



PERGAMON

Developmental and Comparative Immunology 26 (2002) 697–705

**Developmental
& Comparative
Immunology**

www.elsevier.com/locate/devcompimm

IMGT, the international ImMunoGeneTics database: a high-quality information system for comparative immunogenetics and immunology

Marie-Paule Lefranc^{*,1}

*Laboratoire d'ImmunoGénétique Moléculaire, LIGM, Université Montpellier II, UPR CNRS 1142, IGH,
141 rue de la Cardonille, 34396 Montpellier Cedex 5, France*

Received 4 March 2002; accepted 16 April 2002

Abstract

IMGT, the international ImMunoGeneTics database (<http://imgt.cines.fr>), is a high quality integrated information system specializing in Immunoglobulins (IG), T cell Receptors (TR) and Major Histocompatibility Complex (MHC) of human and other vertebrates, created in 1989, by LIGM, at the Université Montpellier II, CNRS, Montpellier, France. IMGT provides a common access to standardized data, which include nucleotide and protein sequences, oligonucleotide primers, gene maps, genetic polymorphisms, specificities, 2D and 3D structures. IMGT includes several databases (IMGT/LIGM-DB, IMGT/HLA-DB, IMGT/3Dstructure-DB), Web resources ('IMGT Marie-Paule page') which comprise IMGT Scientific Chart, IMGT Repertoire, IMGT Bloc-notes, IMGT Education, IMGT Aide-mémoire and IMGT Index, and interactive tools (IMGT/V-QUEST, IMGT/JunctionAnalysis). These expertly annotated data on the genome, proteome, genetics and structure of the IG, TR and MHC are of high value for comparative genome evolution studies of the adaptative immune response. © 2002 Elsevier Science Ltd. All rights reserved.

Keywords: Bioinformatics; Database; Information system; Web resources; Immunogenetics; Immunology; Immunoglobulin; T cell receptor

1. Introduction

The immune system has evolved to protect individuals against pathogenic viruses, micro-organisms and parasites. It is vital therefore that individual have a normal immune system. Normal immune responses depend on the ability to recognize foreign molecules or antigens on the potential pathogens in order to eliminate the source of the antigen. In the vertebrate

species, the molecules involved in the recognition of antigens comprise the Immunoglobulins (IG) and the T cell Receptors (TR). The molecular synthesis and genetics of the IG and TR chains is particularly complex and unique as it 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 human), nucleotide deletions and insertions at the rearrangement junctions (or N-diversity), and somatic hypermutations in the IG loci (for review [1,2]). The number of potential protein forms of IG and TR is almost unlimited. Owing to the complexity and high

* Tel.: +33-4-99-61-99-65; fax: +33-4-99-61-99-01.

E-mail address: lefranc@ligm.igh.cnrs.fr (M.-P. Lefranc).

¹ IMGT, <http://imgt.cines.fr>.

