IMGT®, the international ImMunoGeneTics information system®, the reference in immunogenetics and immunoinformatics

Patrice Duroux, Véronique Giudicelli, Quentin Kaas, Joumana Jabado-Michaloud, Géraldine Folch, Chantal Ginestoux, Xavier Brochet, Jérôme Lane, Laetitia Régnier, Yan Wu, Vijay Phani Garapati, Fattena Bellahcene, Emmanuel-Jean Servier, François Ehrenmann, Gérard Lefranc and Marie-Paule Lefranc*

IMGT®, the international ImMunoGeneTics information system®
Université Montpellier II, Laboratoire d'ImmunoGénétique Moléculaire LIGM
UPR CNRS 1142, Institut de Génétique Humaine IGH,
141 rue de la Cardonille
34396 Montpellier Cedex 5, France
Phone: +33 4 99 61 99 65, Fax: +33 4 99 61 99 01

* corresponding author
E-mail: Marie-Paule.Lefranc@igh.cnrs.fr
Institut Universitaire de France

ABSTRACT

IMGT®, the international ImMunoGeneTics information system® (http://imgt.cines.fr), created in 1989, by the Laboratoire d'ImmunoGénétique Moléculaire LIGM (Université Montpellier II and CNRS) at Montpellier, France, is a high quality integrated knowledge resource specialized in the immunoglobulins (IG), T cell receptors (TR), major histocompatibility complex (MHC) of human and other vertebrates, and related proteins of the immune systems (RPI) that belong to the immunoglobulin superfamily (IgSF) and to the MHC superfamily (MhcSF). IMGT® includes several sequence databases (IMGT/LIGM-DB, IMGT/PRIMER-DB, IMGT/MHC-DB), one genome database (IMGT/GENE-DB) and one 3D structure database (IMGT/3Dstructure-DB), Web resources comprising more than 10,000 HTML pages, and interactive tools. The accuracy and consistency of IMGT® data and the coherence between the different IMGT® components are based on IMGT-ONTOLOGY, the first ontology for immunogenetics and immunoinformatics. IMGT-ONTOLOGY manages the immunogenetics knowledge through diverse facets relying on seven axioms, "IDENTIFICATION", "DESCRIPTION", "CLASSIFICATION", "NUMEROTATION", "LOCALIZATION", "ORIENTATION" and "OBTENTION". These axioms constitute the Formal IMGT-ONTOLOGY, also designated as IMGT-Kaleidoscope. IMGT® tools are particularly useful for the analysis of the IG and TR repertoires in normal physiological and pathological situations. IMGT® is used in medical research (autoimmune diseases, infectious diseases, AIDS, leukemias, lymphomas, myelomas), veterinary research, biotechnology related to antibody engineering (phage displays, combinatorial libraries, chimeric, humanized and human antibodies), diagnostics (clonalities, detection and follow up of residual diseases) and therapeutical approaches (graft, immunotherapy, vaccinology). IMGT® is freely available at http://imgt.cines.fr.

Key words: IMGT, ontology, immunoinformatics, immunogenetics, immune system

INTRODUCTION

Immunogenetics, the science that studies the genetics of the immune responses, has shown a considerable expansion in biomedical fields since the last decades. It has highlighted the complex mechanisms by which B cells and T cells are at the origin of the extreme diversity of antigen receptors that comprise the immunoglobulins (IG) or antibodies and the T cell receptors (TR) (10^{12} different immunoglobulins and 10^{12} different T cell receptors per individual, in humans) [1,2]. These mechanisms include in particular DNA rearrangements and, for the IG, somatic hypermutations. In addition, there is a considerable polymorphism of the major histocompatibility complex (MHC) (human leucocyte antigens HLA, in humans). These particularities of the adaptative immune system specific of the vertebrates, and the mechanisms of the innate immune responses found in any species, allow the immune system to be an excellent model for system biology. The huge amount of immunological experimental data continues to grow exponentially and necessitates to be managed and analysed computationally. This is the goal of immunoinformatics, a new emerging science that implements the bioinformatics methodologies to answer these needs. At the same time, standardized representation of genomic, genetic, proteomic and three-dimensional (3D) structures data is required to organize immunogenetics knowledge towards system biology and for the modelling and a better understanding of the immune system.

IMGT®, the international ImMunoGeneTics information system® (http://imgt.cines.fr) [3], is the international reference in immunogenetics and immunoinformatics. Created in 1989 at the Laboratoire d'ImmunoGénétique Moléculaire (LIGM) by Marie-Paule Lefranc (Université Montpellier II and CNRS) in Montpellier, France, IMGT® provides a high quality integrated knowledge resource, specialized in the IG, TR, MHC of human and other vertebrates, and related proteins of the immune system (RPI), which belong to the immunoglobulin superfamily (IgSF) and to the MHC superfamily (MhcSF) of any species [4-7]. The IMGT® information system consists of databases (three of sequences, one of genes and one of 3D structures) and interactive tools for sequence, genome and 3D structure analysis, which interact together according to genomic, genetic and structural approaches [7] (Fig. 1). Moreover, IMGT® provides Web resources comprising more than 10,000 HTML pages of synthesis (IMGT Repertoire), knowledge (IMGT Scientific chart, IMGT Education, IMGT Index) and external links (IMGT Bloc-notes and IMGT other accesses) [3].

