BRIEFINGS IN FUNCTIONAL GENOMICS AND PROTEOMICS. VOL 6. NO 4. 253–264 doi:10.1093/bfgp/elm032

IG, TR and IgSF, MHC and MhcSF: what do we learn from the IMGT Colliers de Perles?

Quentin Kaas, François Ehrenmann and Marie-Paule Lefranc

Advance Access publication date 21 January 2008

Abstract

The immunoglobulin superfamily (IgSF) comprises the immunoglobulins (IG), T cell receptors (TR) and proteins that have the common feature of having at least one Ig-like domain. The major histocompatibility complex (MHC) superfamily (MhcSF) comprises, in addition to the MHC, proteins which share the common feature of having Mhc-like domains. IMGT®, the international ImMunoGeneTics information system® (http://imgt.cines.fr) has set up a unique numbering system and standardized 2D graphical representations, or IMGT Colliers de Perles, which take into account the structural features of the Ig-like and Mhc-like domains. In this article, we review the IMGT Scientific chart rules for the description of the IgSF (V and C types) and of the MhcSF (G type) domains. These rules are based on the IMGT-ONTOLOGY axioms and concepts and are applicable for the sequence and structure analysis, whatever the species, the IgSF or MhcSF protein, or the chain type. These IMGT Colliers de Perles are particularly useful for antibody engineering, sequence–structure analysis, visualization and comparison of positions for mutations, polymorphisms and contact analysis.

Keywords: *IMGT Collier de Perles; Ig-like; Mhc-like; immunoglobulin superfamily IgSF; major histocompatibility complex superfamily MhcSF; domain*

INTRODUCTION

The common feature of the immunoglobulin superfamily (IgSF) proteins is to have at least one immunoglobulin-like (Ig-like) domain [1, 2]. Despite a large divergence in the amino acid sequences, the Ig-like domains share the structural Ig-fold, which typically consists of about 100 amino acids in anti-parallel β-strands, linked by β-turns or loops and located on two layers maintained by a disulfide bridge [1, 2]. The number of anti-parallel β-strands defines two domain types: nine strands for the V type [which comprises the V-DOMAIN (IMGT labels of the IMGT-ONTOLOGY DESCRIPTION axiom are written in capital letters. Definitions of IMGT labels are available in the IMGT Scientific chart at http://imgt.cines.fr) of the immunoglobulins (IG) and T cell receptors (TR), and the V-LIKE-DOMAIN of the IgSF proteins other than the IG or TR] [1]; and seven strands for the C type (which comprises the C-DOMAIN of the IG and TR, and the C-LIKE-DOMAIN of the IgSF proteins other than the IG or TR) [2]. Thus, the IgSF comprises the IG and TR proteins (each chain has one V-DOMAIN and one or several C-DOMAINs) and the proteins other than IG and TR defined as having at least one V-LIKE-DOMAIN or one C-LIKE-DOMAIN. The IG and TR proteins are involved in antigen recognition, whereas the other IgSF proteins

Corresponding author. 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. Tel: +33 (0)4 99 61 99 65; Fax: +33 (0)4 99 61 99 01; E-mail: Marie-Paule.Lefranc@igh.cnrs.fr

Quentin Kaas, PhD, has developed IMGT/3Dstructure-DB, the database specialized in the analysis of the 3D structures of the IG, TR and MHC. He is presently a PostDoc at the University of Queensland, Brisbane, Australia.

François Ehrenmann, Engineer in bioinformatics, is in charge of the developments of IMGT/3Dstructure-DB, at the Laboratoire d'ImmunoGénétique Moléculaire, Institut de Génétique Humaine CNRS, Montpellier, France.

Marie-Paule Lefranc, PhD, is Professor at the Université Montpellier 2, Head of the Laboratoire d'ImmunoGénétique Moléculaire at the Institut de Génétique Humaine CNRS. She is director of IMGT®, the international ImMunoGeneTics information system® (http://imgt.cines.fr) that she founded in 1989, at Montpellier, France.

are involved in many different functions (in ligand–receptor interactions in development, differentiation, activation, adhesion, regulation, etc.) [3–10].

The common feature of the major histocompatibility complex MHC superfamily (MhcSF) proteins is to have two Mhc-like domains which together contribute to a similar groove 3D structure that consists of one sheet of eight anti-parallel β-strands ('floor' of the groove or platform) and two helical regions ('walls' of the groove) [11, 12]. Each domain made of four anti-parallel β-strands and one helix belongs to the G type (which comprises the G-DOMAIN of the MHC, and the G-LIKE-DOMAIN of the MhcSF proteins other than the MHC) [12]. Thus, the MhcSF comprises the MHC proteins that belong to two classes, MHC class I [one chain has two G-DOMAINs and one C-LIKE-DOMAIN, associated to the β-2-microglobulin (B2M)] and MHC class II (each chain has one G-DOMAIN and one C-LIKE-DOMAIN), and the proteins other than MHC defined as having a groove-like domain made up of two G-LIKE-DOMAINs, associated or not to one C-LIKE-DOMAIN [12]. The MHC proteins are involved in the antigen presentation to the T cells, whereas the other MhcSF proteins are involved in different functions [display of phospholipid antigens for CD1, iron homeostasis for the hemochromatosis protein (HFE), maternal IG transport for the IgG Fc receptor and transporter alpha (FCGRT), stress induced for the MHC class I polypeptide-related sequence A (MICA), etc.] [11–15].

