

IMGT(R), the international ImMunoGeneTics information system(R).

Marie-Paule Lefranc, Véronique Giudicelli, Chantal Ginestoux, Joumana Jabado-Michaloud, Géraldine Folch, Fatena Bellahcene, Yan Wu, Elodie Gemrot, Xavier Brochet, Jérôme Lane, et al.

▶ **To cite this version:**

Marie-Paule Lefranc, Véronique Giudicelli, Chantal Ginestoux, Joumana Jabado-Michaloud, Géraldine Folch, et al.. IMGT(R), the international ImMunoGeneTics information system(R).. Nucleic Acids Research, 2008, 37 (suppl. 1), pp.D1006-D1012. 10.1093/nar/gkn838 . hal-00357158

HAL Id: hal-00357158
https://hal.science/hal-00357158

Submitted on 1 Jun 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

D1006–D1012 *Nucleic Acids Research, 2009, Vol. 37, Database issue* *Published online 31 October 2008*
doi:10.1093/nar/gkn838

IMGT®, the international ImMunoGeneTics information system®

Marie-Paule Lefranc*, Véronique Giudicelli, Chantal Ginestoux, Joumana Jabado-Michaloud, Géraldine Folch, Fatena Bellahcene, Yan Wu, Elodie Gemrot, Xavier Brochet, Jérôme Lane, Laetitia Regnier, François Ehrenmann, Gérard Lefranc and Patrice Duroux

IMGT®, the international ImMunoGeneTics information system®, Université Montpellier 2, Laboratoire d'ImmunoGénétique Moléculaire LIGM, UPR CNRS 1142, Institut de Génétique Humaine IGH, 141 rue de la Cardonille, 34396 Montpellier Cedex 5, France

Received September 13, 2008; Accepted October 14, 2008

ABSTRACT

IMGT®, the international ImMunoGeneTics information system® (http://www.imgt.org), was created in 1989 by Marie-Paule Lefranc, Laboratoire d'ImmunoGénétique Moléculaire LIGM (Université Montpellier 2 and CNRS) at Montpellier, France, in order to standardize and manage the complexity of immunogenetics data. The building of a unique ontology, IMGT-ONTOLOGY, has made IMGT® the global reference in immunogenetics and immunoinformatics. IMGT® is a high-quality integrated knowledge resource specialized in the immunoglobulins or antibodies, T cell receptors, major histocompatibility complex, of human and other vertebrate species, proteins of the IgSF and MhcSF, and related proteins of the immune systems of any species. IMGT® provides a common access to standardized data from genome, proteome, genetics and 3D structures. IMGT® consists of five databases (IMGT/LIGM-DB, IMGT/GENE-DB, IMGT/3Dstructure-DB, etc.), fifteen interactive online tools for sequence, genome and 3D structure analysis, and more than 10 000 HTML pages of synthesis and knowledge. IMGT® is used in medical research (autoimmune diseases, infectious diseases, AIDS, leukemias, lymphomas and myelomas), veterinary research, biotechnology related to antibody engineering (phage displays, combinatorial libraries, chimeric, humanized and human antibodies), diagnostics (clonalities, detection and follow-up of residual diseases) and therapeutical approaches (graft, immunotherapy, vaccinology). IMGT is freely available at http://www.imgt.org.

INTRODUCTION

The number of genomics, genetics, 3D and functional data published in the immunogenetics field is growing exponentially and involves fundamental, clinical, veterinary, and pharmaceutical research. 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 (1,2).

IMGT®, the international ImMunoGeneTics information system® (http://www.imgt.org) (3), was created in 1989 by Marie-Paule Lefranc, Laboratoire d'ImmunoGénétique Moléculaire LIGM (Université Montpellier 2 and CNRS) at Montpellier, France, in order to standardize and manage the complexity of immunogenetics data. IMGT® has reached that goal through the building of a unique ontology, IMGT-ONTOLOGY (4), the first ontology in immunogenetics and immunoinformatics. IMGT-ONTOLOGY has allowed the setting up of the official nomenclature of the IG and TR genes and alleles (5,6), the definition of IMGT standardized labels, and the IMGT unique numbering that bridges the gap between sequences and 3D structures for the variable (V) and constant (C) domains of the IG and TR (7–10) and for the groove (G) domains of the major histocompatibility

*To whom correspondence should be addressed. Tel: +33 4 99 61 99 65; Fax: +33 4 99 61 99 01; Email: marie-paule.lefranc@igh.cnrs.fr

© 2008 The Author(s)
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/2.0/uk/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from https://academic.oup.com/nar/article/37/suppl_1/D1006/1014137 by biuupm Bibliotheque Universitaire UPM user on 01 June 2022

complex (MHC) (11). IMGT® is recognized as the global reference that provides the standards in immunogenetics and immunoinformatics. IMGT® is a high-quality integrated knowledge resource, specialized in the IG, TR, MHC of human and other vertebrates, the proteins that belong to the immunoglobulin superfamily (IgSF) and to the MHC superfamily (MhcSF), and the related proteins of the immune systems (RPI) of any species. IMGT® provides a common access to standardized data from genome, proteome, genetics and 3D structures.