The accuracy and the consistency of the IMGT® data, as well as the coherence between the different IMGT® components (databases, tools and Web resources), are based on IMGT-ONTOLOGY, the first ontology for immunogenetics and immunoinformatics [8]. IMGT-ONTOLOGY provides a semantic specification of the terms to be used in immunogenetics and immunoinformatics and manages the related knowledge [9,10], thus allowing the standardization for immunogenetics data from genome, proteome, genetics and 3D structures [5,7,11,12]. IMGT-ONTOLOGY results from a deep expertise in the domain and an extensive effort of conceptualization. The first standardization step was the identification of the IG and TR nucleotide sequences and the second step their description which led to the creation of IMGT/LIGM-DB [13], the first on-line IMGT® database. The resulting controlled vocabulary comprises a thesaurus of keywords for the sequence identification and a set of labels for the description of the constitutive motifs. The third standardization step was the classification of the IG and TR genes which gave rise to the IMGT gene nomenclature for IG and TR of human and other vertebrates [1,2,14], approved by the Human Genome Organisation (HUGO) Nomenclature Committe HGNC in 1999 [15] and currently used in the generalist genome

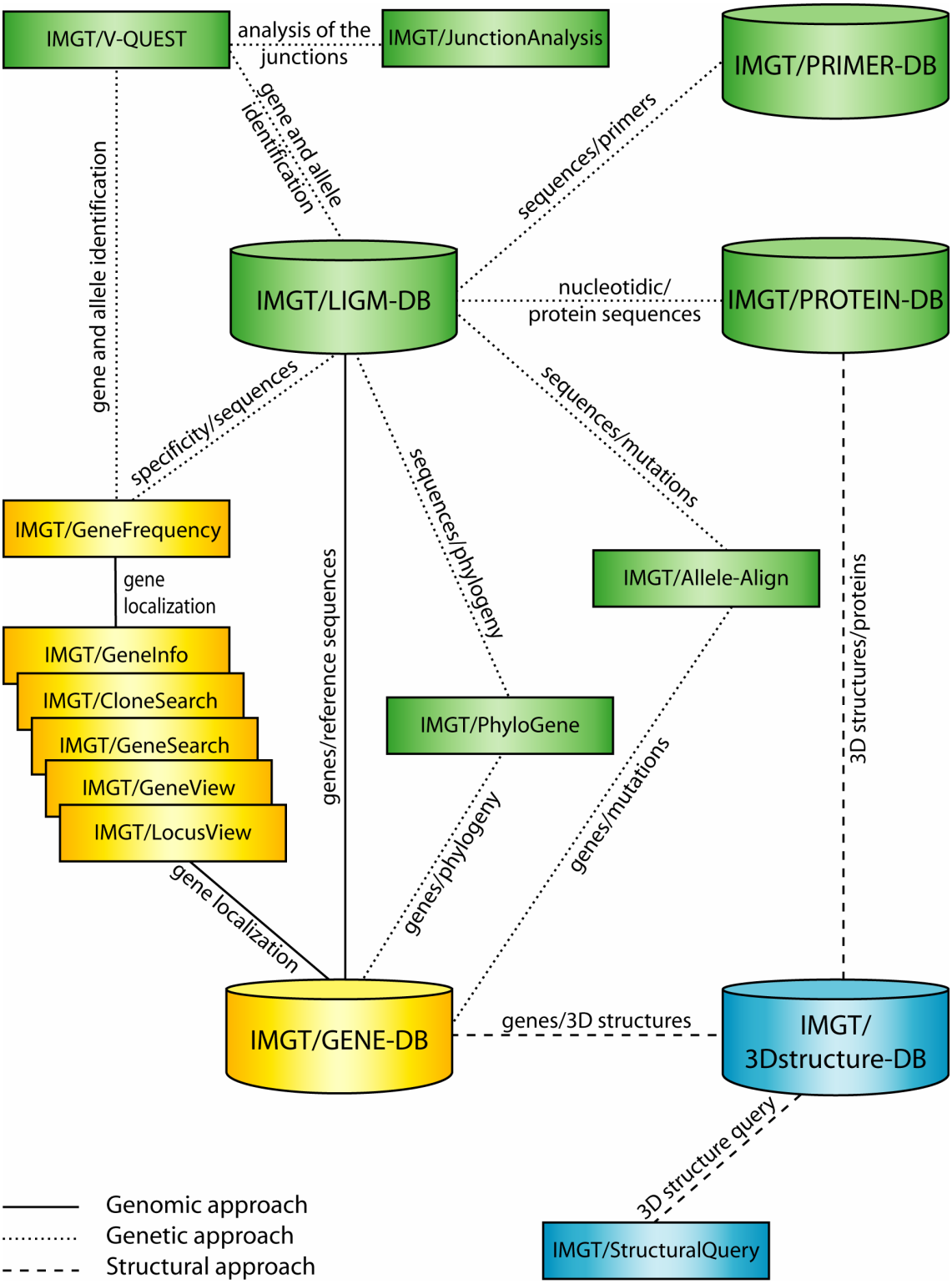

Fig. 1. IMGT®, the international ImMunoGeneTics information system® (http://imgt.cines.fr). The IMGT® databases are shown as cylinders and tools as rectangles (in yellow, green or blue for the genomic, genetic or structural approach, respectively). The IMGT Repertoire and other Web resources are not shown.

databases. The fourth standardization step has been the setting up of the principles for the unique numbering of antigen receptor sequences and structures [16-20].

