]).

9 amino acids (with four gaps 110 to 113), whereas KIR2DL2 [D1] has a FG loop of 14 amino acids with the additional position 112.1 (Figure 5).

2.1.3 Turns

The AB turn (AB-TURN) corresponds to additional positions 15.1, 15.2 and 15.3; the longest AB turns have 3 amino acids. For AB turns shorter than 3 amino acids, gaps are created (missing positions, hatched in the IMGT Colliers de Perles (Fig. 5), or not shown in structural data representations) in an ordinal manner. As an example, KIR2DL2 [D1] has one additional position in 15.1, whereas 15.2 and 15.3 are unoccupied. The DE turn (DE-TURN) comprises positions 84.1 to 84.7 and 85.7 to 85.1, corresponding to 14 amino acids. For DE turns shorter than 14 amino acids, gaps are created in the following order: 85.1, 84.1, 85.2, 84.2, 85.3, 84.3, etc. As an example, VCAM1 [D1] has four additional positions 84.1, 84.2, 85.2 and 85.1. The EF turn (EF-TURN) corresponds to additional positions 96.1 and 96.2 when present, corresponding to 2 amino acids. For EF turns shorter than 2 amino acids, gaps are created in the following order: 96.2, 96.1.

2.2 Characteristics of the KIR2DL2 and VCAM1 C-LIKE-DOMAINs

2.2.1 KIR2DL2

The second killer cell immunoglobulin-like receptor with two domains and a long cytoplasmic tail (KIR2DL2) corresponds to a 348 amino acid type I

transmembrane protein [72,73]. Sequence analysis revealed a structure similar to that described for KIR2DL1, with two extracellular C-LIKE-DOMAINs ([D1] and [D2]), a transmembrane domain, and a long cytoplasmic tail with two immunoreceptor tyrosine-based inhibitory motifs (ITIMs). According to the IMGT unique numbering, divergent positions between KIR2DL2 and KIR2DL1 are localized in the A-STRAND (R12>P), BC-LOOP (R28>M), CD-STRAND (K45.1>M, K45.3>N), D-STRAND (H78>R), E-STRAND (G92>S, P93>R, M95>T), FG-LOOP (L114>V) and G-STRAND (T126>I). KIR2D receptors are divided into two families based on their specificities for different HLA-C allotypes: the KIR2DL1 is specific for HLA-Cw2, 4, 6 and 15, whereas KIR2DL2 is specific for HLA-Cw1, 3, 7 and 8. KIR2DL2/HLA-Cw3 and KIR2DL1/HLA-Cw4 share a common binding mode [74], and a single K45.1>M amino acid change between KIR2DL2 and KIR2DL1 is sufficient to switch allotype specificity [75].

2.2.2 VCAM1

The vascular cell adhesion molecule 1 (VCAM1) is expressed by cytokine-activated endothelium, binds leukocyte integrins and is involved in inflammatory and immune functions [76,77]. This type I membrane protein mediates leukocyte-endothelial cell adhesion and signal transduction, and may play a role in the development of arteriosclerosis and rheumatoid arthritis. VCAM1 is present in single copy in the human genome and contains 9 exons spanning about 25 kb of DNA. At least 2 different VCAM1 precursors can be generated from the human gene as a result of alternative mRNA splicing events, which include or exclude exon 5 [76]. The major form is composed of seven C-LIKE-DOMAINs, of which domains [D1], [D2] and [D3] are strikingly homologous in both structure and function to [D4], [D5] and [D6] domains [77,78]. The functionally important [D1] domain essential for binding to the integrin ligand contains two rather than one pair of cysteine residues; 1st-CYS 23 and 2nd-CYS 104 correspond to the core disulfide bond of the domain between the B and the F strands, whereas C28 and C108 form an additional disulfide bond between the BC and the FG loops. Mutagenesis studies directed at [D1] have identified two sets of residues involved in binding [79,80]. According to the IMGT unique numbering, these amino acids are D45.1 and P45.3 (CD-STRAND), and G95 (EF-LOOP). P45.3 appears to be particularly important, since its limited conformational freedom brings the Cα atoms of T43 (C-STRAND) and L45.4 (CD-STRAND) within 7 Å of each other [81]. G95 is located in the EF loop, in close proximity of the CD loop (Figure 5). It might interact directly with the integrin ligand, or it might play an indirect role by stabilizing the structure of the CD loop. There is an extensive network of hydrogen bonds between the CD and EF loops, some of which involve the side chain of H100. A cyclic peptide that mimics the CD loop inhibits binding of a4b1 integrin-bearing cells to VCAM1. [D2] may have a role in ligand binding [81].

