

# IMGT-Choreography for Immunogenetics and Immunoinformatics

Marie-Paule Lefranc\*, Oliver Clément, Quentin Kaas, Elodie Duprat, Patrick Chastellan, Isabelle Coelho, Kora Combres, Chantal Ginestoux, Véronique Giudicelli, Denys Chaume and Gérard Lefranc

*IMGT, the international ImMunoGeneTics information system*<sup>®</sup>, *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*  
Tel.: +33 4 99 61 99 65; Fax: +33 4 99 61 99 01

Edited by H. Michael; received 11 September 2004; revised and accepted 14 December 2004; published 24 December 2004

**ABSTRACT:** IMGT, the international ImMunoGeneTics information system<sup>®</sup> (<http://imgt.cines.fr>), was created in 1989 at Montpellier, France. IMGT is a high quality integrated knowledge resource specialized in immunoglobulins (IG), T cell receptors (TR), major histocompatibility complex (MHC) of human and other vertebrates, and related proteins of the immune system (RPI) which belong to the immunoglobulin superfamily (IgSF) and MHC superfamily (MhcSF). IMGT provides a common access to standardized data from genome, proteome, genetics and three-dimensional structures. The accuracy and the consistency of IMGT data are based on IMGT-ONTOLOGY, a semantic specification of terms to be used in immunogenetics and immunoinformatics. IMGT-ONTOLOGY has been formalized using XML Schema (IMGT-ML) for interoperability with other information systems. We are developing Web services to automatically query IMGT databases and tools. This is the first step towards IMGT-Choreography which will trigger and coordinate dynamic interactions between IMGT Web services to process complex significant biological and clinical requests. IMGT-Choreography will further increase the IMGT leadership in immunogenetics and immunoinformatics for medical research (repertoire analysis of the IG antibody recognition sites and of the TR recognition sites in autoimmune and infectious diseases, AIDS, leukemias, lymphomas, myelomas), veterinary research (IG and TR repertoires in farm and wild life 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 (detection and follow-up of residual diseases) and therapeutical approaches (grafts, immunotherapy, vaccinology).

IMGT is freely available at <http://imgt.cines.fr>.

**KEYWORDS:** IMGT, ontology, database, information system, knowledge resource, immunoinformatics, immunogenetics, antibody, immunoglobulin, T cell receptor, superfamily, MHC, HLA, Collier de Perles, three-dimensional, 3D structure, polymorphism, choreography, Web service, annotation

## 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

---

\*Corresponding author. Institut Universitaire de France. E-mail: [lefranc@ligm.igh.cnrs.fr](mailto:lefranc@ligm.igh.cnrs.fr).

Table 1  
The IMGT-ONTOLOGY and IMGT-Choreography layers

	Biologist agents	Computer agents
IMGT-Choreography layer	IMGT biological approaches – IMGT genomics approach – IMGT genetics approach – IMGT structural approach IMGT tool diamonds	Web services and IMGT-SOA
IMGT-ONTOLOGY layer	IMGT Scientific chart	IMGT-ML (XML Schema)

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 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 (see FactsBooks [1,2] for review).

IMGT, the international ImMunoGeneTics information system<sup>®</sup> (<http://imgt.cines.fr>) [3,4], was created in 1989, by Marie-Paule Lefranc, at 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. Fifteen years later, IMGT is the international reference in immunogenetics and immunoinformatics, and provides 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 system (RPI) of any species which belong to the immunoglobulin superfamily (IgSF) and to the MHC superfamily (MhcSF) [4–15].

The IMGT information system consists of four sequence databases, one genome database and one three-dimensional (3D) structure database, interactive tools for sequence, genome and 3D structure analysis, Web resources (“IMGT Marie-Paule page”) comprising 8,000 HTML pages of synthesis (IMGT Repertoire), knowledge (IMGT Scientific chart, IMGT Education, IMGT Index) and external links (IMGT Bloc-notes and IMGT other accesses) [4]. Despite the heterogeneity of these different components, all data in the IMGT information system are expertly annotated. The accuracy, the consistency and the integration of the IMGT data, as well as the coherence between the different IMGT components (databases, tools and Web resources) are based on IMGT-ONTOLOGY, which provides a semantic specification of the terms to be used in immunogenetics and immunoinformatics [16]. IMGT-ONTOLOGY, the first ontology in the domain, has allowed the management of knowledge in immunogenetics [4] and provided standardization for immunogenetics data from genome, proteome, genetics and 3D structures [4–15] (IMGT milestones).

IMGT-ONTOLOGY concepts are available, for the biologists and IMGT users, in the IMGT Scientific chart [8], and have been formalized, for the computing scientists, in IMGT-ML which uses XML (Extensible Markup Language) Schema [17,18]. The IMGT Scientific chart (for biologist agents) and IMGT-ML (for computing agents), which are the foundations of the IMGT knowledge and data management system, constitute the “IMGT-ONTOLOGY layer” (Table 1).

In order to allow any IMGT component to be automatically queried and to achieve a higher level of interoperability inside the IMGT information system and with other information systems, we are implementing, on top of the IMGT-ONTOLOGY layer, the “IMGT-Choreography layer”. It comprises the

Table 2

IMGT-ONTOLOGY main concepts, IMGT Scientific chart rules and examples of IMGT expertised data concepts