IMGT-ONTOLOGY [8] manages the immunogenetics knowledge through diverse facets that rely on the axioms of the Formal IMGT-ONTOLOGY or IMGT-Kaleidoscope. These axioms, "IDENTIFICATION", "DESCRIPTION", "CLASSIFICATION", "NUMEROTATION", "LOCALIZATION", "ORIENTATION" and "OBTENTION", postulate that objects, processes and relations have to be identified, described, classified, numerotated, localized, orientated and that the way there are obtained, determined. The axioms and the concepts of identification, description, classification, numerotation, localization, orientation and obtention generated from these axioms have been essential for the conceptualization of the molecular immunogenetics knowledge. IMGT-ONTOLOGY concepts have been formalized, for the biologists and IMGT users, in the IMGT Scientific chart [5,7] and, for programming purpose, in IMGT-ML [21] which is a XML (Extensible Markup Language) Schema.

The IMGT® information system consists of databases, tools and Web resources summarized in Figure 1. Databases include sequence databases (IMGT/LIGM-DB, IMGT/MHC-DB, IMGT/PROTEIN-DB, IMGT/PRIMER-DB), one genome database (IMGT/GENE-DB), and one 3D structure database (IMGT/3Dstructure-DB). Interactive tools are provided for sequence analysis (IMGT/V-QUEST, IMGT/JunctionAnalysis, IMGT/Allele-Align, IMGT/PhyloGene), genome analysis (IMGT/LocusView, IMGT/GeneView, IMGT/GeneSearch, IMGT/CloneSearch, IMGT/GeneInfo, IMGT/GeneFrequency), and 3D structure analysis (IMGT/DomainDisplay, IMGT/DomainGapAlign, IMGT/Collier-de-Perles, IMGT/DomainSuperimpose, IMGT/StructuralQuery). Web resources comprise more than 10,000 HTML pages of synthesis [IMGT Repertoire (for IG and TR, MHC, RPI)], of knowledge [IMGT Scientific chart, IMGT Education (Aide-mémoire, Tutorials, Questions and answers, IMGT Lexique), The IMGT Medical page, The IMGT Veterinary page, The IMGT Biotechnology page, IMGT Index], and external links [IMGT Bloc-notes, The IMGT Immunoinformatics page, and Other accesses (SRS, BLAST, etc.)].

In this chapter, we will briefly present the synthesis of an IG or antibody in humans, as an example of molecular knowledge which necessitates to be managed for genomic, genetic and structural data. We will then describe the IMGT® components (databases, tools and IMGT Repertoire Web resources) that have been developed according to these three main biological approaches - genomic, genetic and structural - to answer these needs. The IMGT genomic approach is gene-centered and mainly orientated towards the study of the genes within their loci and on the chromosomes. The IMGT genetic approach refers to the study of the genes in relation with their sequence polymorphisms and mutations, their expression, their specificity and their evolution. 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 relation with the protein functions, polymorphisms and evolution. IMGT-Choreography, based on the Web service architecture paradigm, will enable significant biological and clinical requests involving every part of the IMGT® information system.

AN EXAMPLE OF KNOWLEDGE MANAGED BY IMGT®

An example of knowledge at the molecular level managed by IMGT® is the immunoglobulin synthesis [1]. An IG or antibody is composed of two identical heavy chains associated with two identical light chains, kappa or lambda. In humans, heavy chain genes (locus IGH), light chain kappa genes (locus IGK) and light chain lambda genes (locus IGL) are located on the chromosomes 14 (14q32.3), 2 (2p11.2) and 22 (22q11.2), respectively. The synthesis of an

immunoglobulin requires rearrangements of the IGH, IGK and IGL genes during the differentiation of the B lymphocytes.

In the human genome (genomic DNA or gDNA), four types of genes code the IG (and TR): variable (V), diversity (D), joining (J) and constant (C). The configuration of the V-gene, D-gene and J-gene is identified as "germline" (Fig. 2), the configuration of the C-gene is "undefined". During the differentiation of the B lymphocytes in the bone marrow, the genomic DNA is rearranged and leads to the junction of a V-gene, a D-gene and a J-gene to form a V-D-J-gene in the IGH locus, and to the junction of a V-gene and a J-gene to form a V-J-gene in the IGK or IGL loci. The configuration of these genes is identified as "rearranged". After transcription and maturation of the pre-messenger by splicing, the messenger RNA (mRNA) L-V-D-J-C-sequence and L-V-J-C-sequence (L for leader) are obtained and then translated into the heavy chain (IG-Heavy_chain) and the light chain (IG-Light_chain) of an IG (or antibody) (Fig.2).