3. IMGT Protein displays

A comparison between the V-set (V-DOMAIN and V-LIKE-DOMAIN) and C-set (C-DOMAIN and C-LIKE-DOMAIN) domains is shown in the IMGT Protein display (Figure 4). Amino acid positions shown on the upper line in the IMGT Protein displays correspond to equivalent positions in both sets, whereas amino acid positions on the lower lines are characteristic of each set. Sixty-five positions are structurally equivalent between the V-set and the C-set, when the strands A to G are compared. They comprise: positions 1-15 (A strand), 16-26 (B strand), 39-45 (C strand), 77-84 (D strand), 85-96 (E strand), 97-104 (F strand), 118 up to at least 121 (G strand). Thirty-five positions are characteristic of the C-set numbering. The positions of these additional positions compared to the V-set numbering are designated by a number followed by a dot and a number: 1.1 to 1.9 (at the N-terminal end), 15.1 to 15.3 (at the AB turn, for example 15.1 in KIR2DL2 [D1], KIR2DL1 [D1] and VCAM1 [D1]), 45.1 to 45.7 (that represent the CD transversal strand), 84.1 to 84.7, 85.7 to 85.1 (at the DE loop; these positions correspond to longer antiparallel D and E strands in the C-set), 96.1 and 96.2 (at the EF turn). Thirty-three positions are missing in the C-set, compared to the V-set. Two of these positions (37 and 38) are missing in the BC loop. The thirty-one other positions (46 to 76) correspond to the two C' and C" strands and to the C’C” loop (CDR2-IMGT), present in the V-set but absent in the C-set. Positions 45.1 to 45.7 are structurally different between the C-set and the V-set [17]. Indeed, as described above, these positions represent a transversal strand between C and D in the C-set, whereas there are two additional C’ and C” strands in the V-set.

In Figure 4, amino acids which result from the splicing with the preceding exon are shown within parentheses. Indeed, the exact delimitations of the V-LIKE-DOMAINs and C-LIKE-DOMAINs can be usually identified when genomic sequences are available [35]. Each domain shown in Figure 4, except CD4 [D1] and MPZ [D], is encoded by a unique exon.

4. IMGT Alignments of alleles

The IMGT unique numbering allows a standardized description of allele polymorphisms and mutations of the V-LIKE-DOMAINs and C-LIKE-DOMAINs. Alleles from the human FCGR3B have recently been described based on these criteria [37]. The mutations and allelic polymorphisms are described per domain and by comparison to the allele *01 from the IMGT reference directory. Based on these criteria, IMGT ‘Alignments of alleles’ [11,12] allow a standardized display according to the IMGT unique numbering and with the strand and loop delimitations [35]. Other features of the amino acid sequences, such as positions of the N-glycosylation site and amino acids involved in ligand-receptor interactions can easily be visualized.

Conclusion

The IMGT unique numbering gives insight in the structural configuration of the V-LIKE-DOMAINs and C-LIKE-DOMAINs belonging to the human IgSF proteins, but also opens interesting views on the evolution of the sequences of the V-set and C-set [16,17]. Indeed, the IMGT unique numbering can be applied to any IgSF V-set or C-set domain. In the absence of available 3D structures, the V-LIKE-DOMAIN and C-LIKE-DOMAIN IMGT Colliers de Perles are particulary useful for comparison with domains of known 3D structures.

The IMGT unique numbering has many advantages. It allows an easy comparison between sequences coding the V-LIKE-DOMAINs and C-LIKE-DOMAINs, whatever the IgSF protein, the chain type or the species. It has allowed to show that the distinction between C1 and C2 becomes unnecessary (discussed in [17]). It allows, by comparison with genomic sequences, to delimit precisely the V-LIKE-DOMAINs and C-LIKE-DOMAINs. Moreover, it allows to determine the lengths of the BC, C'C" and FG loops of the V-LIKE-DOMAINs and those of the BC loop, CD transversal strand and FG loop of the C-LIKE-DOMAINs. The strand and loop lengths (number of codons or of amino acids, that is number of occupied positions) become crucial information characterizing the V-set and C-set domains, and the corresponding genes, cDNAs and proteins [16,17]. IMGT quality assessment of the data is performed at both the sequence level and 3D structure level. Indeed, the delimitations of the domains are based on the location of the splicing sites in genomic sequences. The IMGT unique numbering has allowed standardized analysis and representations of nucleotide and amino acid sequences (Tables of FR and CDR lengths, Tables of alleles, Alignments of alleles, IMGT Protein displays, IMGT Colliers de Perles, 3D structures). The IMGT unique numbering represents, therefore, a major step forward in analysing and comparing the structure and evolution of the proteins belonging to the immunoglobulin superfamily.

