Chapter 1

IMGT-ONTOLOGY, IMGT® Databases, Tools, and Web Resources for Immunoinformatics

Marie-Paule Lefranc

IMGT®, the International ImMunoGeneTics Information System®, Université Montpellier II, Institut Universitaire de France, Laboratoire d'ImmunoGénétique Moléculaire LIGM, UPR CNRS 1142, Institut de Génétique Humaine IGH, 141 rue de la Cardonille, 34396 Montpellier Cedex 5, France, Marie-Paule.Lefranc@igh.cnrs.fr

Abstract. IMGT®, the international ImMunoGeneTics information system®, was created in 1989 as a high-quality integrated knowledge resource specialized in immunoglobulins (IG), T cell receptors (TR), major histocompatibility complexes (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). IMGT® is the international reference in immunogenetics and immunoinformatics. IMGT® combines sequence databases (IMGT/LIGM-DB, IMGT/PRIMER-DB, IMGT/PROTEIN-DB, IMGT/MHC-DB), a genome database (IMGT/ GENE-DB), and a three-dimensional (3D) structure database (IMGT/ 3Dstructure-DB) with interactive analysis tools (IMGT/V-QUEST, IMGT/JunctionAnalysis) and Web resources comprising 8000 HTML pages (IMGT Repertoire). The accuracy and consistency of IMGT data are based on IMGT-ONTOLOGY, available for biologists and IMGT users in the IMGT Scientific chart and for computer scientists in IMGT-ML, in XML format. IMGT® components (databases, tools, and Web resources) have been developed according to three main biological approaches: the genomic approach that is gene centered, the genetic approach that refers to genes in relation to their polymorphisms, expression, specificity, and evolution, and the structural approach that analyses 3D structures in relation to protein function and recognition sites. We are implementing Web services for the IMGT databases and tools. This is the first step toward IMGT-Choreography that will trigger and coordinate dynamic interactions between IMGT Web services in order to process complex significant biological and clinical requests. IMGT® is widely used in fundamental and medical research (repertoire analysis of the IG antibody sites and of the TR recognition sites in autoimmune diseases, infectious diseases, AIDS, leukemias, lymphomas, myelomas), veterinary research situations (IG and TR repertoires in farm and wildlife species), genome diversity and genome evolution studies of the adaptive immune responses, biotechnology related to antibody engineering (single chain Fragment variable (scFv), 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.

1.1 Introduction

Genome and proteome analysis interpretation represents the current great challenge, as a huge quantity of data is produced by many scientific fields, including fundamental, clinical, veterinary, and pharmaceutical research. In particular, the number of sequences

and related data published in the immunogenetics fields is growing exponentially. The 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 10^{12} different 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 (Lefranc and Lefranc 2001a; Lefranc and Lefranc 2001b).

IMGT® (http://imgt.cines.fr), the international ImMunoGeneTics information system® (Lefranc, Giudicelli, Kaas, Duprat, Jabado-Michaloud, Scaviner, Ginestoux, Clément, Chaume, and Lefranc 2005a), 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 the international reference in immunogenetics and immunoinformatics, and represents a high-quality integrated knowledge resource, specialized in the IG, TR, major histocompatibility complex (MHC) of human and other vertebrates, and related proteins of the immune systems (RPI) of any species which belong to the immunoglobulin superfamily (IgSF) and to the MHC superfamily (MhcSF). As such, IMGT® provides a common access to standardized data from genome, proteome, genetics, and three-dimensional (3D) structures.

1.2 The IMGT® Information System

The IMGT® information system consists of databases, tools, and Web resources (Lefranc, Clément, Kaas, Duprat, Chastellan, Coelho, Combres, Ginestoux, Giudicelli, Chaume, and Lefranc 2004a; Lefranc, Giudicelli, Ginestoux, Bosc, Folch, Guiraudou, Jabado-Michaloud, Magris, Scaviner, Thouvenin, Combres, Girod, Jeanjean, Protat, Monod, Duprat, Kaas, Pommié, Chaume, and Lefranc 2004b; Lefranc et al. 2005a). Databases and tools are summarized in Fig. 1.

Databases include several sequence databases (IMGT/LIGM-DB, IMGT/MHC-DB, IMGT/PRIMER-DB, IMGT/PROTEIN-DB), a genome database (IMGT/ GENE-DB), and a 3D structure database (IMGT/3Dstructure-DB). Interactive tools are provided for nucleotide and amino acid sequence analysis (IMGT/V-QUEST, IMGT/JunctionAnalysis, IMGT/Allele-Align, IMGT/PhyloGene, IMGT/DomainDisplay), genome analysis (IMGT/LocusView, IMGT/ GeneView, IMGT/GeneSearch, IMGT/CloneSearch, IMGT/GeneInfo, IMGT/GeneFrequency), and 3D structure analysis (IMGT/StructuralQuery, IMGT/DomainGapAlign, IMGT/Collier-de-Perles, IMGT/DomainSuperimpose). Web resources (IMGT Marie-Paule page) comprise 8000 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