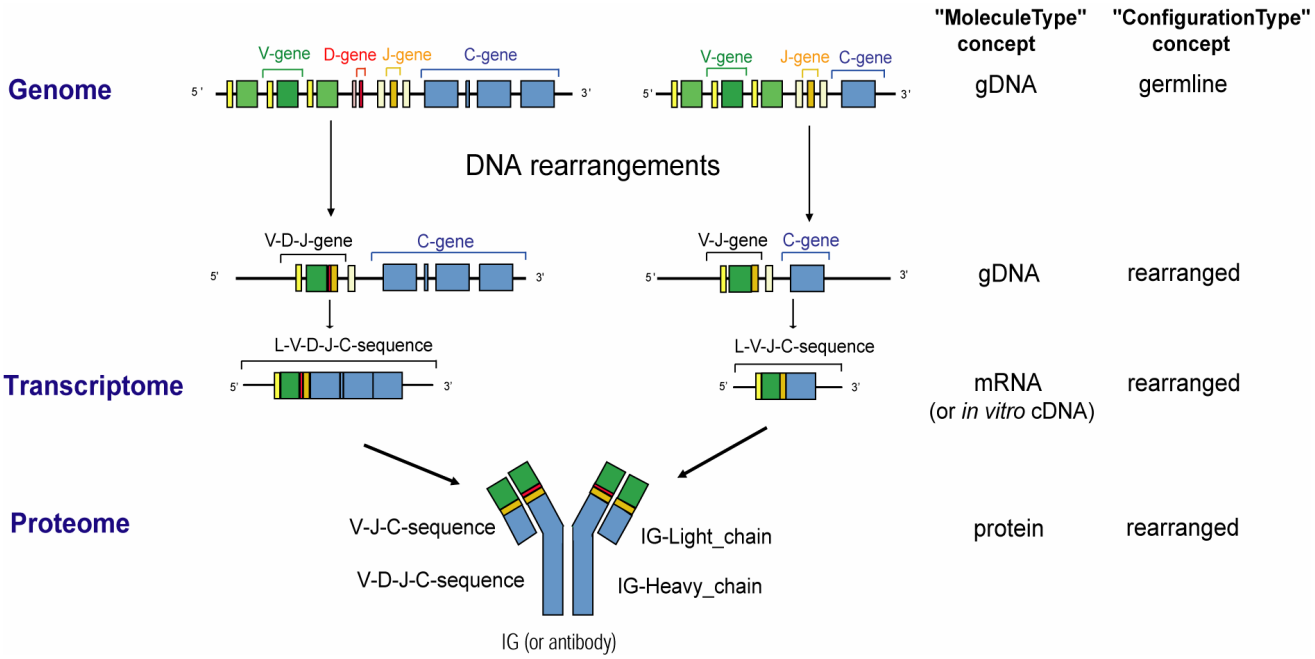

Fig. 2. An example of knowledge at the molecular level: the synthesis of an IG or antibody in humans. A human being may potentially synthesize 10^{12} different antibodies [1].

The variable domains VH and VL are coded by the V-D-J-REGION and the V-J-REGION (Fig. 3). Each domain includes three framework regions (FR) (in pale blue in Fig.3) and three hypervariable loops or complementarity determining regions (CDR). The CDR, and more particularly the CDR3 that result from the junction of the V-D-J genes (VH) and V-J genes (VL), are involved in the antigen recognition. The VH and VL amino acids in contact with the antigen constitute the paratope. The part of the antigen recognized by the antibody is the epitope. The number of potential V-D-J and V-J rearrangements depends on the number of functional V, D and J genes in the genome. Additional mechanisms (N diversity at the V-D-J and V-J junctions and somatic hypermutations) allow to reach 10^{12} different antibodies per individual [1] (IMGT®, http://imgt.cines.fr).

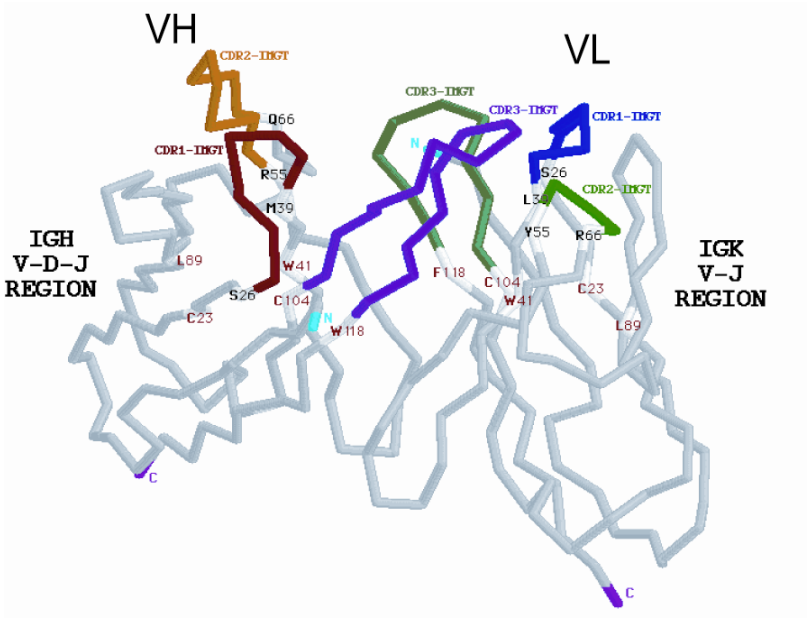

Fig. 3. The variable domains VH and VL of the heavy and light chains of an IG or antibody.