IMGT-ONTOLOGY main concepts [16]	IMGT Scientific chart rules [8]	Examples of IMGT expertised data concepts [9,10]
IDENTIFICATION	Standardized keywords [6]	Species, molecule type, receptor type, chain type, gene type, structure, functionality, specificity [6,7]
DESCRIPTION	Standardized labels and annotations [6]	Core (V-, D-, J-, C-REGION) Prototypes [6] Labels for sequences Labels for 2D and 3D structures
CLASSIFICATION	Reference sequences [8] Standardized IG and TR gene nomenclature (group, subgroup, gene, allele) [8,16]	Nomenclature of the human IG and TR genes (entry in 1999 in GDB, HGNC [22] and LocusLink at NCBI) [1,2] Alignment of alleles [1,2,7] Nomenclature of the IG and TR genes of all vertebrate species [8]
NUMEROTATION	IMGT unique numbering [7,23,24] for: V- and V-LIKE-DOMAINS [25] C- and C-LIKE-DOMAINS [26] G- and G-LIKE-DOMAINS [27]	Protein displays [8] IMGT Colliers de Perles [7,8,28] FR-IMGT and CDR-IMGT delimitations [8] Structural loops and beta strands delimitations [26]
ORIENTATION	Orientation of genomic instances relative to each other	Chromosome orientation Locus orientation Gene orientation DNA strand orientation
OBTENTION	Standardized origin Standardized methodology [4]	

modeling of the three major IMGT biological approaches, genomics, genetics and structural approaches, the analysis of the IMGT components in relation with the concepts (depicted for example, for the tools, as IMGT tool diamonds), and the development of Web services (built using the IMGT-SOA infrastructure). It is the first step towards the implementation of IMGT-Choreography [19], which corresponds to the process of complex immunogenetics knowledge [20,21] and to the connection of treatments performed by the IMGT component Web services.

## THE IMGT-ONTOLOGY LAYER

### *IMGT Scientific Chart*

The IMGT Scientific chart [8] comprises the controlled vocabulary and the annotation rules necessary for the immunogenetics data identification, description, classification and numbering and for knowledge management in the IMGT information system. Standardized keywords, labels and annotation rules, standardized IG and TR gene nomenclature, the IMGT unique numbering, and standardized origin/methodology were defined, respectively, based on the six main concepts of IMGT-ONTOLOGY: IDENTIFICATION, CLASSIFICATION, DESCRIPTION, NUMEROTATION, ORIENTATION and OBTENTION [4,16] (Table 2). 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 expertised data concepts [9,10] derived from the IMGT Scientific chart rules are shown in Table 2.



Table 3  
IMGT components of the genomics approach

IMGT genome database [4]	IMGT genome analysis tools [14]	IMGT Repertoire “Locus and genes” section [10,15,40–60] (1)
IMGT/GENE-DB [31]	IMGT/LocusView IMGT/GeneView IMGT/GeneSearch IMGT/CloneSearch IMGT/GeneInfo [39]	Chromosomal localizations [1,2] Locus representations [1,2] Locus description Gene tables, etc. Potential germline repertoires Lists of genes Correspondence between nomenclatures [1,2]

(1)IMGT Web resources (IMGT Marie-Paule page) also include IMGT Index, IMGT Education (Aide-mémoire, Tutorials, Questions and answers, IMGT Lexique, The IMGT Medical page, The IMGT Veterinary page, The IMGT Biotechnology page), IMGT Bloc-notes (The IMGT Immunoinformatics page, Interesting links, etc.) [4,10] which are not detailed in this paper.

### *IMGT Genomics Approach*

The IMGT genomics approach refers to the study of the genes within their loci and on the chromosomes (Table 3). Genomic data are managed in IMGT/GENE-DB, which is the comprehensive IMGT genome database [31]. In September 2004, IMGT/GENE-DB contained 1,375 genes and 2,204 alleles (673 IG and TR genes and 1,028 alleles from *Homo sapiens*, and 702 IG and TR genes and 996 alleles from *Mus musculus*, *Mus cookii*, *Mus pahari*, *Mus spretus*, *Mus saxicola*, *Mus minuto ides*). All the 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 [1,2] were approved by the HUGO Nomenclature Committee HGNC in 1999 [22], and entered in IMGT/GENE-DB [31], Genome DataBase GDB (Canada) [32], LocusLink at NCBI (USA) [33], and GeneCards [34]. Reciprocal links exist between IMGT/GENE-DB, HGNC, GDB, LocusLink and Entrez at NCBI and GeneCards. All the mouse IG and TR gene names with IMGT reference sequences were provided by IMGT to HGNC and to the Mouse Genome Database MGD [35] 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 [8]. IMGT/GENE-DB interacts dynamically with IMGT/LIGM-DB [36–38] to download and display gene-related sequence data. This is the first example of an interaction between IMGT databases using the CLASSIFICATION concept.

The IMGT genome analysis tools (IMGT/LocusView, IMGT/GeneView, IMGT/GeneSearch and IMGT/CloneSearch) manage the locus organization and 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/LocusView allows to view genes in a locus and to zoom on a given area. IMGT/GeneView allows to view a given gene in a locus. IMGT/GeneSearch allows to search for genes in a locus based on IMGT gene names, functionality or localization on the chromosome. 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 and displays information on the human and mouse TR potential rearrangements [39].

### *IMGT Genetics Approach*

The IMGT genetics approach refers to the study of genes in relation with their polymorphisms, mutations, expression, specificity and evolution (Table 4). 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 names) and on the NUMEROTATION concept (IMGT unique numbering [25,26]).

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 [5,36–38]. The IMGT/LIGM-DB annotations (gene and allele name assignment, labels) allow data retrieval not only from IMGT/LIGM-DB, but also from other IMGT databases. As an example, the IMGT/GENE-DB entries provide the IMGT/LIGM-DB accession numbers of the IG and TR cDNA sequences which contain a given V, D, J or C gene. IMGT/PROTEIN-DB is a new IMGT database which provides IG and TR amino acid sequences for productive rearranged cDNA sequences from IMGT/LIGM-DB. Standardized information on oligonucleotides (or Primers) and combinations of primers (Sets, Couples) for IG and TR are managed in IMGT/PRIMER-DB [61], the IMGT oligonucleotide database on the Web since February 2002. IMGT/MHC-DB [62] hosted at EBI comprises IMGT/HLA for human MHC (or HLA) and IMGT/MHC-NHP for MHC of non-human primates.