IG, TR and MHC data, extended to IgSF and MhcSF, are annotated in IMGT®, the international ImMunoGeneTics information system® (http://imgt.cines.fr), a high quality integrated knowledge resource, which is the international reference in immunogenetics and immunoinformatics [16]. IMGT annotations are according to the IMGT Scientific chart rules, based on the IMGT-ONTOLOGY axioms and concepts [17, 18]. These include standardized IMGT gene and allele names (CLASSIFICATION axiom) [19–22], standardized IMGT labels for the receptors, chains, domains and regions (DESCRIPTION axiom) [23, 24], standardized amino acid positions according to the IMGT unique numbering (NUMEROTATION axiom) [1, 2, 12, 25, 26]. The IMGT standardization is used in the IMGT gene, sequence and structure databases [21, 27, 28], in the IMGT on-line tools [28–30] and in the IMGT knowledge web resources (IMGT Protein displays, IMGT Alignments of alleles, IMGT Colliers de Perles, etc.) [23, 24]. IMGT Colliers de Perles [31–33] are standardized graphical 2D representations, based on the IMGT unique numbering [1, 2, 12]. IMGT Colliers de Perles are available for the IgSF V type domains, based on the IMGT unique numbering for V-DOMAIN and V-LIKE-DOMAIN [1], for the IgSF C type domains, based on the IMGT unique numbering for C-DOMAIN and C-LIKE-DOMAIN [2], and for the MhcSF G type domains, based on the IMGT unique numbering for G-DOMAIN and G-LIKE-DOMAIN [12]. IMGT Colliers de Perles are provided in IMGT/3Dstructure-DB [28] for V type, C type and G type domains for which 3D structures are available. They can also be obtained on-line, starting from V type, C type or G type domain amino acid sequences, using the IMGT/DomainGapAlign and IMGT/Collier-de-Perles tools (http://imgt.cines.fr) [28], or for user V-DOMAIN nucleotide sequences, using the IMGT/V-QUEST tool [29]. IMGT Colliers de Perles provide a standardized delimitation of the strands (framework regions, FR-IMGT) and loops (complementarity determining regions, CDR-IMGT) of the V-DOMAINs and V-LIKE-DOMAINs, of the strands and loops of the C-DOMAINs and C-LIKE-DOMAINs, and of the strands and helices of the G-DOMAINs and G-LIKE-DOMAINs. By taking into account the structural features of the Ig- and Mhc-like domains, the IMGT Colliers de Perles based on the IMGT unique numbering allow to bridge the gap between linear amino acid sequences and 3D structures.

In this article, we review the IMGT Scientific chart rules for the description of the IgSF V type and C type and of the MhcSF G type IMGT Colliers de Perles, which are applicable whatever the species, the protein or the chain type. This standardization is particularly useful in the absence of 3D structural data, for antibody engineering design, for visualization and comparison of positions for mutations, polymorphisms and contact analysis.

IG, TR AND IgSF IMGT COLLIERS DE PERLES

IG, TR and IgSF chains and domains

The IG and TR proteins are antigen receptors formed, for the IG, by four chains (two identical heavy chains and two identical light chains) and, for the TR, by two chains of similar length (α- and β-chains, or γ- and δ-chains, depending on the receptor type) (Figure 1).

