IMGT Colliers de Perles: Standardized Sequence-Structure Representations of the IgSF and MhcSF Superfamily Domains

Quentin Kaas[1] and Marie-Paule Lefranc[*,1,2]

[1]*IMGT®, Laboratoire d'ImmunoGénétique Moléculaire, LIGM, Université Montpellier II, UPR CNRS 1142,* [2]*Institut de Génétique Humaine, IGH, 141 rue de la Cardonille, 34396 Montpellier Cedex 5, Montpellier, France*

Abstract: IMGT®, the international ImMunoGeneTics information system® (http://imgt.cines.fr) provides a common access to expertly annotated data on the genome, proteome, genetics and structure of immunoglobulins (IG), T cell receptors (TR), major histocompatibility complex (MHC) of human and other vertebrates, and related proteins of the immune system (RPI) of any species. RPI include proteins that belong to the immunoglobulin superfamily (IgSF) and MHC superfamily (MhcSF). IMGT has set up a unique numbering system, which takes into account the structural features of the Ig-like and Mhc-like domains. In this paper, we describe the IMGT Scientific chart rules for the description of the IgSF V type and C type and of the MhcSF G type domains. These rules are based on the IMGT-ONTOLOGY concepts and are applicable for the sequence and structure analysis, whatever the species, the IgSF or MhcSF protein, or the chain type. We present examples of IMGT Colliers de Perles of IgSF V type (V-DOMAIN and V-LIKE-DOMAIN), C type (C-DOMAIN and C-LIKE-DOMAIN) and MhcSF G type (G-DOMAIN and G-LIKE-DOMAIN) based on the IMGT unique numbering. These standardized two-dimensional graphical representations are particularly useful for antibody engineering, sequence-structure analysis, visualization and comparison of positions for mutations, polymorphisms and contact analysis.

Keywords: IMGT, immunoglobulin superfamily, major histocompatibility complex superfamily, domain alignment, Ig-like, Mhc-like.

1. INTRODUCTION

IMGT, the international ImMunoGeneTics information system® (http://imgt.cines.fr) [1] is a high quality integrated knowledge resource specialized in immunoglobulins (IG), T cell receptors (TR), major histocompatibility complex (MHC) of human and other vertebrates, and related proteins of the immune system (RPI) of any species [2-7]. The RPI proteins include more particularly proteins other than IG, TR and MHC that belong to the immunoglobulin superfamily (IgSF) and to the MHC superfamily (MhcSF) [1].

The IgSF comprises not only the IG and TR proteins 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.) [7-11]. The common feature of the IgSF proteins is to have at least one immunoglobulin-like (Ig-like) domain [12,13]. 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 [12,13]. The number of antiparallel beta strands defines two domain types: 9 strands for the V type (which comprises the V-DOMAIN[1] of the IG and TR, and the V-LIKE-DOMAIN of the IgSF proteins other than the IG or TR) [12] (Fig. **1A**), and 7 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) [13] (Fig. **2A**). Thus, the IgSF comprises the IG and TR (each chain with 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 MhcSF comprises not only the MHC proteins involved in the antigen presentation to the T cells, but also proteins with a three-dimensional (3D) structure similar to that of MHC and involved in different functions [6,14,15]. The common feature of the 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 antiparallel beta strands ("floor" of the groove or platform) and two helical regions ("walls" of the groove) [6,14] (Fig. **3A**). Each domain made of four antiparallel beta 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) [14]. Thus, the MhcSF comprises the MHC proteins which belong, depending on their structure, to two classes, MHC class I (one chain with two G-DOMAINs and one C-LIKE-DOMAIN, associated to the beta-2-microglobulin (B2M)) and MHC class II (each chain with 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 [14].

IMGT data are described according to the IMGT Scientific chart rules based on the IMGT-ONTOLOGY concepts

*Address correspondence to this author at the International ImMunoGeneTics Information System® (IMGT®), 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

[1]IMGT labels of the IMGT-ONTOLOGY DESCRIPTION concept are written in capital letters. Definitions of IMGT labels are available in the IMGT Scientific chart at http://imgt.cines.fr