IMGT GENOMICS COMPONENTS

IMGT genome database

IMGT/GENE-DB is the comprehensive IMGT genome database, created by LIGM, Montpellier, France, on the Web since January 2003 [14]. All the human and mouse IG and TR genes are available in IMGT/GENE-DB. The human IMGT gene names [1,2] were approved by the Human Genome Organisation (HUGO) Nomenclature Committee (HGNC) in 1999 [15], and entered in IMGT/GENE-DB, Genome DataBase GDB (Canada), LocusLink at NCBI (USA), and GeneCards. Reciprocal links exist between IMGT/GENE-DB and the generalist nomenclature (HGNC Genew) and genome databases (GDB, LocusLink and Entrez Gene at NCBI, and GeneCards). The mouse IG and TR gene names with IMGT reference sequences were provided by IMGT to HGNC and to the Mouse Genome Database MGD in July 2002. Queries in IMGT/GENE-DB can be performed according to IG and TR gene classification criteria and 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, mapped sequence. IMGT/GENE-DB interacts dynamically with IMGT/LIGM-DB to download and display gene-related sequence data. In June 2007, IMGT/GENE-DB contained 1,512 genes and 2,462 alleles (673 IG and TR genes and 1,216 alleles from *Homo sapiens*, and 839 IG and TR genes and 1,246 alleles from *Mus musculus, Mus cookii, Mus pahari, Mus spretus, Mus saxicola, Mus minutoides*).

IMGT genome analysis tools and Web resources

Tools. The IMGT genome analysis tools comprise IMGT/LocusView, IMGT/GeneView, IMGT/GeneSearch, IMGT/CloneSearch, IMGT/GeneInfo and IMGT/GeneFrequency. IMGT/LocusView and IMGT/GeneView manage the locus organization and the gene location and provide the display of physical maps for the human IG, TR and MHC loci and for the mouse TRA/TRD locus. IMGT/GeneSearch and IMGT/CloneSearch allow retrieval of

information concerning genes and clones analysed in IMGT/LocusView. IMGT/GeneInfo, the first tool developed by external collaborators and integrated to IMGT®, provides and displays information on the potential TR rearrangements in human and mouse [22,23]. IMGT/GeneFrequency represents, as vertical bars on locus representations, the number of cDNA sequences found in IMGT/LIGM-DB. The queries can be made for human or mouse genes, per locus and per specificity.

Genome Web resources. The genomic Web resources are compiled in the IMGT Repertoire "Locus and genes" section which includes Chromosomal localizations, Locus representations, Locus description, Gene exon/intron organization, Gene exon/intron splicing sites, Gene tables, Potential germline repertoires, the complete lists of human and mouse IG and TR genes, and the correspondences between nomenclatures [1,2]. The IMGT Repertoire "Probes and RFLP" section provides data on gene insertion/deletion.

IMGT GENETICS COMPONENTS

IMGT sequence databases

IMGT/LIGM-DB. IMGT/LIGM-DB 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 [13]. In June 2007, IMGT/LIGM-DB contained more than 106,500 sequences of 150 vertebrate species. The unique source of data for IMGT/LIGM-DB is EMBL [24] which shares data with the other two generalist databases GenBank and DNA DataBank of Japan (DDBJ). Based on expert analysis, specific detailed annotations are added to IMGT® flat files. The Web interface allows searches according to immunogenetic specific criteria and is easy to use without any knowledge in a computing language. Selection is displayed at the top of the resulting sequences pages, so the users can check their own queries. Users have the possibility to modify their request or consult the results with a choice of nine possibilities. IMGT/LIGM-DB gene and allele name assignment and sequence annotations are performed according to the IMGT Scientific chart rules. These annotations allow retrieval of data from IMGT/LIGM-DB for queries in other IMGT® databases or tools. As an example, the IMGT/LIGM-DB accession numbers of the cDNA expressed sequences for each human and mouse IG and TR gene are available, with direct links to IMGT/LIGM-DB, in the IMGT/GENE-DB entries. IMGT/LIGM-DB data are also distributed by anonymous FTP servers at CINES (ftp://ftp.cines.fr/IMGT/) and EBI (ftp://ftp.ebi.ac.uk/pub/databases/imgt/) and from many Sequence Retrieval System (SRS) sites. IMGT/LIGM-DB can be searched by BLAST or FASTA on different servers (EBI, IGH, Institut Pasteur, etc.).

IMGT/Automat for IMGT/LIGM-DB annotations. IMGT/Automat is an integrated internal IMGT Java tool which automatically performs the annotation of rearranged cDNA sequences that represent the half of the IMGT/LIGM-DB content. The annotation procedure includes the identification of the sequences, the classification of the IG and TR genes and alleles, and the description of all IG and TR specific and constitutive motifs within the nucleotide sequences. Accuracy and reliability of the annotation are mainly estimated by the program itself with the evaluation of the alignment scores, the deduced sequence functionality, and the coherence of the characterized and delimited IG and TR motifs. So far 7,418 human and mouse IG and TR cDNA sequences have been automatically annotated by the IMGT/Automat tool, with annotations being as reliable and accurate as those provided by a human annotator.

Other IMGT sequences databases. IMGT/PRIMER-DB [25] is the IMGT® oligonucleotide primer database for IG and TR, created by LIGM, Montpellier in collaboration with EUROGENTEC S.A., Belgium, on the Web since February 2002. In June 2007, IMGT/PRIMER-DB contained 1,864 entries. IMGT/PRIMER-DB provides standardized information on oligonucleotides (or Primers) and combinations of primers (Sets, Couples) for IG and TR. These primers are useful for combinatorial library constructions, scFv, phage display or microarray technologies. The IMGT Primer cards are linked to the IMGT/LIGM-DB flat files, and to the IMGT Repertoire (IMGT Colliers de Perles and Alignments of alleles of the IMGT/LIGM-DB reference sequence used for the primer description). IMGT/MHC-DB comprises databases hosted at EBI and includes a database of human MHC allele sequences or IMGT/MHC-HLA, developed by Cancer Research UK and maintained by ANRI, London, UK, on the Web since December 1998, and a database of MHC sequences from non human primates IMGT/MHC-NHP, curated by BPRC, The Netherlands, on the Web since April 2002 [26].