Acknowledgements

We are grateful to Géraldine Folch, Chantal Ginestoux, Joumana Jabado-Michaloud, Dominique Scaviner, Vincent Nègre, Oliver Clément and Denys Chaume for their helpful discussions. E.D. is holder of a doctoral grant from the Ministère de l'Education Nationale, de l'Enseignement Supérieur et de la Recherche (MENESR). K.Q. was the recipient of a doctoral grant from the MENESR and is currently supported by a grant from the Association pour la Recherche sur le Cancer (ARC). IMGT is a registered CNRS mark. IMGT is a RIO platform since 2001 (CNRS, INSERM, CEA, INRA). IMGT was funded in part by the BIOMED1 (BIOCT930038), Biotechnology BIOTECH2 (BIO4CT960037) and 5th PCRDT Quality of Life and Management of Living Resources (QLG2-2000-01287) programmes of the European Union and received subventions from ARC and from the Génopole-Montpellier-

Languedoc-Roussillon. IMGT is currently supported by the Centre National de la Recherche Scientifique (CNRS), the MENESR (Université Montpellier II Plan Pluri-Formation, BIOSTIC-LR2004 Région Languedoc-Roussillon and ACI-IMPBIO IMP82-2004). Part of this work was carried out in the frame of the European Science Foundation Scientific Network Myelin Structure and its role in autoimmunity (MARIE).

References

1. Lefranc M-P, Giudicelli V, Kaas Q, Duprat E, Jabado-Michaloud J, Scaviner D, Ginestoux C, Clément O, Chaume D and Lefranc G. IMGT, the international ImMunoGeneTics information system ®. Nucl. Acids Res. 2005, 33: D593-D597.
2. Giudicelli V, Chaume D, Bodmer J, Müller W, Busin C, Marsh S, Bontrop R, Lemaître M, Malik A and Lefranc M-P. IMGT, the international ImMunoGeneTics database. Nucl. Acids Res. 1997; 25:206-211.
3. Lefranc M-P, Giudicelli V, Busin C, Bodmer J, Müller W, Bontrop R, Lemaître M, Malik A and Chaume D. IMGT, the international ImMunoGeneTics database. Nucl. Acids Res. 1998;26: 297-303.
4. Lefranc M-P, Giudicelli V, Ginestoux C, Bodmer J, Müller W, Bontrop R, Lemaître M, Malik A, Barbié V and Chaume D. IMGT, the international ImMunoGeneTics database. Nucl. Acids Res. 1999;27: 209-212.
5. Ruiz M, Giudicelli V, Ginestoux C, Stoehr P, Robinson J, Bodmer J, Marsh S, Bontrop R, Lemaître M, Lefranc G, Chaume D and Lefranc M-P. IMGT, the international ImMunoGeneTics database. Nucl. Acids Res., 2000;28: 219-221.
6. Lefranc M-P. IMGT ImMunoGeneTics Database. International BIOforum, 2000, 4, 98-100.
7. Lefranc M-P. IMGT, the international ImMunoGeneTics database. Nucl. Acids Res. 2001; 29:207-209.
8. Lefranc M-P. IMGT, the international ImMunoGeneTics database: a high-quality information system for comparative immunogenetics and immunology. Dev. Comp. Immunol., 2002, 26: 697-705.
9. Lefranc M-P. IMGT® databases, web resources and tools for immunoglobulin and T cell receptor sequence analysis, http://imgt.cines.fr. Leukemia 2003; 17:260-266.
10. Lefranc M-P. IMGT-ONTOLOGY and IMGT databases, tools and web resources for immunogenetics and immunoinformatics. Mol. Immunol. 2004; 40:647-659.
11. Lefranc M-P, Giudicelli V, Ginestoux C, Bosc N, Folch G, Guiraudou D, Jabado-Michaloud J, Magris S, Scaviener D, Thouvenin V, Combres K, Girod D, Jeanjean S, Protat C, Yousfi Monod M, Duprat E, Kaas Q, Pommié C, Chaume D and Lefranc G. IMGT-ONTOLOGY for Immunogenetics and Immunoinformatics, http://imgt.cines.fr. Epub 2003; 4:0004 http://www.bioinfo.de/isb/2003/04/0004/, 22 November 2003. *In Silico* Biol. 2004; 4:17-29.
12. Lefranc M-P, Clément O, Kaas Q, Duprat E, Chastellan P, Coelho I, Combres K, Ginestoux C, Giudicelli V, Chaume D and Lefranc G. IMGT-Choreography for Immunogenetics and Immunoinformatics. *In Silico* Biol. 2005, 0006. http://www.bioinfo.de/isb/2004/05/0006.