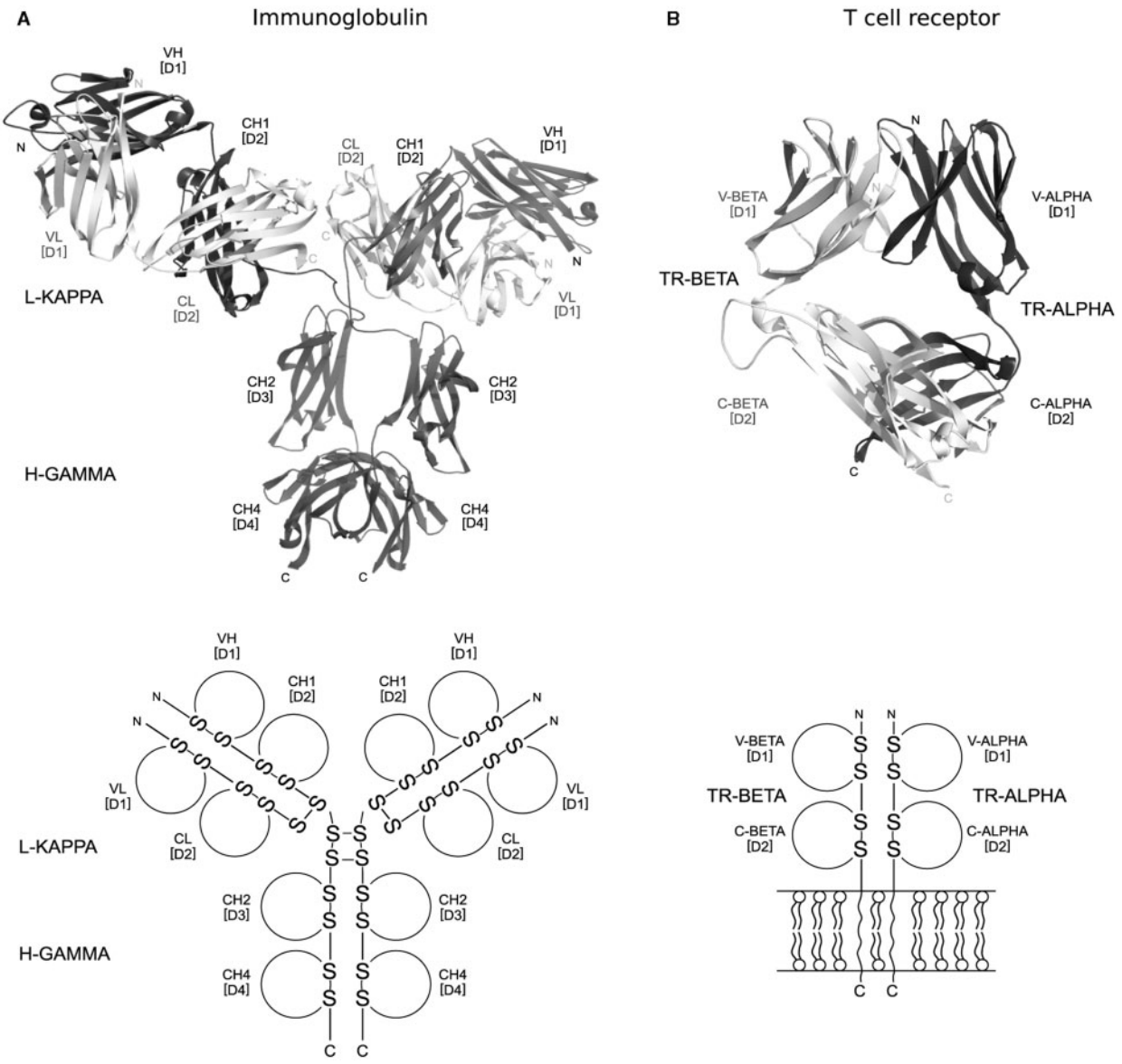

Figure I: The 3D structures and schematic representations of IG and TR proteins. (**A**) 3D structure and representation of an IG. An IG (here an IgGκ) comprises four chains, two identical disulfide-linked heavy chains and two identical light chains, each one disulfide-linked to an heavy chain. The heavy chain comprises a N-terminal V-DOMAIN (VH) and several C-terminal C-DOMAINs (number depending on the heavy chain type, here, CHI, CH2 and CH3 for H-GAMMA). The VH domain results from the junction of three genes (variable V, diversity D and junction J) and corresponds at the sequence level to the V-D-J-REGION [19]. The CH domains are part of the C-REGION that is coded by a constant C gene. The light chain comprises a N-terminal V-DOMAIN (VL, that can be V-KAPPA or V-LAMBDA depending on the light chain type) and one C-terminal C-DOMAIN (CL, that can be C-KAPPA or C-LAMBDA). The VL domain results from the junction of two genes (V and J) and corresponds to the V-J-REGION. The CL domain correspond to the C-REGION and is coded by a C gene. (**B**) 3D structure and representation of a TR. A TR comprises two chains, TR-ALPHA and TR-BETA (disulfide-linked) for the TR-ALPHA.BETA, TR-GAMMA and TR-DELTA (disulfide-linked or not) for the TR-GAMMA.DELTA. The TR chains comprise a N-terminal V-DOMAIN and one C-terminal C-DOMAIN. The V-ALPHA and V-GAMMA result from the junction of two genes (V and J) and correspond to the V-J-REGION, whereas the V-BETA and V-DELTA result from the junction of three genes (V, D and J) and correspond to the V-D-J-REGION [20]. The C-ALPHA, C-BETA, C-GAMMA and C-DELTA are part of the C-REGION, that also comprises the connecting region CO, transmembrane region TM, cytoplasmic region CY (not present in the 3D structures), and which is coded by a C gene [20]. [DI], [D2], . . . , indicate the positions of the domains from the N-terminal end of the chains.

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,. . .). 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 (for review, see [19, 20] and IMGT Scientific chart, http://imgt.cines.fr 'Correspondence between labels for IG and TR domains').

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 [3, 10]. As examples, the MOG and MPZ (or P0) proteins have a unique N-terminal V-LIKE-DOMAIN, the CEA family proteins have a single N-terminal V-LIKE-DOMAIN followed by a variable number of C-LIKE-DOMAINs and the VCAM1 protein is composed of seven C-LIKE-DOMAINs [3, 34–36]. IgSF proteins with diverse V type domain and C type domain combinations, interspersed or not with domains belonging to other types, are continuously described [10].

IMGT Colliers de Perles for V-DOMAIN and V-LIKE-DOMAIN

The IMGT Colliers de Perles for V-DOMAIN (IG and TR) and V-LIKE-DOMAIN (IgSF other than IG and TR) are based on the IMGT unique numbering for V-DOMAIN and V-LIKE-DOMAIN [1]. Indeed, the 3D structure of a V-LIKE-DOMAIN is very similar to that of an IG and TR V-DOMAIN (Figure 2A, Table 1). Both domains are made of nine anti-parallel β-strands (A, B, C, C′, C″, D, E, F and G) linked by β-turns (AB, CC′, C″D, DE and EF) or loops (BC, C′C″ and FG), forming a sandwich of two sheets. The sheets are closely packed against each other through hydrophobic interactions giving a hydrophobic core and joined together by a disulfide bridge between the B-STRAND in the first sheet and the F-STRAND in the second sheet [7, 28]. 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 amino acid 89 and cysteine 104 (2nd-CYS).

The hydrophobic amino acids of the framework regions are also found in conserved positions [1]. 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 and the sequence divergence of the IgSF domains. Indeed, the V-LIKE-DOMAIN is usually encoded by a unique exon, whereas the IG and TR V-DOMAIN results from the rearrangement of two (V, J) or three (V, D, J) genes [19, 20]. The V-LIKE-DOMAIN is usually, as the IG and TR V-DOMAIN, the most N-terminal (and extracellular) domain of the protein. 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 protein and interspersed with C-LIKE-DOMAINs or with domains of other superfamilies.

The anti-parallel β-strands of the V-LIKE-DOMAIN correspond to the conserved framework regions (FR-IMGT) described in the IG and TR V-DOMAIN, whereas the BC, C′C″ and FG loops correspond to the CDR-IMGT [1] (Table 1). The loop length (number of amino acids or by extrapolation number of codons that is number of occupied positions) is a crucial and original concept of IMGT-ONTOLOGY [17]. The lengths of the BC (CDR1-IMGT), C′C″ (CDR2-IMGT) and FG (CDR3-IMGT) loops characterize the V-DOMAIN and V-LIKE-DOMAIN. Thus, the length of the three loops BC, C′C″ and FG is shown, in number of amino acids (or codons), into brackets and separated by dots. In Figure 2A, the CDR-IMGT lengths are [8.8.12].

IMGT Colliers de Perles for C-DOMAIN and C-LIKE-DOMAIN

The IMGT Colliers de Perles for C-DOMAIN (IG and TR) and C-LIKE-DOMAIN (IgSF other than IG and TR) is based on the IMGT unique numbering for C-DOMAIN and C-LIKE-DOMAIN [2]. This numbering is itself derived from the IMGT unique numbering first described for the V-REGION [25, 26] and for the V-DOMAIN [1]. Indeed, the sandwich β-sheet of the C type (C-DOMAIN and C-LIKE-DOMAIN) has the same topology and similar 3D structure than the V type (V-DOMAIN and V-LIKE-DOMAIN), but they differ by the number of strands (Figure 2B, Table 2). The C-DOMAIN and C-LIKE-DOMAIN are made of seven β-strands