1574-8936/07 $50.00+.00

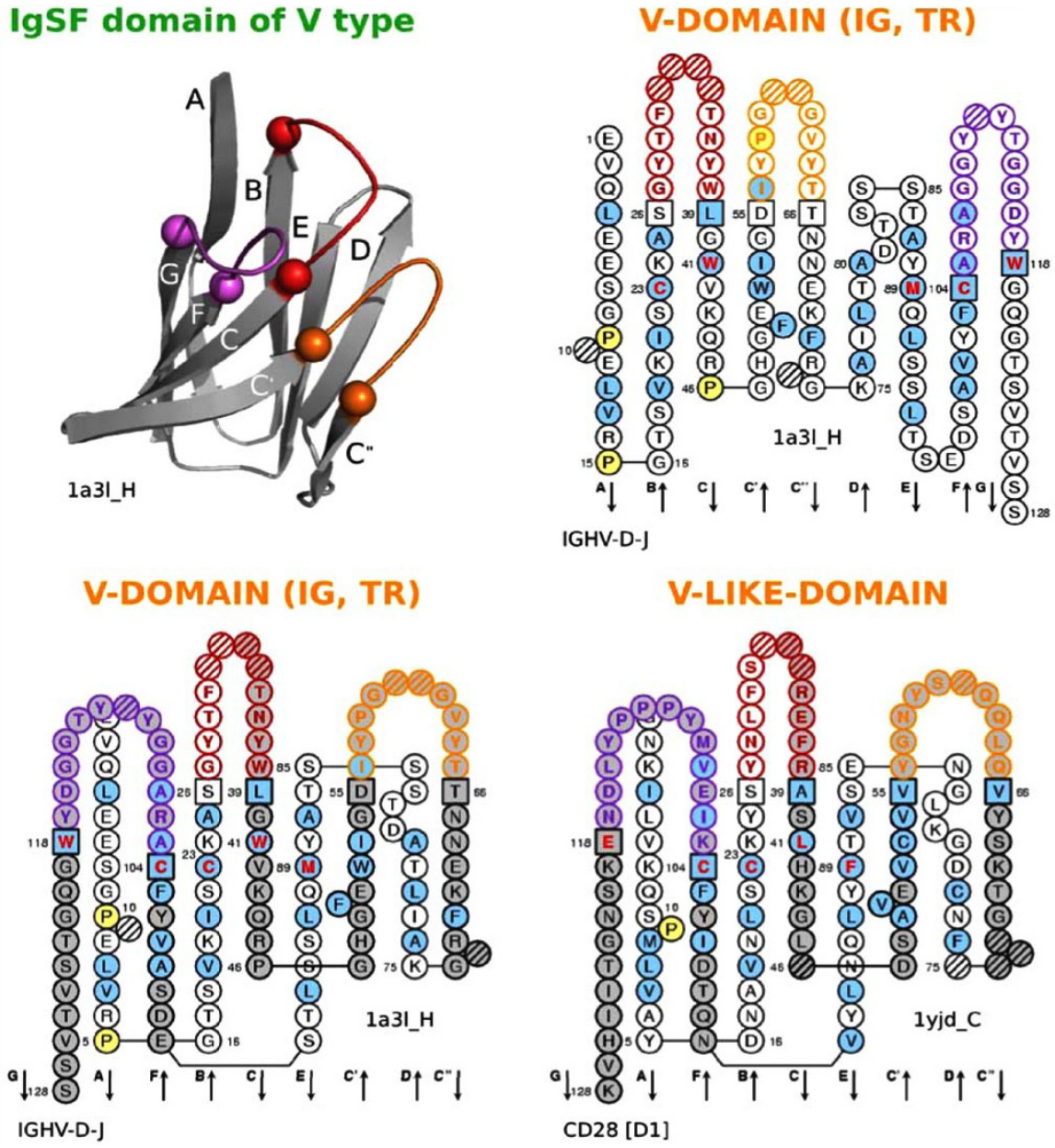

Fig. (1). IMGT Colliers de Perles of V-DOMAIN and V-LIKE-DOMAIN. (A) IgSF domain of V type. The ribbon representation is that of an IG V-DOMAIN taken as an example (PDB and IMGT/3Dstructure-DB: 1a3l_H). A similar topology and 3D structure characterize a V-LIKE-DOMAIN of IgSF proteins other than IG and TR. A V type domain (V-DOMAIN or V-LIKE-DOMAIN) consists of about one hundred amino acids in nine antiparallel beta strands, linked by beta turns or loops, and located on two layers maintained by a disulfide bridge. (B) and (C) V-DOMAIN (IG, TR). *Mus musculus* VH [8.8.12] (PDB and IMGT/3Dstructure-DB: 1a3l_H), on one layer and on two layers, respectively. (D) V-LIKE-DOMAIN on two layers. *Homo sapiens* CD28 [9.9.13] (PDB and IMGT/3Dstructure-DB: 1yjd_C). 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). 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 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 for V-DOMAIN and V-LIKE-DOMAIN [12].

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.