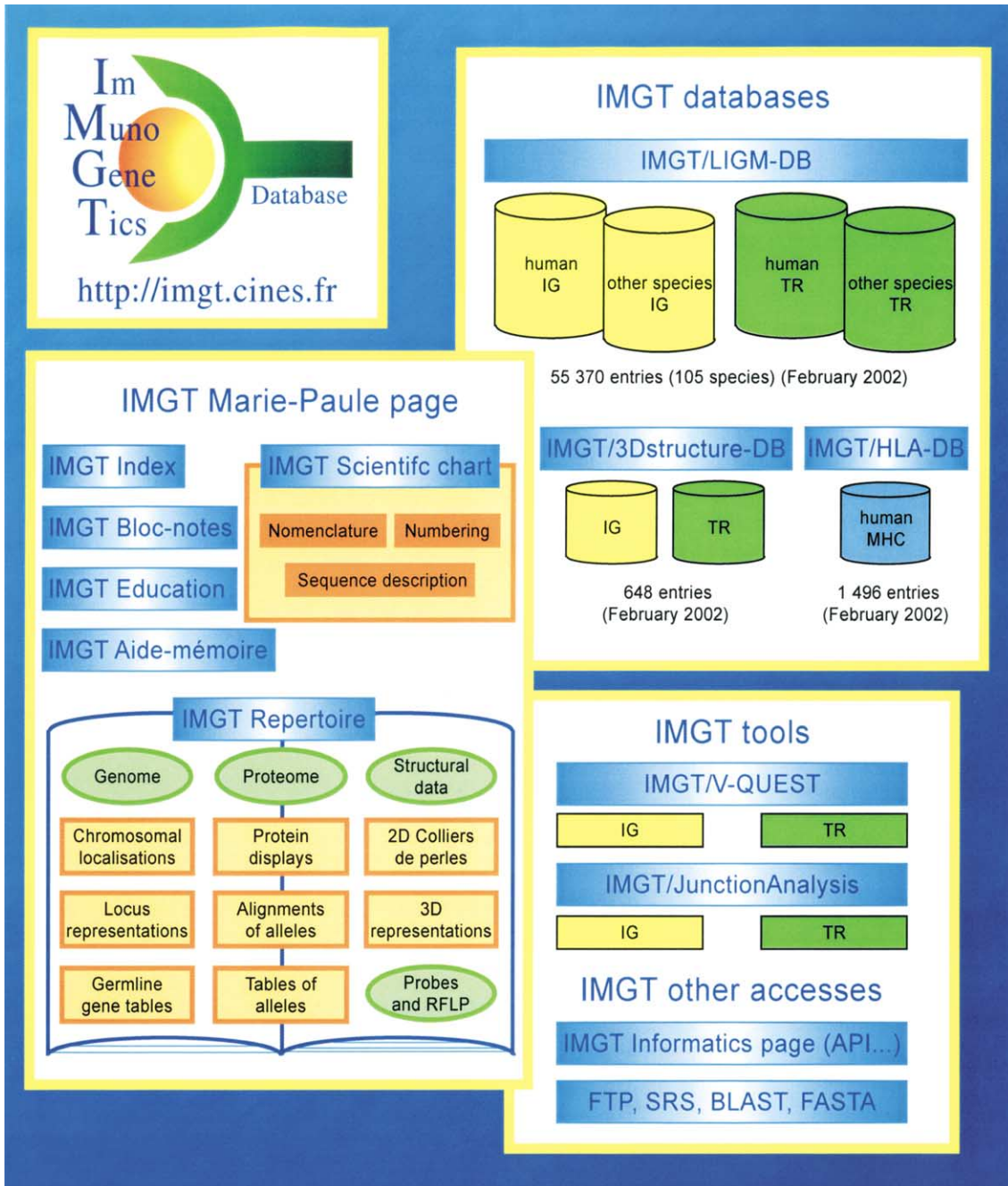


Fig. 1. Overview of the IMGT information system (IMGT, the international ImMunoGeneTics database, <http://imgt.cines.fr>).

number of published sequences, data control and classification and detailed annotations are a very difficult task for the generalist databanks, such as EMBL, GenBank, and DDBJ [3–5].

These observations were the starting point of IMGT, the international ImMunoGeneTics database (<http://imgt.cines.fr>) [6], created in 1989, by LIGM, at the Université Montpellier II, CNRS, Montpellier, France. IMGT is a high quality integrated information system specializing in IG, TR and Major Histocompatibility Complex (MHC) of human and other vertebrates which consists of several databases (IMGT/LIGM-DB, IMGT/HLA-DB, IMGT/3D structure-DB), Web resources (IMGT Marie-Paule page) which comprise IMGT Scientific chart, IMGT Repertoire, IMGT Bloc-notes, IMGT Education, IMGT Aide-mémoire and IMGT Index, and interactive tools (IMGT/V-QUEST, IMGT/Junction Analysis) (Fig. 1). IMGT provides a common access to expertly annotated data on the genome, proteome, genetics and structural data of the IG, TR and MHC (Fig. 1). By its high quality and its easy data distribution, IMGT has important implications in medical research (repertoire in autoimmune diseases, AIDS, leukemias, lymphomas, myelomas), therapeutic approaches, biotechnology related to antibody engineering, veterinary research, genome diversity and genome evolution studies of the adaptive immune response. IMGT is freely available at <http://imgt.cines.fr>.

2. IMGT databases

The IMGT databases comprise three databases (Fig. 1): (i) IMGT/LIGM-DB is a comprehensive database of IG and TR nucleotide sequences from human and other vertebrate species, with translation for fully annotated sequences, created in 1989 by LIGM, Laboratoire d'ImmunoGénétique Moléculaire, Montpellier, France, on the Web since July 1995 [6–10]. In February 2002, IMGT/LIGM-DB contained 55370 nucleotide sequences of IG and TR from 105 species. (ii) IMGT/HLA-DB is a database of the human MHC allele sequences, developed by ICRF, Imperial Cancer Research Fund, Oxford, and ANRI, Anthony Nolan Research Institute, London, UK, on the Web since December 1998 [11]. In February 2002,

IMGT/HLA-DB contained 1496 entries. (iii) IMGT/3Dstructure-DB is a database, which provides the IMGT gene and allele identification and Colliers de Perles of IG and TR with known 3D structures, created by LIGM, on the Web since November 2001. In February 2002, IMGT/3Dstructure-DB contained 648 entries [12]. The two specialized databases IMGT/HLA-DB and IMGT/3Dstructure-DB have been described elsewhere [11,12].