The IMGT tools for the genetics approach comprise IMGT/V-QUEST [9,63] for the identification of the V, D and J genes and of their mutations, IMGT/JunctionAnalysis [64] 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, and IMGT/Phylogene [65] for gene evolution analyses. IMGT/V-QUEST (V-QUERy and STandardization) is an integrated software for IG and TR [9,63]. 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 delimitations of the FR-IMGT and CDR-IMGT based on the IMGT unique numbering, the protein translation of the input sequence, the identification of the JUNCTION, and the two-dimensional (2D) IMGT Collier de Perles representation of the V-REGION (“IMGT/V-QUEST output” in IMGT/V-QUEST Documentation). IMGT/JunctionAnalysis [64] is a tool, complementary to IMGT/V-QUEST, which provides a thorough analysis of the V-J and V-D-J junction of IG and TR rearranged genes. 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 (“IMGT/JunctionAnalysis output results” in IMGT/JunctionAnalysis Documentation). Several hundreds of junction sequences can be analysed simultaneously. IMGT/Allele-Align allows the comparison of two alleles highlighting the nucleotide and amino acid differences. IMGT/Phylogene [65] is an easy to use tool for phylogenetic analysis of 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 [65] (IMGT/PhyloGene Documentation).

IMGT/Automat [66] is an internal Java tool which implements IMGT/V-QUEST and IMGT/Junction Analysis to automatically perform the annotation of rearranged cDNA sequences in IMGT/LIGM-DB. IMGT/GeneFrequency is a new IMGT interactive tool which dynamically computes cartographic bar-chart representations of the different IG or TR loci. Histograms represent the contribution of individual V, D and J genes in sets of expressed 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. It will be available from the IMGT Home page at <http://imgt.cines.fr> within the year 2005. In a next step, dynamic interactions will be developed with IMGT/GENE-DB and the genome analysis tools, and with the IMGT/PROTEIN-DB and IMGT/PRIMER-DB [61] sequence databases.

Table 4  
IMGT components of the genetics approach

IMGT sequence databases [4]	IMGT sequence analysis tools [14]	IMGT Repertoire “Proteins and alleles” section [10,15,40–60]
IMGT/LIGM-DB [5,36] IMGT/PROTEIN-DB IMGT/PRIMER-DB [61] IMGT/MHC-DB [62]	IMGT/V-QUEST [63] IMGT/JunctionAnalysis [64] IMGT/Allele-Align IMGT/PhyloGene [65] IMGT/Automat [66] IMGT/GeneFrequency	Alignments of alleles IG and TR [1,2,7] Alignments of alleles RPI Protein displays IG and TR [1,2,8,46,53] Protein displays MHC Protein displays RPI Tables of alleles IG and TR Tables of alleles RPI Allotypes [67,68] Isotypes [69], etc.

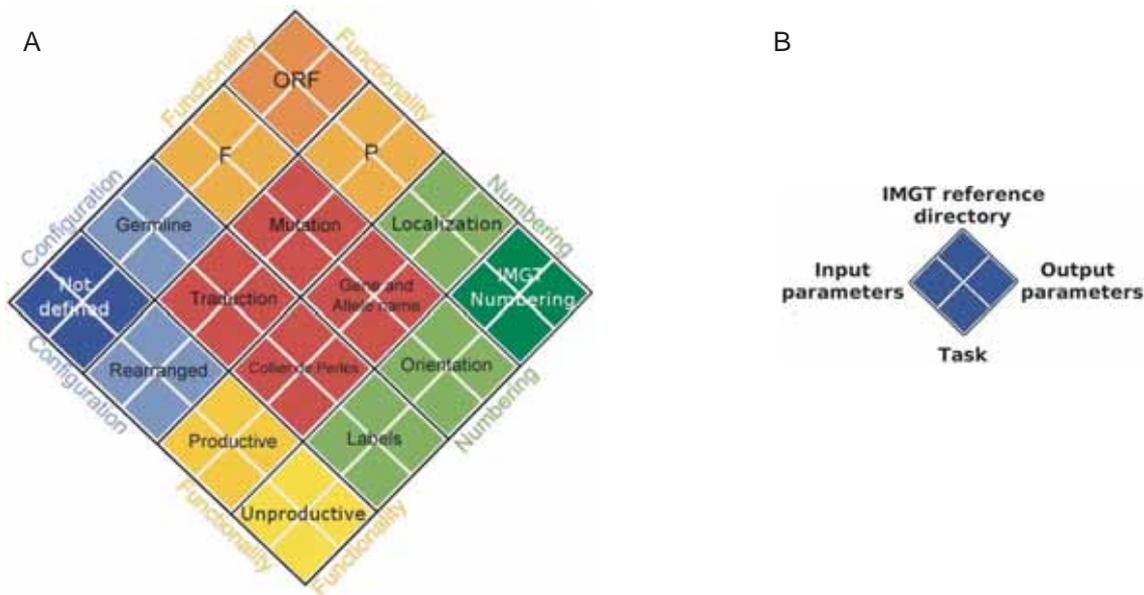


Fig. 2. IMGT tool diamond. A. The 16 modules of an IMGT tool diamond. B. An individual module. (A). Each diamond is constituted of 16 modules related to the IMGT-ONTOLOGY concepts. The four core modules (red) correspond to the major concepts of the tool. The outer modules correspond to the configuration (west pole, blue), to the functionality (north pole, orange, and south pole, yellow) (IDENTIFICATION concept), and to the labels (DESCRIPTION concept), IMGT numbering (NUMEROTATION concept), localization and orientation (ORIENTATION concept) (east pole, green). (B). Each module comprises 4 facets: input parameters, task, IMGT reference directory and output parameters.