The IMGT® information system consists of databases, tools and Web resources (3). IMGT® databases include one genome database, several sequence databases and one 3D structure database. Fifteen IMGT® interactive online tools are provided for genome, sequence and 3D structure analysis. IMGT® Web resources comprise more than 10 000 HTML pages of synthesis and knowledge (IMGT Scientific chart, IMGT Repertoire, The IMGT Medical page, The IMGT Veterinary page, The IMGT Biotechnology page, IMGT Education, IMGT Lexique, IMGT Aide-Mémoire, Tutorials, IMGT Index), external links (IMGT Bloc-notes, The IMGT Immunoinformatics page) and IMGT other accesses (SRS, MRS). Despite the heterogeneity of these different components, all data in IMGT® 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 the IMGT-ONTOLOGY axioms and concepts (4,12).

IMGT-ONTOLOGY

Formal IMGT-ONTOLOGY

The Formal IMGT-ONTOLOGY, also designated as IMGT Kaleidoscope (12), comprises seven axioms: IDENTIFICATION, DESCRIPTION, CLASSIFICATION, NUMEROTATION, ORIENTATION, LOCALIZATION and OBTENTION that postulate that objects, processes and relations have to be identified, described, classified, numerotated, localized, orientated, and that the way they are obtained has to be determined. IMGT-ONTOLOGY concepts derived from these axioms are available, for the biologists and IMGT® users, in the IMGT Scientific chart, and have been formalized, for the computing scientists, in IMGT-ML which is an XML Schema (http://www.w3.org/TR/xmlschema-0/). In order to formalize the semantic relations between concepts and instances that are essential for high-quality data processing and coherence control, IMGT-ONTOLOGY is currently designed with Protégé (13) and OBO-Edit (http://oboedit.org/).

IMGT Scientific chart

The IMGT Scientific chart is constituted by controlled vocabulary and annotation rules for data and knowledge management of the IG, TR, MHC, IgSF, MhcSF and RPI. All IMGT® data are expertly annotated according to the IMGT Scientific chart rules.

Keywords and labels. IMGT standardized keywords (concepts of identification) are assigned to all entries in the IMGT® databases. More than 500 IMGT standardized labels (concepts of description) were necessary to describe all structural and functional subregions that compose IG and TR (221 labels for sequences and 285 for 3D structures). Interestingly, 64 IMGT specific labels defined for nucleotide sequences have been entered in the newly created Sequence Ontology (SO) (14).

Gene and allele nomenclature. All the human IMGT standardized gene names (5,6) (concepts of classification) were approved by the Human Genome Organisation (HUGO) Nomenclature Committee (HGNC) in 1999 (15), and entered in IMGT/GENE-DB (16), and in Entrez Gene NCBI (17), and more recently on the Ensembl server (18) at the European Bioinformatics Institute (EBI) in 2006, and in the Vega (19) database at the Wellcome Trust Sanger Institute in 2008. All the mouse IMGT® gene and allele names and the corresponding IMGT reference sequences were provided to HGNC and to the Mouse Genome Informatics Mouse Genome Database (20) in July 2002 and were presented by IMGT® at the 19th International Mouse Genome Conference, IMGC 2005, in Strasbourg, France, and entered in IMGT/GENE-DB. IMGT reference sequences have been defined for each allele of each gene based on one or, whenever possible, several of the following criteria: germline sequence, first sequence published, longest sequence, mapped sequence.

IMGT unique numbering. The IMGT unique numbering (concepts of numerotation) (7–11) is, with its 2D graphical representation or IMGT Collier de Perles (21,22), the flagship of IMGT® that allows to bridge the gap between sequences, genes and 3D structures in the IMGT® databases, tools and Web resources (23). Structural and functional domains of the IG and TR chains comprise the V-DOMAIN (9-strand β-sandwich) which corresponds to the V-J-REGION or V-D-J-REGION and is encoded by two or three genes (1,2), and the constant domain or C-DOMAIN (7-strand β-sandwich). The IMGT unique numbering initially defined for the IG and TR domains has been extended to the V-LIKE-DOMAIN and C-LIKE-DOMAIN of IgSF proteins other than IG and TR (9,10,22). The IMGT unique numbering for the MHC G-DOMAIN (four β-strand and one α-helix) has been extended to the G-LIKE-DOMAIN of MhcSF proteins other than MHC (11,22).