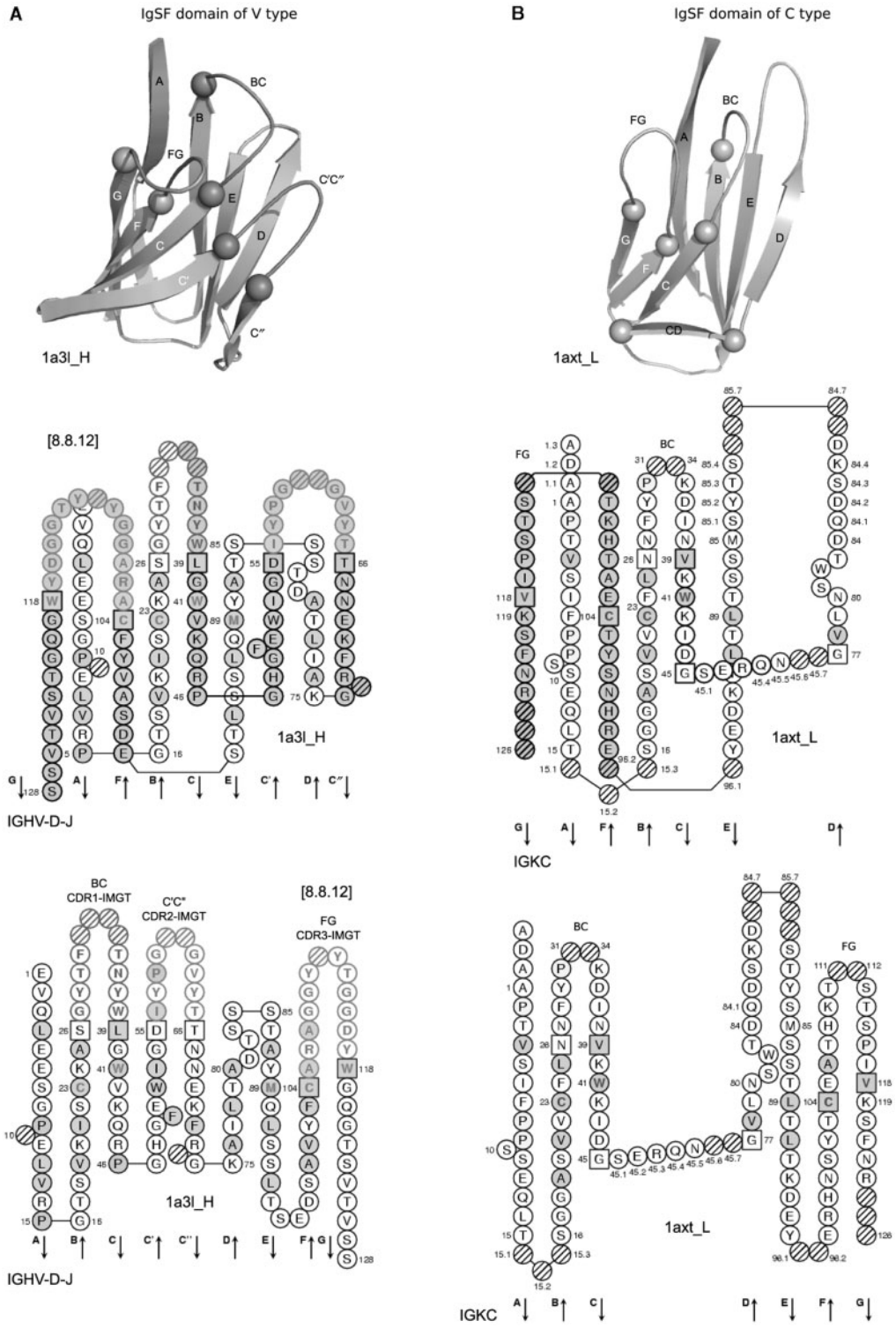

Figure 2: Ribbon representations and IMGT Colliers de Perles for V type and C type domains. (**A**) IgSF domain of V type, based on the IMGT unique numbering for V-DOMAIN and V-LIKE-DOMAIN [1]. (**B**) IgSF domain of C type, based on the IMGT unique numbering for C-DOMAIN and C-LIKE-DOMAIN [2]. For the V type and C type domains, the 3D structure ribbon representation (upper part), IMGT Collier de Perles on two layers (middle part) and IMGT Collier de Perles on one layer (bottom part) are shown. Squares indicate positions that belong to strands and represent anchor positions for BC-LOOP, C′C″-LOOP and FG-LOOP of the V type, and for BC-LOOP, CD-STRAND and FG-LOOP of the C type. Amino acids are shown in the one-letter abbreviation. Hatched circles correspond to missing positions according to the IMGT unique numbering.

Table I: V type domain (V-DOMAIN and V-LIKE-DOMAIN)

Strands and loops	IMGT positions[a]	Lengths[b]	Characteristic positions	FR-IMGT and CDR-IMGT in V-DOMAIN
A-STRAND	1–15	15 (14 if gap at 10)		FR1-IMGT
B-STRAND	16–26	11	1st-CYS 23	
BC-LOOP	27–38	12 (or less)		CDR1-IMGT
C-STRAND	39–46	8	CONSERVED-TRP 41	FR2-IMGT
C′-STRAND	47–55	9		
C′C″-LOOP	56–65	10 (or less)		CDR2-IMGT
C″-STRAND	66–74	9 (or 8 if gap at 73)		FR3-IMGT
D-STRAND	75–84	10 (or 8 if gaps at 81 to 82)		
E-STRAND	85–96	12	hydrophobic 89	
F-STRAND	97–104	8	2nd-CYS 104	
FG-LOOP	105–117	13 (or less, or more)		CDR3-IMGT
G-STRAND	118–128	11	(1)	FR4-IMGT

[a]Based on the IMGT unique numbering for V-DOMAIN and V-LIKE-DOMAIN [1].
[b]In number of amino acids (or codons).
(1) 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.

Table 2: C type domain (C-DOMAIN and C-LIKE-DOMAIN)

Strands, loops and turns	IMGT positions[a]	Lengths[b]	Characteristic positions
A-STRAND	1–15	15	
AB-TURN	15.1–15.3	0–3	
B-STRAND	16–26	11	1st-CYS 23
BC-LOOP	27–38	10 (or less)	no 32, 33[c]
C-STRAND	39–45	7	CONSERVED-TRP 41 no 46[c]
CD-STRAND	45.1–45.9	1–9	
D-STRAND	77–84	8	no 75, 76[c]
DE-TURN	84.1–84.7, 85.7–85.1	0–14	
E-STRAND	85–96	12	hydrophobic 89
EF-TURN	96.1–96.2	0–2	
F-STRAND	97–104	8	2nd-CYS 104
FG-LOOP	105–117	13 (or less, or more)	
G-STRAND	118–128	11 (or less)	

[a]Based on the IMGT unique numbering for C-DOMAIN and C-LIKE-DOMAIN [2].
[b]In number of amino acids (or codons).
[c]Compared to V type.

linked by β-turns or loops, and arranged so that four strands form one sheet and three strands form a second sheet [2]. A characteristic transversal CD-STRAND links the two sheets; depending on the CD-STRAND length, the D-STRAND is in the first or in the second sheet [2].

The C′-STRAND, C′C″-LOOP and C″-STRAND are missing in the C type and are replaced by the characteristic transversal CD-STRAND [2]. Additional positions in the C type define the AB-TURN, DE-TURN and EF-TURN [2].

MHC AND MhcSF IMGT COLLIERS DE PERLES

MHC and MhcSF chains and domains

The MHC proteins belong to two classes: MHC-I and MHC-II (Figure 3). The MHC-I proteins, expressed on the cell surface of most cells, are formed by the association of a transmembrane heavy chain (I-ALPHA chain) and a non-covalently linked light chain β-2-microglobulin (B2M) [12]. The MHC-II proteins, expressed on the cell surface of professional antigen presenting cells (APC), are heterodimers formed by the association of two transmembrane chains, an α-chain (II-ALPHA chain) and a β-chain (II-BETA chain) [12].

