2

IMGT®, the International ImmunoGeneTics Information System® for Immunoinformatics

Methods for Querying IMGT® Databases, Tools, and Web Resources in the Context of Immunoinformatics

Marie-Paule Lefranc

Summary

IMGT®, the international ImMunoGeneTics information system® (http://imgt.cines.fr), was created in 1989 by the Laboratoire d'ImmunoGénétique Moléculaire (LIGM) (Université Montpellier II and CNRS) at Montpellier, France, in order to standardize and manage the complexity of immunogenetics data. *IMGT®* is recognized as the international reference in immunogenetics and immunoinformatics. *IMGT®* is a high quality integrated knowledge resource, specialized in (i) the immunoglobulin (IG), T cell receptors (TR), major histocompatibility complex (MHC) of human and other vertebrates; (ii) proteins that belong to the immunoglobulin superfamily (IgSF) and to the MHC superfamily (MhcSF); and (iii) related proteins of the immune systems (RPI) of any species. *IMGT®* provides a common access to standardized data from genome, proteome, genetics, and three-dimensional (3D) structures for the IG, TR, MHC, IgSF, MhcSF, and RPI. *IMGT®* interactive on-line tools are provided for genome, sequence, and 3D structure analysis. *IMGT®* Web resources comprise 8,000 HTML pages of synthesis and knowledge (*IMGT* Scientific chart, *IMGT* Repertoire, *IMGT* Education, etc.) and external links (*IMGT* Bloc-notes and *IMGT* other accesses).

Key Words: IMGT; ontology; immunoglobulin; T cell receptor; MHC; IgSF; MhcSF

1. Introduction

The number of genomics, genetics, three-dimensional (3D), and functional data published in the immunogenetics field is growing exponentially and involves fundamental, clinical, veterinary, and pharmaceutical research. The

From: *Methods in Molecular Biology, vol. 409: Immunoinformatics: Predicting Immunogenicity In Silico*
Edited by: D. R. Flower © Humana Press Inc., Totowa, NJ

number of potential protein forms of the antigen receptors, immunoglobulins (IG), and T cell receptors (TR) is almost unlimited. The potential repertoire of each individual is estimated to comprise about 10^{12} different IG (or antibodies) and TR, and the limiting factor is only the number of B and T cells that an organism is genetically programmed to produce. This huge diversity is inherent to the particularly complex and unique molecular synthesis and genetics of the antigen receptor chains. This includes biological mechanisms such as DNA molecular rearrangements in multiple loci (three for IG and four for TR in humans) located on different chromosomes (four in humans), nucleotide deletions and insertions at the rearrangement junctions (or N-diversity), and somatic hypermutations in the IG loci (for review, see ref. ***1***,***2***).

IMGT®, the international ImMunoGeneTics information system® (http://imgt.cines.fr) ***(3***,***4)***, was created in 1989 by the Laboratoire d'ImmunoGénétique Moléculaire (LIGM) (Université Montpellier II and CNRS) at Montpellier, France, in order to standardize and manage the complexity of the immunogenetics data. *IMGT*® is as the international reference in immunogenetics and immunoinformatics. *IMGT*® is a high quality integrated knowledge resource, specialized in (i) the IG, TR, major histocompatibility complex (MHC) of human and other vertebrates, (ii) proteins that belong to the immunoglobulin superfamily (IgSF) and to the MHC superfamily (MhcSF), and (iii) related proteins of the immune systems (RPI) of any species. *IMGT*® provides a common access to standardized data from genome, proteome, genetics and 3D structures for the IG, TR, MHC, IgSF, MhcSF and RPI ***(3***,***4)***.

The *IMGT*® information system consists of databases, tools, and Web resources ***(3)***. *IMGT*® databases include one genome database, three sequence databases, and one 3D structure database. *IMGT*® interactive on-line tools are provided for genome, sequence, and 3D structure analysis. *IMGT*® Web resources comprise 8,000 HTML pages of synthesis and knowledge (*IMGT* Scientific chart, *IMGT* Repertoire, *IMGT* Education, *IMGT* Index, etc.) and external links (*IMGT* Bloc-notes and *IMGT* other accesses) ***(4)***. Despite the heterogeneity of these different components, all data in the *IMGT*® information system are expertly annotated. The accuracy, the consistency, and the integration of the *IMGT*® data, as well as the coherence between the different *IMGT*® components (databases, tools, and Web resources), are based on *IMGT-ONTOLOGY* ***(5)***, the first ontology in immunogenetics and immunoinformatics. *IMGT-ONTOLOGY* provides a semantic specification of

the terms to be used in the domain and, thus, allows the management of immunogenetics knowledge for all vertebrate species.

2. Standardization: *IMGT-ONTOLOGY* and *IMGT* Scientific Chart

IMGT-ONTOLOGY concepts are available, for the biologists and *IMGT®* users, in the *IMGT* Scientific chart ***(4)*** and have been formalized, for the computing scientists, in *IMGT-ML* ***(6,7)***. The *IMGT* Scientific chart ***(4)*** comprises the controlled vocabulary and the annotation rules necessary for the immunogenetics data identification, description, classification, and numbering and for knowledge management in the *IMGT®* information system. All *IMGT®* data are expertly annotated according to the *IMGT* Scientific chart rules. Standardized keywords, labels and annotation rules, standardized IG and TR gene nomenclature, the *IMGT* unique numbering, and standardized origin/methodology were defined, respectively, based on the six main concepts of *IMGT-ONTOLOGY* ***(5)*** (Table 1). The *IMGT* Scientific chart is available as a section of the *IMGT®* Web resources. Examples of *IMGT®* expertised data concepts derived from the *IMGT* Scientific chart rules are summarized in Table 1.

2.1. IDENTIFICATION concept: standardized keywords

IMGT® standardized keywords for IG and TR include the following: (i) general keywords—indispensable for the sequence assignments, they are described in an exhaustive and non-redundant list, and are organized in a tree structure and (ii) specific keywords—they are more specifically associated to particularities of the sequences (orphon, transgene, etc.). The list is not definitive and new specific keywords can easily be added if needed. *IMGT/LIGM-DB* standardized keywords have been assigned to all entries.

2.2. DESCRIPTION concept: standardized sequence annotation

Two hundred and fifteen feature labels are necessary to describe all structural and functional subregions that compose IG and TR sequences, whereas only seven of them are available in *EMBL*, *GenBank* or *DDBJ* ***(14–16)***. Levels of annotation have been defined, which allow the users to query sequences in *IMGT/LIGM-DB* even though they are not fully annotated. Prototypes represent the organizational relationship between labels and give information on the order and expected length (in number of nucleotides) of the labels. This provides rules to verify the manual annotation and to design automatic annotation tool. One hundred and seventy-two additional feature labels have been defined for the 3D structures. Annotation of sequences and 3D structures with these labels

Table 1
***IMGT-ONTOLOGY* main concepts, *IMGT* Scientific chart rules, and examples of *IMGT*® expertised data concepts**

IMGT-ONTOLOGY main concepts ***(5)***	*IMGT* Scientific chart rules ***(4)***	Examples of *IMGT*® expertised data concepts
IDENTIFICATION	Standardized keywords	Species, molecule type, receptor type, chain type, gene type, structure, functionality, specificity
DESCRIPTION	Standardized labels and annotations	Core (V-, D-, J-, C-REGION) Prototypes Labels for sequences Labels for 2D and 3D structures
CLASSIFICATION	Reference sequences Standardized IG and TR gene nomenclature (group, subgroup, gene, and allele)	Nomenclature of the human IG and TR genes ***(1,2)*** [entry in 1999 in GDB ***(8)***, HGNC ***(9)***, and LocusLink and Entrez Gene at NCBI] Alignment of alleles Nomenclature of the IG and TR genes of all vertebrate species
NUMEROTATION	IMGT unique numbering for: V- and V-LIKE-DOMAINs ***(10)*** C- and C-LIKE-DOMAINs ***(11)*** G- and G-LIKE-DOMAINs ***(12)***	Protein displays Colliers de Perles ***(13)*** FR-IMGT and CDR-IMGT delimitations Structural loops and beta strands delimitations
ORIENTATION	Orientation of genomic instances relative to each other	Chromosome orientation Locus orientation Gene orientation DNA strand orientation
OBTENTION	Standardized origin and methodology	

(in capital letters) constitutes the main part of the expertise. Interestingly, 65 *IMGT*®-specific labels have been entered in the newly created Sequence Ontology ***(17)***.

2.3. CLASSIFICATION concept: standardized IG and TR gene nomenclature

The objective is to provide immunologists and geneticists with a standardized nomenclature per locus and per species which allows extraction and comparison of data for the complex B-cell and T-cell antigen receptor molecules. The concepts of classification have been used to set up a unique nomenclature of human IG and TR genes, which was approved by the Human Genome Organization (HUGO) Nomenclature Committee *HGNC* in 1999 ***(9)***. All the human IG and TR genes ***(1,2,18,19)*** have been entered by the *IMGT* Nomenclature Committee in *Genome Database GDB* ***(8)***, *LocusLink* and *Entrez Gene* at NCBI, USA, and in *IMGT/GENE-DB* ***(20)***. *IMGT* reference sequences have been defined for each allele of each gene based on one or, whenever possible, several of the following criteria: germline sequence, first sequence published, longest sequence, and mapped sequence. They are listed in the germline gene tables of the *IMGT* Repertoire. The *IMGT* Protein displays show the translated sequences of the alleles *01 of the functional or ORF genes ***(1,2)***.

2.4. NUMEROTATION concept: the IMGT unique numbering

A uniform numbering system for IG and TR sequences of all species has been established to facilitate sequence comparison and cross-referencing between experiments from different laboratories whatever the antigen receptor (IG or TR), the chain type or the species ***(21,22)***.