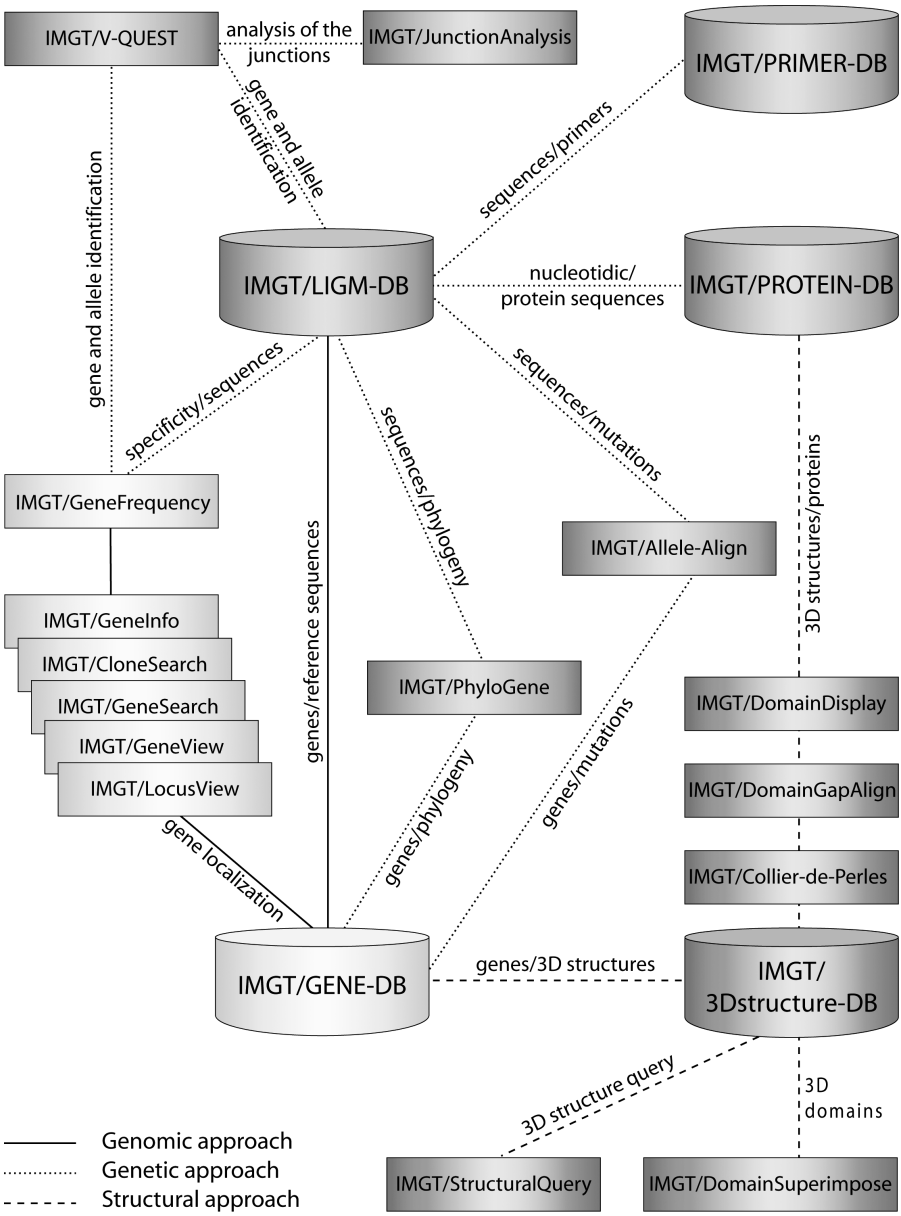

Fig. 1. IMGT®, the international ImMunoGeneTics information system® (http://imgt.cines.fr) databases and tools. The IMGT Repertoire and other IMGT Web resources are not shown. Examples of interactions between the databases (cylinders) and tools (rectangles) in the genomic, genetic and structural approaches are represented respectively by continuous, dotted and broken lines. (A color version of this figure appears between pages 76 and 77.)

Veterinary page, The IMGT Biotechnology page, IMGT Index], and external links [IMGT Bloc-notes (The IMGT Immunoinformatics page, Interesting links), and Other accesses (SRS, BLAST)]. 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 (Giudicelli and Lefranc 1999), the first ontology in the domain, which provides a semantic specification of the terms to be used in immunogenetics and immunoinformatics, and thus allows the management of immunogenetic knowledge for all vertebrate species. IMGT-ONTOLOGY comprises seven main concepts: IDENTIFICATION, CLASSIFICATION, DESCRIPTION, NUMEROTATION, LOCALIZATION, ORIENTATION, and OBTENTION (Giudicelli and Lefranc 1999; Lefranc et al. 2004a; Lefranc et al. 2004b; Lefranc et al. 2005a). Standardized keywords, standardized IG and TR gene nomenclature, standardized labels, the IMGT unique numbering, annotation rules, and standardized origin/methodology were defined, respectively, based on these seven main concepts.

IMGT-ONTOLOGY concepts are available for biologists and IMGT® users in the IMGT Scientific chart and formalized for computer scientists in IMGT-ML using XML (Extensible Markup Language) Schema. The IMGT Scientific chart (Lefranc, Giudicelli, Ginestoux, Bodmer, Müller, Bontrop, Lemaitre, Malik, Barbié, and Chaume 1999) comprises controlled vocabulary and annotation rules for data and knowledge management of the IG, TR, and MHC of vertebrate species, and of the RPI of any species, that belong to IgSF and MhcSF. All IMGT® data are expertly annotated according to the IMGT Scientific chart rules. The IMGT Scientific chart is available as a section of the IMGT Web resources (IMGT Marie-Paule page). These HTML pages are devoted to biologists, IMGT users, and IMGT annotators. Examples of IMGT expert data concepts, derived from the IMGT Scientific chart rules, correspond to section titles and subtitles in IMGT Repertoire (Lefranc et al. 2004a; Lefranc et al. 2004b).