The I-ALPHA chain of the MHC-I, and the II-ALPHA and II-BETA chains of the MHC-II proteins, comprise an extracellular region made of three domains for the MHC-I chain and of two domains for each MHC-II chain, a connecting region, a transmembrane region and an intracytoplamic region. The I-ALPHA chain comprises two groove domains (G-DOMAINs), the G-ALPHA1 [D1] and G-ALPHA2 [D2] domains, and one C-LIKE-DOMAIN [D3] [12]. The II-ALPHA chain and the II-BETA chain each comprises two domains, the G-ALPHA [D1] and one

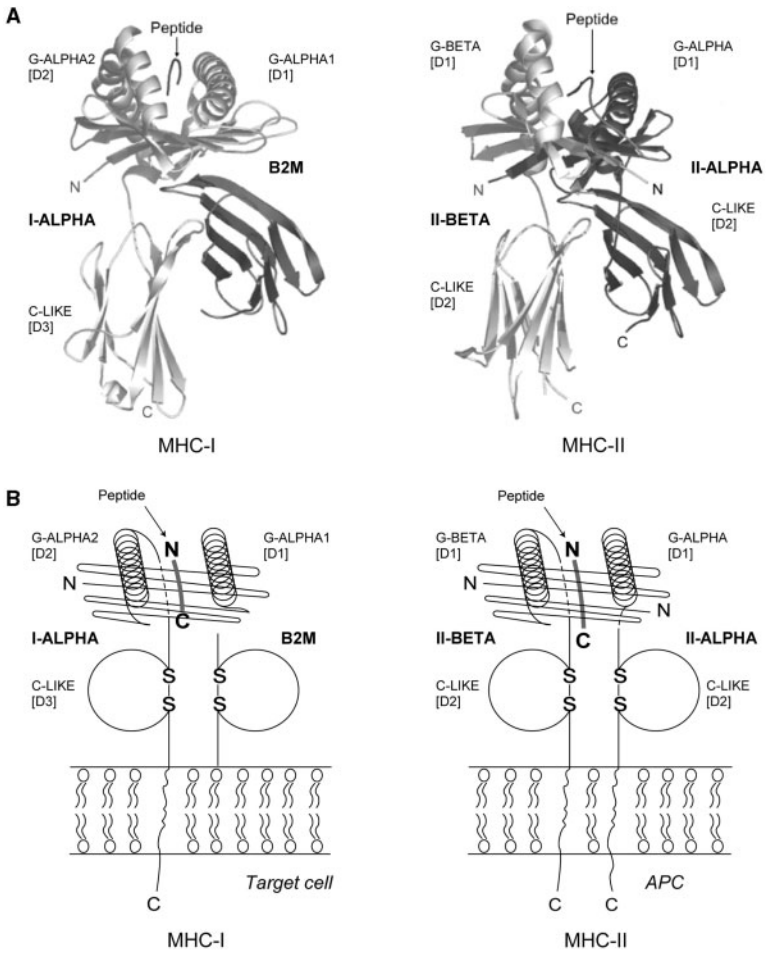

Figure 3: The 3D structures and schematic representations of the MHC-I and MHC-II proteins. (**A**) 3D structures of MHC-I and MHC-II. The MHC-I comprises the I-ALPHA and the B2M chains. The I-ALPHA chain is shown with its extracellular domains (G-ALPHAI, G-ALPHA2 and C-LIKE) [I2]. The MHC-II comprises the II-ALPHA and II-BETA chains that are shown with their extracellular domains (G-ALPHA and C-LIKE for the II-ALPHA chain, G-BETA and C-LIKE for the II-BETA chain). (**B**) Schematic representations of the MHC-I and MHC-II proteins. The MHC-I and MHC-II are shown as transmembrane proteins, at the surface of a target cell and of an APC, respectively. Complete MHC-I and MHC-II chains comprise the extracellular domains (shown in A) and the connecting, transmembrane and cytoplamic regions (not present in 3D structures). [DI], [D2] and [D3] indicate the position of the domains from the N-terminal end of the chains. Arrows indicate the peptide localization in the MHC groove (the N-terminal end of the peptide is in the back).

C-LIKE-DOMAIN [D2], and the G-BETA [D1] and one C-LIKE-DOMAIN [D2], respectively. The four G-DOMAINs, G-ALPHA1 and G-ALPHA2 of the MHC-I proteins, and G-ALPHA and G-BETA of the MHC-II proteins have a similar groove 3D structure [12]. Interestingly, this groove is found in the 'classical' MHC (MHC-Ia and MHC-IIa) proteins that present peptides to the T cells (the groove is part of the cleft that is the peptide binding site), and also in 'non-classical' MHC (MHC-Ib and MHC-IIb) proteins with more specific functions or which do not present peptides to the T cells [12].

This groove has also been found in proteins other than MHC, which are said to belong to the MhcSF, but so far, only MHC-I-like chains have been identified in the MhcSF [14, 15, 37, 38]. These chains either include a C-LIKE-DOMAIN and are bound to the B2M (e. g. CD1, FCGRT, HFE, MR1) or not (MIC, AZGP1), or do not even include a C-LIKE-DOMAIN (EPCR, RAE) [14]. The G-ALPHA1-LIKE [D1] and G-ALPHA2-LIKE [D2] domains of these proteins show a striking structural homology with the MHC G-ALPHA1 and G-ALPHA2 domains and this, despite a high sequence divergence [12].

IMGT Colliers de Perles for G-DOMAIN and G-LIKE-DOMAIN

The IMGT Colliers de Perles for G-DOMAIN (MHC) and G-LIKE-DOMAIN (MhcSF other than MHC) is based on the IMGT unique numbering for G-DOMAIN and G-LIKE-DOMAIN [12] (Figure 4, Table 3).

For each G type domain (G-DOMAIN and G-LIKE-DOMAIN), the positions that contribute to the groove floor comprise positions 1–49, with four strands linked by turns [12]. The numbering of the G type helix starts at position 50 and ends at position 92. Interestingly, we showed that, despite the high sequence divergence, only five additional positions (54A, 61A, 61B, 72A and 92A) are necessary to align any G-DOMAIN and G-LIKE-DOMAIN [12, 33]. It is worthwhile to note that position 54A in G-ALPHA1-LIKE is the only additional position needed to extend the IMGT numbering for G-DOMAIN to the G-LIKE-DOMAINs.

The helix (positions 50–92) seats on the beta sheet and its axis forms an angle of about 40° with the β-strands. Two cysteines, CYS-11 (in strand A) and CYS-74 (in the helix) are well conserved in the G-ALPHA2 and G-BETA domains, where they participate to a disulfide bridge that fastens the helix on the groove floor. The IMGT Colliers de Perles allow to describe specific features (detailed in [12, 33]). As an example, the G-ALPHA1 and G-ALPHA domains have a conserved *N*-glycosylation site at position 86 (N-X-S/T, where *N* is asparagine, X any amino acid except proline, S is serine and T is threonine).