This numbering results from the analysis of more than 5,000 IG and TR variable region sequences of vertebrate species from fish to human. It takes into account and combines the definition of the framework (FR) and complementarity determining region (CDR) ***(23)***, structural data from X-ray diffraction studies ***(24)***, and the characterization of the hypervariable loops ***(25)***. In the *IMGT* unique numbering, conserved amino acids from FR always have the same number whatever the IG or TR variable sequence and whatever the species they come from, for example cysteine 23 (in FR1-IMGT), tryptophan 41 (in FR2-IMGT), leucine (or other hydrophobic amino acid) 89, and cysteine 104 (in FR3-IMGT). Tables and two-dimensional (2D) graphical representations designated as *IMGT* Colliers de Perles are available on the *IMGT*® Web site

at http://imgt.cines.fr and in the works of M.-P. Lefranc and G. Lefranc ***(1,2)***. The *IMGT* Collier de Perles of a variable domain or V-DOMAIN of an IG light chain is shown, as an example, in Fig. 1.

This *IMGT* unique numbering has several advantages:

1. It has allowed the redefinition of the limits of the FR and CDR of the IG and TR variable domains. The FR-IMGT and CDR-IMGT lengths become in themselves crucial information, which characterize variable regions belonging to a group, a subgroup, and/or a gene.
2. FR amino acids (and codons) located at the same position in different sequences can be compared without requiring sequence alignments. This also holds for amino acids belonging to CDR-IMGT of the same length.
3. The unique numbering is used as the output of the *IMGT/V-QUEST* alignment tool. The aligned sequences are displayed according to the *IMGT* unique numbering and with the FR-IMGT and CDR-IMGT delimitations.
4. The unique numbering has allowed a standardization of the description of mutations and the description of IG and TR allele polymorphisms ***(1,2)***. The mutations and allelic polymorphisms of each gene are described by comparison to the *IMGT* reference sequences of the allele *01 ***(1,2)***.
5. The unique numbering allows the description and comparison of somatic hypermutations of the IG variable domains.

By facilitating the comparison between sequences and by allowing the description of alleles and mutations, the *IMGT* unique numbering represents a big step forward in the analysis of the IG and TR sequences of all vertebrate species. Moreover, it gives insight into the structural configuration of the domains and opens interesting views on the evolution of these sequences, as this numbering can be used for all sequences belonging to the V-set and C-set of the IgSF. Structural and functional domains of the IG and TR chains comprise the V-DOMAIN (9-strand beta-sandwich) (Fig. 2), which corresponds to the V-J-REGION or V-D-J-REGION and is encoded by two or three genes ***(1,2)***, and the constant domain or C-DOMAIN (7-strand beta-sandwich) (Fig. 2). The *IMGT* unique numbering has been initially defined for the V-DOMAINs of the IG and TR and for the V-LIKE-DOMAINs of IgSF proteins other than IG and TR, for example in vertebrates human CD4 and *Xenopus* CTXg1 and in invertebrates *Drosophila* amalgam and *Drosophila* fasciclin II. ***(10,26)***. It has been extended to the C-DOMAINs of the IG and TR and to the C-LIKE-DOMAINs of IgSF proteins other than IG and TR ***(11,26,27)***. More recently, the IMGT unique numbering has also been defined for the groove domain or G-DOMAIN (four beta-strand and one alpha-helix) (Fig. 2) of the MHC class I and class II chains and for the G-LIKE-DOMAINs of MhcSF proteins other than MHC, for example MICA ***(12,28)***.

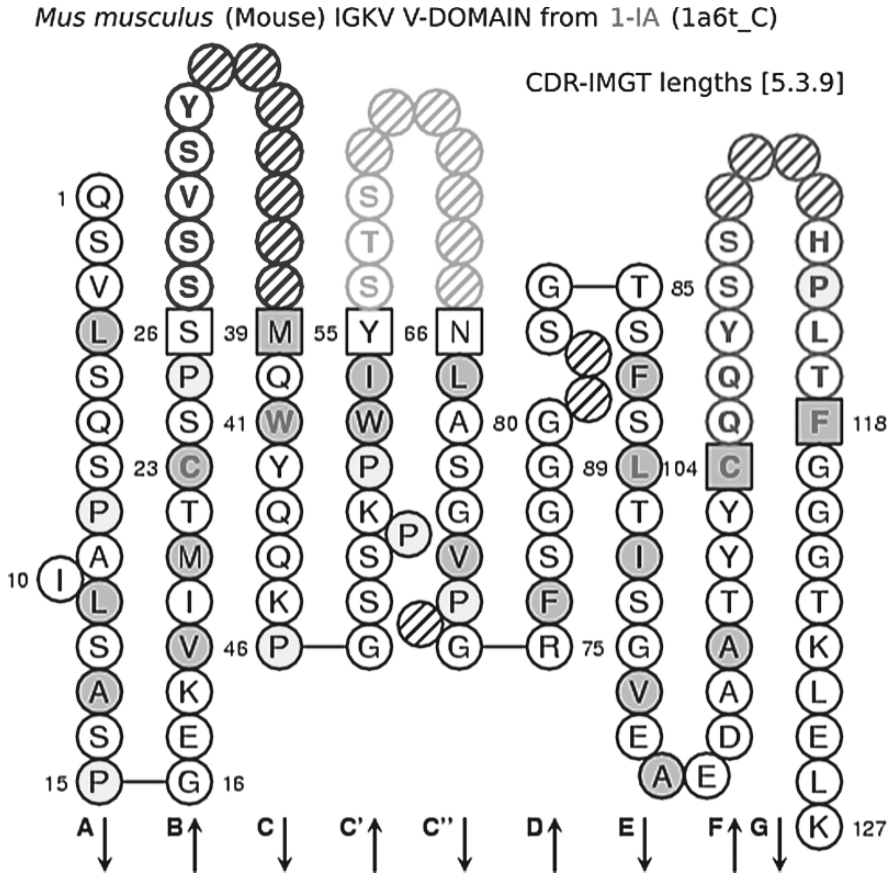
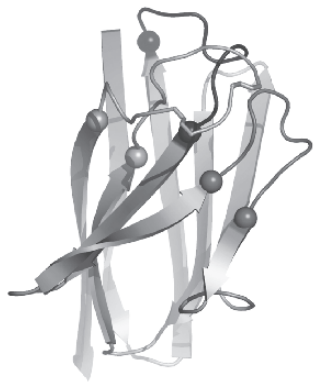
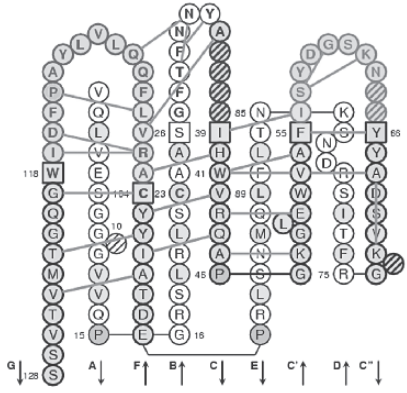


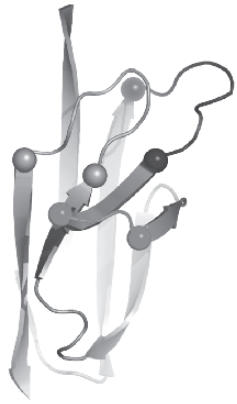
Fig. 1. *IMGT* Collier de Perles of a V-DOMAIN. The *IMGT* Collier de Perles of V-DOMAIN is based on the *IMGT* unique numbering for V-DOMAIN and V-LIKE-DOMAIN ***(10)***. Amino acids are shown in the one-letter abbreviation. The CDR-IMGT are limited by amino acids shown in squares, which belong to the neighbouring FR-IMGT. The CDR3-IMGT extend from position 105 to position 117. Hatched circles correspond to missing positions according to the IMGT unique numbering for V-DOMAIN and V-LIKE-DOMAIN ***(10)***. Arrows indicate the direction of the nine beta strands that form the two beta sheets of the immunoglobulin (IG) fold ***(1,2)***. Positions at which hydrophobic amino acids (hydropathy index with positive value: I, V, L, F, C, M and A) and tryptophan (W) are found in more than 50% of analysed IG and TR sequences are shown in blue. All proline (P) are shown in yellow. The V-DOMAIN chosen as an example is a murine IG light kappa domain or V-KAPPA (*IMGT/3Dstructure-DB*: 1a6t_C). CDR-IMGT regions (for a IG light kappa or lambda, or a TR alpha or gamma V-DOMAIN) are coloured as follows: CDR1-IMGT (blue), CDR2-IMGT (bright green) and CDR3-IMGT (dark green). (*See* Color Plate 1 following p. 32.)

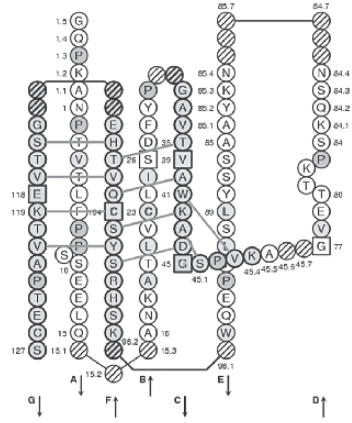
(A) **V-DOMAIN (IG, TR)**



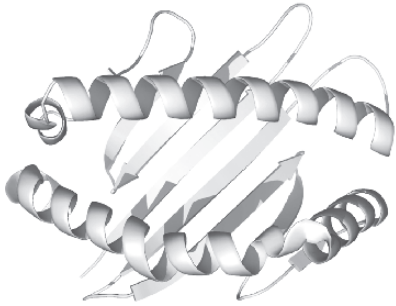


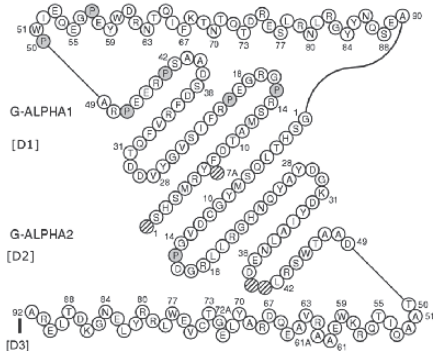
(B) **C-DOMAIN (IG, TR)**





(C) **G-DOMAIN (MHC)**





2.5. ORIENTATION concept: orientation of genomic instances relative to each other

The ORIENTATION concept allows to set up genomic orientation (for chromosome, locus, and gene) and DNA strand orientation. It is particularly useful in large genomic projects to localize a gene in a locus and/or a sequence (or a clone) in a contig or on a chromosome.