IMGT-ML (Chaume, Giudicelli, and Lefranc 2001; Chaume, Giudicelli, Combres, and Lefranc 2003; Chaume, Giudicelli, Combres, Ginestoux, and Lefranc 2005) is the formalization of IMGT-ONTOLOGY using XML Schema for interoperability with other information systems. IMGT® components (databases, tools, and IMGT Repertoire Web resources) have been developed according to three main biological approaches. The IMGT® genomic approach is gene-centered focusing on 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 to sequence polymorphisms and mutations and their expression, specificity, and evolution. The IMGT® structural approach refers to the study of 2D and 3D structures of the IG, TR, MHC, and RPI, and to the antigen or ligand binding characteristics in relation to protein functions, polymorphisms, and evolution. IMGT-Choreography, based on the Web service architecture paradigm, will enable significant biological and clinical requests addressing the entire IMGT® information system.

1.3 IMGT-ONTOLOGY Concepts and IMGT® Components for Genomics

1.3.1 IMGT® Genome Database

The IMGT® genomic approach refers to the study of the genes within their loci and on the chromosomes. Genomic data are managed in IMGT/GENE-DB, which is the comprehensive IMGT® genome database, created by LIGM, Montpellier, France, on the Web since January 2003 (Giudicelli, Chaume, and Lefranc 2005).

In March 2007, IMGT/GENE-DB contained 1512 genes and 2461 alleles (673 IG and TR genes and 1215 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 minutoïdes*). All human and mouse IG and TR genes are available in IMGT/GENE-DB. Based on the IMGT® CLASSIFICATION concept, all the human IMGT® gene names (Lefranc and Lefranc 2001a; Lefranc and Lefranc 2001b; Lefranc 2000a; Lefranc 2000b; Lefranc 2000c; Lefranc 2000d) were approved by the HUman Genome Organisation (HUGO) Nomenclature Committee HGNC in 1999 (Wain, Bruford, Lovering, Lush, Wright, and Povey 2002), and entered in IMGT/GENE-DB (Giudicelli et al. 2005), Genome DataBase GDB, Canada (Letovsky, Cottingham, Porter, and Li 1998), LocusLink at NCBI, USA (Pruitt and Maglott 2001), and GeneCards (Safran, Chalifa-Caspi, Shmueli, Olender, Lapidot, Rosen, Shmoish, Peter, Glusman, Feldmesser, Adato, Peter, Khen, Atarot, Groner, and Lancet 2003). Reciprocal links exist between IMGT/GENE-DB and the generalist nomenclature (HGNC) and genome databases (GDB, Entrez Gene at NCBI, and GeneCards). The mouse IG and TR gene names (Martinez and Lefranc 1998; Bosc and Lefranc 2000; Bosc, Contet, and Lefranc 2001; Martinez, Folch, and Lefranc 2001; Bosc and Lefranc 2003) with IMGT reference sequences were provided by IMGT® to HGNC and to the Mouse Genome Database MGD (Blake, Richardson, Bult, Kadin, Eppig, and Mouse Genome Database Group 2003) 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 (Lefranc et al. 1999). IMGT/GENE-DB interacts dynamically with IMGT/LIGM-DB (Giudicelli, Ginestoux, Folch, Jabado-Michaloud, Chaume, and Lefranc 2006) to download and display gene-related sequence data. This is the first example of an interaction between IMGT® databases using the CLASSIFICATION concept.

1.3.2 IMGT® Genome Analysis Tools

The IMGT® genome analysis on-line 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 display of physical maps for the human IG, TR, and MHC loci and for the mouse TRA/TRD locus. IMGT/LocusView allows

users to view genes at their loci and then zoom in on a selected area. IMGT/GeneView allows users to directly view a given gene at a locus. IMGT/GeneSearch and IMGT/CloneSearch allow retrieval of information concerning genes and clones, respectively, analysed in IMGT/LocusView. IMGT/GeneSearch allows searching for genes at a locus, based on IMGT gene names, functionality, or chromosomal localization. IMGT/CloneSearch provides information on the clones that were used to build the locus contigs displayed in IMGT/LocusView (accession numbers are from IMGT/LIGM-DB, gene names from IMGT/GENE-DB, and clone position and orientation, and overlapping clones from IMGT/LocusView). IMGT/GeneInfo provides information on potential human and mouse TR rearrangements (Baum, Hierle, Pascal, Bellahcene, Chaume, Lefranc, Jouvin-Marche, Marche, and Demongeot 2006). IMGT/GeneFrequency is an IMGT interactive tool that dynamically computes histograms which represent the contribution of individual V, D, and J genes in sets of expressed IG and TR rearranged V-(D)-J sequences in IMGT/LIGM-DB. IMGT/GeneFrequency results are obtained by querying IMGT/LIGM-DB for sequences which are selected, for example, on the specificity criteria.

1.3.3 IMGT® Genome Web Resources

The IMGT® genomic Web resources are compiled in the IMGT Repertoire "Locus and genes" section that 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 (Lefranc and Lefranc 2001a; Lefranc and Lefranc 2001b). The IMGT Repertoire "Probes and RFLP" section provides data on probes used in Southern analysis and on gene insertion/deletion polymorphisms (Osipova, Posukh, Wiebe, Miyazaki, Matsumoto, Lefranc, and Lefranc 1999; Dard, Lefranc, Osipova, and Sanchez-Mazas 2001; Elemento, Gascuel, and Lefranc 2002; Lefranc and Lefranc 2004).