IMGT sequence analysis tools and genetics Web resources

IMGT/V-QUEST and IMGT/JunctionAnalysis. The first tool, IMGT/V-QUEST (V-QUEry and STandardization), is an integrated software for IG and TR [27], used for the identification of the V, D and J genes and of their mutations. 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 alignments by comparison with sequences from the IMGT reference directory, the FR-IMGT and CDR-IMGT delimitations based on the IMGT unique numbering, the translation of the input sequence, the display of nucleotide and amino acid mutations compared to the closest IMGT reference sequence, the identification of the JUNCTION and results from IMGT/JunctionAnalysis (default option), and the V-REGION IMGT Collier de Perles [27-29]. IMGT/JunctionAnalysis [30] is a tool, complementary to IMGT/V-QUEST, which provides a thorough analysis of the V-J and V-D-J junctions which confer the antigen receptor specificity to IG and TR rearranged genes. IMGT/JunctionAnalysis identifies the D-GENEs and alleles 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 [30,31]. Several hundreds of junction sequences can be analysed simultaneously.

Other sequence analysis tools. IMGT/Allele-Align is used for the detection of polymorphisms. It allows the comparison of two alleles highlighting the nucleotide and amino acid differences. IMGT/Phylogene [32] is an easy to use tool for phylogenetic analysis of IG and TR variable region (V-REGION) and constant domain (C-DOMAIN) sequences. This tool is particularly useful in developmental and comparative immunology. The users can analyse their own sequences by comparison with the IMGT standardized reference sequences for human and mouse IG and TR.

Genetics Web resources. The genetics Web resources are compiled in the IMGT Repertoire "Proteins and alleles" section which includes Protein displays, Alignments of alleles, Tables of alleles, Allotypes, Isotypes, etc.

IMGT STRUCTURAL COMPONENTS

IMGT structural database

IMGT/3Dstructure-DB. IMGT/3Dstructure-DB is the IMGT® database for 3D structures, created by LIGM, on the Web since November 2001 [33]. In June 2007, IMGT/3Dstructure-DB contained 1265 atomic coordinate files. IMGT/3Dstructure-DB comprises IG, TR, MHC and RPI with known 3D structures. Coordinate files extracted from the Protein Data Bank (PDB) (http://www.rcsb.org/pdb/) [34] are renumbered according to the standardized IMGT unique numbering [16-18]. The IMGT/3Dstructure-DB cards provide IMGT annotations (assignment of IMGT genes and alleles, IMGT chain and domain labels, IMGT Colliers de Perles [1,2,35,36] on one layer and two layers), downloadable renumbered IMGT/3Dstructure-DB flat files, vizualisation tools, and external links. The IMGT/3Dstructure-DB residue cards provide detailed information on the inter- and intra-domain contacts of each residue position. An IMGT/3Dstructure-DB card provides receptor and chain description, IMGT gene and allele names, domain delimitations, amino acid positions according to the IMGT unique numbering. Structural and functional domains of the IG and TR chains comprise the variable domain or V-DOMAIN (9-strand beta-sandwich) which corresponds to the V-J-REGION or V-D-J-REGION and is encoded by two or three genes, the constant domain or C-DOMAIN (7-strand beta-sandwich), and, for the MHC chains, the groove domain or G-DOMAIN (4 beta-strand and one alpha-helix) [18-20]. The IMGT unique numbering has been extended to the V-LIKE-DOMAINs and C-LIKE-DOMAINs of IgSF proteins other than IG and TR [18,19,37-41], and to the G-LIKE-DOMAINs of MhcSF proteins other than MHC [20,42,43].

IMGT structural analysis tool and Web resources

IMGT structural analysis tools. Five IMGT® structural analysis tools are available on-line at http://imgt.cines.fr: IMGT/DomainDisplay, IMGT/DomainGapAlign, IMGT/Collier-de-Perles, IMGT/DomainSuperimpose and IMGT/StructuralQuery. IMGT/DomainDisplay allows to query the IMGT Domain directory that contains domains of the IgSF and MhcSF superfamilies and to visualize the resulting amino acid sequences. Queries can be made by domain type (V, C or G). IMGT/DomainGapAlign aligns the user domain amino acid sequence with the closest sequence from the IMGT Domain directory and inserts gaps in the user sequence according to the IMGT unique numbering. IMGT/Collier-de-Perles allows to obtain the IMGT Collier de Perles for a domain amino acid sequence in which the gaps according to the IMGT unique numbering have been inserted (for instance using IMGT/DomainGapAlign). The IMGT Colliers de Perles can be obtained on one layer or two layers and with amino acid properties (hydropathy, volume, physico-chemical characteristics [44]). IMGT/DomainSuperimpose allows to superimpose 3D structures of two domains from IMGT/3Dstructure-DB. The IMGT/StructuralQuery tool [33] analyses the interactions of the residues of the antigen receptors IG and TR), MHC, RPI, antigens and ligands. The contacts are described per domain (intra- and inter-domain contacts) and annotated in term of IMGT labels (chains, domains), 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, CDR-IMGT lengths. IMGT/StructuralQuery is currently available for the V-DOMAINs.