WHAT DO WE LEARN FROM IMGT COLLIERS DE PERLES?

Any domain represented by an IMGT Collier de Perles is characterized by the length of its strands, loops and turns and, for the G type, by the length of its helix [1, 2, 12]. The strand, loop, turn or helix lengths (the number of amino acids or codons, that is the number of occupied positions) become crucial information which characterizes the domains. This first feature of the IMGT standardization based on the IMGT unique numbering allowed, for instance, to show that the distinction between the C1, C2, I1 and I2 domain types found in the literature and in the databases to describe the IgSF C type domains is unnecessary and moreover unapplicable when dealing with sequences for which no structural data are known (discussed in [2]).

A second feature of the IMGT standardization is the comparison of cDNA and/or amino acid sequences with genomic sequences, and the identification of the splicing sites, to delimit precisely the domains: a V-LIKE-DOMAIN, a C-DOMAIN, a C-LIKE-DOMAIN, a G-DOMAIN or a G-LIKE-DOMAIN is frequently encoded by a unique exon [1, 2, 12]. This IMGT standardization for the domain delimitation explains the discrepancies observed with the generalist UniProt/Swiss-Prot database, which identifies domains based on amino acid sequences and does not take into account the genomic information. The IMGT Colliers de Perles also put the question of the leader region. Indeed, the N-terminal end of the first domain of an IgSF or MhcSF chain depends on the proteolytic cleavage site of the leader region (peptide signal), which is rarely determined experimentally. When this site is not known, the IMGT Colliers de Perles start with the first amino acid resulting from the splicing ('Splicing sites' in IMGT Aide-mémoire, http://imgt.cines.fr). For IG and TR V-DOMAIN the leader proteolytic site is known (or is extrapolated) and the IMGT Colliers de Perles start with the first amino acid of the V-REGION [19, 20].

The IMGT Colliers de Perles allow a precise visualization of the inter-species differences for the IgSF V and C type domain strands and loops, and MhcSF G type domain strands and helix, even in the absence of 3D structures. This has been applied to the teleost CD28 family members and their B7 family ligands and to the B and T lymphocyte associated (BTLA) protein, which belong to the IgSF by their V type and/or C type domains [9]. The IMGT Colliers de Perles are particularly useful in molecular engineering and antibody humanization design based on CDR grafting. Indeed they allow to precisely define the CDR-IMGT and to easily compare the amino acid sequences of the four FR-IMGT (FR1-IMGT: positions 1–26, FR2-IMGT: 39–55, FR3-IMGT: 66–104 and FR4-IMGT: 118–128) between the murine and the closest human V-DOMAINs. A recent analysis performed on humanized antibodies used in oncology underlines the importance of a correct delimitation of the CDR regions to be grafted [39].

The IMGT Colliers de Perles also allow a comparison to the IMGT Collier de Perles statistical profiles for the human expressed IGHV, IGKV and

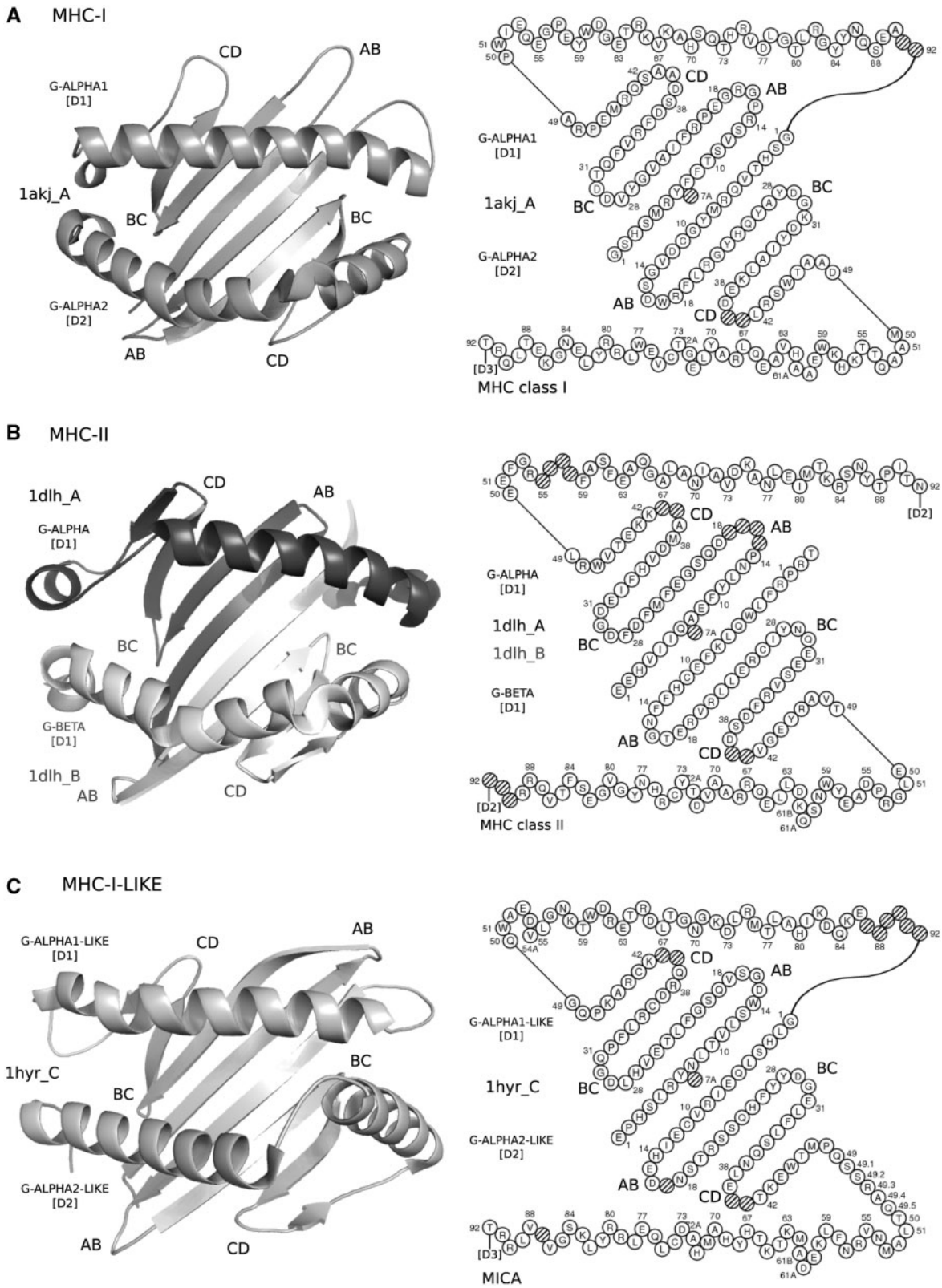

Figure 4: Ribbon representations and IMGT Colliers de Perles for G type domains. (**A**) MHC-I, (**B**) MHC-II, (**C**) MHC-I-like. The MhcSF domain of G type is based on the IMGT unique numbering for G-DOMAIN and G-LIKE-DOMAIN [12]. Note that the N-terminal end of a peptide in the cleft of MHC-I and MHC-II G-DOMAINs would be on the left hand side. Amino acids are shown in the one-letter abbreviation. Hatched circles correspond to missing positions according to the IMGT unique numbering.