2.1. IMGT/LIGM-DB data

IMGT/LIGM-DB sequence data are identified by the EMBL/GenBank/DDBJ accession number. The unique source of data for IMGT/LIGM-DB is EMBL, which shares data with the other two generalist databases GenBank and DDBJ. Once the sequences are allowed by the authors to be made public, LIGM automatically receives IG and TR sequences by e-mail from EMBL. After control by LIGM curators, data are scanned to store sequences, bibliographical references and taxonomic data, and standardized IMGT/LIGM-DB keywords are assigned to all entries. Based on expert analysis, specific detail annotations are added to IMGT flat files in a second step [7].

Since August 1996, the IMGT/LIGM-DB content closely follows the EMBL one for the IG and TR, with the following advantages: IMGT/LIGM-DB does not contain sequences which have been wrongly assigned earlier to IG and TR; conversely, IMGT/LIGM-DB contains IG and TR entries which have disappeared from the generalist databases (as examples: the L36092 accession number which encompasses the complete human TRB locus is still present in IMGT/LIGM-DB, whereas it has been deleted from EMBL/GenBank/DDBJ due to its too large size (684973 bp); in 1999, IMGT detected the disappearance of 20 IG and TR sequences which inadvertently had been lost by GenBank, and allowed the recuperation of these sequences in the generalist databases).

2.2. IMGT/LIGM-DB interface and data distribution

One of the major objectives of IMGT was to provide the immunologists with a user friendly interface. The Web interface allows searches according to immunogenetic specific criteria and is easy to

use without any knowledge in a computing language. The interface allows the users to get easily connected from any type of platform (PC, Macintosh, workstation) using freeware such as Netscape. All IMGT/LIGM-DB information is available through search criteria [9]: catalogue, accession number, mnemonic, definition, length, etc.; taxonomy, nucleic acid type, loci, genes or chains, functionality, structure, specificity, etc.; keywords; annotation labels; references.

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. They can (i) add new conditions to increase or decrease the number of resulting sequences, (ii) view details concerning the selected sequences and choose among nine possibilities: annotations, IMGT flat file, coding regions with protein translation, catalogue and external references, sequence in dump format, sequence in FASTA format, sequence with three reading frames, EMBL flat file, IMGT/V-QUEST, or (iii) search for sequence fragments corresponding to a particular label [9].

IMGT/LIGM-DB data are also distributed by EBI (distribution of CD-ROM, network fileserver: netserv@ebi.ac.uk, and anonymous FTP server, <ftp://ftp.ebi.ac.uk/pub/databases/imgt/>), by the CINES anonymous FTP server (<ftp://ftp.cines.fr/pub/IMG/>), and from many Sequence Retrieval System (SRS) sites.

IMGT/LIGM-DB can be searched by BLAST or FASTA on different servers (EBI, INFOBIOGEN, Institut Pasteur, etc.).

3. IMGT Web resources

IMGT Web resources (IMGT Marie-Paule page) comprise the following sections: 'IMGT Scientific chart', 'IMGT Repertoire', 'IMGT Bloc-notes', 'IMGT Education', 'IMGT Aide-mémoire' and 'IMGT Index'.

3.1. IMGT Scientific chart

The IMGT Scientific chart provides the controlled vocabulary and the annotation rules and concepts defined by IMGT [13] for the identification, the

description, the classification and the numerotation of the IG and TR data of all vertebrate species.

3.1.1. Concept of identification: standardized keywords

IMGT standardized keywords for IG and TR include the following (i) *general keywords*: indispensable for the sequence assignments, they are described in an exhaustive and non-redundant list, and are organized in a tree structure. (ii) *specific keywords*: they are more specifically associated to particularities of the sequences (orphon, transgene, etc.) or to diseases (leukemia, lymphoma, myeloma, etc.) [7]. The list is not definitive and new specific keywords can easily be added if needed. IMGT/LIGM-DB standardized keywords have been assigned to all entries.