1.4 IMGT-ONTOLOGY Concepts and IMGT® Components for Genetics

1.4.1 IMGT® Sequence Databases

The IMGT® genetic approach refers to the study of genes in relation to their polymorphisms, mutations, expression, specificity, and evolution. The IMGT® genetics approach heavily relies on the DESCRIPTION concept (and particularly on the V-REGION, D-REGION, J-REGION, and C-REGION core concepts for the IG and TR), on the CLASSIFICATION concept (gene and allele concepts), and on the NUMEROTATION concept (IMGT unique numbering) (Lefranc 1997; Lefranc 1999; Ruiz and Lefranc 2002; Duprat and Lefranc 2003; Lefranc, Pommié, Ruiz, Giudicelli, Foulquier, Truong, Thouvenin-Contet, and Lefranc 2003; Lefranc, Pommié, Kaas,

Duprat, Bosc, Guiraudou, Jean, Ruiz, Da Piedade, Rouard, Foulquier, Thouvenin, and Lefranc 2005b; Lefranc, Duprat, Kaas, Tranne, Thiriot, and Lefranc 2005c).

1.4.1.1 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, and available on the Web since July 1995 (Lefranc, Giudicelli, Busin, Malik, Mougenot, Déhais, and Chaume 1995; Giudicelli et al. 2006). In March 2007, IMGT/LIGM-DB contained 107,737 sequences of 150 vertebrate species. The unique source of data for IMGT/LIGM-DB is EMBL (Kulikova, Aldebert, Althorpe, Baker, Bates, Browne, van den Broek, Cochrane, Duggan, Eberhardt, Faruque, Garcia-Pastor, Harte, Kanz, Leinonen, Lin, Lombard, Lopez, Mancuso, McHale, Nardone, Silventoinen, Stoehr, Stoesser, Tuli, Tzouvara, Vaughan, Wu, Zhu, and Apweiler 2004) which shares data with the other two generalist databases GenBank and DNA Data Bank of Japan (DDBJ). Based on expert analysis, specific detailed annotations are added to IMGT® flat files. The Web interface allows searches according to specific immunogenetic criteria and is easy to use without any programming language knowledge. 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 (Lefranc 2003; Lefranc et al. 2004b). 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 (EBI Hinxton UK, Institut Pasteur Paris, DKFZ Heidelberg Germany, Columbia University New York USA, IUBio Indiana University USA, DDBJ Japan, etc.). IMGT/LIGM-DB can be searched by BLAST or FASTA on different servers (EBI Hinxton UK, Institut Pasteur Paris).

1.4.1.2 IMGT/Automat for IMGT/LIGM-DB Annotations

IMGT/Automat (Giudicelli, Protat, and Lefranc 2003) is an integrated internal IMGT® Java tool which automatically performs the annotation of rearranged cDNA sequences that represent 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 9890 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.

1.4.1.3 Other IMGT® IG and TR Sequence Databases

IMGT/PRIMER-DB (Folch, Bertrand, Lemaitre, and Lefranc 2004) 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 (http://www3.oup.co.uk/nar/database/summary/505). In March 2007, IMGT/PRIMER-DB contained 1864 entries and 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/PROTEIN-DB is a new IMGT® database related to immunoglobulin and T-cell receptor amino acid sequences. The database will be available on the IMGT® Web site in 2007.

1.4.1.4 IMGT® MHC Sequence Databases

IMGT/MHC-DB comprises databases hosted at EBI and includes a database of human MHC allele sequences, IMGT/MHC-HLA (or IMGT/HLA), developed by Cancer Research, UK and maintained by the Anthony Nolan Research Institute ANRI, London, UK, on the Web since December 1998, and a database of MHC sequences from nonhuman primates IMGT/MHC-NHP, curated by the Biomedical Primate Research Centre BPRC, The Netherlands, on the Web since April 2002 (Robinson, Waller, Parham, de Groot, Bontrop, Kennedy, Stoehr, and Marsh 2003).

1.4.2 IMGT® Sequence Analysis Tools

The IMGT® tools for the genetics approach comprise IMGT/V-QUEST (Lefranc 2003; Giudicelli, Chaume, and Lefranc 2004), for the identification of the V, D, and J genes and of their mutations, IMGT/JunctionAnalysis (Yousfi Monod, Giudicelli, Chaume, and Lefranc 2004) for the analysis of the V-J and V-D-J junctions which confer the antigen receptor specificity, IMGT/Allele-Align for the detection of polymorphisms, IMGT/Phylogene (Elemento and Lefranc 2003) for gene evolution analyses, and IMGT/DomainDisplay for amino acid sequences.

1.4.2.1 IMGT/V-QUEST

IMGT/V-QUEST (V-QUEry and STandardization) is an integrated software for IG and TR (Lefranc 2003, Giudicelli et al. 2004), used for the identification of the V, D, and J genes and of their mutations. This tool is easy to use for the analysis of input IG or TR germline or rearranged variable nucleotide sequences. 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 sequences, the display of nucleotide and amino acid mutations compared to the closest IMGT reference sequences, the identification of the JUNCTION and results from IMGT/JunctionAnalysis (default option), and the V-REGION IMGT Colliers de Perles. IMGT/V-QUEST provides a synthetic view of the results when several sequences (up to 50) are analysed in the same run.

1.4.2.2 IMGT/JunctionAnalysis

IMGT/JunctionAnalysis (Yousfi Monod et al. 2004) 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. Several hundred junction sequences can be analysed simultaneously.

1.4.2.3 IMGT/Allele-Align

IMGT/Allele-Align is used for the detection of polymorphisms. It allows the comparison of two alleles highlighting the nucleotide and amino acid differences.

1.4.2.4 IMGT/PhyloGene

IMGT/PhyloGene (Elemento and Lefranc 2003) is an easy 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.

1.4.2.5 IMGT/DomainDisplay

IMGT/DomainDisplay provides a display of amino acid sequences per domain (V, C, or G type domain) for IG, TR, MHC and for RPI (that include IgSF proteins other than IG and TR, and MhcSF proteins other than MHC), based on the IMGT unique numbering (Lefranc et al. 2003; Lefranc et al. 2005b; Lefranc et al. 2005c).