13. Giudicelli V and Lefranc M-P. Ontology for Immunogenetics: IMGT-ONTOLOGY. Bioinformatics 1999;12:1047-1054.
14. Lefranc M-P. Unique database numbering system for immunogenetic analysis. Immunol. Today 1997;18:509.
15. Lefranc M-P. The IMGT unique numbering for Immunoglobulins, T cell receptors and Ig-like domains. The Immunologist 1999;7:132-136.
16. Lefranc M-P, Pommié C, Ruiz M, Giudicelli V, Foulquier E, Truong L, Thouvenin-Contet V and Lefranc G. IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains. Dev. Comp. Immunol., 2002, 27: 55-77.
17. Lefranc M-P, Pommié C, Kaas Q, Duprat E, Bosc N, Guiraudou D, Jean C, Ruiz M, Da Piédade I, Rouard M, Foulquier E, Thouvenin V and Lefranc G. IMGT unique numbering for immunoglobulin and T cell receptor constant domains and Ig superfamily C-like domains. Dev. Comp. Immunol. 2005, 29:185-203.
18. Lefranc M-P and Lefranc G. The Immunoglobulin FactsBook. London, UK: Academic Press, 458 pages, ISBN:012441351X, 2001.
19. Lefranc M-P and Lefranc G. The T cell receptor FactsBook. London, UK: Academic Press, 398 pages, ISBN:0124413528, 2001.
20. Pallarès N, Frippiat JP, Giudicelli V and Lefranc M-P. The human immunoglobulin lambda variable (IGLV) genes and joining (IGLJ) segments. Exp. Clin. Immunogenet. 1998;15:8-18.
21. Barbié V and Lefranc M-P. The human immunoglobulin kappa variable (IGKV) genes and joining (IGKJ) segments. Exp. Clin. Immunogenet. 1998;15:171-183.
22. Pallarès N, Lefebvre S, Contet V, Matsuda F and Lefranc M-P. The human immunoglobulin heavy variable (IGHV) genes. Exp. Clin. Immunogenet. 1999;16:36-60.
23. Scaviner D, Barbié V, Ruiz M and Lefranc M-P. Protein displays of the human immunoglobulin heavy, kappa and lambda variable and joining regions. Exp. Clin. Immunogenet. 1999;16:234-240.
24. Folch G, Scaviner D, Contet V and Lefranc M-P. Protein displays of the human T cell Receptor alpha, beta, gamma and delta variable and joining regions. Exp. Clin. Immunogenet. 2000; 17:205-215.
25. Folch G and Lefranc M-P. The human T cell receptor beta variable (TRBV) genes. Exp. Clin. Immunogenet. 2000;17:42-54.
26. Scaviner D and Lefranc M-P. The human T cell receptor alpha variable (TRAV) genes. Exp. Clin. Immunogenet. 2000;17:83-96.
27. Bosc N and Lefranc M-P. The mouse (*Mus musculus*) T cell Receptor Beta Variable (TRBV), Diversity (TRBD), and Joining (TRBJ) Genes. Exp. Clin. Immunogenet. 2000;17:216-228.
28. Artero S and Lefranc M-P. The Teleostei immunoglobulin heavy IGH genes. Exp. Clin. Immunogenet. 2000;17:148-161.
29. Artero S and Lefranc M-P: The Teleostei immunoglobulin light IGL1 and IGL2 V, J and C genes. Exp. Clin. Immunogenet. 2000;17:162-172.
30. Bosc N, Contet V and Lefranc M-P. The mouse (*Mus musculus*) T cell Receptor Delta Variable (TRDV), Diversity (TRDD), and Joining (TRDJ) Genes. Exp. Clin. Immunogenet. 2001;18:51-58.