3.1.2. Concept of description: standardized sequence annotation

One hundred and seventy seven feature labels are necessary to describe all structural and functional subregions that compose IG and TR sequences [7], whereas only seven of them are available in EMBL, GenBank or DDBJ. Annotation of sequences with these labels constitutes the main part of the expertise. Levels of annotation have been defined, which allow the users to query sequences in IMGT/LIGM-DB even though they are not fully annotated [7].

Prototypes represent the organizational relationship between labels and give information on the order and expected length (in number of nucleotides) of the labels [7,9]. Prototypes can apply to general configuration of IG or TR, independently of the chain type, the species or any other parameters like functionality. However, prototypes may also be established for very precise cases when sequence characteristics are clearly established.

3.1.3. Concept of classification: standardized IG and TR gene nomenclature

The objective is to provide immunologists and geneticists with a standardized nomenclature per locus and per species which will allow extraction and comparison of data for the complex B and T cell antigen receptor molecules.

The concepts of classification have been used to set up a unique nomenclature of human IG and TR genes,

which was approved by HGNC, the HUGO (Human Genome Organization) Nomenclature Committee in 1999 [6]. The complete list of the human IG and TR gene names [1,2,14–20] has been entered by the IMGT Nomenclature Committee in GDB, Toronto, and LocusLink, NCBI, USA, and is available from the IMGT site [6]. 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 [9,21]. They are listed in the germline gene tables of the IMGT Repertoire [22–29]. The protein displays show translated sequences of the alleles (*01) of the functional or ORF genes [1,2,30,31]. This is currently extended to mouse [32–34], teleostei [35,36] and other species (see ‘Creations and updates’, at the IMGT Home page <http://imgt.cines.fr>).

3.1.4. Concept of numerotation: the IMGT unique numbering

A uniform numbering system for IG and TR sequences of all species has been established to facilitate sequence comparison and cross-referencing between experiments from different laboratories whatever the antigen receptor (IG or TR), the chain type, or the species [37,38].

This numbering results from the analysis of more than 5000 IG and TR variable region sequences of vertebrate species from fish to human. It takes into account and combines the definition of the framework (FR) and complementarity determining region (CDR) [39], structural data from X-ray diffraction studies [40], and the characterization of the hypervariable loops [41]. In the IMGT numbering, conserved amino acids from frameworks always have the same number whatever the IG or TR variable sequence, and whatever the species they come from. As examples: Cysteine 23 (in FR1), Tryptophan 41 (in FR2), Leucine 89 and Cysteine 104 (in FR3). Tables and graphs are available on the IMGT Web site at <http://imgt.cines.fr> and in Refs. [1,2].

This IMGT unique numbering has several advantages:

- It has allowed redefinition of the limits of the FR and CDR of the IG and TR variable domains. The FR-IMGT and CDR-IMGT

lengths become in themselves crucial information, which characterize variable regions belonging to a group, a subgroup and/or a gene.

- Framework amino acids (and codons) located at the same position in different sequences can be compared without requiring sequence alignments. This also holds for amino acids belonging to CDR-IMGT of same length.
- The unique numbering is used as the output of the IMGT/V-QUEST alignment tool. The aligned sequences are displayed according to the IMGT numbering and with the FR-IMGT and CDR-IMGT delimitations.
- The unique numbering has allowed a standardization of the description of mutations and the description of IG and TR allele polymorphisms [1,2]. These mutations and allelic polymorphisms are described by comparison to the IMGT reference sequences of the alleles *01 [8,9].
- The unique numbering allows the description and comparisons of somatic hypermutations of the IG IMGT variable domains.

By facilitating the comparison between sequences and by allowing the description of alleles and mutations, the IMGT unique numbering represents a big step forward in the analysis of the IG and TR sequences of all vertebrate species. Moreover, it gives insight into the structural configuration of the variable domain and opens interesting views on the evolution of these sequences, since this numbering has been applied with success to all the sequences belonging to the V-set of the immunoglobulin superfamily, including non-rearranging sequences in vertebrates (human CD4, *Xenopus* CTXg1, etc.) and in invertebrates (*drosophila* amalgam, *drosophila* fasciclin II, etc.) [8,9,37,38].