*IMGT Structural Approach*

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 (Table 5). 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 [25,26]). 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 [1,2], the constant domain or C-DOMAIN (7-strand

Table 5  
 IMGT components of the structural approach

IMGT structural database [4]	IMGT structural analysis tool [14]	IMGT Repertoire “2D and 3D structures” section [10,15,40–60]
IMGT/3Dstructure-DB [70]	IMGT/StructuralQuery [70]	2D Colliers de Perles [28] IG and TR (1), 2D Colliers de Perles MHC, 2D Colliers de Perles RPI IMGT classes for amino acid characteristics [72] IMGT Colliers de Perles reference profiles [72] 3D representations (1)

(1) Cover of the Nucleic Acids Research 1999 database issue ([http://imgt.cines.fr/textes/IMGTinformation/Couv\\_NAR99.jpg](http://imgt.cines.fr/textes/IMGTinformation/Couv_NAR99.jpg)).

beta-sandwich), and, for the MHC chains, the groove domain or G-DOMAIN (4 beta-strand and one alpha-helix). The IMGT unique numbering, initially defined for the V-DOMAINS of the IG and TR and for the V-LIKE-DOMAINS of IgSF proteins other than IG and TR [25], has been extended to the C-DOMAINS of the IG and TR and to the C-LIKE-DOMAINS of IgSF proteins other than IG and TR [26]. An IMGT unique numbering has also been implemented for the groove domain (G-DOMAIN) of the MHC class I and II chains, and for the G-LIKE-DOMAINS of MhcSF proteins other than MHC [27].

Structural data are compiled and annotated in IMGT/3Dstructure-DB, the IMGT 3D structure database, on the Web since November 2001 [70]. IMGT/3Dstructure-DB comprises IG, TR, MHC and RPI with known 3D structures. Coordinate files extracted from the Protein Data Bank (PDB) [71] are renumbered according to the standardized IMGT unique numbering [25,26]. The IMGT/3Dstructure-DB cards provide IMGT annotations (assignment of IMGT genes and alleles, IMGT chain and domain labels, IMGT Colliers de Perles on one layer and two layers), downloadable renumbered IMGT/3Dstructure-DB flat files, visualization tools and external links. IMGT/3Dstructure-DB residue cards provide detailed information on the inter- and intra-domain contacts of each residue position (IMGT/3Dstructure-DB Documentation).

The IMGT/StructuralQuery tool [70] analyses the intramolecular interactions for the V-DOMAINS. The contacts are described per domain (intra- and inter-domain contacts) and annotated in term of IMGT labels (chains, 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 [25].

In order to appropriately analyse the amino acid resemblances and differences between IG, TR, MHC and RPI chains, eleven IMGT classes were defined for the ‘chemical characteristics’ amino acid properties and used to set up IMGT Colliers de Perles reference profiles [72]. The IMGT Colliers de Perles reference profiles allow to easily compare amino acid properties at each position whatever the domain, the chain, the receptor or the species. The IG and TR variable and constant domains represent a privileged situation for the analysis of amino acid properties in relation with 3D structures, by the conservation of their 3D structure despite divergent amino acid sequences, and by the considerable amount of genomic (IMGT Repertoire), structural (IMGT/3Dstructure-DB) and functional data available. These data are not only useful to study mutations and allele polymorphisms, but are also needed to establish correlations between amino acids in the protein sequences and 3D structures and to determine amino acids potentially involved in the immunogenicity.

Table 6  
 IMGT/V-QUEST output and IMGT/JunctionAnalysis input/output

IMGT-ONTOLOGY concept	IMGT tool diamond module	IMGT/V-QUEST module output parameters IMGT/JunctionAnalysis module input parameters	IMGT/JunctionAnalysis module output parameters
CLASSIFICATION	Gene and allele name	V-GENE, J-GENE and allele name	D-GENE and allele name
NUMEROTATION	IMGT Numbering	2nd-CYS 104, J-PHE or J-TRP 108	CDR3-IMGT length and numbering
DESCRIPTION	Labels	3' V-REGION, 5' J-REGION	P-REGION, N-REGION, D-REGION

### *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” (Fig. 2) were developed to obtain tool profiles and to compare the state of the art of each tool in relation with the IMGT ontological concepts. Each IMGT tool diamond is composed of 16 modules (Fig. 2A) and each module comprises 4 facets: input parameters, task, IMGT reference directory and output parameters (Fig. 2B). For a given module (that is a given concept), each facet acts as a Boolean switch and indicates whether input parameters are necessary or not, whether a task is performed or not, whether an expertised IMGT reference directory is needed or not, and whether output parameters are provided or not, respectively.

The four modules at the core of the IMGT tool (red) diamond correspond to the major concepts of the tool and are usually supported by specific tasks [39,63–65,70]. The 12 outer modules correspond to concepts usually shared with other tools: those of the west pole (blue) correspond to the gene configuration (germline, rearranged or not defined), those of the north pole (orange) to the functionality of the germline sequences (Functional (F), Open Reading Frame (ORF), Pseudogenes (P)), those of the south pole (yellow) to the functionality of the rearranged sequences (productive, unproductive) (IDENTIFICATION concept), and those of the east pole (green) include the labels (DESCRIPTION concept), IMGT unique numbering (NUMEROTATION concept) and the localization and the orientation (ORIENTATION concept) (Fig. 2A).