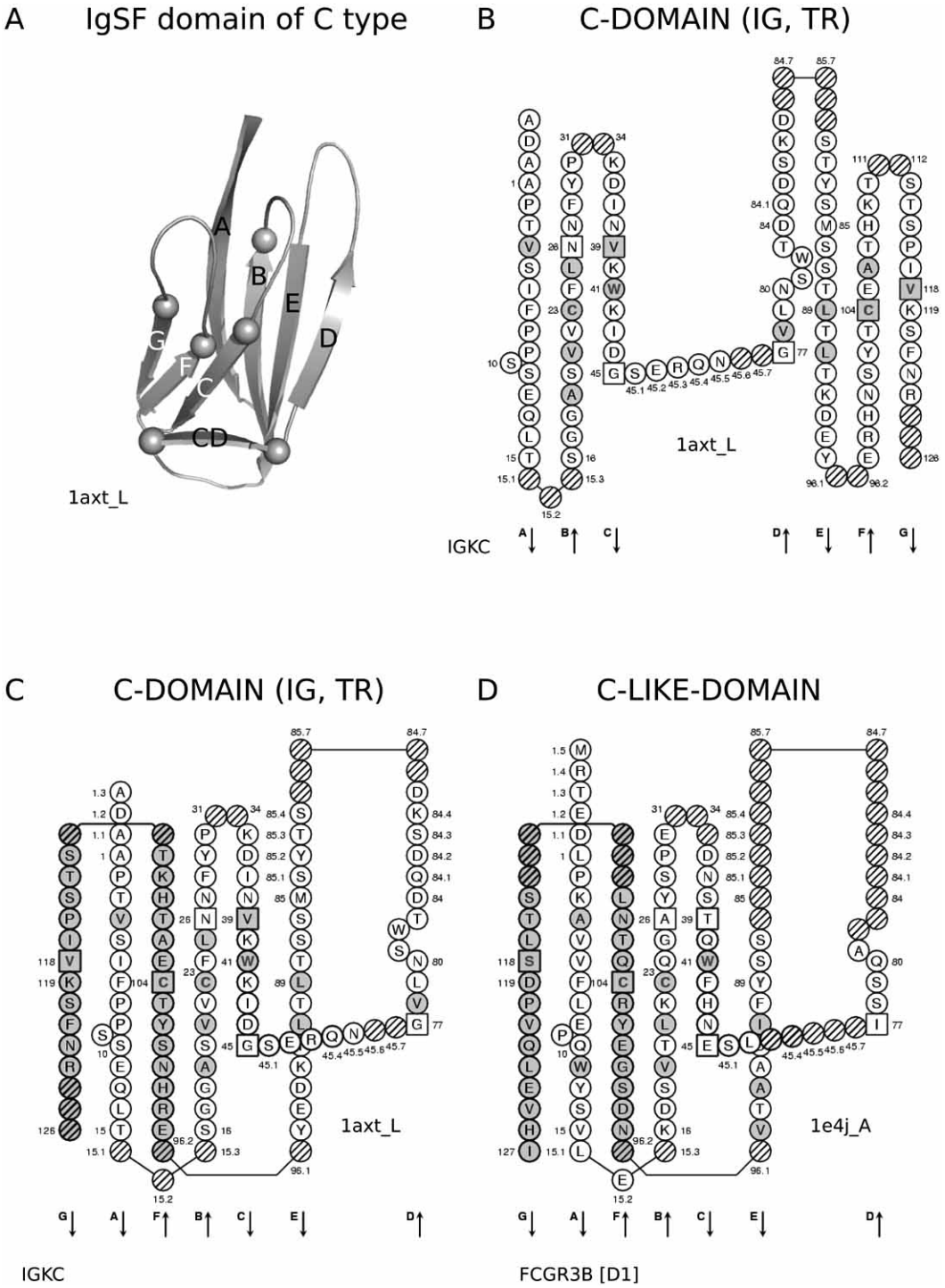

Fig. (2). (A) IMGT Colliers de Perles of C-DOMAIN and C-LIKE-DOMAIN. (A) IgSF domain of C type. The ribbon representation is that of a IG C-DOMAIN taken as an example (PDB and IMGT/3Dstructure-DB: 1axt_L). A similar topology and 3D structure characterize a C-LIKE-DOMAIN of IgSF proteins other than IG and TR. A C type domain (C-DOMAIN or C-LIKE-DOMAIN) consists of about one hundred amino acids in seven antiparallel and one transversal beta strands, linked by beta turns or loops, and located on two layers maintained by a disulfide bridge. (B) and (C) C-DOMAIN (IG, TR). *Mus musculus* C-KAPPA (PDB and IMGT/3Dstructure-DB: 1axt_L) on one layer and on two layers, respectively. (D) C-LIKE-DOMAIN on two layers. *Homo sapiens* FCGR3B [D1] (PDB and IMGT/3Dstructure-DB: 1e4j_A). Amino acids are shown in the one-letter abbreviation. Position at which 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 gray. The positions 26, 39 and 104 are shown in squares by homology with the corresponding positions in the V type (V-DOMAIN and V-LIKE-DOMAIN). Positions 45 and 77 which delimit the characteristic CD-STRAND of the C type (C-DOMAIN and C-LIKE-DOMAIN), and position 118 which corresponds structurally to J-PHE or J-TRP of the IG and TR J-REGION [2,3], are also shown in squares. Hatched circles correspond to missing positions according to the IMGT unique numbering for C-DOMAIN and C-LIKE-DOMAIN [13]. 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 show, on the forefront, the GFC strands and, on the back, the ABE strands.

[16]. This includes standardized IMGT gene and allele names (CLASSIFICATION concept) [2,3,17], standardized IMGT labels for the receptors, chains, domains and regions (DESCRIPTION concept) [18,19], standardized amino acid positions according to the IMGT unique numbering (NUMEROTATION concept) [12-14,20,21]. The IMGT standardization is used in the IMGT gene, sequence and structure databases [17,22,23], IMGT on-line tools [23-25] and in the IMGT knowledge web resources (IMGT Protein displays, IMGT Alignments of alleles, IMGT Colliers de Perles) [2-5]. IMGT Colliers de Perles [26,27] are standardized graphical two-dimensional (2D) representations, based on the IMGT unique numbering [12-14]. 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 [12], for the IgSF C type domains, based on the IMGT unique numbering for C-DOMAIN and C-LIKE-DOMAIN [13], and for the MhcSF G type domains, based on the IMGT unique numbering for G-DOMAIN and G-LIKE-DOMAIN [14]. IMGT Colliers de Perles are provided in IMGT/3Dstructure-DB [23] 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/Domain GapAlign and IMGT/Collier-de-Perles tools (http://imgt.cines.fr) [23], or for user V-DOMAIN nucleotide sequences, using the IMGT/V-QUEST tool [24]. 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 [2,3,6,7,12,26,27], of the strands and loops of the C-DOMAINs and C-LIKE-DOMAINs [7,13,28,29], and of the strands and helices of the G-DOMAINs and G-LIKE-DOMAINs [6,14,29]. By taking into account the structural features of the Ig-like and Mhc-like domains, the IMGT Colliers de Perles based on the IMGT unique numbering allow to fill in the gap between linear amino acid sequences and 3D structures.

In this paper, we review the IMGT Scientific chart rules for the description of the IgSF V type and C type and of the Mhc G type IMGT Colliers de Perles, which are applicable, for the IgSF and MhcSF, 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 in the Ig-like and Mhc-like domains.

2. IgSF AND MhcSF CHAINS AND DOMAINS

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 (alpha and beta chains, or gamma and delta chains, depending on the receptor type). An IG or TR chain comprises of 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 [12,13] (for review, see [2] and [3]). 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 [7]. As examples, the MOG and MPZ (or P0) proteins have a unique N-terminal V-LIKE-DOMAIN [30-37], the CEA family proteins have a single N-terminal V-LIKE-DOMAIN followed by a variable number of C-LIKE-DOMAINs [38-39], VCAM1 is composed of seven C-LIKE-DOMAINs [40-44]. IgSF proteins with diverse V type domain and C type domain combinations interspersed with domains belonging to other types are described constantly.

MhcSF Chains and Domains

The MHC proteins belong to two classes: the MHC-I proteins, expressed on the cell surface of most cells, are formed with an association of a transmembrane heavy chain (I-ALPHA chain) and a non-covalently linked light chain beta-2-microglobulin (B2M) [14]. The MHC-II proteins, expressed on the cell surface of professional antigen presenting cells (APC), are heterodimers formed with an association of two transmembrane chains, an alpha chain (II-ALPHA chain) and a beta chain (II-BETA chain) [14].

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 intracytoplasmic region. The I-ALPHA chain comprises of two groove domains (G-DOMAINs), the G-ALPHA1 [D1] and G-ALPHA2 [D2] domains, and one C-LIKE-DOMAIN [D3] [14]. The II-ALPHA chain and the II-BETA chain each comprises of two domains, the G-ALPHA [D1] and one C-LIKE-DOMAIN [D2], or 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 (Fig. **3B** and Fig. **3C**). 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 "nonclassical" MHC (MHC-Ib and MHC-IIb) proteins with more specific functions or which do not present peptides to the T cells [14]. This groove has also been found in proteins other than MHC, but so far, only MHC-I-like chains have been identified in the MhcSF [45-56]. These chains either include a C-LIKE-DOMAIN and are B2M bound (e.g. CD1, FCGRT, HFE, MR1) or B2M unbound (MIC, AZGP1), or do not include a C-LIKE-DOMAIN (EPCR, RAE) [45]. 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 [14] (Fig. **3D**).