Table 3: G type domain (G-DOMAIN and G-LIKE-DOMAIN)

Strands, turns and helix	IMGT positions[a]	Lengths[b]	Characteristic positions[c]
A-STRAND	1–14	14	7A, CYS-11
AB-TURN	15–17	3 (or 0)	
B-STRAND	18–28	11	
BC-TURN	29–30	2	
C-STRAND	31–38	8	
CD-TURN	39–41	3 (or 1)	
D-STRAND	42–49	8	
HELIX	50–92	43–48	54A, 61A, 61B, 72A, CYS-74, 92A

[a]Based on the IMGT unique numbering for G-DOMAIN and G-LIKE-DOMAIN [12].
[b]In number of amino acids (or codons).
[c]For more details, see [12].

IGLV repertoires [32]. These statistical profiles are based on the definition of eleven IMGT amino acid physicochemical characteristics classes which take into account the hydropathy, volume and chemical characteristics of the 20 common amino acids [32] ('Amino acids' in IMGT Aide-mémoire, http://imgt.cines.fr). The statistical profiles identified positions that are conserved for the physicochemical characteristics: 41 FR-IMGT positions for the human IGHV and 59 FR-IMGT positions for the human IGKV and IGLV at >80% threshold (see Plate 3 in [32]). After assignment of the IMGT Colliers de Perles amino acids to the IMGT amino acid physicochemical classes, comparison can be made with the statistical profiles of the human expressed repertoires. This comparison is useful to identify potential immunogenic residues at given positions in chimeric or humanized antibodies [39] or to evaluate immunogenicity of primate antibodies [40].

IMGT Colliers de Perles are also of interest when 3D structures are available. In IMGT/3Dstructure-DB [28], 'IMGT Collier de Perles on two layers' are displayed with hydrogen bonds for V type and C type domains. Clicking on a residue in 'IMGT Collier de Perles on one layer' gives access to the corresponding IMGT Residue@Position card, which provides the atom contact types and atom contact categories for that amino acid. IMGT Colliers de Perles display the IMGT pMHC contact sites for 3D structures with peptide/MHC (pMHC) complexes [11], which can be compared with the pMHC contact sites available in IMGT/3Dstructure-DB.

The IMGT Colliers de Perles for the V type, C type and G type, based on the IMGT unique numbering, therefore represent a major step forward for the comparative analysis of the sequences and structures of the IgSF and MhcSF domains, for the study of their evolution and for the applications in antibody engineering, IG and TR repertoires in autoimmune diseases and leukemias [41], pMHC contact analysis, and more generally ligand–receptor interactions involving V type, C type and/or G type domains.

Key Points

- The IMGT Colliers de Perles based on the IMGT unique numbering allow to compare V, C and G type domains of the IgSF and MhcSF proteins, whatever the species, the receptor or the chain type may be.
- The IMGT Colliers de Perles are used in antibody humanization design based on CDR grafting, to precisely define the CDR-IMGT to be grafted.
- The IMGT Colliers de Perles statistical profiles for the human expressed IGHV, IGKV and IGLV repertoires help to identify potential immunogenic residues at given positions in chimeric or humanized antibodies.
- The IMGT Colliers de Perles gives access, in IMGT/3D structure-DB, to the IMGT Residue@Position cards, which provide the atom contact types and atom contact categories.
- The IMGT Colliers de Perles bridge the gap between linear amino acid sequences and 3D structures, as illustrated by the display of hydrogen bonds (for V and C type domains) and pMHC contact sites (for G type domains) in IMGT/3Dstructure-DB.

Acknowledgements

We are grateful to Gérard Lefranc for helpful discussion and to the IMGT team for its expertise and constant motivation. This work was funded by Centre National de la Recherche Scientifique CNRS; Ministère de l'Enseignement Supérieur et de la Recherche MESR (ACI-IMPBIO IMP82-2004, Université Montpellier 2 Plan Pluri-Formation); Région Languedoc-Roussillon; Agence Nationale de la Recherche ANR BYOSIS (ANR-06-BYOS-0005-01); European Community ImmunoGrid project (IST-2004-0280069).

References

1. Lefranc M-P, Pommié C, Ruiz M, *et al*. IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains. *Dev Comp Immunol* 2003;**27**:55–77.
2. Lefranc M-P, Pommié C, Kaas Q, *et al*. IMGT unique numbering for immunoglobulin and T cell receptor

constant domains and Ig superfamily C-like domains. *Dev Comp Immunol* 2005;**29**:185–203.