2.6. OBTENTION concept: controlled vocabulary for biological origin and experimental methodology

The OBTENTION concept, that is still in development, will be particularly useful for clinical data integration. This will help us to compare the repertoires of the IG antibody recognition sites and of the TR recognition sites in normal and pathological situations (autoimmune diseases, infectious diseases, leukemias, lymphomas, and myelomas).

3. *IMGT*® Genomics, Genetics, and Structural Approaches

In order to extract knowledge from *IMGT*® standardized immunogenetics data, three main *IMGT*® biological approaches have been developed: genomics, genetics, and structural approaches (Table 2). The *IMGT*® genomics approach

Fig. 2. Three-dimensional structures and *IMGT* Collier de Perles of a V-DOMAIN, a C-DOMAIN and G-DOMAINs. (**A**) V-DOMAIN. The *IMGT* Collier de Perles is based on the *IMGT* unique numbering for V-DOMAIN and V-LIKE-DOMAIN ***(10)***. The V-DOMAIN chosen as an example is a human immunoglobulin (IG) variable heavy domain or VH (*IMGT/3Dstructure-DB*: 1aqk_H). CDR-IMGT regions (for a IG heavy, or a TR beta or delta V-DOMAIN) are colored as follows: CDR1-IMGT (red), CDR2-IMGT (orange) and CDR3-IMGT (purple). Arrows indicate the direction of the nine beta strands of the V-DOMAIN that form the two beta sheets of the IG fold ***(1,2)***. Hydrogen bonds of the [GFCC′C′] sheet are shown with green lines. (**B**) C-DOMAIN. The *IMGT* Collier de Perles is based on the *IMGT* unique numbering for C-DOMAIN and C-LIKE-DOMAIN ***(11)***. The C-DOMAIN chosen as an example is a human IG constant light lambda domain or C-LAMBDA (*IMGT/3Dstructure-DB*: 1mcd_B). Arrows indicate the direction of the seven beta strands of the C-DOMAIN that form the two beta sheets of the IG fold ***(1,2)***. Hydrogen bonds of the [GFC] sheet are shown with green lines. (**C**) G-DOMAINs. The *IMGT* Colliers de Perles are based on the *IMGT* unique numbering for G-DOMAIN and G-LIKE-DOMAIN ***(12)***. The G-DOMAINs chosen as examples are human major histocompatibility complex (MHC) class I alpha groove domains or G-ALPHA1 and G-ALPHA2 (*IMGT/3Dstructure-DB*: 1agb_A). Amino acids are shown in the one-letter abbreviation. Hatched circles correspond to missing positions according to the IMGT unique numbering ***(10–12)***. (*See* Color Plate 2 following p. 32.)

Table 2
***IMGT*® databases, tools and Web resources for genomics, genetics and structural approaches**

Approaches	Databases	Tools	Web resources[a]
Genomics	*IMGT/GENE-DB* ***(16)***	*IMGT/GeneView* *IMGT/LocusView* *IMGT/CloneSearch* *IMGT/GeneSearch* *IMGT/GeneInfo* ***(29)***	*IMGT* Repertoire "Locus and genes" section: • Chromosomal localizations ***(1,2)*** • Locus representations ***(1,2)*** • Locus description • Gene tables, etc. • Potential germline repertoires • Lists of genes • Correspondence between nomenclatures ***(1,2)***
Genetics	*IMGT/LIGM-DB* ***(30)*** *IMGT/PRIMER-DB* ***(31)*** *IMGT/MHC-DB* ***(32)***	*IMGT/V-QUEST* ***(33)*** *IMGT/JunctionAnalysis* ***(34)*** *IMGT/Allele-Align* *IMGT/PhyloGene* ***(35)***	*IMGT* Repertoire "Proteins and alleles' section: • Alignments of alleles • Protein displays • Tables of alleles, etc.
Structural	*IMGT/3Dstructure-DB* ***(36)***	*IMGT/StructuralQuery* ***(36)***	*IMGT* Repertoire "2D and 3D structures" section: • *IMGT* Colliers de Perles (2D representations on one layer or two layers) • *IMGT*® classes for amino acid characteristics ***(37)*** • *IMGT* Colliers de Perles reference profiles ***(37)*** • 3D representations

[a] Only Web resources examples from the *IMGT* Repertoire section are shown.

is gene-centered and mainly orientated towards the study of the genes within their loci and on the chromosomes. The *IMGT®* genetics approach refers to the study of the genes in relation with their sequence polymorphisms and mutations, their expression, their specificity, and their evolution. The genetics approach heavily relies on the DESCRIPTION concept (and particularly on the V-, D-, J- and C-REGION core concepts for the IG and TR), on the CLASSIFICATION concept (*IMGT®* gene and allele names) and on the NUMEROTATION concept [*IMGT* unique numbering ***(10–12)***]. The *IMGT®* structural approach refers to the study of the 2D and 3D structures of the IG, TR, MHC, and RPI and to the antigen- or ligand-binding characteristics in relationship with the protein functions, polymorphisms and evolution. The structural approach relies on the CLASSIFICATION concept (*IMGT®* gene and allele names), DESCRIPTION concept (receptor and chain description and domain delimitations), and NUMEROTATION concept [amino acid positions according to the *IMGT* unique numbering ***(10–12)***].

For each approach, *IMGT®* provides databases [one genome database (*IMGT/GENE-DB*), three sequence databases (*IMGT/LIGM-DB*, *IMGT/MHC-DB*, and *IMGT/PRIMER-DB*), one 3D structure database (*IMGT/3Dstructure-DB*)], interactive tools (ten on-line tools for genome, sequence and 3D structure analysis), and *IMGT* Repertoire Web resources (providing an easy-to-use interface to carefully and expertly annotated data on the genome, proteome, and polymorphism and structural data of the IG and TR, MHC and RPI) (Table 2). These databases, tools, and Web resources are detailed in the following Sections 4–6. Other *IMGT®* Web resources include:

1. *IMGT* Bloc-notes (Interesting links, etc.) provides numerous hyperlinks towards the Web servers specializing in immunology, genetics, molecular biology, and bioinformatics (associations, collections, companies, databases, immunology themes, journals, molecular biology servers, resources, societies, tools, etc.) ***(38)***.
2. *IMGT* Lexique.
3. The *IMGT* Immunoinformatics page.
4. The *IMGT* Medical page.
5. The *IMGT* Veterinary page.
6. The *IMGT* Biotechnology page.
7. *IMGT* Education (Aide-mémoire, Tutorials, Questions and answers, etc.) provides useful biological resources for students and includes figures and tutorials (in English and/or in French) in immunogenetics.
8. *IMGT* Aide-mémoire provides an easy access to information such as genetic code, splicing sites, amino acid structures, and restriction enzyme sites.
9. *IMGT* Index is a fast way to access data when information has to be retrieved from different parts of the *IMGT* site. For example, “allele” provides links to the

IMGT Scientific chart rules for the allele description and to the *IMGT* Repertoire "Alignments of alleles" and "Tables of alleles" (http://imgt.cines.fr).

4. *IMGT*® Databases, Tools, and Web Resources for Genomics

Genomic data are managed in *IMGT/GENE-DB*, which is the comprehensive IMGT® genome database ***(20)***. In February 2007, *IMGT/GENE-DB* contained 1,512 IG and TR genes and 2,461 alleles from human and mouse IG and TR genes. Based on the *IMGT*® CLASSIFICATION concept, all the human *IMGT*® gene names ***(1,2)***, approved by the HUGO Nomenclature Committee HGNC in 1999, are available in *IMGT/GENE-DB* ***(20)*** and in Entrez Gene at NCBI (USA). All the mouse *IMGT*® gene and allele names and the corresponding IMGT reference sequences were provided to Mouse Genome Informatics MGI Mouse Genome Database MGD in July 2002 and were presented by *IMGT*® at the 19th International Mouse Genome Conference IMGC 2005, in Strasbourg, France. *IMGT-GENE-DB* allows a query per gene and allele name. *IMGT/GENE-DB* interacts dynamically with *IMGT/LIGM-DB* ***(30)*** to download and display human and mouse gene-related sequence data. This is the first example of an interaction between *IMGT*® databases using the CLASSIFICATION concept.

The *IMGT*® genome analysis tools manage the locus organization and gene location and provide the display of physical maps for the human and mouse IG, TR, and MHC loci. They allow to view genes in a locus (*IMGT/GeneView* and *IMGT/LocusView*) to search for clones (*IMGT/CloneSearch*), to search for genes in a locus (*IMGT/GeneSearch* and *IMGT/GeneInfo*) based on *IMGT*® gene names, functionality or localization on the chromosome, to provide information on the clones that were used to build the locus contigs (accession numbers are from *IMGT/LIGM-DB* and gene names from *IMGT/GENE-DB*) or to display information on the human and mouse IG and TR potential rearrangements.

The *IMGT* Repertoire genome data include chromosomal localizations, locus representations, locus description, germline gene tables, potential germline repertoires, lists of IG and TR genes and links between *IMGT*, *HGNC*, *GDB*, *Entrez Gene*, and *OMIM*, and correspondence between nomenclatures ***(1,2)***.

5. *IMGT*® Databases, Tools, and Web Resources for Genetics

IMGT/LIGM-DB ***(30)*** is the comprehensive *IMGT*® database of IG and TR nucleotide sequences from human and other vertebrate species, with translation for fully annotated sequences, created in 1989 by LIGM, Montpellier, France, on the Web since July 1995. IMGT/LIGM-DB is the first and the largest *IMGT*® database. In February 2007, *IMGT/LIGM-DB* contained 105,188

nucleotide sequences of IG and TR from 150 species. The unique source of data for *IMGT/LIGM-DB* is *EMBL* that shares data with the other two generalist databases *GenBank* and *DDBJ*. *IMGT/LIGM-DB* sequence data are identified by the *EMBL/GenBank/DDBJ* accession number. Based on expert analysis, specific detailed annotations are added to *IMGT* flat files.