1.4.3 IMGT® Genetics Web Resources

The IMGT® genetic Web resources are compiled in the IMGT Repertoire "Proteins and alleles" section which includes Protein displays, Alignments of alleles, Tables of alleles, Allotypes, and Isotypes (Osipova et al. 1999; Dard et al. 2001; Lefranc and Lefranc 2004).

1.5 IMGT-ONTOLOGY Concepts and IMGT® Components for 2D and 3D Structures

1.5.1 IMGT® Structural Database

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 to 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, domain delimitations), and NUMEROTATION concept (amino acid positions according to the IMGT unique numbering) (Lefranc et al. 2003; Lefranc et al. 2005b; Lefranc et al. 2005c). Structural and functional domains of the IG and TR chains comprise the variable domain or V-DOMAIN (nine-strand β-sandwich) that corresponds to the V-J-REGION or V-D-J-REGION and is encoded by two or three genes (Lefranc and Lefranc 2001a; Lefranc and Lefranc 2001b; Lefranc et al. 2003), the constant domain or C-DOMAIN (seven-strand β-sandwich) (Lefranc et al. 2005b), and, for the MHC chains, the groove domain or G-DOMAIN (four β-strand and one α-helix) (Lefranc et al. 2005c). The IMGT unique numbering has been extended to the V-LIKE-DOMAINs (Lefranc et al. 2003) and C-LIKE-DOMAINs (Lefranc et al. 2005b) of IgSF proteins other than IG and TR, and to the G-LIKE-DOMAINs (Lefranc et al. 2005c) of MhcSF proteins other than MHC.

Structural data are compiled and annotated in IMGT/3Dstructure-DB. IMGT/3Dstructure-DB is the IMGT® 3D structure database for IG, TR, MHC, and RPI, created by LIGM, on the Web since November 2001 (Kaas, Ruiz, and Lefranc 2004). In March 2007, IMGT/3Dstructure-DB contained 1221 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 (Berman, Westbrook, Feng, Gilliland, Bhat, Weissig, Shindyalov, and Bourne 2000) (http://www.rcsb.org/pdb/) are renumbered according to the standardized IMGT unique numbering (Lefranc et al. 2003; Lefranc et al. 2005b; Lefranc et al. 2005c). The IMGT/3Dstructure-DB cards provide IMGT annotations (assignment of IMGT genes and alleles, IMGT chain and domain labels, IMGT Colliers de Perles for V, C, and G type domains (Ruiz and Lefranc 2002; Kaas and Lefranc 2005; Kaas and Lefranc 2007), downloadable renumbered IMGT/3Dstructure-DB flat files, visualization 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, and amino acid positions according to the IMGT unique numbering. Standardized IMGT pMHC contact sites have been defined for peptide/MHC complexes (Kaas and Lefranc 2005).

1.5.2 IMGT® Structure Analysis Tools

Several on-line IMGT® structure analysis tools are available for the analysis of 2D and 3D structures, and particularly for the comparison of V, C, and G domains. The IMGT/StructuralQuery tool (Kaas et al. 2004) 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) with backbone or side-chain implication. IMGT/StructuralQuery allows users to retrieve the IMGT/3Dstructure-DB entries, based on specific structural characteristics: ϕ (phi) and ψ (psi) angles, accessible surface area (ASA), amino acid type, distance in angstroms between amino acids, CDR-IMGT lengths. IMGT/StructuralQuery is currently available for the V-DOMAINs. IMGT/DomainGapAlign aligns users' amino acid sequences against the closest reference sequences from the IMGT domain sequence directory. IMGT/DomainGapAlign also provides the IMGT gaps and thus allows users to graphically represent their domain sequences with the IMGT/Collier-de-Perles tool. IMGT/DomainSuperimpose allows superimposing two 3D structures of domains from IMGT/3Dstructure-DB.

1.5.3 IMGT® Structural Web Resources

The IMGT® structural Web resources are compiled in the IMGT Repertoire "2D and 3D structures" section which includes 2D representations or IMGT Colliers de Perles (Lefranc et al. 2003; Lefranc et al. 2005b; Lefranc et al. 2005c; Ruiz and Lefranc 2002; Kaas and Lefranc 2007), 3D representations, FR-IMGT and CDR-IMGT lengths, and amino acid physico-chemical characteristic profiles (Pommié, Sabatier, Lefranc, and Lefranc 2004).

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 "chemical characteristics" amino acid properties and used to set up IMGT Colliers de Perles reference profiles (Pommié et al. 2004). The IMGT Colliers de Perles reference profiles allow one to easily compare amino acid properties at each position whatever the domain, the chain, the receptor, or the species. The IG and TR variable and constant domains represent a privileged situation for the analysis of amino acid properties in relation to 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 and to determine amino acids potentially involved in the immunogenicity.

1.6 IMGT-Choreography

The goal of IMGT-Choreography is to orchestrate dynamic procedure calls between IMGT® databases querying and analysis tools, using IMGT's biological approaches (Lefranc et al. 2004a). Major existing or potential "conversation nodes" can be identified between IMGT® tools, by an analysis of their profiles (IMGT tool diamonds; Lefranc et al. 2004a). IMGT-Choreography is based on the Web service architecture paradigm (see W3C; http://www.w3.org/). Conversations between Web services are expressed using the sole IMGT-ML language both for queries and for result fetches.

1.6.1 IMGT Tool Diamonds

In order to enhance the interoperability between the IMGT® components, IMGT® tools were analysed for input and output parameters, performed tasks, and accompanying databases (IMGT reference directories). Graphical diamond-shaped representations, designated as "IMGT tool diamonds" (Lefranc et al. 2004a), were developed to obtain tool profiles and to compare the state of the art of each tool in relation to the IMGT ontological concepts. Each IMGT tool diamond is composed of modules that correspond to different IMGT-ONTOLOGY concepts and each module comprises four facets: input parameters, task, IMGT reference directory, and output parameters (Lefranc et al. 2004a).