31. Lefranc M-P. Nomenclature of the human immunoglobulin heavy (IGH) genes. Exp. Clin. Immunogenet. 2001; 18:100-116.
32. Lefranc M-P. Nomenclature of the human immunoglobulin kappa (IGK) genes. Exp. Clin. Immunogenet. 2001;18:161-174.
33. Lefranc M-P. Nomenclature of the human immunoglobulin lambda (IGL) genes. Exp. Clin. Immunogenet. 2001;18:242-254.
34. Ruiz M and Lefranc M-P. IMGT gene identification and Colliers de Perles of human immunoglobulin with known 3D structures. Immunogenetics 2002;53:857-883.
35. Bosc N and Lefranc M-P. The mouse (Mus musculus) T cell receptor alpha (TRA) and delta (TRD) variable genes. Dev. Comp. Immunol. 2003; 27:465-97.
36. Pommié C, Levadoux S, Sabatier R, Lefranc G and Lefranc M-P. IMGT standardized criteria for statistical analysis of immunoglobulin V-REGION amino acid properties. J. Mol. Recognition 2004; 17: 17-32.
37. Bertrand G, Duprat E, Lefranc M-P, Marti J and Coste J. Characterization of human FCGR3B*02 (HNA-1b, NA2) cDNAs and IMGT standardized description of FCGR3B alleles. Tissue Antigens 2004; 64:119-31.
38. Williams AF and Barclay AN. The immunoglobulin superfamily-domains for cell surface recognition. Annu Rev Immunol 1988;6:381-405.
39. Bork P, Holm L and Sander C. The immunoglobulin fold. Structural classification, sequence patterns and common core. J Mol Biol 1994;242:309-320.
40. Hunkapiller T and Hood L. Diversity of the immunoglobulin gene superfamily. Adv. Immunol. 1989, 44:1-63.
41. Jones EY. The immunoglobulin superfamily. Curr. Opin. Struct. Biol. 1993; 3:846-852.
42. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN and Bourne PE. The Protein Data Bank. Nucl. Acids Res. 2000; 28:235-242.
43. Hinoda Y, Neumaier M, Hefta SA, Drzeniek Z, Wagener C, Shively L, Hefta LJ, Shively JE and Paxton RJ. Molecular cloning of a cDNA coding biliary glycoprotein I: primary structure of a glycoprotein immunologically crossreactive with carcinoembryonic antigen. Proc. Natl. Acad. Sci. 1988; 85(18):6959-6963.
44. DiSanto JP, Smith D, De Bruin D, Lacy E and Flomenberg N. Transcriptional diversity at the duplicated human CD8 beta loci. Eur. J. Immunol. 1993; 23(2):320-326.
45. Schrewe H, Thompson J, Bona M, Hefta LJ, Maruya A, Hassauer M, Shively JE, von Kleist S and Zimmermann W. Cloning of the complete gene for carcinoembryonic antigen: analysis of its promoter indicates a region conveying cell type-specific expression. Mol. Cell Biol. 1990; 10(6):2738-2748.
46. Chaume D, Giudicelli V and Lefranc M-P. IMGT/LIGM-DB. In: The Molecular Biology Database Collection. Nucl Acids Res 2004;32. http://www3.oup.co.uk/ nar/ database/summary/504
47. Kulikova T, Aldebert P, Althorpe N, Baker W, Bates K, Browne P, Van den Broek A, Cochrane G, Duggan K, Eberhardt R, Faruque N, Garcia-Pastor M, Harte N, Kanz C, Leinonen R, Lin Q, Lombard V, Lopez R, Mancuso R, McHale M, Nardone F, Silventoinen V, Stoehr P, Stoesser G, Tuli MA, Tzouvara K, Vaughan R, Wu D, Zhu W and Apweiler R. The EMBL Nucleotide Sequence Database. Nucl Acids Res 2004;32:D27-D30.

48. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J and Wheeler DL. GenBank: update. Nucl Acids Res 2004;32:D23-D26.
49. Miyazaki S, Sugawara H, Ikeo K, Gojobori T and Tateno Y. DDBJ in the stream of various biological data. Nucl Acids Res 2004;32:D31-D34.
50. Wain HM, Bruford EA, Lovering RC, Lush MJ, Wright MW and Povey S. Guidelines for human gene nomenclature. Genomics 2002;79:464-470.
51. Brunner C, Lassmann H, Waehneldt TV, Matthieu JM and Linington C. Differential ultrastructural localization of myelin basic protein, myelin/oligodendroglial glycoprotein, and 2',3'-cyclic nucleotide 3'-phosphodiesterase in the CNS of adult rats. J. Neurochem. 1989; 52:296-304.
52. Birling MC, Roussel G, Nussbaum F and Nussbaum JL. Biochemical and immunohistochemical studies with specific polyclonal antibodies directed against bovine myelin/oligodendrocyte glycoprotein. Neurochem. Res. 1993; 18:937-45.
53. Gardinier MV, Amiguet P, Linington C and Matthieu JM. Myelin/Oligodendrocyte glycoprotein is a unique member of the immunoglobulin superfamily. J. Neurosci. Res. 1992; 33:177-187.
54. Steinman L. Connections between the immune system and the nervous system. Proc. Nat. Acad. Sci. 1993; 90:7912-7914.
55. Sun J, Link H, Olsson T, Xiao BG, Andersson G, Ekre HP, Linington C and Diener P. T and B cell responses to myelin-oligodendrocyte glycoprotein in multiple sclerosis. J. Immunol. 1991; 146:1490-1495.
56. Kerlero de Rosbo N, Milo R, Lees MB, Burger D, Bernard CCA and Ben-Nun A. Reactivity to myelin antigens in multiple sclerosis: peripheral blood lymphocytes respond predominantly to myelin oligodendrocyte glycoprotein. J. Clin. Invest. 1993; 92:2602-2608.
57. Pham-Dinh D, Della Gaspera B, Kerlero de Rosbo N and Dautigny A. Structure of the human myelin/oligodendrocyte glycoprotein gene and multiple alternative spliced isoforms. Genomics 1995; 29:345-352.
58. Jack LJW and Mather IH. Cloning and analysis of cDNA encoding bovine butyrophilin, an apical glycoprotein expressed in mammary tissue and secreted in association with the milk-fat globule membrane during lactation. J. Biol. Chem. 1990; 265:14481-14486.
59. Kaufman J and Salomonsen J. B-G: We know what it is, but what does it do? Immunol. Today 1992; 13:1-3.
60. Pourquié O, Corbel C, Le Caer JP, Rossier J and Le Douarin N. BEN, a surface glycoprotein of the immunoglobulin superfamily, is expressed in a variety of developing systems. Proc. Nat. Acad. Sci. 1992; 89:5261-5265.
61. Pham-Dinh D, Mattei M-G, Nussbaum J-L, Roussel G, Pontarotti P, Roeckel N, Mather IH, Artzt K, Lindahl KF and Dautigny A. Proc. Nat. Acad. Sci. 1993; 90:7990-7994.
62. Roth M-P, Malfroy L, Offer C, Sevin J, Enault G, Borot N, Pontarroti P and Coppin H. The human myelin oligodendrocyte glycoprotein (MOG) gene: complete nucleotide sequence and structural characterization. Genomics 1995; 28:241-250.
63. Breithaupt C, Schubart A, Zander H, Skerra A, Huber R, Linington C and Jacob U. Structural insights into the antigenicity of myelin oligodendrocyte glycoprotein. Proc. Natl. Acad. Sci. 2003; 100(16):9446-9451.