3.2. IMGT Repertoire

IMGT Repertoire is the global Web Resource in ImMunoGeneTics for the IG and TR of human and other vertebrates, based on the IMGT Scientific chart. IMGT Repertoire provides an easy-to-use interface to carefully and expertly annotated data on the genome, proteome, polymorphism and structural data of the IG and TR [6] (Fig. 1). Only titles of this large section are quoted here. *Genome* data include chromosomal

localizations, locus representations, locus description, germline gene tables, potential germline repertoires, lists of IG and TR genes and links between IMGT, HUGO, GDB, LocusLink and OMIM, correspondence between nomenclatures [1,2]. *Proteome* and *polymorphism* data are represented by protein displays, alignments of alleles, tables of alleles, allotypes, particularities in protein designations, IMGT reference directory in FASTA format, correspondence between IG and TR chain and receptor IMGT designations [1,2]. *Structural data* comprise 2D graphical representations designated as Colliers de Perles [1,2,6,8,9], FR-IMGT and CDR-IMGT lengths, and 3D representations of IG and TR variable domains [10,12]. This visualization permits rapid correlation between protein sequences and 3D data retrieved from the Protein Data Bank (PDB). Other data comprise: (i) phages, (ii) probes used for the analysis of IG and TR gene rearrangements and expression, and Restriction Fragment Length Polymorphism (RFLP) studies, (iii) data related to gene regulation and expression: promoters, primers, cDNAs, reagent monoclonal antibodies, etc. (iv) genes and clinical entities: translocations and inversions, humanized antibodies, monoclonal antibodies with clinical indications, (v) taxonomy of vertebrate species present in IMGT/LIGM-DB, (vi) immunoglobulin superfamily: gene exon–intron organization, protein displays, Colliers de Perles and 3D representations of V-LIKE and C-LIKE domains.

3.3. IMGT Bloc-notes

The IMGT Bloc-notes provides numerous hyperlinks towards the Web servers specializing in immunology, genetics, molecular biology and bioinformatics (associations, collections, companies, databases, immunology themes, journals, molecular biology servers, resources, societies, tools, etc.) [42].

3.4. IMGT Education

IMGT Education is a new section which provides useful biological resources for students. It includes figures and tutorials (in English and/or in French) on the IG and TR variable and constant domain 3D structures, the molecular genetics of IG, the regulation

of IG gene transcription, B cell differentiation and activation, etc.

3.5. IMGT Aide-mémoire and IMGT index

IMGT Aide-mémoire provides an easy access to information such as genetic code, splicing sites, amino acid structures, restriction enzyme sites, etc.

IMGT Index is a fast way to access data when information has to be retrieved from different parts of the IMGT site. For example, ‘allele’ provides links to the IMGT Scientific chart rules for the allele description, and to the IMGT Repertoire Alignments of alleles and Tables of alleles.

4. IMGT Interactive tools

4.1. IMGT/V-QUEST

IMGT/V-QUEST (V-QUERy and STandardization) is an integrated software for IG and TR [6]. This tool, easy to use, analyses an input IG or TR variable nucleotide sequence and provides the nucleotide alignment by comparison with 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 Collier de Perles representation of the V-REGION.

IMGT/V-QUEST is particularly useful for the analysis of the rearranged variable genes: it allows to identify the V-GENE and J-GENE and alleles involved in the IG and TR rearrangements, and to delimit the JUNCTION. Searches can be done related to IG and to TR of human, mouse and teleostei. The set of sequences used for IMGT/V-QUEST can be downloaded in FASTA format from the IMGT site.

IMGT/V-QUEST can also be used for the analysis of functional or ORF germline variable genes, from other species, to delimit the FR-IMGT and CDR-IMGT, provided that the similarity with human, mouse or teleostei sequences is sufficiently high.

4.2. IMGT/JunctionAnalysis

IMGT/JunctionAnalysis 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-GENE and allele involved in the IGH, TRB and TRD V–D–J rearrangements by comparison with the IMGT reference directory, and delimits precisely the P, N and D regions [1,2]. Results from IMGT/JunctionAnalysis are more accurate than those given by IMGT/V-QUEST regarding the D-GENE identification. Indeed, IMGT/JunctionAnalysis works on shorter sequences (JUNCTION), and with a higher constraint since the identification of the V-GENE and J-GENE and alleles is a prerequisite to perform the analysis. Several hundreds of junction sequences can be analyzed simultaneously.