Structural Web resources. The structural Web resources are compiled in the IMGT Repertoire "2D and 3D structures" section which includes 2D representations or IMGT Colliers de Perles [1,2,33,35-43], 3D representations, FR-IMGT and CDR-IMGT lengths, amino acid chemical characteristics profiles [44], etc.

IMGT-Choreography

IMGT-Choreography [7,45] is based on the Web service architecture paradigm (see W3C http://www.w3.org/). Its goal is to orchestrate dynamic procedure calls between IMGT databases querying and analysis tools. Conversations between Web services are expressed using the sole IMGT-ML language both for queries and result fetches. This ensures semantic consistency between exchanged messages as IMGT-ML (available at IMGT Index>IMGT-ML) is an XML Schema formalization of the IMGT-ONTOLOGY concepts. IMGT Web services are developed using the JAVA programming language and deployed using the Apache Axis (http://ws.apache.org/axis/) Web services development framework. Composition and chaining of IMGT Web services through IMGT-Choreography will enable processing of complex significant biological and clinical requests involving every part of the IMGT® information system.

CONCLUSION

Since July 1995, IMGT® has been available on the Web at the IMGT® Home page http://imgt.cines.fr (Montpellier, France). IMGT has an exceptional response with more than 150,000 requests a month. IMGT is the international reference in immunogenetics and immunoinformatics and provides a common access to all standardized data which include nucleotide and protein sequences, oligonucleotide primers, gene maps, genetic polymorphisms, specificities, 2D and 3D structures, based on IMGT-ONTOLOGY. The information is of much value to clinicians and biological scientists in general [46-49]. IMGT databases and tools are extensively queried and used by scientists, from both academic and industrial laboratories, who are equally distributed between the United States (1/3), Europe (1/3) and the remaining world (1/3). IMGT is used in very diverse domains: (i) fundamental research and medical research (repertoire analysis of the IG antibody sites and of the TR recognition sites in normal and pathological situations such as autoimmune diseases, infectious diseases, AIDS, leukemias, lymphomas, myelomas), (ii) veterinary research (IG and TR repertoires in farm and wild life 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, vaccinology). By its high quality and its data distribution based on IMGT-ONTOLOGY, IMGT® has an important role to play in the development of immunogenetics Web services. The design of IMGT-Choreography and the creation of dynamic interactions between the IMGT® databases and tools, using Web services and IMGT-ML, represent novel and major developments of IMGT®, the international reference in immunogenetics and immunoinformatics.

ACKNOWLEDGEMENTS

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 LivingResources (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 registered CNRS mark. IMGT® is a National Bioinformatics RIO 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. and Lefranc, G. (2001). The Immunoglobulin FactsBook. Academic Press, London, UK, 458 pages.
2. Lefranc, M.-P. and Lefranc, G. (2001). The T cell receptor FactsBook. Academic Press, London, UK, 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®. *Nucl. Acids Res*. **33**, D593-D597.
4. Lefranc, M.-P. (2003). IMGT® databases, web resources and tools for immunoglobulin and T cell receptor sequence analysis, http://imgt.cines.fr. *Leukemia*, **17**, 260-266.
5. 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. and Lefranc, G. (2004). IMGT-ONTOLOGY for Immunogenetics and Immunoinformatics. *In Silico Biology*, **4**, 17-29.
6. Lefranc, M.-P. (2004). IMGT-ONTOLOGY and IMGT databases, tools and web resources for immunogenetics and immunoinformatics. *Mol. Immunol.,* **40**, 647-659.
7. 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. *In Silico Biology*, **5**, 45-60.
8. Giudicelli, V. and Lefranc, M.-P. (1999). Ontology for Immunogenetics: the IMGT-ONTOLOGY. *Bioinformatics*, **12**, 1047-1054.
9. 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.
10. Lefranc, M.-P. (2005). IMGT, the international ImMunoGeneTics information system®: a standardized approach for immunogenetics and immunoinformatics. *Immunome Research*, **1**, 3.
11. Giudicelli, V. and Lefranc, M.-P. (2005). Interactive IMGT on-line tools for the analysis of immunoglobulin and T cell receptor repertoires. In: *New Research on Immunology* (B.A. Veskler, ed.), Nova Science, pp. 77-105.