The IMGT tool diamonds are particularly useful for the IMGT Web service developers, as they allow to control and to enhance the coherence inside and between the IMGT tools in the frame of IMGT-Choreography. Indeed, the comparison of two IMGT tool diamonds allows to identify, among the facets which are “switched on”, those relevant to both tools, and then to analyse the expertised concepts which are involved. Thus, in the example in Fig. 3, three modules that are relevant to both the IMGT/V-QUEST [63] and IMGT/JunctionAnalysis [64] tools were selected for analysis: “Gene and allele name” (core, red), “IMGT numbering” and “Labels” (east pole, green). IMGT/V-QUEST “Gene and allele name” module (V-GENE, J-GENE and allele names) (Table 6) are the necessary input parameters of the IMGT/JunctionAnalysis “Gene and allele name” module (Table 6, same column) that identifies the D-GENE and allele name (IMGT/JunctionAnalysis output). The output parameters of the IMGT/V-QUEST “IMGT numbering” module (identification of 2nd-CYS at position 104, and J-PHE or J-TRP at position 118) are the necessary input parameters of the IMGT/JunctionAnalysis “IMGT numbering” module that characterizes the CDR3-IMGT length and numbering (Table 6). In the same way, the output parameters of the IMGT/V-QUEST “Labels” module (3' V-REGION and 5' J-REGION) are the necessary input parameters of the IMGT/JunctionAnalysis “Labels” module that identifies the P, N and D-REGION (Table 6) [64].

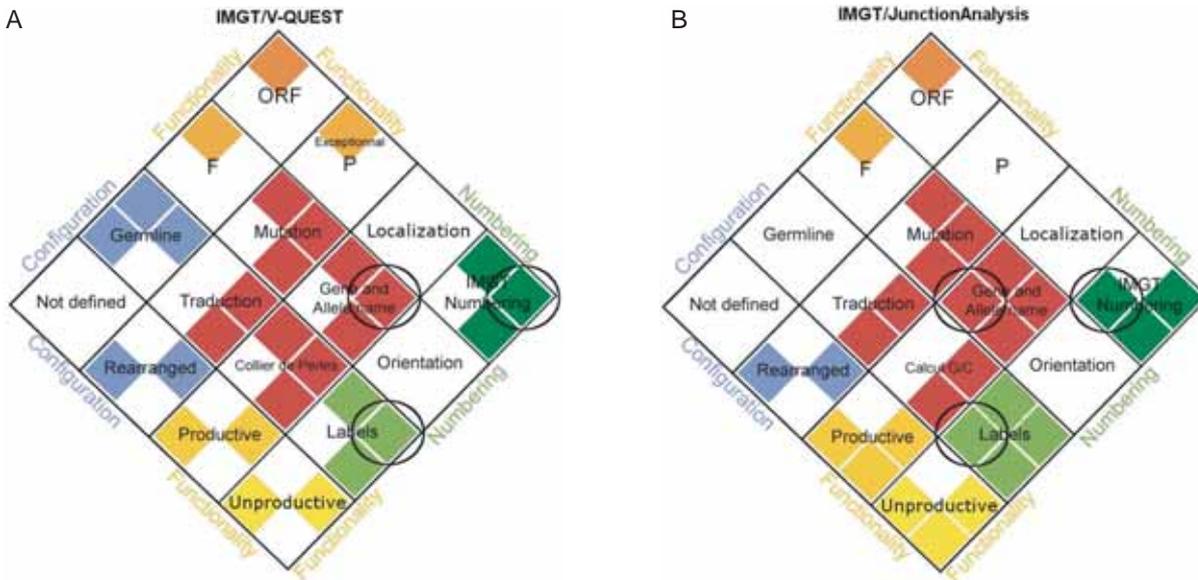


Fig. 3. IMGT tool diamond profiles of the IMGT/V-QUEST [63] and IMGT/JunctionAnalysis [64] sequence analysis tools. A. IMGT/V-QUEST. B. IMGT/JunctionAnalysis. The output parameters of the IMGT/V-QUEST “Gene and allele name” (core, red), “IMGT numbering” and “Labels” (east pole, green) modules (circled in (A)) are the necessary input parameters of the IMGT/JunctionAnalysis “Gene and allele name”, “IMGT numbering” and “Labels” modules, respectively (circled in (B)). Note that, in contrast to IMGT/JunctionAnalysis, IMGT/V-QUEST does not require input parameters for these modules (empty facets).

### Web Services and IMGT-SOA

Web services have been chosen as the mean to create dynamic interactions between IMGT databases and tools. The choice of the Web services to be developed in priority is based on the major existing or potential “conversation nodes” detected in the IMGT biological approaches or with the IMGT tool diamonds.

The Web Service paradigm considers as a service any application accessible over Internet fulfilling the requirements of interoperability, weak-coupling and platform independence between applications by making extensive use of open standards, based for example on XML, and existing networking protocols. Precisely, Service Oriented Architectures (SOA) use the Web Services Description Language (WSDL) (<http://www.w3.org/TR/wsdl>) for the description of new services, the Simple Object Access Protocol (SOAP) (<http://www.w3.org/TR/soap/>) ensures communication between services, and the Universal Description, Discovery and Integration (UDDI) protocol (<http://www.uddi.org/about.html>) enables applications to quickly, easily, and dynamically find and use Web services over the Internet. However, this framework does not specify the underlying semantics of communications. IMGT-SOA introduces a semantic layer by imposing that messages, that are exchanged between service providers and consumers, be 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. IMGT-ML is the unique language used for both services inputs and outputs. Clients and providers for these services can be written using any SOAP-capable programming language (i.e. the 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

**Request :**

```

<querySeqData>
  <seqData>
    <simpleCatalogEntry numacc="M26678" />
  </seqData>
</querySeqData>

```

**Response :**

```

<seqDataList>
  <seqData>
    <catalogEntry id="M26678" >
      <simpleCatalogEntry numacc="M26678" name="MMIGKIZZ" />
    </catalogEntry>
    <identification>
      <partIdent moleculeType="DNA" configuration="rearranged">
        <taxon taxonName="Mus musculus" />
      </partIdent>
    </identification>
    <classification>
      <group name="IGKV">
        <subgroup name="IGKV8" />
      </group>
    </classification>
    ....
  </seqData>
</seqDataList>

```

Fig. 4. IMGT/LIGM-DB Web service request and response example. The IMGT/LIGM-DB Web service uses IMGT-ML compliant XML messages for both request and response. In this example, IMGT/LIGM-DB is queried for a “catalog” entry which has “M26678” as accession identifier. The corresponding response lists the instances of the IMGT-ONTOLOGY concepts (identification, classification, etc.) which are related to that particular entry.

the Apache Axis (<http://ws.apache.org/axis/>) Web services development framework. Apache Axis is an implementation of the SOAP submission to W3C.