Since August 1996, the *IMGT/LIGM-DB* content closely follows the *EMBL* one for the IG and TR, with the following advantages: *IMGT/LIGM-DB* does not contain sequences that have previously been wrongly assigned to IG and TR; conversely, *IMGT/LIGM-DB* contains IG and TR entries that have disappeared from the generalist databases [for example, the L36092 accession number that encompasses the complete human TRB locus is still present in *IMGT/LIGM-DB*, whereas it has been deleted from *EMBL/GenBank/DDBJ* due to its too large size (684,973 bp); in 1999, *IMGT/LIGM-DB* detected the disappearance of 20 IG and TR sequences that inadvertently had been lost by *GenBank*, and allowed the recuperation of these sequences in the generalist databases].

The *IMGT/LIGM-DB* annotations (gene and allele name assignment, labels) allow data retrieval not only from *IMGT/LIGM-DB* but also from other *IMGT®* databases. For example, the *IMGT/GENE-DB* entries provide the *IMGT/LIGM-DB* accession numbers of the IG and TR cDNA sequences that contain a given V, D, J or C gene. The automatic annotation of rearranged human and mouse cDNA sequences in *IMGT/LIGM-DB* is performed by *IMGT/Automat **(39)***, an internal Java tool that implements *IMGT/V-QUEST* and *IMGT/ JunctionAnalysis.*

Standardized information on oligonucleotides (or primers) and combinations of primers (Sets and Couples) for IG and TR are managed in *IMGT/PRIMER-DB **(31)***, the *IMGT®* oligonucleotide database on the Web since February 2002. *IMGT/MHC-DB **(32)*** hosted at EBI comprises *IMGT/HLA* for human MHC (or HLA) and *IMGT/MHC-NHP* for MHC of non-human primates.

The *IMGT®* tools for the genetics approach comprise *IMGT/V-QUEST **(33**, **40)*** for the identification of the V, D, and J genes and of their mutations, *IMGT/JunctionAnalysis **(34**,**40)*** for the analysis of the V-J and V-D-J junctions that confer the antigen receptor specificity, *IMGT/Allele-Align* for the detection of polymorphisms, and *IMGT/Phylogene **(35)*** for gene evolution analyses. *IMGT/V-QUEST* (V-QUEry and STandardization) (http://imgt.cines.fr) is an integrated software for IG and TR ***(33**,**40)***. This tool, easy to use, analyses an input IG or TR germline or rearranged variable nucleotide sequence. *IMGT/V-QUEST* results comprise the identification of the V, D, and J genes and alleles and the nucleotide alignment by comparison with sequences from the *IMGT* reference directory, the delimitations of the FR-IMGT and CDR-IMGT based

on the *IMGT* unique numbering, the protein translation of the input sequence, the identification of the JUNCTION, the description of the mutations and amino acid changes of the V-REGION, and the 2D *IMGT* Collier de Perles representation of the V-REGION or V-DOMAIN. The set of sequences from the *IMGT* reference directory, used for *IMGT/V-QUEST*, can be downloaded in *FASTA* format from the *IMGT*® site.

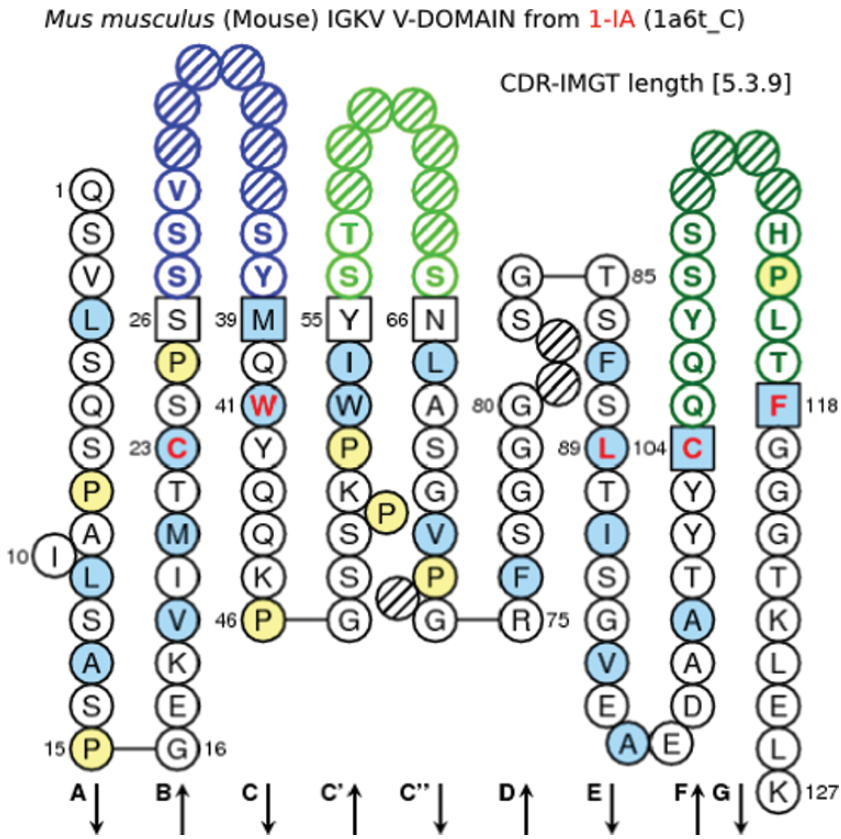
IMGT/JunctionAnalysis ***(34,40)*** is a tool developed by LIGM, complementary to *IMGT/V-QUEST*, which provides a thorough analysis of the V-J and V-D-J junction of IG and TR rearranged genes. The JUNCTION extends from 2nd-CYS 104 to J-PHE or J-TRP 118 inclusive. J-PHE or J-TRP are easily identified for in-frame rearranged sequences when the conserved Phe/Trp-Gly-X-Gly motif of the J-REGION is present. The length of the CDR3-IMGT of rearranged V-J-GENEs or V-D-J-GENEs is a crucial piece of information. It is the number of amino acids or codons from position 105 to 117 (J-PHE or J-TRP non-inclusive). CDR3-IMGT amino acid and codon numbers are according to the IMGT unique numbering for V-DOMAIN ***(10)***. *IMGT/JunctionAnalysis* identifies the D-GENE and allele involved in the IGH, TRB, and TRD V-D-J rearrangements by comparison with the *IMGT* reference directory and delimits precisely the P, N, and D regions ***(1,2)***. Results from *IMGT/JunctionAnalysis* are more accurate than those given by *IMGT/V-QUEST* regarding the D-GENE identification. Indeed, *IMGT/JunctionAnalysis* works on shorter sequences (JUNCTION) and with a higher constraint because the identification of the V-GENE and J-GENE and alleles is a prerequisite to perform the analysis. Several hundreds of junction sequences can be analysed simultaneously.

Other *IMGT*® Tools for sequence analysis comprise *IMGT/Allele-Align* that allows the comparison of two alleles highlighting the nucleotide and amino acid differences and *IMGT/PhyloGene* ***(35)***, an easy-to-use tool for phylogenetic analysis of IMGT standardized reference sequences.

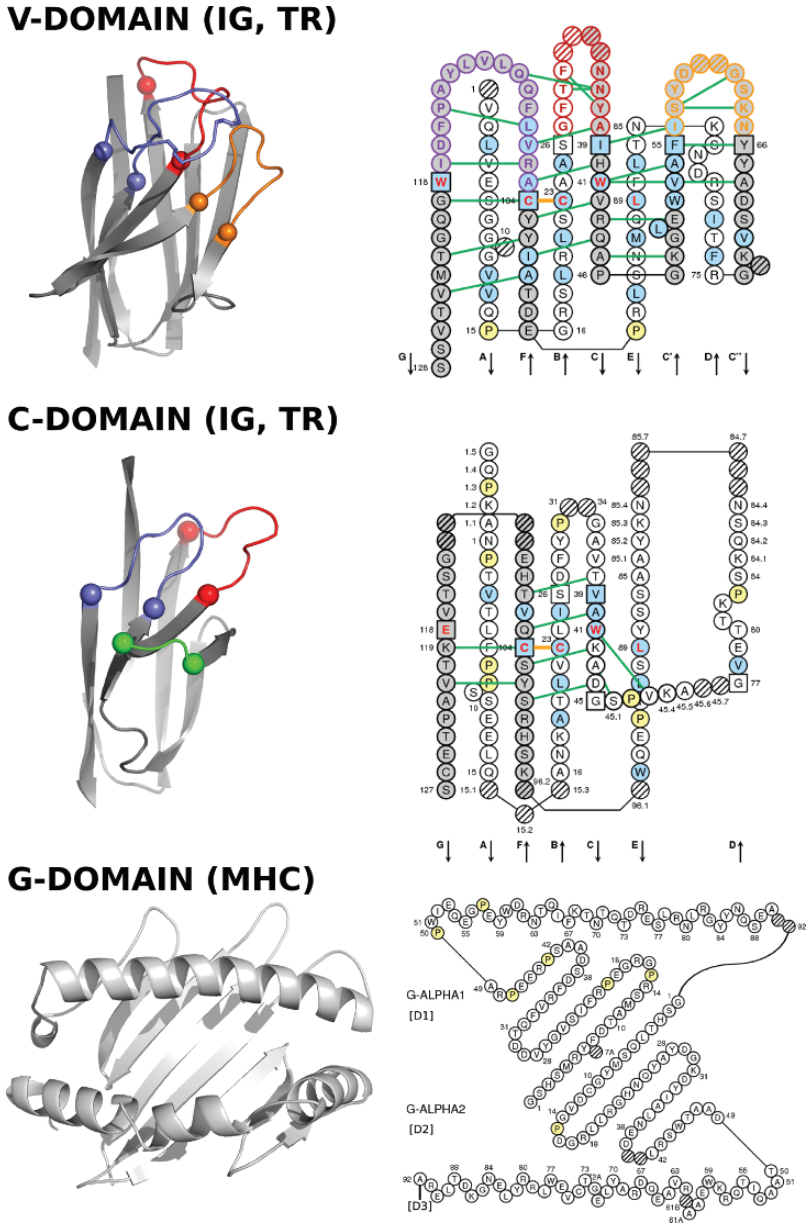
The *IMGT* Repertoire polymorphism data are represented by "Alignments of alleles," "Tables of alleles,"' "Allotypes,", "Protein displays," particularities in protein designations, *IMGT* reference directory in *FASTA* format, correspondence between IG and TR chain and receptor *IMGT* designations ***(1,2)***.