3. IMGT Colliers de Perles FOR V-DOMAIN AND V-LIKE-DOMAIN

V Type Characteristics

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

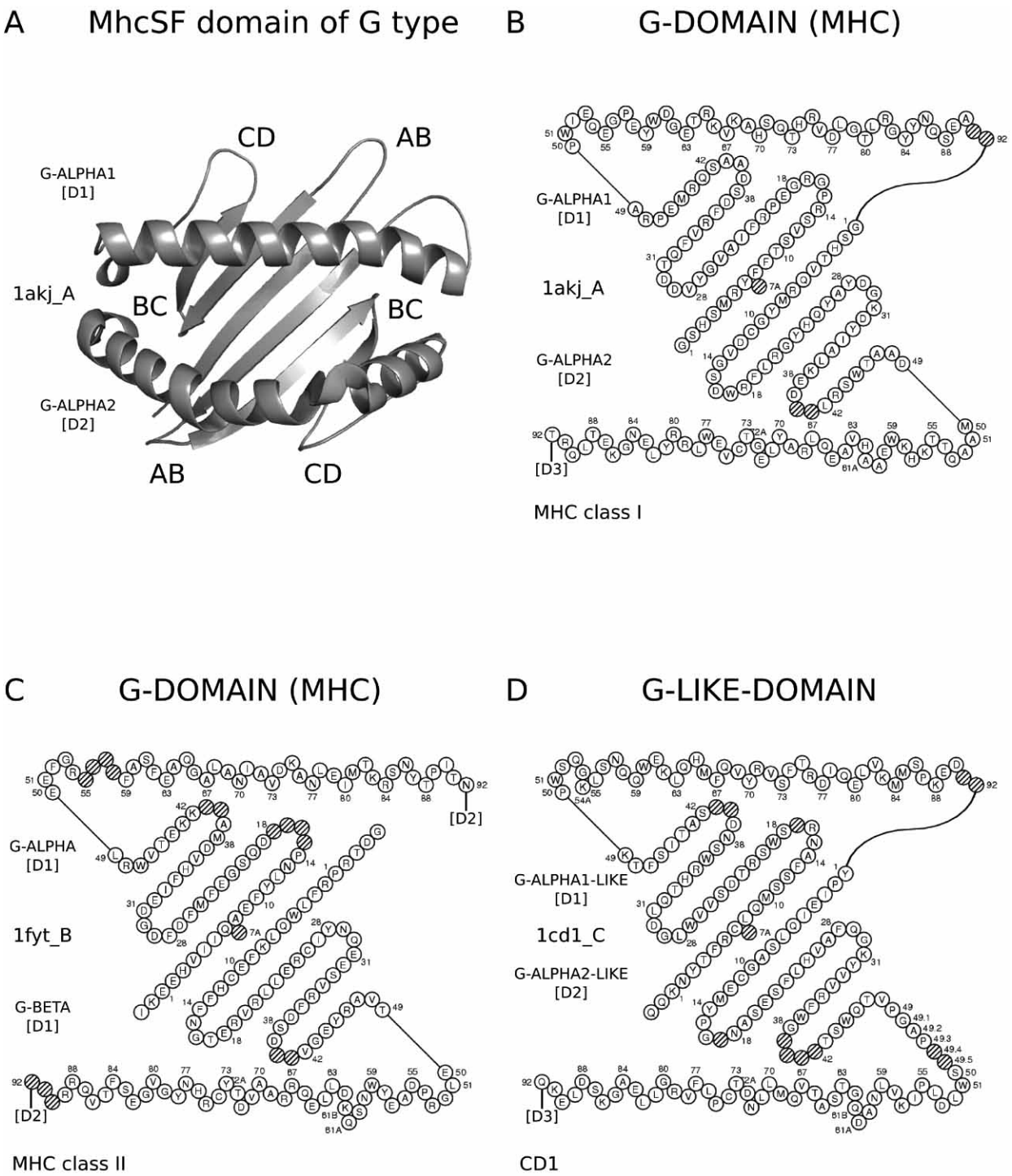

Fig. (3). IMGT Colliers de Perles of G-DOMAIN and G-LIKE-DOMAIN. (A) MhcSF domain of G type. The ribbon representation is that of two MHC G-DOMAINs taken as an example (PDB and IMGT/3Dstructure-DB: 1akj_A). A G type domain (G-DOMAIN or G-LIKE-DOMAIN) consists of four antiparallel beta strands linked by beta turns or loops and one alpha helix which form, with the beta strands and helix of a second G type domain, the characteristic MhcSF groove. A similar topology and 3D structure characterize the G-LIKE-DOMAINs in MhcSF proteins other than MHC. (B) G-DOMAIN (MHC) MHC class I. G-ALPHA1 [D1] and G-ALPHA2 [D2] of *Homo sapiens* HLA-A*0201 (PDB and IMGT/3Dstructure-DB: 1akj_A). (C) G-DOMAIN (MHC) MHC class II. G-ALPHA [D1] and G-BETA [D1] of *Homo sapiens* HLA-DRA*0101 and HLA-DRB1*0101 (PDB and IMGT/3Dstructure-DB: 1fyt_B). (C) G-LIKE-DOMAIN MHC-I-like. G-ALPHA1-LIKE [D1] and G-ALPHA2-LIKE [D2] of *Mus musculus* CD1D1 (PDB and IMGT/3Dstructure-DB: 1cd1_C). Amino acids are shown in the one-letter abbreviation. Hatched circles correspond to missing positions according to the IMGT unique numbering for G-DOMAIN and G-LIKE-DOMAIN. In IMGT Colliers de Perles, position 7A is only displayed in the G-ALPHA, G-ALPHA1 and G-ALPHA1-LIKE domains, and positions 61A and 61B in the G-BETA, G-ALPHA2 and G-ALPHA2-LIKE domains. Position 54A is only occupied in G-ALPHA1-LIKE of MHC-I-like proteins. Position 92A is only added for MHC-DMA and H2-DMA IMGT Colliers de Perles. Note that the N-terminal end of a peptide in the cleft would be on the left hand side.

V-LIKE-DOMAIN [12]. Indeed, the 3D structure of a V-LIKE-DOMAIN is very similar to that of an IG and TR V-DOMAIN (Fig. **1**). Both domains 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 (Fig. **1**). 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 [11,23,57]. 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, cysteine 104 (2nd-CYS). The hydrophobic amino acids of the framework regions are also found in conserved positions [12]. 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 [2,3] 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 (for review [2,3]). 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.