12. Kaas, Q. and Lefranc, M.-P. (2005). T cell receptor/peptide/MHC molecular characterization and standardized pMHC contact sites in IMGT/3Dstructure-DB. *In Silico Biology* **5,** 505-528.
13. Giudicelli, V., 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. *Nucl. Acids Res.,* **34**, D781-D784.
14. 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. *Nucl. Acids Res*. **33**, D256-D261.
15. 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.
16. Lefranc, M.-P. (1997). Unique database numbering system for immunogenetic analysis. *Immunology Today* **18**, 509.
17. Lefranc, M.-P. (1999). The IMGT unique numbering for Immunoglobulins, T cell receptors and Ig-like domains. *The Immunologist*, **7**, 132-136 .
18. 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.
19. 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.
20. 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.
21. Chaume, D., Giudicelli, V. and Lefranc, M.-P. (2001). IMGT-ML, a XML language for IMGT-ONTOLOGY and IMGT/LIGM-DB data. In: *Proceedings of NETTAB 2001*, *Networks tools and applications in Biology*, Genoa, Italy, May 17-18, 2001, pp. 71-75.
22. 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. *Nucl. Acids Res*., **32**, D51-D54.
23. Baum, T.P., Hierle, V., Pascal, N., Bellahcene, F., Chaume, D., Lefranc, M.-P., Jouvin-Marche, E., Marche, P.N. and Demongeot, J. (2006). IMGT/GeneInfo: T cell receptor gamma TRG and delta TRD genes in database give access to all TR potential V(D)J recombinations. *BMC Bioinformatics,* **7**: 224.
24. Kulikova, T., Aldebert, P., Althorpe, N., Baker, W., Bates, K., Browne, P., van den Broek, A., Cochrane, G., Duggan, K., Eberhardt, R., Faruque, N., Garcia-Pastor, M., Harte, N., Kanz, C., Leinonen, R., Lin, Q., Lombard, V., Lopez, R., Mancuso, R., McHale, M., Nardone, F., Silventoinen, V., Stoehr, P., Stoesser, G., Tuli, M.A., Tzouvara, K., Vaughan, R., Wu, D., Zhu, W. and Apweiler, R. (2004). The EMBL Nucleotide Sequence Database. *Nucl. Acids Res.*, **32**, D27-D30.
25. Folch, G., Bertrand, J., Lemaitre, M. and Lefranc, M.-P.(2004). IMGT/PRIMER-DB. *In*: The Molecular Biology Database Collection: 2004 update, Galperin M.Y. (ed.). Database listing. *Nucl. Acids Res*., **32**, 3-22.
26. 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. *Nucl. Acids Res.*, **31**, 311-314.

27. 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. *Nucl. Acids Res*., **32**, W435-W440.
28. Lefranc, M.-P. IMGT, the international ImMunoGeneTics information system®, http://imgt.cines.fr (2004). In : Antibody engineering methods and protocols, 2nd edition (Lo B.K.C., ed). Humana Press, Totowa, NJ, USA. *Methods Mol. Biol.* **248,** 27-49.
29. Lefranc, M.-P. (2006). Using bioinformatics tools for the sequence analysis of the immunoglobulins and T cell receptors. 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.1W.1-A.1W.15.
30. 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**, i379-i385.
31. Bleakley, K., Giudicelli, V., Wu, Y., Lefranc, M.-P. and Biau, G. (2006). IMGT standardization for statistical analyses of T cell receptor junctions: the TRAV-TRAJ example. *In Silico Biol.*, Epub 2006, 6, 0051, http://www.bioinfo.de/isb/2006/06/0051/
32. 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.
33. 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. *Nucl. Acids Res.*, **32**, D208-D210.
34. 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. *Nucl. Acids Res.,* **28**, 235-242.
35. 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.
36. Lefranc, M.-P. and Lefranc, G. (2004). Immunoglobulin lambda (IGL) genes of human and mouse. In: *Molecular Biology of B cells* (Honjo T., Alt F.W. and Neuberger M.S. eds), Academic Press, Elsevier Science, pp. 37-59.
37. Kaas, Q. and Lefranc, M.-P. (2007). IMGT Colliers de Perles: standardized sequence-structure representations of the IgSF and MhcSF superfamily domains. *Current Bioinformatics* **2**, 21-30.
38. 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**, pp. 111-136.
39. Garapati, V.P. and Lefranc, M.-P. (2007). IMGT Colliers de Perles and IgSF domain standardization for T cell costimulatory activatory (CD28, ICOS) and inhibitory (CTLA4, PDCD1 and BTLA) receptors. *Dev. Comp. Immunol*., **31**, 1050-1072.
40. Bernard, D., Hansen, J.D., du Pasquier, L., Lefranc, M.-P., Benmansour, A. and Boudinot, P. (2007). Costimulatory receptors in jawed vertebrates: conserved CD28, odd CTLA4 and multiple BTLAs. *Dev. Comp. Immunol*., **31**, 255-271.
41. 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.
42. 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**, pp. 95-145.

43. 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.
44. 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.
45. Chaume, D., Giudicelli, V., Combres, K., Ginestoux, C. and Lefranc, M.-P. (2005). IMGT-Choreography: processing of complex immunogenetics knowledge. Computational Methods in Systems Biology: International Conference CMSB 2004, Paris, France, May 26-28, 2004 (Danos V. and Schachter V., eds). In: *Lecture Notes in Computer Science*, Springer-Verlag GmbH Berlin Heidelberg, pp. 73-84.
46. Chassagne, S., Laffly, E., Drouet, E., Herodin, F., Lefranc, M.-P. and Thullier, P. (2004). A high affinity macaque antibody Fab with human-like framework regions obtained from a small phage display immune library. *Mol. Immunol*. **41**, 539-546.
47. Laffly, E., Danjou, L., Condemine, F., Vidal, D., Drouet, E., Lefranc, M.-P., Bottex, C. and Thullier, P. (2005). 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*., **49**, 3414-3420.
48. Pelat T., Hust M., Laffly E., Condemine F., Bottex C., Vidal D., Lefranc M.-P., Dubel S. and Thullier P. (2007). A high affinity, human-like antibody fragment (scFv) neutralising the lethal factor (LF) of *Bacillus anthracis* by inhibiting PA-LF complex formation. *Antimicrob. Agents Chemother*., **51**, 2758-2764.
49. Magdelaine-Beuzelin, C., Kaas, Q., Wehbi, V., Ohresser, M., Jefferis, R., Lefranc, M.-P. and Watier, H. (2007). Structure-function relationships of the variable domains of monoclonal antibodies approved for cancer treatment. *Critical reviews in Oncology and Haematology*, **64**, 210-225.