6. *IMGT*® Databases, Tools, and Web Resources for Structural Analysis

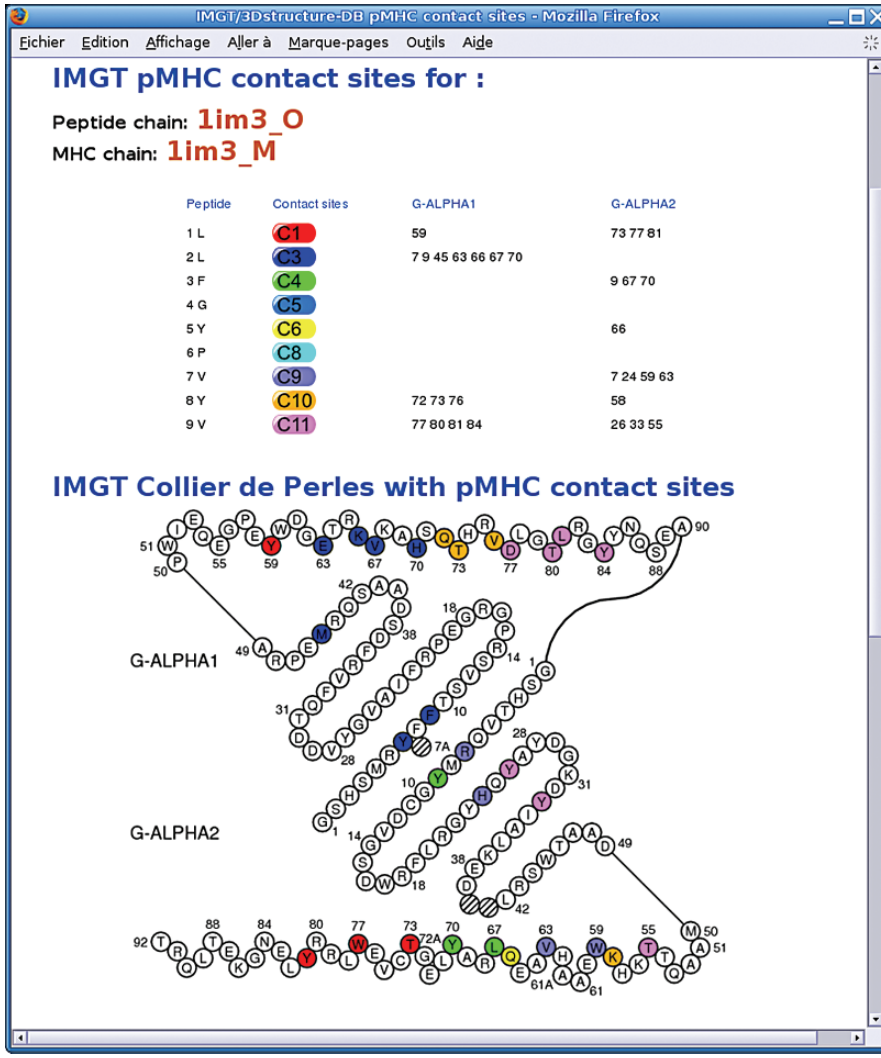
Structural data are compiled and annotated in *IMGT/3Dstructure-DB* ***(36)***, the *IMGT*® 3D structure database, created by LIGM, on the Web since November 2001. *IMGT/3Dstructure-DB* comprises IG, TR, MHC, and RPI with known 3D structures. In February 2007, *IMGT/3Dstructure-DB* contained 1,221



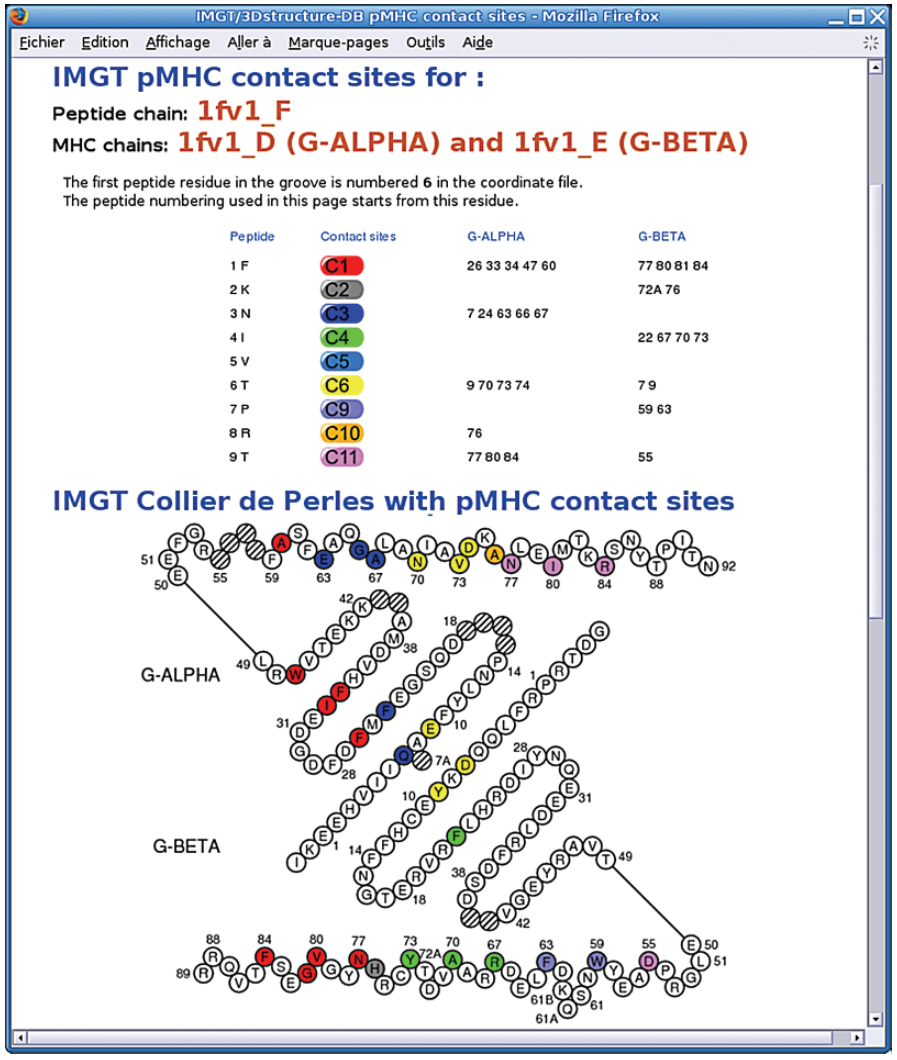
Color Plate 1, *IMGT* Collier de Perles of a V-DOMAIN. The *IMGT* Collier de Perles of V-DOMAIN is based on the *IMGT* unique numbering for V-DOMAIN and V-LIKE-DOMAIN (***10***) (Chapter 2, Fig. 1; *see* full caption on p. 25 and discussion on p. 24.)



Color Plate 2, Three-dimensional structures and *IMGT* Collier de Perles of a V-DOMAIN, a C-DOMAIN and G-DOMAINs. (Chapter 2, Fig. 2; *see* full caption on p. 27 and discussion on p. 24.)



Color Plate 3, *IMGT* pMHC contact sites of human HLA-A*0201 MHC-I and a 9-amino acid peptide side chains (*IMGT/3Dstructure-DB*: 1im3). (Chapter 2, Fig. 4; *see* full caption on p. 36 and discussion on p. 35.)