V Type Strands

The antiparallel beta strands of the V-LIKE-DOMAIN (Fig. **1D**) correspond to the conserved regions or framework (FR-IMGT) described in the IG and TR V-DOMAIN [12] (Fig. **1B**, **1C**). The A-STRAND of fifteen (or fourteen if gap at position 10) amino acids (positions 1 to 15) and the B-STRAND of eleven amino acids (positions 16 to 26) with the first conserved cysteine (1st-CYS) at position 23 correspond to the FR1-IMGT. The C-STRAND of eight amino acids (positions 39 to 46) with the tryptophan (CONSERVED-TRP) at position 41 and the C'-STRAND of nine amino acids, (positions 47 to 55) correspond to FR2-IMGT. The C"-STRAND of nine (or eight if gap at position 73) amino acids (positions 66 to 74), the D-STRAND of ten (or eight if gaps at positions 81,82) amino acids (positions 75 to 84), the E-STRAND of twelve amino acids (positions 85 to 96) with a conserved hydrophobic amino acid at position 89 and the F-STRAND of eight amino acids (positions 97 to 104) with the second conserved cysteine (2nd-CYS) at position 104 correspond to the FR3-IMGT. The G-STRAND of eleven amino acids (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 Fig. **1** correspond to missing positions according to the IMGT unique numbering. In the IMGT Protein display [2-5,7,12], unoccupied positions according to the IMGT unique numbering are shown by dots.

V Type 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-DOMAIN [12]. The BC-LOOP comprises positions 27 to 38; the longest BC loops have 12 amino acids. The C'C"-LOOP comprises positions 56 to 65; the longest C'C" loops have 10 amino acids. For BC loops or for C'C" loops shorter than 12 or 10 amino acids, respectively, gaps are created at the apex (missing positions, hatched in IMGT Collier de Perles (Fig. **1**), 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. The FG-LOOP comprises positions 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, etc. For FG loops longer than 13 amino acids, additional positions are created, between positions 111 and 112 at the top of the FG loop, in the following order: 112.1, 111.1, 112.2, 111.2, 112.3, etc. [12].

The loop length (number of codons or amino acids that is number of occupied positions) is a crucial and original concept of IMGT-ONTOLOGY [16]. 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 codons (or amino acids), into brackets and separated by dots. For examples *Homo sapiens* MOG [D] [9.6.9] mean that in the human MOG [D] domain, the BC, C'C" and FG loops have a length of 9, 6 and 9 codons, respectively [7]. In Fig. (**1**), the CDR-IMGT lengths are [8.8.12] for the V-DOMAIN and [9.9.13] for the V-LIKE-DOMAIN.

4. IMGT Colliers de Perles FOR C-DOMAIN AND C-LIKE-DOMAIN

C Type Characteristics

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 [13]. This numbering is itself derived from the IMGT unique numbering first described for the V-REGION [20,21] and for the V-DOMAIN [12]. Indeed, the sandwich beta 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 (Figs. **1** and **2**). The C-DOMAIN and C-LIKE-DOMAIN are 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 [13]. 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 [13].

The C-DOMAIN and C-LIKE-DOMAIN are composed by the A-STRAND of fifteen amino acids (positions 1 to 15), the B-STRAND of eleven amino acids (positions 16 to 26) with the 1st-CYS at position 23, the BC-LOOP (positions 27 to 36), the C-STRAND of seven amino acids (positions 39 to 45) with the CONSERVED-TRP at position 41, the CD-STRAND of one to nine amino acids (positions 45.1

to 45.9), the D-STRAND of eight amino acids (positions 77 to 84), the DE-TURN (positions 84.1 to 84.7 and 85.7 to 85.1, see below), the E-STRAND of twelve amino acids (positions 85 to 96) with a conserved hydrophobic amino acid at position 89, the F-STRAND of eight amino acids (positions 97 to 104) with the 2nd-CYS at position 104, the FG-LOOP (positions 105 to 117, these positions corresponding to a FG loop of 13 amino acids), and the G-STRAND of eleven (or less) amino acids (positions 118 to 128).

C Type and V Type Comparison

The A-STRAND and B-STRAND of the C type (C-DOMAIN and C-LIKE-DOMAIN) are similar to those of the V type (V-DOMAIN and V-LIKE-DOMAIN) [13]. The C-STRAND and the D-STRAND of the C type are shorter of one position and two positions, respectively, compared to those of the V type. As previously described [13], the C' and C" strands are missing in the C type and are replaced by the characteristic transversal CD-STRAND. The E-STRAND, F-STRAND and G-STRAND of the C type are similar to those of the V type.

The longest BC-LOOP of the C type have 10 amino acids (gaps at positions 32,33), instead of 12 amino acids in the V type. For BC loops shorter than 10 amino acids, gaps are created from the apex in the following order 34, 31, 35, 30, 36, etc. The FG-LOOP of the C type is similar to that of the V type. 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 type.

Additional positions in the C type define the AB-TURN, DE-TURN and EF-TURN. The AB-TURN corresponds to additional positions 15.1, 15.2 and 15.3; the longest AB-TURN has 3 amino acids. For AB-TURN shorter than 3 amino acids, gaps are created (missing positions, hatched in IMGT Colliers de Perles, or not shown in structural data representations) in an ordinal manner. For example, in Fig. **2D**, there are two additional positions in 15.1 and 15.2, whereas 15.3 is unoccupied. The additional positions defining the DE-TURN comprise positions 84.1 to 84.7 and 85.7 to 85.1, corresponding to 14 amino acids. For DE-TURN shorter than 14 amino acids, gaps are created in the following order: 85.7, 84.7, 85.6, 84.6, 85.5, etc. In Fig **2B**, there are 9 additional positions in the DE-TURN of the C-DOMAIN, whereas in Fig. **2D** positions of the DE-TURN are unoccupied, and the D and E strands are shorter than usual. The EF-TURN corresponds to additional positions 96.1 and 96.2, corresponding to 2 amino acids. For EF-TURN shorter than 2 amino acids, gaps are created in the following order: 96.2, 96.1.

5. 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 [14].

G Type Strands

For each G type domain (G-DOMAIN and G-LIKE-DOMAIN), the positions that contribute to the groove floor comprise positions 1 to 49, with the A-STRAND of fourteen amino acids (positions 1 to 14), the AB-TURN of three or less amino acids (positions 15 to 17), the B-STRAND of eleven amino acids (positions 18 to 28), the BC-TURN of two amino acids (positions 29 and 30), the C-STRAND of eight amino acids (positions 31 to 38), the CD-TURN of three or less amino acids (positions 39 to 41) and the D-STRAND of eight amino acids (positions 42 to 49) [14] (Fig. **3**).

The gaps are localized in the turns. The AB-TURN (positions 15 to 17) comprises three amino acids in the G-ALPHA1, G-ALPHA2 and G-BETA domains but these positions are unoccupied in the G-ALPHA domains (as well as position 18 of B-STRAND) (Fig. **3C**). The BC-TURN (positions 29 and 30) comprises two positions that are occupied in all G-DOMAINs. The CD-TURN (positions 39 to 41) is occupied by three amino acids in the G-ALPHA1 domains and only one in the other domains (G-ALPHA2, G-ALPHA and G-BETA) (Fig. **3**).