1.6.2 IMGT-ML

IMGT-ML (Chaume et al. 2001; Chaume et al. 2003) (available at IMGT Index>IMGT-ML, http://imgt.cines.fr) represents the specification of the main IMGT-ONTOLOGY concepts (Giudicelli and Lefranc 1999), formalized through a markup language defined in-house, based on Extensible Markup Language (XML) (http://www.w3.org/XML/) and constrained through XML Schema (http://www.w3.org/XML/Schema). Messages that are exchanged between service providers and consumers are encoded using valid IMGT-ML streams. IMGT-ML can be seen as a kind of Rosetta stone since it extends the ease of interconnection between IMGT Web services and is the unique language used for both services inputs and outputs. This ensures semantic consistency between exchanged messages as IMGT-ML is an XML schema formalization of the IMGT-ONTOLOGY concepts (Chaume et al. 2003).

1.6.3 IMGT® Web Services

Web services have been chosen as the means to create dynamic interactions between IMGT® databases and tools. Clients and providers for these services can be written using any SOAP-capable programming language such as SOAP::lite (http://www.soaplite.com/), development library for Perl or webMethods Glue for JAVA, thus facilitating the conversion of legacy applications to services. 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.

The IMGT/LIGM-DB Web service is the first Web service currently developed and implemented with Axis (Lefranc et al. 2004a). It includes the "queryKnowledge" and "querySeqData" services. The queryKnowledge service provides the lists of instances for the IMGT-ONTOLOGY concepts, for example the list of chain types, functionalities, specificities defined in the IDENTIFICATION concept, the lists of groups and subgroups defined in the CLASSIFICATION concept, or the list of labels defined in the DESCRIPTION concept. The querySeqData service allows the retrieval of any sequence-related data that are identified, classified, and described in IMGT/LIGM-DB according to the IMGT concepts. The querySeqData input has the form of an incomplete IMGT-ML data entry in which the given values are used as criteria to query IMGT/LIGM-DB. The result is a list of data entries, in IMGT-ML format, sharing these given values. Other Web services are developed to automatically query IMGT® databases and tools.

1.6.4 Perspectives

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. IMGT-Choreography has for goal to combine and join the IMGT® database queries and analysis tools.

In order to keep only significant approaches, a rigorous analysis of the scientific standards of the biologist research (Giudicelli and Lefranc 1999; Lefranc and Lefranc 2001a; Lefranc and Lefranc 2001b; Osipova et al. 1999; Dard et al. 2001; Chardes, Chapal, Bresson, Bes, Giudicelli, Lefranc, and Peraldi-Roux 2002;; Chassagne, Laffly, Drouet, Herodin, Lefranc, and Thullier 2004; Bertrand, Duprat, Lefranc, Marti, and Coste 2004) and of the clinician's needs (Ghia, Stamatopoulos, Belessi, Moreno, Stella, Giuda, Michel, Crespo, Laoutaris, Montserrat, Anagnostopoulos, Dighiero, Fassas, Caligaris-Cappio, and Davi 2005; Stamatopoulos, Belessi, Papadaki, Kalagiakou, Stavroyianni, Douka, Afendaki, Saloum, Parasi, Anagnostou, Laoutaris, Fassas, and Anagnostopoulos 2004) has been undertaken in parallel with the modelling of interactions between the IMGT® components (databases, tools, and Web resources). To increase interoperability with other biological information systems and ontologies, IMGT-ONTOLOGY is currently being implemented with Protégé (http://protege.stanford.edu/) (Noy, Fergerson, and Musen 2000).

1.7 Conclusions

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 140,000 requests a month. IMGT® is the international reference in immunogenetics and immunoinformatics and provides a common access to standardized data which include nucleotide and protein sequences, oligonucleotide primers, gene maps, genetic polymorphisms, specificities, 2D and 3D structures, based on IMGT-ONTOLOGY. Although the IMGT® genome, sequence, and 3D structure databases, IMGT® analysis tools, and IMGT Repertoire Web resources, were initially imple-

mented for the IG, TR, and MHC of human and other vertebrates, data and knowledge management standardization has now been extended to the IgSF proteins other than IG or TR (Williams and Barclay 1988) and to the MhcSF proteins other than MHC (Maenaka and Jones 1999), of any species (IMGT Repertoire (RPI)). Thus, standardization in IMGT® contributed to data enhancement of the system and new expertised data concepts were readily incorporated.

The IMGT® information is of much value to clinicians and biological scientists in general. 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, Europe, and the rest of the world. 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 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, 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.

1.8 Citing IMGT®

Users are requested to cite this article, and to quote the IMGT® home page URL, http://imgt.cines.fr. Individual IMGT® databases, tools, and Web resources should also be quoted where relevant: IMGT/GENE-DB (Giudicelli et al. 2005), IMGT/GeneInfo (Baum et al. 2006), IMGT/LIGM-DB (Giudicelli et al. 2006), IMGT/MHC-DB (Robinson et al. 2003), IMGT/PRIMER-DB (Folch et al. 2004), IMGT/V-QUEST (Giudicelli et al. 2004), IMGT/JunctionAnalysis (Yousfi Monod et al. 2004), IMGT/PhyloGene (Elemento and Lefranc 2003), IMGT/3Dstructure-DB (Kaas et al. 2004), IMGT/StructuralQuery (Kaas et al. 2004), and IMGT/Collier-de-Perles (Kaas and Lefranc 2005; Kaas and Lefranc 2007).