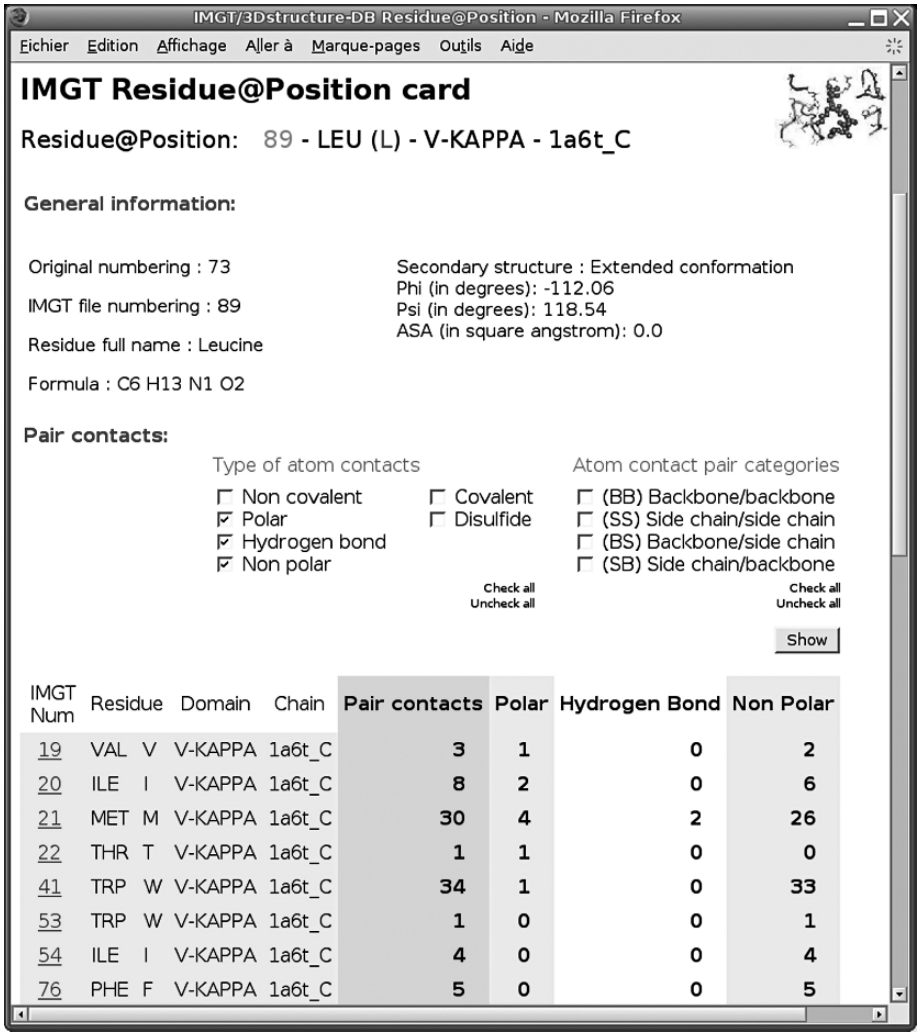
Color Plate 4, *IMGT* pMHC contact sites of human HLA-DRA*0101 and HLA-DRB5*0101 MHC-II and the peptide side chains (9 amino acids located in the groove) (*IMGT/3Dstructure-DB*: 1fv1). (Chapter 2, Fig. 5; *see* full caption on p. 37 and discussion on p. 35.)

atomic coordinate files. These coordinate files, extracted from the *Protein Data Bank* (*PDB*) ***(41)***, are renumbered according to the standardized *IMGT* unique numbering ***(10–12)***. The *IMGT/3Dstructure-DB* cards provide *IMGT*® annotations (assignment of *IMGT*® genes and alleles, *IMGT*® chain and domain labels, and *IMGT* Colliers de Perles on one layer and two layers), downloadable renumbered *IMGT/3Dstructure-DB* flat files, visualization tools and external links. *IMGT/3Dstructure-DB* residue cards provide detailed information on the inter- and intra-domain contacts of each residue position (Fig. 3).

The *IMGT/StructuralQuery* tool ***(36)*** analyses the intramolecular interactions for the V-DOMAINs. The contacts are described per domain (intra- and inter-domain contacts) and annotated in term of *IMGT*® labels (chains and domain), positions (*IMGT* unique numbering), backbone or side-chain implication. *IMGT/StructuralQuery* allows to retrieve the *IMGT/3Dstructure-DB* entries, based on specific structural characteristics: phi and psi angles, accessible surface area (ASA), amino acid type, distance in angstrom between amino acids, and CDR-IMGT lengths ***(36)***.

In order to appropriately analyse the amino acid resemblances and differences between IG, TR, MHC, and RPI chains, 11 *IMGT*® classes were defined for the amino acid "chemical characteristics" properties and used to set up *IMGT* Colliers de Perles reference profiles ***(37)***. The *IMGT* Colliers de Perles reference profiles allow to easily compare amino acid properties at each position whatever the domain, the chain, the receptor or the species ***(37)***. The IG and TR variable and constant domains and the MHC groove domains represent a privileged situation for the analysis of amino acid properties in relation with 3D structures, by the conservation of their 3D structure despite divergent amino acid sequences and by the considerable amount of genomic (*IMGT* Repertoire), structural (*IMGT/3Dstructure-DB*) and functional data available. These data are not only useful to study mutations and allele polymorphisms but are also needed to establish correlations between amino acids in the protein sequences and 3D structures, to analyse the IgSF and MhcSF domain interactions ***(42)*** and to determine amino acids potentially involved in the immunogenicity. One of the key elements in the adaptive immune response is the presentation of peptides by the MHC to the TR at the surface of T cells. The characterization of the TR/peptide/MHC trimolecular complexes (TR/pMHC) is crucial to the fields of immunology, vaccination, and immunotherapy. In *IMGT/3Dstructure-DB*, TR/pMHC molecular characterization, and pMHC contact analysis have been standardized, based on the *IMGT* unique numbering for G-DOMAIN, and 11 *IMGT* pMHC contact sites (C1–C11) have been defined ***(43)***. The *IMGT* pMHC contact sites represent the MHC amino acid positions that have contacts

with the peptide side chains. They are particularly useful to compare pMHC interactions whatever the MHC classes or chains, whatever the species and whatever the peptide sequence or length ***(43)***. There are no C2, C7, and C8 contact sites for MHC-I with 8-amino acid peptides and no C2 and C7 for MHC-I 3D structures with 9-amino acid peptides. In contrast, for MHC-II, C2 is present but there are no C7 and C8 ***(43)***. The *IMGT* pMHC contact sites are provided dynamically for the pMHC and the TR/pMHC 3D structures



IMGT Num	Residue		Domain	Chain	Pair contacts	Polar	Hydrogen Bond	Non Polar
19	VAL	V	V-KAPPA	1a6t_C	3	1	0	2
20	ILE	I	V-KAPPA	1a6t_C	8	2	0	6
21	MET	M	V-KAPPA	1a6t_C	30	4	2	26
22	THR	T	V-KAPPA	1a6t_C	1	1	0	0
41	TRP	W	V-KAPPA	1a6t_C	34	1	0	33
53	TRP	W	V-KAPPA	1a6t_C	1	0	0	1
54	ILE	I	V-KAPPA	1a6t_C	4	0	0	4
76	PHE	F	V-KAPPA	1a6t_C	5	0	0	5

available in *IMGT/3Dstructure-DB*. For example, the *IMGT* pMHC contact sites of a MHC-I (human HLA-A*0201) and a 9-amino acid peptide side chains are shown in Fig. 4 (*IMGT/3Dstructure-DB*: 1im3), and the *IMGT* pMHC contact sites of a MHC-II (human HLA-DRA*0101 and HLA-DRB5*0101) binding 9 amino acids of the peptide in the groove are shown in Fig. 5 (*IMGT/3Dstructure-DB*: 1fv1).

The *IMGT* Repertoire Structural data comprise *IMGT* Colliers de Perles ***(1,2,10–12)***, FR-IMGT and CDR-IMGT lengths, and 3D representations of IG and TR variable domains. This visualization permits rapid correlation between protein sequences and 3D data retrieved from the *PDB*.

7. Conclusion

Since July 1995, *IMGT®* has been available on the Web at http://imgt.cines.fr. *IMGT®* has an exceptional response with more than 140,000 requests a month. The information is of much value to clinicians and biological scientists in general. *IMGT®* databases, tools, and Web resources are extensively queried and used by scientists from both academic and industrial laboratories, who are equally distributed between the United States, Europe, and the remaining world. *IMGT®* is used in very diverse domains: (i) fundamental and medical research (repertoire analysis of the IG antibody recognition sites and of the TR recognition sites in normal and pathological situations such as autoimmune diseases, infectious diseases, AIDS, leukemias, lymphomas, and

Fig. 3. *IMGT* Residue@Position card. The identification of a "*IMGT* Residue@Position" comprises the position number according to the *IMGT* unique numbering ***(10–12)***, the residue name (with three letters and eventually one letter abbreviation), the domain description and the *IMGT/3Dstructure-DB* chain ID. The example shows the contacts of position 89, occupied by a leucine LEU (L), in the V-KAPPA domain of the 1a6t_C chain. The original number in the *PDB* file is indicated. The secondary structure, the phi and psi angles (in degrees) and accessible surface area (ASA) (in square angstroms) are provided. The user can select, for the result display, the types of contacts (non covalent, polar, hydrogen bond, non polar, covalent bond or disulfide bond) and the atom contact pair categories (backbone/backbone, side chain/side chain, backbone/side chain and side chain/backbone atoms). The results are shown as a table with a list of the *IMGT* Residue@Position which are in contact with the *IMGT* Residue@Position at the top of the card, and for each of them, the total number of atom pair contacts and the detailed description of the contacts as selected by the user are also indicated.