64. Clements CS, Reid HH, Beddoe T, Tynan FE, Perugini MA, Johns TG, Bernard CC and Rossjohn J. The crystal structure of myelin oligodendrocyte glycoprotein, a key autoantigen in multiple sclerosis. Proc. Natl. Acad. Sci. 2003; 100(19):11059-11064.
65. Thompson JA, Grunert F and Zimmermann W. Carcinoembryonic antigen gene family: molecular biology and clinical perspectives. J. Clin. Lab. Anal. 1991; 5:344-366.
66. Neumaier M, Paululat S, Chan A, Matthaes P and Wagener C. Biliary glycoprotein, a potential human cell adhesion molecule, is down-regulated in colorectal carcinomas. Proc. Nat. Acad. Sci. 1993; 90:10744-10748.
67. Pham-Dinh D, Fourbil Y, Blanquet F, Mattei M-G, Roeckel N, Latour P, Chazot G, Vandenberghe A and Dautigny A. The major peripheral myelin protein zero gene: structure and localization in the cluster of Fc-gamma receptor genes on human chromosome 1q21.3-q23. Hum. Molec. Genet. 1993; 2:2051-2054.
68. Hayasaka K, Himuro M, Wang Y, Takata M, Minoshima S, Shimizu N, Miura M, Uyemura K and Takada G. Structure and chromosomal localization of the gene encoding the human myelin protein zero (MPZ). Genomics 1993; 17:755-758.
69. Kulkens T, Bolhuis PA, Wolterman RA, Kemp S, te Nijenhuis S, Valentijn LJ, Hensels GW, Jennekens FGI, de Visser M, Hoogendijk JE and Baas F. Deletion of the serine 34 codon from the major peripheral myelin protein P(0) gene in Charcot-Marie-Tooth disease type 1B. Nature Genet. 1993; 5:35-39.
70. Tachi N, Ishikawa Y and Minami R. Two cases of congenital hypomyelination neuropathy. Brain Dev. 1984; 6:560-565.
71. Watanabe M, Yamamoto N, Ohkoshi N, Nagata H, Kohno Y, Hayashi A, Tamaoka A and Shoji S. Corticosteroid-responsive asymmetric neuropathy with a myelin protein zero gene mutation. Neurology 2002; 59:767-769.
72. Wagtmann N, Biassoni R, Cantoni C, Verdiani S, Malnati MS, Vitale M, Bottino C, Moretta L, Moretta A and Long EO. Molecular clones of the p58 NK cell receptor reveal immunoglobulin-related molecules with diversity in both the extra- and intracellular domains. Immunity 1995; 2:439-449.
73. Dohring C, Samaridis J and Colonna M. Alternatively spliced forms of human killer inhibitory receptors. Immunogenetics 1996; 44:227-230.
74. Boyington JC and Sun PD. A structural perspective on MHC class I recognition by killer cell immunoglobulin-like receptors. Mol. Immunol. 2001; 38:1007-1021.
75. Winter CC and Long EO. A single amino acid in the p58 killer cell inhibitory receptor controls the ability of natural killer cells to discriminate between the two groups of HLA-C allotypes. J. Immunol. 1997; 158:4026-4028.
76. Cybulsky MI, Fries JWU, Williams AJ, Sultan P, Eddy R, Byers M, Shows T, Gimbrone MA and Collins T. Gene structure, chromosomal location, and basis for alternative mRNA splicing of the human VCAM1 gene. Proc. Nat. Acad. Sci. 1991; 88:7859-7863.
77. Osborn L, Hession C, Tizard R, Vassallo C, Luhowskyj S, Chi-Rosso G and Lobb R. Direct expression cloning of vascular cell adhesion molecule 1, a cytokine-induced endothelial protein that binds to lymphocytes. Cell 1989; 59:1203-1211.
78. Hession C, Tizard R, Vassallo C, Schiffer SB, Goff D, Moy P, Chi-Rosso G, Luhowskyj S, Lobb R and Osborn L. Cloning of an alternate form of vascular cell adhesion molecule-1 (VCAM1). J. Biol. Chem. 1991; 266:6682-6685.

79. Osborn L, Vassallo C, Browning BG, Tizard R, Haskard DO, Benjamin CD, Dougas I and Kirchhausen T. Arrangement of domains, and amino acid residues required for binding of vascular cell adhesion molecule-1 to its counter-receptor VLA-4 (alpha 4 beta 1). J. Cell Biol. 1994; 124:601-608.
80. Vonderheide RH, Tedder TF, Springer TA and Staunton DE. Residues within a conserved amino acid motif of domains 1 and 4 of VCAM-1 are required for binding to VLA-4. J. Cell Biol. 1994; 125:215-22.
81. Wang JH, Pepinsky RB, Stehle T, Liu JH, Karpusas M, Browning B and Osborn L. The crystal structure of an N-terminal two-domain fragment of vascular cell adhesion molecule 1 (VCAM-1): a cyclic peptide based on the domain 1 C-D loop can inhibit VCAM-1-alpha 4 integrin interaction. Proc. Nat. Acad. Sci. 1995; 92:5714-5718.