Acknowledgements

I am very grateful to the IMGT® team for its expertise and its constant motivation. I thank Véronique Giudicelli, Joumana Jabado-Michaloud, Chantal Ginestoux, Géraldine Folch, Patrice Duroux, François Ehrenmann, Xavier Brochet, and Gérard Lefranc for helpful discussion. IMGT® is a registered Centre National de la Recher-

che Scientifique (CNRS) mark. IMGT® has been a National Bioinformatics Platform RIO (CNRS, INSERM, CEA, INRA) since 2001. 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) programs of the European Union and received subventions from the Association pour la Recherche sur le Cancer (ARC) and from the Génopole-Montpellier-Languedoc-Roussillon. 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, the Réseau National des Génopoles (RNG), BIOSTIC-LR2004, ACI-IMPBIO IMP82-2004, GIS AGENAE, Région Languedoc-Roussillon, Agence Nationale de la Recherche ANR (BIOSYS06_ 135457), and the ImmunoGrid project (IST-2004-028069) of the 6th framework program of the European Union.

References

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.

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.

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.

Blake, J.A., Richardson, J.E., Bult, C.J., Kadin, J.A., Eppig, J.T., and Mouse Genome Database Group (2003) MGD: The Mouse Genome Database. Nucleic Acids Res. 31:193-195.

Bosc, N., and Lefranc, M.-P. (2000) The Mouse (*Mus musculus*) T cell receptor β variable (TRBV), diversity (TRBD) and joining (TRBJ) genes. Exp. Clin. Immunogenet. 17:216-228.

Bosc, N., Contet, V. and Lefranc, M.-P. (2001) The mouse (*Mus musculus*) T cell receptor delta variable (TRDV), diversity (TRDD) and joining (TRDJ) genes. Exp. Clin. Immunogenet. 18:51-58.

Bosc, N., and Lefranc, M.-P. (2003) IMGT Locus in focus: The mouse (*Mus musculus*) T cell receptor α (TRA) and delta (TRD) variable genes. Dev. Comp. Immunol. 27:465-497.

Chardes, T., Chapal, N., Bresson, D., Bes, C., Giudicelli, V., Lefranc, M.-P., and Peraldi-Roux, S. (2002) The human anti-thyroid peroxidase autoantibody repertoire in Graves' and Hashimoto's autoimmune thyroid diseases. Immunogenetics 54:141-157.

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.

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, pp. 71-75.

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, pp. 22-23.

Chaume, D., Giudicelli, V., Combres, K., Ginestoux, C., and Lefranc, M.-P. (2005) IMGT-Choreography: Processing of complex immunogenetics knowledge. In: *Computational Methods in Systems Biology CMSB 2004. Lecture Notes in Bioinformatics LNBI,* Springer, pp. 73-84.

Dard, P., Lefranc, M.-P., Osipova, L., and Sanchez-Mazas, A. (2001) DNA sequence variability of IGHG3 genes associated to the main G3m haplotypes in human populations. Eur. J. Hum. Genet. 9:765-772.

Duprat, E., and Lefranc, M.-P. (2003) IMGT standardization and analysis of V-LIKE-, C-LIKE- and G-LIKE-DOMAINs. In: *Proceedings of the European Conference on Computational Biology ECCB''2003,* PS-32, pp. 223-224.

Elemento, O., Gascuel, O., and Lefranc, M.-P. (2002) Reconstructing the duplication history of tandemly repeated genes. Mol. Biol. Evol. 19:278-288.

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.

Folch, G., Bertrand, J., Lemaitre, M., and Lefranc, M.-P. (2004) IMGT/PRIMER-DB. In: M.Y. Galperin (Ed.), The Molecular Biology Database Collection: 2004 update. Nucleic Acids Res. 32:3-22.

Ghia, P., Stamatopoulos, K., Belessi, C., Moreno, C., Stella, S., Giuda, G., Michel, A., Crespo, M., Laoutaris, N., Montserrat, E., Anagnostopoulos, A., Dighiero, G., Fassas, A., Caligaris-Cappio, F., and Davi, F. (2005) Geographical patterns and pathogenetic implications of IGHV gene usage in chronic lymphocytic leukemia: The lesson of the IGHV3-21 gene. Blood 105:1678-1685.

Giudicelli, V., and Lefranc, M.-P. (1999) Ontology for Immunogenetics: The IMGT-ONTOLOGY. Bioinformatics 12:1047-1054.

Giudicelli, V., Protat, C., and Lefranc, M.-P. (2003) The IMGT strategy for the automatic annotation of IG and TR cDNA sequences: IMGT/Automat. In: *Proceedings of the European Conference on Computational Biology ECCB'2003,* DKB-31, pp. 103-104.

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.

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.

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. Nucleic Acids Res. 34**:**D781-D784.

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.

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* Biol. 5:505-528.

Kaas, Q., and Lefranc, M.-P. (2007) IMGT Colliers de Perles: Standardized sequence-structure representations of the IgSF and MhcSF superfamily domains. Curr. Bioinformatics 2:21-30.

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. Nucleic Acids Res. 32:D27-D30.

Lefranc, M.-P., Giudicelli, V., Busin, C., Malik, A., Mougenot, I., Déhais, P., and Chaume, D. (1995) LIGM-DB/IMGT: An integrated database of Ig and TcR, part of the Immunogenetics database. Ann. N. Y. Acad. Sci. 764:47-49.

Lefranc, M.-P. (1997) Unique database numbering system for immunogenetic analysis. Immunol. Today 18:509.

Lefranc, M.-P. (1999) The IMGT unique numbering for immunoglobulins, T cell receptors and Ig-like domains. The Immunologist 7:132-136.