The IMGT/LIGM-DB Web service is the first Web service currently developed and implemented with Axis. 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, identified, classified, described according to the IMGT concepts, such as the nucleotide sequence, the description labels, the literature references, the metadata, etc. The querySeqData input has the form of an incomplete IMGT-ML data entry (Fig. 4). The given values are used as criteria to query the database. The result is then 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.

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 [1,2,73–78], of the biologist requests [79–82] and of the clinician needs [83–85] has been undertaken in the three main biological approaches: genomics, genetics and structural approaches. The detailed interactions between IMGT components ([3,86], and this paper) are currently being carefully modelled in UML [87].

## CONCLUSION

The design of IMGT-Choreography and the creation of dynamic interactions between the IMGT databases and tools, using the Web services and IMGT-ML, represent novel and major developments of IMGT, the international reference in immunogenetics and immunoinformatics. IMGT-Choreography enhances the dynamic interactions between the IMGT components to answer complex biological and clinical requests.

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

## CITING IMGT

If you use IMGT databases, tools and/or Web resources, please cite [4] and this paper as references, and quote the IMGT Home page URL address, <http://imgt.cines.fr>.

## ACKNOWLEDGEMENTS

We are grateful to Mehdi Yousfi Monod, Joumana Jabado-Michaloud and Vincent Negre for helpful discussion. We thank our “2004” students Nabil Belkebir, Alain Carnec, Nathalie Clavert, Laurent Douchy, Fadhel Fattoum, Valérie Garelle, Guillaume Gauby, Sandra Ghayad, Bertrand Monnier, Elise Parrod, Erwan Rondeau, Fabrice Sarniguet, Thomas Spiesser and Guillaume Tournour for their motivation. E.D. is the recipient of a doctoral grant from the Ministère de l’Education Nationale, de l’Enseignement Supérieur et de la Recherche (MENESR). K.Q. received a doctoral grant from the MENESR and is currently supported by the Association pour la Recherche sur le Cancer (ARC). O.C. is supported in the frame of the BIOSTIC programme. The ORIEL (Online Research Information Environment for the Life Sciences) project is funded by the European Union IST programme (ST-2001-32688). IMGT is funded by the Centre National de la Recherche Scientifique (CNRS), the MENESR (Université Montpellier II Plan Pluri-Formation, BIOSTIC-LR2004 Région Languedoc-Roussillon and ACI-IMPBIO IMP82-2004).

## 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., Clement, O., Chaume, D. and Lefranc, G. (2005). IMGT, the international ImMunoGeneTics information system®. *Nucleic Acids Res.* **33**, D593-D597.
- [4] 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. (2003). IMGT-ONTOLOGY for Immunogenetics and Immunoinformatics (<http://imgt.cines.fr>). *In Silico Biol.* **4**, 0004.
- [5] Lefranc, M.-P., Giudicelli, V., Busin, C., Malik, A., Mougnot, I., Déhais, P. and Chaume, D. (1995). LIGM-DB/IMGT: an integrated database of Ig and TcR, part of the Immunogenetics database. *Annals of the New York Academy of Sciences* **764**, 47-49.
- [6] Giudicelli, V., Chaume, D., Bodmer, J., Müller, W., Busin, C., Marsh, S., Bontrop, R., Lemaitre, M., Malik, A. and Lefranc, M.-P. (1997). IMGT, the international ImMunoGeneTics database. *Nucleic Acids Res.* **25**, 206-211.
- [7] Lefranc, M.-P., Giudicelli, V., Busin, C., Bodmer, J., Müller, W., Bontrop, R., Lemaitre, M., Malik, A. and Chaume, D. (1998). IMGT, the international ImMunoGeneTics database. *Nucleic Acids Res.* **26**, 297-303.
- [8] 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.
- [9] Ruiz, M., Giudicelli, V., Ginestoux, C., Stoehr, P., Robinson, J., Bodmer, J., Marsh, S. G., Bontrop, R., Lemaitre, M., Lefranc, G., Chaume, D. and Lefranc, M.-P. (2000). IMGT, the international ImMunoGeneTics database. *Nucleic Acids Res.* **28**, 219-221.
- [10] Lefranc, M.-P. (2001). IMGT, the international ImMunoGeneTics database. *Nucleic Acids Res.* **29**, 207-209.
- [11] Lefranc, M.-P. (2002). IMGT, the international ImMunoGeneTics database: a high-quality information system for comparative immunogenetics and immunology. *Dev. Comp. Immunol.* **26**, 697-705.
- [12] Lefranc, M.-P. (2003). IMGT® databases, web resources and tools for immunoglobulin and T cell receptor sequence analysis. *Leukemia* **17**, 260-266.
- [13] Lefranc, M.-P. (2004). IMGT-ONTOLOGY and IMGT databases, tools and web resources for immunogenetics and immunoinformatics. *Molecul. Immunol.* **40**, 647-659.
- [14] Lefranc, M.-P. (2003). IMGT, the international ImMunoGeneTics information system® *In: Antibody Engineering: Methods and Protocols*. 2nd edition, Lo B.K.C. (ed.). *Methods in Molecular Biology*. Humana Press, Totowa, NJ, USA **248**, 27-49.
- [15] Warr, G. W., Clem, L. W. and Söderhall, K. (2003). Editorial: The International ImMunoGeneTics database IMGT. *Dev. Comp. Immunol.* **27**, 1-2.
- [16] Giudicelli, V. and Lefranc, M.-P. (1999). Ontology for Immunogenetics: the IMGT-ONTOLOGY. *Bioinformatics* **12**, 1047-1054.
- [17] 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, 71-75.
- [18] 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*, 22-23.
- [19] Chaume, D., Giudicelli, V., Combres, K., Ginestoux, C. and Lefranc, M.-P. (in press) IMGT-Choreography: processing of complex immunogenetics knowledge. *Computational Methods in Systems Biology CMSB 2004. Lecture Notes in Bioinformatics LNBI*, Springer.
- [20] Giudicelli, V. and Lefranc, M.-P. (2003). IMGT-ONTOLOGY: gestion et découverte de connaissances au sein d'IMGT. *In: Extraction et gestion des connaissances EGC 2003. Extraction des connaissances et apprentissage. Revue des Sciences et Technologies de l'Information RSTI*, Ed. Lavoisier, Series RIA-ECA, **17**, 13-23.
- [21] Lefranc, M.-P., Giudicelli, V., Ginestoux, C. and Chaume, D. (2003). IMGT, the international ImMunoGeneTics information system®, <http://imgt.cines.fr>: the reference in immunoinformatics. *In: Proceeding of the Medical Informatics Europe MIE 2003*, IOS Press, Amsterdam, 74-79.
- [22] 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.
- [23] Lefranc, M.-P. (1997). Unique database numbering system for immunogenetic analysis. *Immunol. Today* **18**, 509.
- [24] Lefranc, M.-P. (1999). The IMGT unique numbering for Immunoglobulins, T cell receptors and Ig-like domains. *The Immunologist* **7**, 132-136.