5. IMGT-ONTOLOGY and IMGT interoperability

5.1. IMGT-ONTOLOGY

IMGT distributes high quality data with an important incremental value added by the IMGT expert annotations, according to the rules described in the IMGT Scientific chart. IMGT has developed a formal specification of the terms to be used in the domain of immunogenetics and bioinformatics to ensure accuracy, consistency and coherence in IMGT. This has been the basis of the IMGT-ONTOLOGY [13], the first ontology in the domain, which allows the management of the immunogenetics knowledge for all vertebrate species. Control of coherence in IMGT combines data integrity control and biological data evaluation [43,44].

5.2. IMGT interoperability

Since July 1995, IMGT has been available on the Web at <http://imgt.cines.fr>. IMGT provides the biologists with an easy to use and friendly interface. From January 1996 to February 2002, the IMGT WWW server at Montpellier was accessed by more than 156,000 sites. IMGT has an exceptional response with more than 100,000 requests a month. Two thirds of the visitors are equally distributed between the European Union and the United States. To facilitate the integration of IMGT data into applications

developed by other laboratories, we have built an Application Programming Interface to access the database and its software tools (see ‘IMGT Informatics page (API...)’) (Fig. 1) [43]. This API includes: a set of URL links to access biological knowledge data (keywords, labels, functionalities, list of gene names, etc.), a set of URL links to access all data related to one given sequence, a set of JAVA™ class packages to select and retrieve data from an appropriate IMGT server using an Object Oriented approach.

5.3. Internal cross-references and data integrity

IMGT contains sequences in different biological forms (e.g. alleles, germline, rearranged or cDNA). These sequences of different biological states are cross-referenced to each other [8]. Accession numbers of associated chains belonging to a same receptor are also cross-referenced to each other as ‘partner’ sequences.

Control of data coherence or data integrity has been introduced step by step in IMGT/LIGM-DB according to the semantic evolution of the data model [8]. Data integrity is important for the database management especially when data input is from more than one site (the core data come from EMBL, the annotations from the authors or from LIGM). It is checked, for example, that each entry is associated to IMGT standardized keywords, and that there is coherence between keywords of an annotated entry and the labels of its subregions. This step is essential to maintain the quality of the database since biological knowledge is always improving. When new rules for describing the data appear, the existing rules are updated accordingly. This allows new entries to be correctly annotated. For the backlog, we implement automatic procedures that regularly verify the coherence of the data and select pools of sequences to be updated. This avoids the introduction of data discrepancies [8].

6. Conclusion

The information provided by IMGT is of much value to clinicians and biological scientists in general. Tools for the analysis of genetic and phylogenetic data (IMGT/PhyloGene) and the display of physical maps

(IMGT/GeneView, IMGT/LocusView) and new specific databases (IMGT/PROTEIN-DB, IMGT/MHC-NHP, IMGT/PRIMER-DB) are currently in development and will be integrated into IMGT. IMGT/PROTEIN-DB, a protein database for IG and TR, will contain translations of potentially functional and ORF sequences from IMGT/LIGM-DB, and protein data from Kabat [39] and PDB. IMGT/MHC-NHP is a database of non-human primate MHC developed by BPRC (Rijswijk, The Netherlands) and IMGT/PRIMER-DB is an oligonucleotide primer database for IG, TR, and MHC, developed in collaboration with EUROGENTEC S.A. (Seraing, Belgium). IMGT is designed to allow a common access to all immunogenetics data, and a particular attention is given to the establishment of cross-referencing links to other databases pertinent to the users of IMGT.

7. Citing IMGT

Authors who make use of the information provided by IMGT should cite [6] as a general reference for the access to and content of IMGT, and quote the IMGT home page URL, <http://imgt.cines.fr>.

Acknowledgments

I thank Véronique Giudicelli, Denys Chaume and Gérard Lefranc for helpful discussion, Chantal Ginestoux for the figure and Nora Bonnet for typing the manuscript. I am deeply grateful to the IMGT team for its expertise and constant motivation. IMGT is funded by the European Union's Fifth PCRDT programme (QLG2-2000-01287), the CNRS (Centre National de la Recherche Scientifique), the Ministère de l'Éducation Nationale, and the Ministère de la Recherche. Subventions have been received from ARC (Association pour la Recherche sur le Cancer), and the Région Languedoc-Roussillon.