Lefranc, M.-P., Giudicelli, V., Ginestoux, C., Bodmer, J., Müller, W., Bontrop, R., Lemaitre, M., Malik, A., Barbié, V., and Chaume, D. (1999) IMGT, the international ImMunoGeneTics database. Nucleic Acids Res. 27:209-212.

Lefranc, M.-P. (2000a) Nomenclature of the human T cell receptor genes. In: *Current Protocols in Immunology,* John Wiley & Sons, New York, Supplement 40, A.1O.1-A.1O.23.

Lefranc, M.-P. (2000b) Nomenclature of the human immunoglobulin genes. In: *Current Protocols in Immunology,* John Wiley & Sons, New York, Supplement 40, A.1P.1-A.1P.37.

Lefranc, M.-P. (2000c) Locus maps and genomic repertoire of the human T cell receptor genes. The Immunologist 8:72-79.

Lefranc, M.-P. (2000d) Locus maps and genomic repertoire of the human immunoglobulin genes. The Immunologist 8:80-88.

Lefranc, M.-P., and Lefranc, G. (2001a) *The Immunoglobulin FactsBook.* Academic Press, London, 458.

Lefranc, M.-P., and Lefranc, G. (2001b) *The T cell receptor FactsBook.* Academic Press, London, 398.

Lefranc, M.-P. (2003) IMGT, the international ImMunoGeneTics information system® (http://imgt.cines.fr). In: B.K.C. Lo (Ed.), *Antibody Engineering: Methods and Protocols.* 2nd edition. *Methods in Molecular Biology.* Humana Press, USA, 248, pp. 27-49.

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.

Lefranc, M.-P., and Lefranc, G. (2004) Immunoglobulin lambda (IGL) genes of human and mouse. In: T. Honjo, F.W. Alt, and M.S. Neuberger (Eds.), *Molecular Biology of B Cells.* Academic Press, Elsevier Science, pp. 37-59.

Lefranc, M.-P., Clément, O., Kaas, Q., Duprat, E., Chastellan, P., Coelho, I., Combres, K., Ginestoux, C., Giudicelli, V., Chaume, D., and Lefranc, G. (2004a) IMGT-Choreography for Immunogenetics and Immunoinformatics. *In Silico* Biol. 5:6.

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., Monod, M.Y., Duprat, E., Kaas, Q., Pommié, C., Chaume, D., and Lefranc, G. (2004b) IMGT-ONTOLOGY for Immunogenetics and Immunoinformatics (http://imgt.cines.fr). *In Silico* Biol. 4:17-29.

Lefranc, M.-P., Giudicelli, V., Kaas, Q., Duprat, E., Jabado-Michaloud, J., Scaviner, D., Ginestoux, C., Clément, O., Chaume, D., and Lefranc, G. (2005a) IMGT, the international ImMunoGeneTics information system®. Nucleic Acids Res. 33**:**D593-D597.

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. (2005b) IMGT unique numbering for immunoglobulin and T cell receptor constant domains and Ig superfamily C-like domains. Dev. Comp. Immunol. 29:185-203.

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

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.

Maenaka, K., and Jones, E.Y. (1999) MHC superfamily structure and the immune system. Curr. Opin. Struct. Biol. 9:745-753.

Martinez, C., and Lefranc, M.-P. (1998) The mouse (*Mus musculus*) immunoglobulin kappa variable (IGKV) genes and joining (IGKJ) segments. Exp. Clin. Immunogenet. 15: 184-193.

Martinez, C., Folch, G., and Lefranc, M.-P. (2001) Nomenclature and overview of the mouse (*Mus musculus* and *Mus sp.*) immunoglobulin kappa (IGK) genes. Exp. Clin. Immunogenet. 18:255-279.

Noy, N.F., Fergerson, R.W., and Musen, M.A. (2000) The knowledge model of Protege-2000: Combining interoperability and flexibility. *2th International Conference on Knowledge Engineering and Knowledge Management (EKAW'2000), Juan-les-Pins, France.*

Osipova, L.P., Posukh, O.L., Wiebe, V.P., Miyazaki, T., Matsumoto, H., Lefranc, G., and Lefranc, M.-P. (1999) *Bam*HI-*Sac*I RFLP and Gm analysis of the immunoglobulin IGHG genes in Northern Selkups (West Siberia): New haplotypes with deletion, duplication and triplication. Hum. Genet. 105:530-541.

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.

Pruitt, K.D., and Maglott, D.R. (2001) RefSeq and LocusLink: NCBI gene-centered resources. Nucleic Acids Res. 29:137-140.

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.

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.

Safran, M., Chalifa-Caspi, V., Shmueli, O., Olender, T., Lapidot, M., Rosen, N., Shmoish, M., Peter, Y., Glusman, G., Feldmesser, E., Adato, A., Peter, I., Khen, M., Atarot, T., Groner, Y., and Lancet, D. (2003) Human Gene-Centric Databases at the Weizmann Institute of Science: GeneCards, UDB, CroW 21 and HORDE. Nucleic Acids Res. 31:142-146.

Stamatopoulos, K., Belessi, C., Papadaki, T., Kalagiakou, E., Stavroyianni, N., Douka, V., Afendaki, S., Saloum, R., Parasi, A., Anagnostou, D., Laoutaris, N. Fassas, A., and Anagnostopoulos, A. (2004) Immunoglobulin heavy and light chain repertoire in splenic marginal zone lymphoma. Mol Med. 10:89-95.

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.

Williams, A.F., and Barclay, A.N. (1988) The immunoglobulin family: Domains for cell surface recognition. Annu. Rev. Immunol. 6:381-440.

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.