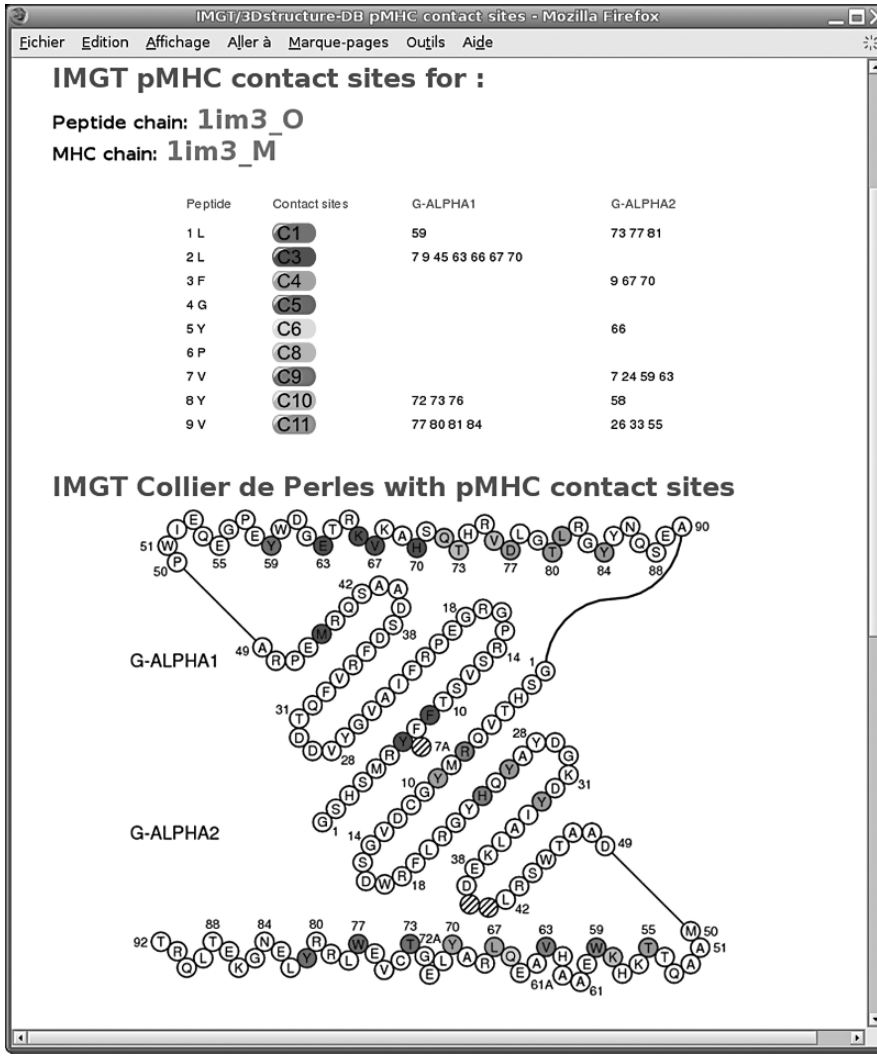


Fig. 4. IMGT peptide major histocompatibility complex (pMHC) contact sites of human HLA-A*0201 MHC-I and a 9-amino acid peptide side chains (IMGT/3Dstructure-DB: 1im3). The numbers 1–9 refer to the numbering of the peptide amino acids P1–P9. C1–C11 refer to the 11 pMHC contact sites defined by IMGT® ***(43)***. There are no C2 and C7 in MHC-I 3D structures with 9-amino acid peptides. There are no C5 and C8 in this 3D structure as P4 and P6 do not contact MHC amino acids. The view of the IMGT Collier de Perles is from above the cleft, with G-ALPHA1 on top and G-ALPHA2 on bottom of the figure. (*See* Color Plate 3 following p. 32.)

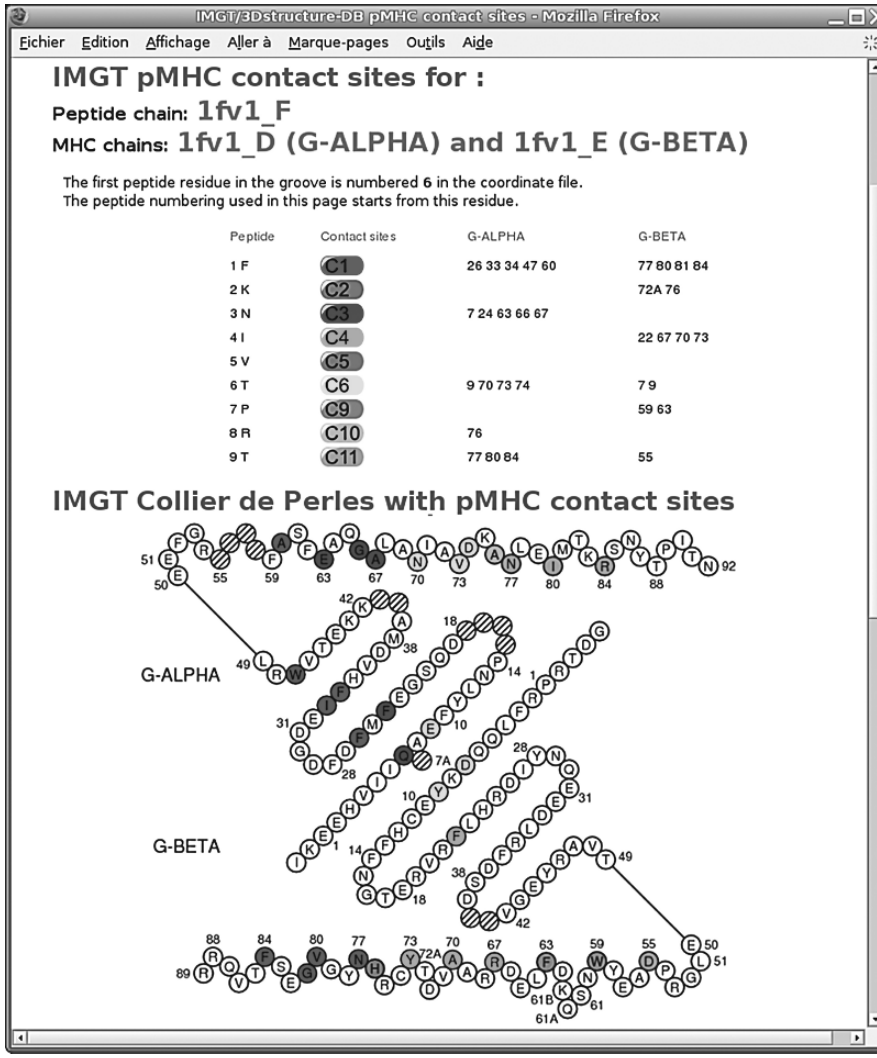


Fig. 5. IMGT peptide major histocompatibility complex (pMHC) contact sites of human HLA-DRA*0101 and HLA-DRB5*0101 MHC-II and the peptide side chains (9 amino acids located in the groove) (IMGT/3Dstructure-DB: 1fv1). The numbers 1–9 refer to the numbering of the peptide amino acids 1–9 located in the groove. C1–C11 refer to the 11 pMHC contact sites defined by IMGT® ***(43)***. There are no C7 and C8 in MHC-II 3D structures with peptide of 9 amino acids located in the groove. There is no C5 in this 3D structure as 5 does not contact MHC amino acids. The view of the IMGT Collier de Perles is from above the cleft, with G-ALPHA on top and G-BETA on bottom of the figure. (*See* Color Plate 4 following p. 32.)

myelomas), (ii) veterinary research (IG and TR repertoires in farm and wildlife species), (iii) genome diversity and genome evolution studies of the adaptive immune responses, (iv) structural evolution of the IgSF and MhcSF proteins, (v) biotechnology related to antibody engineering [single chain Fragment variable (scFv), phage displays, combinatorial libraries, chimeric, humanized, and human antibodies], (vi) diagnostics (clonalities, detection, and follow-up of residual diseases) and (vii) therapeutical approaches (grafts, immunotherapy, and vaccinology). Owing to its high quality and data distribution based on *IMGT-ONTOLOGY*, *IMGT®* has an important role to play in the development of immunogenetics Web services. The creation of dynamic interactions between the *IMGT®* databases and tools, using Web services and *IMGT-ML*, and the design of *IMGT-Choreography* ***(4)***, represents novel and major developments of *IMGT®*, the international reference in immunogenetics and immunoinformatics.

8. Citing *IMGT®*

Authors who make use of the information provided by *IMGT®* should cite **ref.** ***3*** as a general reference for the access to and content of *IMGT®* and quote the *IMGT®* home page URL, http://imgt.cines.fr.

Acknowledgments

I thank Véronique Giudicelli, Patrice Duroux, Quentin Kaas, Joumana Jabado-Michaloud, Géraldine Folch, Chantal Ginestoux, Denys Chaume, and Gérard Lefranc for helpful discussion. I am deeply grateful to the *IMGT®* team for its expertise and constant motivation. *IMGT®* is a registered mark of the Centre National de la Recherche Scientifique (CNRS). *IMGT®* has received the National Bioinformatics Platform RIO label since 2001 (CNRS, INSERM, CEA, and 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 (EU). *IMGT-ML* was developed in collaboration with the EU Online Research Information Environment for the Life Sciences, ORIEL project (IST-2001-32688). 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, Institut Universitaire de France, ACI-IMPBIO IMP82-2004), the EU ImmunoGrid (IST-028069) programme, GIS AGENAE, Réseau National des Génopoles and the Région Languedoc-Roussillon BIOSTIC-LR2004.

References

1. Lefranc, M.-P. and Lefranc, G. (2001) *The Immunoglobulin FactsBook*. Academic Press, London, UK, ISBN:012441351X, 458 pages.
2. Lefranc, M.-P. and Lefranc, G. (2001) *The T cell Receptor FactsBook*. Academic Press, London, UK, ISBN:0124413528, 398 pages.
3. Lefranc, M.-P., Giudicelli, V., Kaas, Q., Duprat, E., Jabado-Michaloud, J., Scaviner, D., Ginestoux, C., Clément, O., Chaume, D., and Lefranc G. (2005) IMGT, the International ImMunoGeneTics information system. *Nucleic Acids Res.* **33**, D593–D597.
4. Lefranc, M.-P., Clément, O., Kaas, Q., Duprat, E., Chastellan, P., Coelho, I., Combres, K., Ginestoux, C., Giudicelli, V., Chaume, D., and Lefranc, G. (2005) IMGT-Choreography for immunogenetics and immunoinformatics. Epub *In Silico Biol.* (Reference for Epub 5 0006) http://www.bioinfo.de/isb/2004/05/0006/24 December 2004 *In Silico Biol.* **5**, pp. 45–60.
5. Giudicelli, V. and Lefranc, M.-P. (1999) Ontology for immunogenetics: the IMGT-ONTOLOGY. *Bioinformatics* **12**, 1047–1054.
6. Chaume, D., Giudicelli, V., and Lefranc, M.-P. (2001) IMGT-ML a language for IMGT-ONTOLOGY and IMGT/LIGM-DB data. In: *CORBA and XML: Towards a Bioinformatics Integrated Network Environment, Proceedings of NETTAB 2001, Network tools and Applications in Biology,* May 17–18, Gchoa, Italy, pp. 71–75.
7. Chaume, D., Giudicelli, V., Combres, K., and Lefranc, M.-P. (2003) IMGT-ONTOLOGY and IMGT-ML for Immunogenetics and immunoinformatics. In: *Abstract book of the Sequence Databases and Ontologies Satellite Event,* European Congress in Computational Biology ECCB'2003, September 27–30, Paris, France, pp. 22–23.
8. Letovsky, S.I., Cottingham, R.W., Porter, C.J., and Li, P.W. (1998) GDB: the human genome database. *Nucleic Acids Res.* **26**, 94–99.
9. Wain, H.M., Bruford, E.A., Lovering, R.C., Lush, M.J., Wright, M.W., and Povey, S. (2002) Guidelines for human gene nomenclature. *Genomics* **79**, 464–470.
10. Lefranc, M.-P., Pommié, C., Ruiz, M., Giudicelli, V., Foulquier, E., Truong, L., Thouvenin-Contet, V., and Lefranc, G. (2003) IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains. *Dev. Comp. Immunol.* **27**, 55–77.
11. Lefranc, M.-P., Pommié, C., Kaas, Q., Duprat, E., Bosc, N., Guiraudou, D., Jean C., Ruiz M., Da Piedade, I., Rouard, M., Foulquier, E., Thouvenin, V., and Lefranc, G. (2005) IMGT unique numbering for immunoglobulin and T cell receptor constant domains and Ig superfamily C-like domains. *Dev. Comp. Immunol.* **29**, 185–203.
12. Lefranc, M.-P., Duprat, E., Kaas, Q., Tranne, M., Thiriot, A., and Lefranc, G. (2005) IMGT unique numbering for MHC groove G-DOMAIN and MHC superfamily (MhcSF) G-LIKE-DOMAIN. *Dev. Comp. Immunol.* **29**, 917–938.