- [25] 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.
- [26] 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.
- [27] 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, 223-224.
- [28] 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.
- [29] Williams, A. F. and Barclay A. N. (1988). The immunoglobulin superfamily-domains for cell surface recognition. *Annu. Rev. Immunol.* **6**, 381-405.
- [30] Maenaka, K. and Jones E. Y. (1999). MHC superfamily structure and the immune system. *Curr. Opin. Struct. Biol.* **9**, 745-753.
- [31] 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.
- [32] 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.
- [33] Pruitt, K. D. and Maglott, D. R. (2001). RefSeq and LocusLink: NCBI gene-centered resources. *Nucleic Acids Res.* **29**, 137-140.
- [34] 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.
- [35] 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.
- [36] Chaume, D., Giudicelli, V. and Lefranc, M.-P. (2004). IMGT/LIGM-DB. *In: The Molecular Biology Database Collection: 2004 update*, Galperin M.Y. (ed.). *Nucleic Acids Res.* **32**, <http://www3.oup.co.uk/nar/database/summary/504>.
- [37] Giudicelli, V., Chaume, D., Mennessier, G., Althaus, H.-H., Muller, W., Bodmer, J., Malik, A. and Lefranc M.-P. (1998). IMGT, the international ImMunoGeneTics database: a new design for immunogenetics data access. *In: MEDINFO'1998*. Cesnik, B., Mc Gray, A. T. and Scherrer, J. R. (eds.), IOS Press, Amsterdam, 351-355.
- [38] Giudicelli, V., Chaume, D. and Lefranc M.-P. (1998). IMGT/LIGM-DB: A systematized approach for ImMunoGeneTics database coherence and data distribution improvement. *In: Proceedings Sixth International Conference on Intelligent Systems for Molecular Biology ISBM 98*, 59-68.
- [39] Baum, T. P., Pasqual, N., Thuderoz, F., Hierle, V., Chaume, D., Lefranc, M.-P., Jouvin-Marche, E., Marche, P. N. and Demongeot, J. (2004). IMGT/GeneInfo: enhancing V(D)J recombination database accessibility. *Nucleic Acids Res.* **32**, D51-D54.
- [40] Lefranc, M.-P. (1998). IMGT Locus in Focus: a new section of Experimental and Clinical Immunogenetics. *Exp. Clin. Immunogenet.* **15**, 1-7.
- [41] Pallarès, N., Fripiat, J.-P., Giudicelli, V. and Lefranc, M.-P. (1998). The human immunoglobulin lambda variable (IGLV) genes and joining (IGLJ) segments. *Exp. Clin. Immunogenet.* **15**, 8-18.
- [42] Barbié, V. and Lefranc, M.-P. (1998). The human immunoglobulin kappa variable (IGKV) genes and joining (IGKJ) segments. *Exp. Clin. Immunogenet.* **15**, 171-183.
- [43] Martínez, 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.
- [44] Pallarès, N., Lefebvre, S., Contet, V., Mastuda, F. and Lefranc, M.-P. (1999). The human immunoglobulin heavy variable (IGHV) genes. *Exp. Clin. Immunogenet.* **16**, 36-60.
- [45] Ruiz, M., Pallarès, N., Contet, V., Barbié, V. and Lefranc, M.-P. (1999). The human immunoglobulin heavy diversity (IGHD) and joining (IGHJ) segments. *Exp. Clin. Immunogenet.* **16**, 173-184.
- [46] Scaviner, D., Barbié, V., Ruiz, M. and Lefranc, M.-P. (1999). Protein displays of the human immunoglobulin heavy, kappa and lambda variable and joining regions. *Exp. Clin. Immunogenet.* **16**, 234-240.
- [47] Folch, G. and Lefranc, M.-P. (2000). The human T cell Receptor Beta Variable (TRBV) genes. *Exp. Clin. Immunogenet.* **17**, 42-54.
- [48] Scaviner, D. and Lefranc, M.-P. (2000). The human T cell Receptor Alpha Variable (TRAV) Genes. *Exp. Clin. Immunogenet.* **17**, 83-96.
- [49] Scaviner, D. and Lefranc, M.-P. (2000). The human T cell Receptor Alpha joining (TRAJ) Genes. *Exp. Clin. Immunogenet.* **17**, 97-106.