References

- [1] Lefranc M-P, Lefranc G. The immunoglobulin FactsBook. London: Academic Press, 2001. 458p, ISBN: 012441351X.
- [2] Lefranc M-P, Lefranc G. The T cell receptor FactsBook. London: Academic Press, 2001. 398p, ISBN: 0124413528.
- [3] Stoesser G, Baker W, Van den Broek A, Camon E, Garcia-Pastor M, Kanz C, Kulikova T, Lombard V, Lopez R, Parkinson H, Redaschi N, Sterk P, Stoehr P, Tuli M-A. The EMBL nucleotide sequence database. *Nucl Acids Res* 2001; 29:17–21.
- [4] Benson DA, Karsch-Mizrachi I, Lipman D, Ostell J, Rapp BA, Wheeler DL. GenBank *Nucl Acids Res* 2000;15–18.
- [5] Tatenno Y, Miyazaki S, Ota M, Sugawara H, Gojobori T. DNA Data Bank of Japan (DDBJ) in collaboration with mass sequencing teams. *Nucl Acids Res* 2000;28:24–6.
- [6] Lefranc M-P. IMGT, the international ImMunoGeneTics database. *Nucl Acids Res* 2001;29:207–9.
- [7] Giudicelli V, Chaume D, Bodmer J, Müller W, Busin C, Marsh S, Bontrop R, Lemaitre M, Malik A, Lefranc M-P. IMGT, the international ImMunoGeneTics database. *Nucl Acids Res* 1997;25:206–11.
- [8] Lefranc M-P, Giudicelli V, Busin C, Bodmer J, Müller W, Bontrop R, Lemaitre M, Malik A, Chaume D. IMGT, the international ImMunoGeneTics database. *Nucl Acids Res* 1998;26:297–303.
- [9] Lefranc M-P, Giudicelli V, Ginestoux C, Bodmer J, Müller W, Bontrop R, Lemaitre M, Malik A, Barbié V, Chaume D. IMGT, the international ImMunoGeneTics database. *Nucl Acids Res* 1999;27:209–12.
- [10] Ruiz M, Giudicelli V, Ginestoux C, Stoehr P, Robinson J, Bodmer J, Marsh SG, Bontrop R, Lemaitre M, Lefranc G, Chaume D, Lefranc M-P. IMGT, the international ImMunoGeneTics database. *Nucl Acids Res* 2000;28:219–21.
- [11] Robinson J, Malik A, Parham P, Bodmer JG, Marsh SGE. IMGT/HLA Database—a sequence database for the human major histocompatibility complex. *Tissue Antigens* 2000;55: 280–7.
- [12] Ruiz M, Lefranc M-P. IMGT gene identification and Colliers de Perles of human immunoglobulin with known 3D structures. *Immunogenetics* DOI 10.1007/s00251-001-0408-6. *Immunogenetics* 2002;53:857–83.
- [13] Giudicelli V, Lefranc M-P. Ontology for Immunogenetics: IMGT-ONTOLOGY. *Bioinformatics* 1999;12:1047–54.
- [14] Lefranc M-P. Nomenclature of the human immunoglobulin genes. *Current Protocols in Immunology*, Suppl. 40. New York: Wiley; 2000. p. A1P1–A1P37.
- [15] Lefranc M-P. Nomenclature of the human T cell receptor genes. *Current Protocols in Immunology*, Suppl. 40. New York: Wiley; 2000. p. A1O1–A1O23.
- [16] Lefranc M-P. Locus maps and genomic repertoire of the human Ig genes. *Immunologist* 2000;8:80–7.
- [17] Lefranc M-P. Locus maps and genomic repertoire of the human T-cell receptor genes. *Immunologist* 2000;8:72–9.
- [18] Lefranc M-P. Nomenclature of the human immunoglobulin heavy (IGH) genes. *Exp Clin Immunogenet* 2001;18:100–16.
- [19] Lefranc M-P. Nomenclature of the human immunoglobulin kappa (IGK) genes. *Exp Clin Immunogenet* 2001;18:161–74.
- [20] Lefranc M-P. Nomenclature of the human immunoglobulin lambda (IGL) genes. *Exp Clin Immunogenet* 2002;18: 242–54.