The additional position 7A represents a bulge in 3D structures and is present in some G-ALPHA domains, for instance, those of the human HLA-DQA1 and HLA-DOA, and mouse H2-AA and H2-DOA chains [14]. Additional positions at the N-terminus of A-STRAND or at the C-terminus of D-STRAND can be added if necessary. Thus, two additional positions (1.2 and 1.1) are added at the N-terminus of the A-STRAND of the G-ALPHA domains as the presence of these two amino acids was demonstrated by protein sequencing of the HLA-DRA, however the proteolytic cleavage site of the leader peptide (L-REGION) needs to be confirmed experimentally for the other G-ALPHA [14]. In each [D1] G-DOMAIN (except the G-ALPHA domain of human HLA-DMA and mouse H2-DMA discussed below), the amino acid at position 1 (shown within parentheses in IMGT Protein displays, http://imgt.cines.fr) is encoded by the codon that results from the splicing between the first exon (EX1) that encodes the L-REGION, and the second exon (EX2) that encodes [D1] [14]. Ten amino acids have been added at positions 1.10 to 1.1 of HLA-DMA and H2-DMA but it is necessary to confirm experimentally if they belong, or not, to the mature protein. Four additional positions (49.1 to 49.4) are also observed at the C-terminus of D-STRAND of the G-BETA domain of HLA-DMB and H2-DMB1 [14].

G Type Helix

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 [14]. Three of them (61A, 61B, 72A) characterize the G-ALPHA2 and/or G-BETA domains. Indeed, positions 61A and 72A are occupied in the G-ALPHA2 domain, whereas positions 61A, 61B and 72A are occupied in G-BETA domains (except for the mouse H2-AB chain). The position 92A is only occupied in the HLA-DMA and H2-DMA G-ALPHA domains. 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 of the Mhc-I-like proteins.

The helix (positions 50 to 92) seats on the beta sheet and its axis forms an angle of about 40 degrees with the beta strands. The helix is split into two parts separated by a kink, positions 58 of G-ALPHA1, 61 of G-ALPHA2, 63 of G-ALPHA, and 62 of G-BETA being the "highest" points on the groove floor [6,14].

G Type Comparison

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 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). An N-glycosylation site is also found at position 86 in the G-ALPHA2 domain of the mouse MHC-Ia chains (H2-D1, H2-K1 and H2-L). The G-BETA domains (except for the human HLA-DMB and mouse H2-DMB1 chains) have a conserved potential N-glycosylation site at position 15 (AB-TURN).

Interestingly, the G-ALPHA domains of the HLA-DMA and H2-DMA chains have specific features compared to the other G-ALPHA domains and share common characteristics with the G-ALPHA2 and G-BETA domains: there is a conserved CYS-11_CYS-74 disulfide bridge, positions 61A and 61B are occupied (as in the G-BETA domains) and there is no N-glycosylation site at position 86 [14] (IMGT Protein display in IMGT Repertoire, http://imgt.cines.fr).

Practically, the IMGT unique numbering for positions 1-39 and 73-92 of the G-ALPHA2 domains can be obtained very easily by subtracting 90 from the mature protein numbering (91-129 and 163-182). Between these positions, the two gaps (at positions 40 and 41) and the two insertions (at positions 61A and 72A) are necessary, in the IMGT unique numbering, to allow meaningful sequence and structure alignment and comparison between the G-ALPHA1 and G-ALPHA2 sequences.

6. PERSPECTIVES

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 [7,12-14]. The strand, loop, turn or helix lengths (the number of codons or amino acids 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 inapplicable when dealing with sequences for which no structural data are known (discussed in [13]).

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 [12-14]. 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 Collier de Perles also puts 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 (usually a splicing frame 1) ('Splicing sites' in IMGT Aide-mémoire, http://imgt.cines.fr). For an 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 [2,3].

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 BTLA protein which belong to the IgSF by their V type and/or C type domains [58]. The IMGT Collier 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 to 26, FR2-IMGT: 39 to 55, FR3-IMGT: 66 to 104, and FR4-IMGT: 118 to 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 [59].

The IMGT Colliers de Perles also allow a comparison to the IMGT Colliers de Perles statistical profiles for the human expressed IGHV, IGKV and IGLV repertoires [27]. 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 [27] ('Amino acids' in IMGT Aide-mémoire, http://imgt.cines.fr). The statistical profiles identified positions which 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 [27]). After the assignment of the IMGT Collier 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 [59] or to evaluate immunogenicity of primate antibodies [60].

IMGT Colliers de Perles are also of interest when 3D structures are available [6,26]. In IMGT/3Dstructure-DB [23], 'IMGT Collier de Perles on 2 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, which can be compared with the IMGT reference pMHC contact sites available in IMGT/3Dstructure-DB [6].

The IMGT Colliers de Perles for the V type, C type and G type, based on the IMGT unique numbering, represent therefore 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 leukaemia [61], pMHC contact analysis [6], and more generally ligand-receptor interactions involving V type, C type and/or G type domains.

ACKNOWLEDGEMENTS

We thank Vijay Garapati for the CD28 IMGT Collier de Perles. We are grateful to the IMGT team for helpful discussion. IMGT was funded by the Centre National de la Recherche Scientifique (CNRS), the BIOMED1 (BIOCT 930038), Biotechnology BIOTECH2 (BIO4CT960037) and 5th PCRDT Quality of Life and Management of Living Resources (QLG2-2000-01287) programmes of the European Union. IMGT received subventions from Association pour la Recherche sur le Cancer (ARC), Génopole-Montpellier-Languedoc-Roussillon. IMGT is a National RIO Bioinformatics Platform since 2001 (CNRS, INSERM, CEA, INRA). IMGT is currently supported by the CNRS, the Ministère de l'Education Nationale, de l'Enseignement Supérieur et de la Recherche MENESR (Université Montpellier II Plan Pluri-Formation, BIOSTIC-LR2004 Région Languedoc-Roussillon and ACI-IMPBIO IMP82-2004), the Réseau National des Génopoles RNG, GIS-AGENAE, Agence Nationale de la Recherche ANR (BIOSYS06_135457) and the European Union ImmunoGrid project (IST-2004-0280069).