3. Duprat E, Kaas Q, Garelle V, *et al*. IMGT standardization for alleles and mutations of the V-LIKE-DOMAINs and C-LIKE-DOMAINs of the immunoglobulin superfamily. *Recent Res Devel Human Genet* 2004;**2**: 111–36.
4. Williams AF, Barclay AN. The immunoglobulin superfamily-domains for cell surface recognition. *Annu Rev Immunol* 1988;**6**:381–405.
5. Hunkapiller T, Hood L. Diversity of the immunoglobulin gene superfamily. *Adv Immunol* 1989;**44**:1–63.
6. Jones EY. The immunoglobulin superfamily. *Curr Opin Struct Biol* 1993;**3**:846–52.
7. Bork P, Holm L, Sander C. The immunoglobulin fold. Structural classification, sequence patterns and common core. *J Mol Biol* 1994;**242**:309–20.
8. Bertrand G, Duprat E, Lefranc M-P, *et al*. Characterization of human FCGR3B*02 (HNA-1b, NA2) cDNAs and IMGT standardized description of FCGR3B alleles. *Tissue Antigens* 2004;**64**:119–31.
9. Bernard D, Hansen JD, du Pasquier L, *et al*. Costimulatory receptors in non mammalian vertebrates: conserved CD28, odd CTLA4 and multiple BTLAs. *Dev Comp Immunol* 2007; **31**:255–71.
10. Garapati VP, Lefranc M-P. IMGT Colliers de Perles and IgSF domain standardization for T cell costimulatory activatory (CD28, ICOS) and inhibitory (CTLA4, PDCD1 and BTLA) receptors. *Dev Comp Immunol* 2007; **31**:1050–72.
11. Kaas Q, Lefranc M-P. T cell receptor/peptide/MHC molecular characterization and standardized pMHC contact sites in IMGT/3Dstructure-DB. *In Silico Biol* 2005; **5**:505–28.
12. Lefranc M-P, Duprat E, Kaas Q, *et al*. IMGT unique numbering for MHC groove G-DOMAIN and MHC superfamily (MhcSF) G-LIKE-DOMAIN. *Dev Comp Immunol* 2005;**29**:917–38.
13. Maenaka K, Jones EY. MHC superfamily structure and the immune system. *Curr Opin Struct Biol* 1999;**9**:745–53.
14. Duprat E, Lefranc M-P, Gascuel O. A simple method to predict protein binding from aligned sequences - application to MHC superfamily and beta2-microglobulin. *Bioinformatics* 2006;**22**:453–9.
15. Frigoul A, Lefranc M-P. MICA: standardized IMGT allele nomenclature, polymorphisms and diseases. *Recent Res Devel Human Genet* 2005;**3**:95–105.
16. Lefranc M-P, Giudicelli V, Kaas Q, *et al*. IMGT, the international ImMunoGeneTics information system®. *Nucl Acids Res* 2005;**33**:D593–7.
17. Giudicelli V, Lefranc M-P. Ontology for immunogenetics: IMGT-ONTOLOGY. *Bioinformatics* 1999;**15**:1047–54.
18. Duroux P, Kaas Q, Brochet X, *et al*. IMGT-Kaleidoscope, the formal IMGT-ONTOLOGY paradigm. *Biochimie* 2007 Sep 11; [Epub ahead of print].
19. Lefranc M-P, Lefranc G. *The Immunoglobulin FactsBook*. London: Academic Press, 2001:1–458.
20. Lefranc M-P, Lefranc G. *The T cell receptor FactsBook*. London: Academic Press, 2001:1–398.
21. Giudicelli V, Chaume D, Lefranc M-P. IMGT/GENE-DB: a comprehensive database for human and mouse immunoglobulin and T cell receptor genes. *Nucl Acids Res* 2005;**33**:D256–61.
22. Lefranc M-P. WHO-IUIS Nomenclature Subcommittee for Immunoglobulins and T cell receptors report. *Dev Comp Immunol* 2007 Nov 6; [Epub ahead of print].
23. Lefranc M-P, Giudicelli V, Ginestoux C, *et al*. IMGT-ONTOLOGY for immunogenetics and immunoinformatics (http://imgt.cines.fr). *In Silico Biol* 2004;**4**:17–29.
24. Lefranc M-P, Clément O, Kaas Q, *et al*. IMGT-Choreography for immunogenetics and immunoinformatics (http://imgt.cines.fr). *In Silico Biol* 2005;**29**:185–203.
25. Lefranc M-P. Unique database numbering system for immunogenetic analysis. *Immunol. Today* 1997;**18**:509.
26. Lefranc M-P. The IMGT unique numbering for immunoglobulins, T cell receptors and Ig-like domains. *Immunologist* 1999;**7**:132–6.
27. Giudicelli V, Ginestoux C, Folch G, *et al*. IMGT/LIGM-DB, the IMGT® comprehensive database of immunoglobulin and T cell receptor nucleotide sequences. *Nucleic Acids Res* 2006;**34**:D781–4.
28. Kaas Q, Ruiz M, Lefranc M-P. IMGT/3Dstructure-DB and IMGT/StructuralQuery, a database and a tool for immunoglobulin, T cell receptor and MHC structural data. *Nucleic Acids Res* 2004;**32**:D208–10.
29. Giudicelli V, Chaume D, Lefranc M-P. IMGT/V-QUEST, an integrated software for immunoglobulin and T cell receptor V-J and V-D-J rearrangement analysis. *Nucleic Acids Res* 2004;**32**:W435–40.
30. Yousfi Monod M, Giudicelli V, Chaume D, *et al*. IMGT/JunctionAnalysis: the first tool for the analysis of the immunoglobulin and T cell receptor complex V-J and V-D-J JUNCTIONs. *Bioinformatics* 2004;**20**:i379–85.
31. Ruiz M, Lefranc M-P. IMGT gene identification and Colliers de Perles of human immunoglobulin with known 3D structures. *Immunogenetics* 2002;**53**:857–83.
32. Pommié C, Levadoux S, Sabatier R, *et al*. IMGT standardized criteria for statistical analysis of immunoglobulin V-REGION amino acid properties. *J Mol Recognit* 2004; **17**:17–32.
33. Kaas Q, Lefranc M-P. IMGT Colliers de Perles: standardized sequence-structure representations of the IgSF and MhcSF superfamily domains. *Curr Bioinformatics* 2007; **2**:21–30.
34. Gardinier MV, Amiguet P, Linington C, *et al*. Myelin/Oligodendrocyte glycoprotein is a unique member of the immunoglobulin superfamily. *J Neurosci Res* 1992;**33**:177–87.
35. Schrewe H, Thompson J, Bona M, *et al*. 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**:2738–48.
36. Cybulsky MI, Fries JWU, Williams AJ, *et al*. Gene structure, chromosomal location, and basis for alternative mRNA splicing of the human VCAM1 gene. *Proc Natl Acad Sci USA* 1991;**88**:7859–63.
37. Wilson IA, Bjorkman PJ. Unusual MHC-like molecules: CD1, Fc receptor, the hemochromatosis gene product, and viral homologs. *Curr Opin Immunol* 1998;**10**:67–73.
38. Braud VM, Allan DS, McMichael AJ. Functions of nonclassical MHC and non-MHC-encoded class I molecules. *Curr Opin Immunol* 1999;**11**:100–8.

39. Magdelaine-Beuzelin C, Kaas Q, Wehbi V, *et al*. Structure-function relationships of the variable domains of monoclonal antibodies approved for cancer treatment. *Crit Rev Oncol/Hematol* 2007;**64**:210–25.

40. Laffly E, Danjou L, Condemine F, *et al*. Selection of a macaque Fab with human-like framework regions, high affinity, and that neutralizes the protective antigen (PA) of *Bacillus anthracis*. *Antimicrob Agents Chemother* 2005; **49**:3414–20.

41. Belessi CJ, Davi FB, Stamatopoulos KE, *et al*. IGHV gene insertions and deletions in chronic lymphocytic leukemia: 'CLL-biased' deletions in a subset of cases with stereotyped receptors. *Eur J Immunol* 2006;**36**: 1963–74.