13. Ruiz, M. and Lefranc, M.-P. (2002) IMGT gene identification and Colliers de Perles of human immunoglobulins with known 3D structures. *Immunogenetics* **53**, 857–883.
14. Cochrane, G., Aldebert, P., Althorpe, N., Andersson, M., Baker, W., Baldwin, A., Bates, K., Bhattacharyya, S., Browne, P., van den Broek, A., Castro, M., Duggan, K., Eberhardt, R., Faruque, N., Gamble, J., Kanz, C., Kulikova, T., Lee, C., Leinonen, R., Lin, Q., Lombard, V., Lopez, R., McHale, M., McWilliam, H., Mukherjee, G., Nardone, F., Garcia Pastor, M.P., Sobhany, S., Stoehr, P., Tzouvara, K., Vaughan, R., Wu, D., Zhu, W., and Apweiler, R.(2006) EMBL nucleotide sequence database: developments in 2005. *Nucleic Acids Res.* **34**, D10–D15.
15. Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., and Wheeler, D.L. (2006) GenBank. *Nucleic Acids Res.* **34**, D16–D20.
16. Okubo, K., Sugawara, H., Gojobori, T., and Tateno, Y. (2006) DDBJ in preparation for overview of research activities behind data submissions. *Nucleic Acids Res.* **34**, D6–D9.
17. Eilbeck, K., Lewis, S.E., Mungall, C.J., Yandell, M., Stein, L., Durbin, R., and Ashburner, M. (2005) The sequence ontology: a tool for the unification of genome annotations. *Genome Biol.* **6** (5), R44. Epub 29 Apr 2005.
18. Lefranc, M.-P. (2000) Nomenclature of the human immunoglobulin genes. In: *Current Protocols in Immunology* (Coligan, J.E., Bierer, B.E., Margulies, D.E., Shevach, E.M. and Strober W., eds.), John Wiley and Sons, Hoboken, N.J., A.1P.1–A.1P.37.
19. Lefranc, M.-P. (2000) Nomenclature of the human T cell receptor genes. In: *Current Protocols in Immunology* (Coligan, J.E., Bierer, B.E., Margulies, D.E., Shevach, E.M. and Strober, W., eds.), John Wiley and Sons, Hoboken, N.J., A.1O.1–A.1O.23.
20. Giudicelli, V., Chaume, D., and Lefranc, M.-P. (2005) IMGT/GENE-DB: a comprehensive database for human and mouse immunoglobulin and T cell receptor genes. *Nucleic Acids Res*. **33**, D256–D261.
21. Lefranc, M.-P. (1997) Unique database numbering system for immunogenetic analysis. *Immunol. Today* **18,** 509.
22. Lefranc, M.-P. (1999) The IMGT unique numbering for immunoglobulins, T cell receptors and Ig-like domains. *The Immunologist* **7,** 132–136.
23. Kabat, E.A., Wu, T.T., Perry, H.M., Gottesman, K.S., and Foeller, C. (1991) Sequences of proteins of immunological interest. National Institute of Health Publications, Washington D.C., USA, Publication no. 91-3242.
24. Satow, Y., Cohen, G.H., Padlan, E.A., and Davies, D.R. (1986) Phosphocholine binding immunoglobulin Fab McPC603. *J. Mol. Biol.* **190**, 593–604.
25. Chothia, C. and Lesk, A.M. (1987) Canonical structures for the hypervariable regions of immunoglobulins. *J. Mol. Biol.* **196,** 901–917.

26. Duprat, E., Kaas, Q., Garelle, V., Lefranc, G., and Lefranc, M.-P. (2004) IMGT standardization for alleles and mutations of the V-LIKE-DOMAINs and C-LIKE-DOMAINs of the immunoglobulin superfamily. *Recent Research Developments in Human Genetics* (Pandalai, S.G., ed.) Research Signpost, Trivandrum, Kerala, India, **2**, 111–136.
27. Bertrand, G., Duprat, E., Lefranc, M.-P., Marti, J., and Coste, J. (2004) Characterization of human FCGR3B*02 (HNA-1b, NA2) cDNAs and IMGT standardized description of FCGR3B alleles. *Tissue Antigens* **64**, 119–131.
28. Frigoul, A., and Lefranc M.-P. (2005) MICA: standardized IMGT allele nomenclature, polymorphisms and diseases. *Recent Research Developments in Human Genetics* (Pandalai, S.G., ed.) Research Signpost, Trivandrum, Kerala, India, **3**, 95–145.
29. Baum, T.P., Pasqual, N., Thuderoz, F., Hierle, V., Chaume, D., Lefranc, M.-P., Jouvin-Marche, E., Marche, P.N., and Demongeot, J. (2004) IMGT/GeneInfo: enhancing V(D)J recombination database accessibility. *Nucleic Acids Res.* **32**, D51–D54.
30. Giudicelli, V., Duroux, P., Ginestoux, C., Folch, G., Jabado-Michaloud, J., Chaume, D., and Lefranc, M.-P. (2006) IMGT/LIGM-DB, the IMGT® comprehensive database of immunoglobulin and T cell receptor nucleotide sequences. *Nucleic Acids Res.* **34**, D781–D784.
31. Folch G., Bertrand J., Lemaitre M., and Lefranc M.-P. (2004) IMGT/PRIMER-DB. In: *Database Listing* (Galperin, M.Y., ed.), The Molecular Biology Database Collection: 2004 update. *Nucleic Acids Res*. **32**, D3–D22.
32. Robinson, J., Waller, M.J., Parham, P., de Groot, N., Bontrop, R., Kennedy, L. J., Stoehr, P., and Marsh, S.G. (2003) IMGT/HLA and IMGT/MHC sequence databases for the study of the major histocompatibility complex. *Nucleic Acids Res*. **31**, 311–314.
33. Giudicelli, V., Chaume, D., and Lefranc, M.-P. (2004) IMGT/V-QUEST, an integrated software program for immunoglobulin and T cell receptor V-J and V-D-J rearrangement analysis. *Nucleic Acids Res.* **32**, W435–W440.
34. Yousfi Monod, M., Giudicelli, V., Chaume, D., and Lefranc, M.-P. (2004) IMGT/JunctionAnalysis: the first tool for the analysis of the immunoglobulin and T cell receptor complex V-J and V-D-J JUNCTIONs. *Bioinformatics* **20**, 379–385.
35. Elemento, O., and Lefranc, M.-P. (2003) IMGT/PhyloGene: an on-line tool for comparative analysis of immunoglobulin and T cell receptor genes. *Dev. Comp. Immunol.* **27**, 763–779.
36. Kaas, Q., Ruiz, M., and Lefranc, M.-P. (2004) IMGT/3Dstructure-DB and IMGT/StructuralQuery, a database and a tool for immunoglobulin, T cell receptor and MHC structural data. *Nucleic Acids Res.* **32**, D208–D210.
37. Pommié, C., Sabatier, S., Lefranc, G., and Lefranc, M.-P. (2004) IMGT standardized criteria for statistical analysis of immunoglobulin V-REGION amino acid properties. *J. Mol. Recognit.* **17**, 17–32.

38. Lefranc, M.-P. (2006) Web sites of interest to immunologists. *In*: *Current Protocols in Immunology* (Coligan, J.E., Bierer, B.E., Margulies, D.E., Shevach, E.M., and Strober, W., eds.), John Wiley and Sons, Hoboken N.J. pp. A.1J.1–A.1J.74.
39. Giudicelli V., Chaume D., Jabado-Michaloud J., and Lefranc M.-P. (2005) Immunogenetics sequence annotation: the strategy of IMGT based on IMGT-ONTOLOGY. *Stud. Health Technol. Inform.* **116**, 3–8.
40. Lefranc, M.-P. (2004) IMGT, The International ImMunoGeneTics Information System®, http://imgt.cines.fr. In: *Antibody engineering: Methods and Protocols* (Lo, B.K.C., ed.), Humana, Totowa, N.J., *Methods Mol. Biol.* **248**, 27–49.
41. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000) The Protein Data Bank. *Nucleic Acids Res*. **28**, 235–242.
42. Duprat, E., Lefranc, M.-P., and Gascuel, O. (2006) A simple method to predict protein binding from aligned sequences – application to MHC superfamily and beta2-microglobulin. *Bioinformatics*, **22**, 453–459.
43. Kaas, Q., and Lefranc, M.-P. (2005) T cell receptor/peptide/MHC molecular characterization and standardized pMHC contact sites in IMGT/3Dstructure-DB. Epub *In Silico Biol.*, (5 0046 refers to Epub) 0046 20 Oct 2005 In Silico Biol. **5**, 505–528.