REFERENCES

[1] Lefranc M-P, Giudicelli V, Kaas Q, Duprat E, Jabado-Michaloud J, Scaviner D, Ginestoux C, Clément O, Chaume D, Lefranc G. IMGT, the international ImMunoGeneTics information system®. *Nucl Acids Res* **2005**; 33: D593-D597.

[2] Lefranc M-P, Lefranc G. The Immunoglobulin FactsBook. Academic Press, London, UK. 458 pages, ISBN:012441351X, **2001**.

[3] Lefranc M-P, Lefranc G. The T cell receptor FactsBook. Academic Press, London, UK. 398 pages, ISBN:0124413528, **2001**.

[4] Scaviner D, Barbié V, Ruiz M, Lefranc M-P. Protein displays of the human immunoglobulin heavy, kappa and lambda variable and joining regions. *Exp Clin Immunogenet* **1999**; 16: 234-240.

[5] Folch G, Scaviner D, Contet V, 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.

[6] Kaas Q, Lefranc M.-P. T cell receptor/peptide/MHC molecular characterization and standardized pMHC contact sites in IMGT/3Dstructure-DB. Epub 2005, 5, 0046, http://www.bioinfo.de/isb/2005/05/0046/, 20 October 2005. *In Silico Biology* **2005**; 5: 505-528.

[7] Duprat E, Kaas Q, Garelle V, Giudicelli V, Lefranc G, Lefranc M-P. 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: 111-136.

[8] Williams AF, Barclay AN. The immunoglobulin superfamily-domains for cell surface recognition. *Annu Rev Immunol* **1988**; 6: 381-405.

[9] Hunkapiller T, Hood L. Diversity of the immunoglobulin gene superfamily. *Adv Immunol* **1989**; 44: 1-63.

[10] Jones EY. The immunoglobulin superfamily. *Curr Opin Struct Biol* **1993**; 3: 846-852.

[11] Bork P, Holm L, Sander C. The immunoglobulin fold. Structural classification, sequence patterns and common core. *J Mol Biol* **1994**; 242: 309-320.

[12] Lefranc M-P, Pommié C, Ruiz M, Giudicelli V, Foulquier E, Truong L, Thouvenin-Contet V, 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.

[13] 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, 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.

[14] Lefranc M-P, Duprat E, Kaas Q, Tranne M, Thiriot A, Lefranc G. IMGT unique numbering for MHC groove G-DOMAIN and MHC superfamily (MhcSF) G-LIKE-DOMAIN. *Dev Comp Immunol* **2005**; 29: 917-938.

[15] Maenaka K, Jones EY. MHC superfamily structure and the immune system. *Curr Opin Struct Biol* **1999**; 9: 745-753.

[16] Giudicelli, V, Lefranc, M-P. Ontology for Immunogenetics: IMGT-ONTOLOGY. *Bioinformatics* **1999**; 15: 1047-1054.

[17] 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-D261.

[18] Lefranc M-P, Giudicelli V, Ginestoux C, Bosc N, Folch G, Guiraudou D, Jabado-Michaloud J, Magris S, Scaviner D, Thouvenin V, Combres K, Girod D, Jeanjean S, Protat C, Yousfi Monod M, Duprat E, Kaas Q, Pommié C, Chaume D, Lefranc G. IMGT-ONTOLOGY for Immunogenetics and Immunoinformatics (http://imgt.cines.fr). Epub 2003; 4:0004, http://www.bioinfo.de/isb/2003/04/0004/. *In Silico Biol* **2004**; 4: 17-29.

[19] Lefranc M-P, Clément O, Kaas Q, Duprat E, Chastellan P, Coelho I, Combres K, Ginestoux C, Giudicelli V, Chaume D, Lefranc G. IMGT-Choreography for Immunogenetics and Immunoinformatics (http://imgt.cines.fr). *In Silico* Biol. 2005, 0006, 2004. Epub <http://www.bioinfo.de/isb/2004/05/0006/>. *In Silico Biol* **2005**; 29: 185-203.

[20] Lefranc M-P. Unique database numbering system for immunogenetic analysis. *Immunol Today* **1997**; 18: 509.

[21] Lefranc M-P. The IMGT unique numbering for Immunoglobulins, T cell receptors and Ig-like domains. *Immunologist* **1999**; 7: 132-136.

[22] Giudicelli V, Ginestoux C, Folch G, Jabado-Michaloud J, Chaume D, Lefranc M-P. IMGT/LIGM-DB, the IMGT® comprehensive database of immunoglobulin and T cell receptor nucleotide sequences. *Nucl Acids Res* **2006**; 34: D781-D784.

[23] 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. *Nucl Acids Res* **2004**; 32: D208-D210.

[24] 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. *Nucl Acids Res* **2004**; 32: W435-W440.

[25] Yousfi Monod M, Giudicelli V, Chaume D, Lefranc M-P. 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-I385.

[26] Ruiz M, Lefranc M-P. IMGT gene identification and Colliers de Perles of human immunoglobulin with known 3D structures. *Immunogenetics* **2002**; 53: 857-883.

[27] Pommié C, Levadoux S, Sabatier R, Lefranc G, Lefranc M-P. IMGT standardized criteria for statistical analysis of immunoglobulin V-REGION amino acid properties. *J Mol Recognition* **2004**; 17: 17-32.

[28] Bertrand G, Duprat E, Lefranc M-P, Marti J, Coste J. Characterization of human FCGR3B*02 (HNA-1b, NA2) cDNAs and IMGT standardized description of FCGR3B alleles. *Tissue Antigens* **2004**; 64: 119-131.

[29] Frigoul A, Lefranc M-P. MICA: standardized IMGT allele nomenclature, polymorphisms and diseases. *Recent Res Devel Human Genet* **2005**; 3: 95-105.

[30] Gardinier MV, Amiguet P, Linington C, Matthieu JM. Myelin/Oligodendrocyte glycoprotein is a unique member of the immunoglobulin superfamily. *J Neurosci Res* **1992**; 33: 177-187.

[31] Pham-Dinh D, Della Gaspera B, Kerlero de Rosbo N, Dautigny A. Structure of the human myelin/oligodendrocyte glycoprotein gene and multiple alternative spliced isoforms. *Genomics* **1995**; 29: 345-352.

[32] Roth M-P, Malfroy L, Offer C, Sevin J, Enault G, Borot N, Pontarroti P, Coppin H. The human myelin oligodendrocyte glycoprotein

(MOG) gene: complete nucleotide sequence and structural characterization. *Genomics* **1995**; 28: 241-250.