- [50] Folch, G. and Lefranc, M.-P. (2000). The human T cell Receptor Beta Diversity (TRBD) and Beta Joining (TRBJ) Genes. *Exp. Clin. Immunogenet.* **17**, 107-114.
- [51] Artero, S. and Lefranc, M.-P. (2000). The Teleostei Immunoglobulin Heavy IGH Genes. *Exp. Clin. Immunogenet.* **17**, 148-161.
- [52] Artero, S. and Lefranc, M.-P. (2000). The Teleostei Immunoglobulin Light IGL1 and IGL2 V, J and C Genes. *Exp. Clin. Immunogenet.* **17**, 162-172.
- [53] Folch, G., Scaviner, D., Contet, V. and Lefranc, M.-P. (2000). Protein displays of the Human T cell Receptor Alpha, Beta, Gamma and Delta Variable and Joining Regions. *Exp. Clin. Immunogenet.* **17**, 205-215.
- [54] Bosc, N. and Lefranc, M.-P. (2000). The Mouse (*Mus musculus*) T-cell receptor beta variable (TRBV), diversity (TRBD) and joining (TRBJ) genes. *Exp. Clin. Immunogenet.* **17**, 216-228.
- [55] 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.
- [56] Lefranc, M.-P. (2001). Nomenclature of the human immunoglobulin heavy (IGH) genes. *Exp. Clin. Immunogenet.* **18**, 100-116.
- [57] Lefranc, M.-P. (2001). Nomenclature of the human immunoglobulin kappa (IGK) genes. *Exp. Clin. Immunogenet.* **18**, 161-174.
- [58] Lefranc, M.-P. (2001). Nomenclature of the human immunoglobulin lambda (IGL) genes. *Exp. Clin. Immunogenet.* **18**, 242-254.
- [59] Martinez-Jean, 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.
- [60] Bosc, N. and Lefranc, M.-P. (2003). IMGT Locus in Focus: The mouse (*Mus musculus*) T cell receptor alpha (TRA) and delta (TRD) variable genes. *Dev. Comp. Immunol.* **27**, 465-497.
- [61] 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.). *Nucleic Acids Res.* **32**.
- [62] 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.
- [63] 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.
- [64] 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.
- [65] 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.
- [66] 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, 103-104.
- [67] Osipova, L. P. Posukh, O. L., Wiebe, V. P., Miyazaki, T., Matsumoto, H., Lefranc, G. and Lefranc, M.-P. (1999). BamHI-SacI RFLP and Gm analysis of the immunoglobulin IGHG genes in Northern Selkups (West Siberia): new haplotypes with deletion, duplication and triplication. *Human Genetics* **105**, 530-541.
- [68] 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. *European Journal of Human Genetics* **9**, 765-772.
- [69] 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.
- [70] 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.
- [71] 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.
- [72] Pomié, 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.
- [73] Lefranc, M.-P. (2000). Nomenclature of the human T cell Receptor genes. *In: Current Protocols in Immunology*, J. Wiley and Sons, New York, Supplement 40, A.10.1-A.10.23.
- [74] Lefranc, M.-P. (2000). Nomenclature of the human immunoglobulin genes. *In: Current Protocols in Immunology*, J. Wiley and Sons, New York, Supplement 40, A.1P.1-A.1P.37.
- [75] Lefranc, M.-P. (2000). Locus Maps and Genomic Repertoire of the Human T cell Receptor Genes. *The Immunologist* **8/3**, 72-79.
- [76] Lefranc, M.-P. (2000). Locus Maps and Genomic Repertoire of the Human Immunoglobulin Genes. *The Immunologist* **8/3**, 80-88.

- [77] Ruiz, M. and Lefranc, M.-P. (2000). IMGT sequence profile: a standardized visualization for the immunoglobulin and T cell receptor V-REGIONS applicable to other protein alignments. *In: Currents in Computational Molecular Biology, Frontiers Science Series n°30*, Universal Academy Press, Tokyo, pp 126-127.
- [78] Elemento, O., Gascuel, O. and Lefranc, M.-P. (2002). Reconstructing the duplication history of tandemly repeated genes. *Mol. Biol. Evol.* **19**, 278-288.
- [79] Lefranc, M.-P. (2000). IMGT ImMunoGeneTics Database. *International BIOforum* **4**, 98-100.
- [80] Lefranc, M.-P. (2003). IMGT, the international ImMunoGeneTics information system®, <http://imgt.cines.fr>. *In: Immunoinformatics: Bioinformatic strategies for better understanding of immune function*, Bock, G and Goode, J. (eds.), Novartis Foundation Symposium 254, John Wiley and sons Chichester, UK, 126-136, discussion 136-142, 216-222, 250-252.
- [81] Lefranc, M.-P., Giudicelli, V., Ginestoux, C. and Chaume, D. (2003). IMGT, the international ImMunoGenetics information system®, <http://imgt.cines.fr>: the reference in immunoinformatics. *Stud. Health Technol. Inform.* **95**, 74-79.
- [82] Lefranc, M.-P. (2000). Web sites of Interest to Immunologists. *In: Current Protocols in Immunology*, J. Wiley and Sons, New York, Supplement 35, A.1J.1-A.1J.33.
- [83] Charde, 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.
- [84] 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.
- [85] 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.
- [86] Lefranc, M.-P. (2003). IMGT, the international ImMunoGeneTics database®. *Nucleic Acids Res.* **31**, 307-310.
- [87] Cranefield, S. (2001). UML and the Semantic Web. *Proceedings of the International Semantic Web Working Symposium (SWWS)*.