- [21] Lefranc M-P. IGHV (ImMunoGeneTics) locus on focus. A new section of experimental and clinical immunogenetics. *Exp Clin Immunogenet* 1998;15:1–7.
- [22] Pallarès N, Fripiat JP, Giudicelli V, Lefranc M-P. The human immunoglobulin lambda variable (IGLV) genes and joining (IGLJ) segments. *Exp Clin Immunogenet* 1998;15:8–18.
- [23] Barbié V, Lefranc M-P. The human immunoglobulin kappa variable (IGKV) genes and joining (IGKJ) segments. *Exp Clin Immunogenet* 1998;15:171–83.
- [24] Pallarès N, Lefebvre S, Contet V, Matsuda F, Lefranc M-P. The human immunoglobulin heavy variable (IGHV) genes. *Exp Clin Immunogenet* 1999;16:36–60.
- [25] Ruiz M, Pallarès N, Contet V, Barbié V, Lefranc M-P. The human immunoglobulin heavy diversity (IGHD) and joining (IGHJ) segments. *Exp Clin Immunogenet* 1999;16:173–84.
- [26] Folch G, Lefranc M-P. The human T cell receptor beta variable (TRBV) genes. *Exp Clin Immunogenet* 2000;17:42–54.
- [27] Scaviner D, Lefranc M-P. The human T cell receptor alpha variable (TRAV) genes. *Exp Clin Immunogenet* 2000;17:83–96.
- [28] Scaviner D, Lefranc M-P. The human T cell receptor alpha joining (TRAJ) genes. *Exp Clin Immunogenet* 2000;17:97–106.
- [29] Folch G, Lefranc M-P. The human T cell receptor beta diversity (TRBD) and beta joining (TRBJ) genes. *Exp Clin Immunogenet* 2000;17:107–14.
- [30] Scaviner D, Barbié V, Ruiz M, Lefranc M-P. Protein displays of the human immunoglobulin heavy, kappa and lambda variable and joining regions. *Exp Clin Immunogenet* 1999;16:234–40.
- [31] Folch G, Scaviner D, Contet V, Lefranc M-P. Protein displays of the human T cell receptor alpha, beta, gamma and delta variable and joining regions. *Exp Clin Immunogenet* 2000;17:205–15.
- [32] Martínez C, Lefranc M-P. The mouse (*Mus musculus*) immunoglobulin kappa variable (IGKV) genes and joining (IGKJ) segments. *Exp Clin Immunogenet* 1998;15:184–93.
- [33] Bosc N, Lefranc M-P. The mouse (*Mus musculus*) T cell receptor beta variable (TRBV), diversity (TRBD), and joining (TRBJ) genes. *Exp Clin Immunogenet* 2000;17:216–28.
- [34] Bosc N, Contet V, Lefranc M-P. The mouse (*Mus musculus*) T cell receptor delta variable (TRDV), diversity (TRDD), and joining (TRDJ) Genes. *Exp Clin Immunogenet* 2001;18:51–8.
- [35] Artero S, Lefranc M-P. The Teleostei immunoglobulin heavy IGH genes. *Exp Clin Immunogenet* 2000;17:148–61.
- [36] Artero S, Lefranc M-P. The teleostei immunoglobulin light IGL1 and IGL2 V, J and C genes. *Exp Clin Immunogenet* 2000;17:162–72.
- [37] Lefranc M-P. Unique database numbering system for immunogenetic analysis. *Immunol Today* 1997;18:509.
- [38] Lefranc M-P. The IGHV unique numbering for Immunoglobulins, T cell receptors and Ig-like domains. *Immunologist* 1999;7:132–6.
- [39] Kabat EA, Wu TT, Perry HM, Gottesman KS, Foeller C. Sequences of proteins of immunological interest. Washington DC: National Institute of Health Publications; 1991. p. 91–3242.
- [40] Satow Y, Cohen GH, Padlan EA, Davies DR. Phosphocholine binding immunoglobulin Fab McPC603. *J Mol Biol* 1986;190:593–604.
- [41] Chothia C, Lesk AM. Canonical structures for the hypervariable regions of immunoglobulins. *J Mol Biol* 1987;196:901–17.
- [42] Lefranc M-P. Web sites of interest to immunologists. *Current protocols in immunology*, New York: Wiley, 2000. p. A1J1–A1J33.
- [43] Giudicelli V, Chaume D, Lefranc M-P. IGHV/LIGM-DB: a systematized approach for ImMunoGeneTics database coherence and data distribution improvement. *Proceedings of the Sixth International Conference on Intelligent Systems for Molecular Biology, ISBM-98*; 1998. p. 59–68.
- [44] Giudicelli V, Chaume D, Mennessier G, Althaus HH, Müller W, Bodmer J, Malik A, Lefranc M-P. IGHV, the international ImMunoGeneTics database: a new design for immunogenetics data access. In: Cesnik B, editor. *Proceedings of the Ninth World Congress on Medical Informatics, MEDINFO'98*, Amsterdam: IOS Press, 1998. p. 351–5.