[33] Breithaupt C, Schubart A, Zander H, Skerra A, Huber R, Linington C, Jacob U. Structural insights into the antigenicity of myelin oligodendrocyte glycoprotein. *Proc Natl Acad Sci USA* **2003**; 100: 9446-9451.

[34] Clements CS, Reid HH, Beddoe T, Tynan FE, Perugini MA, Johns TG, Bernard CC, Rossjohn J. The crystal structure of myelin oligodendrocyte glycoprotein, a key autoantigen in multiple sclerosis. *Proc Natl Acad Sci USA* **2003**; 100: 11059-11064.

[35] Hayasaka K, Himuro M, Wang Y, Takata M, Minoshima S, Shimizu N, Miura M, Uyemura K, Takada G. Structure and chromosomal localization of the gene encoding the human myelin protein zero (MPZ). *Genomics* **1993**; 17: 755-758.

[36] Pham-Dinh D, Fourbil Y, Blanquet F, Mattei M-G, Roeckel N, Latour P, Chazot G, Vandenberghe A, 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 Mol Genet* **1993**; 2: 2051-2054.

[37] Kulkens T, Bolhuis PA, Wolterman RA, Kemp S, te Nijenhuis S, Valentijn LJ, Hensels GW, Jennekens FGI, de Visser M, Hoogendijk JE, 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.

[38] Schrewe H, Thompson J, Bona M, Hefta LJ, Maruya A, Hassauer M, Shively JE, von Kleist S, 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: 2738-2748.

[39] Thompson JA, Grunert F, Zimmermann W. Carcinoembryonic antigen gene family: molecular biology and clinical perspectives. *J Clin Lab Anal* **1991**; 5: 344-366.

[40] Cybulsky MI, Fries JWU, Williams AJ, Sultan P, Eddy R, Byers M, Shows T, Gimbrone MA, Collins T. Gene structure, chromosomal location, and basis for alternative mRNA splicing of the human VCAM1 gene. *Proc Natl Acad Sci USA* **1991**; 88: 7859-7863.

[41] Hession C, Tizard R, Vassallo C, Schiffer SB, Goff D, Moy P, Chi-Rosso G, Luhowskyj S, Lobb R, Osborn L. Cloning of an alternate form of vascular cell adhesion molecule-1 (VCAM1). *J Biol Chem* **1991**; 266: 6682-6685.

[42] Osborn L, Vassallo C, Browning BG, Tizard R, Haskard DO, Benjamin CD, Dougas I, 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.

[43] Vonderheide RH, Tedder TF, Springer TA, 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.

[44] Wang JH, Pepinsky RB, Stehle T, Liu JH, Karpusas M, Browning B, 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 Natl Acad Sci USA* **1995**; 92: 5714-5718.

[45] 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: 453-459.

[46] 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.

[47] Braud VM, Allan DS, McMichael AJ. Functions of nonclassical MHC and non-MHC-encoded class I molecules. *Curr Opin Immunol* **1999**; 11: 100-108.

[48] Zeng Z-H, Castano AR, Segelke BW, Stura EA, Peterson PA, Wilson IA. Crystal structure of mouse CD1: an MHC-like fold with a large hydrophobic binding groove. *Science* **1997**; 277: 339-345.

[49] West AP, Bjorkman, PJ. Crystal structure and immunoglobulin G binding properties of the human major histocompatibility complex-related Fc receptor. *Biochemistry* **2000**; 39; 9698-9708.

[50] Feder JN, Penn Fed y DM, Irrinki A, Lee VK, Lebron JA, Watson N, Tsuchihashi Z, Sigal E, Bjorkman PJ, Schatzman RC. The hemochromatosis gene product complexes with the transferrin receptor and lowers its affinity for ligand binding. *Proc Natl Acad Sci USA* **1998**; 95: 1472-1477.

[51] Miley MJ, Truscott SM, Yu YY, Gilfillan S, Fremont DH, Hansen TH, Lybarger L. Biochemical features of the MHC-related protein 1 consistent with an immunological function. *J Immunol* **2003**; 170: 6090-6098.

[52] Holmes MA, Li Hol P, Petersdorf EW, Strong RK. Structural studies of allelic diversity of the MHC class I homolog MIC-B, a stress-inducible ligand for the activating immunoreceptor NKG2D. *J Immunol* **2002**; 169: 1395-1400.

[53] Sanchez LM, Chirino AJ, Bjorkman PJ. Crystal structure of human ZAG, a fat-depleting factor related to MHC molecules. *Science* **1999**; 283: 1914-1919.

[54] Delker SL, West AP Jr, McDermott L, Kennedy MW, Bjorkman PJ. Crystallographic studies of ligand binding by Zn-alpha2-glycoprotein. *J Struct Biol* **2004**; 148: 205-213.

[55] Simmonds RE, Lane DA. Structural and functional implications of the intron/exon organization of the human endothelial cell protein C/activated protein C receptor (EPCR) gene: comparison with the structure of CD1/major histocompatibility complex alpha1 and alpha2 domains. *Blood* **1999**; 94: 632-641.

[56] Li P, McDermott G, Strong RK. Crystal structures of RAE-1 beta and its complex with the activating immunoreceptor NKG2D. *Immunity* **2002**; 16: 77-86.

[57] Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. *Nucl Acids Res* **2000**; 28: 235-242.

[58] Bernard D, Hansen JD, du Pasquier L, Lefranc M-P, Benmansour A, Boudinot P. Costimulatory receptors in non mammalian vertebrates: conserved CD28, odd CTLA4 and multiple BTLAs. *Dev Comp Immunol* (in press).

[59] Magdelaine-Beuzelin C, Kaas Q, Wehbi V, Ohresser M, Jefferis R, Lefranc M-P, Watier H. Structure-function relationships of the variable domains of monoclonal antibodies approved for cancer treatment. *Critical Reviews in Oncology/Hematology* (submitted).

[60] Laffly E, Danjou L, Condemine F, Vidal D, Drouet E, Lefranc M-P, Bottex C, Thullier P. 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: 3414-3420.

[61] Belessi CJ, Davi FB, Stamatopoulos KE, Degano M, Andreou TM, Moreno C, Merle-Béral H, Crespo M, Laoutaris NP, Montserrat E, Caligaris-Cappio F, Anagnostopoulos AZ, Ghia P. 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: 1963-1974.

Received: August 1, 2006 Revised: October 27, 2006 Accepted: Ocotber 28, 2006