IMGT Choreography

In order to extract knowledge from IMGT® standardized immunogenetics data, three main IMGT® biological approaches have been developed: genomic, genetic and structural approaches (Figure 1). The IMGT® genomic approach is gene-centered and mainly orientated towards the study of the genes within their loci and on the chromosomes. The IMGT® genetic approach refers to the study of the genes in relation with their sequence polymorphisms and mutations, their expression, their specificity and their evolution. The IMGT® structural

Downloaded from https://academic.oup.com/nar/article/37/suppl_1/D1006/1014137 by biuupm Bibliotheque Universitaire UPM user on 01 June 2022

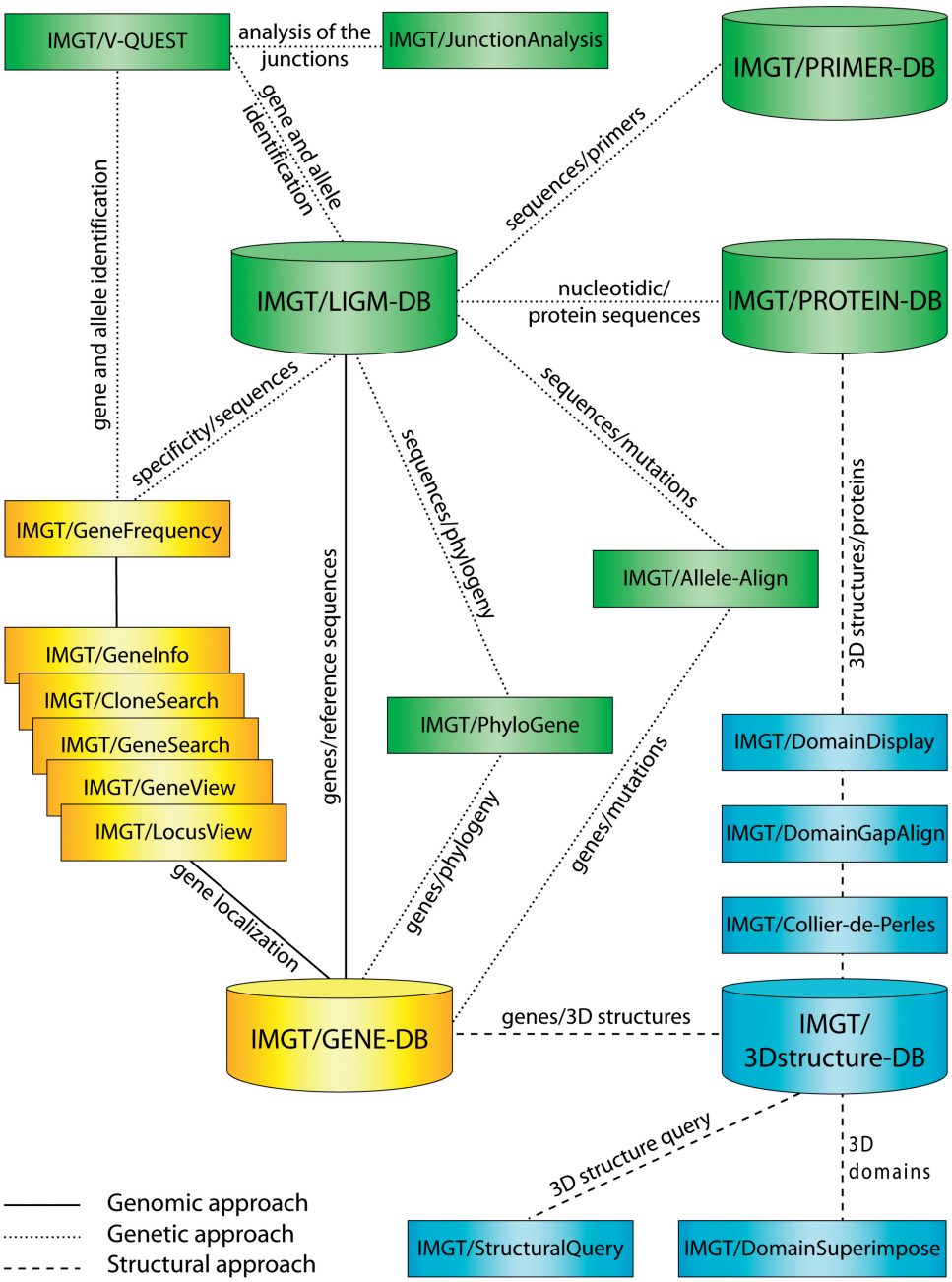


Figure 1. IMGT®, the international ImMunoGeneTics information system® (http://www.imgt.org). Genomic, genetic and structural components (databases and tools) are in yellow, green and blue, respectively. The IMGT Repertoire and other Web resources are not shown. Interactions in the genomic, genetic and structural approaches are represented with continuous, dotted and broken lines, respectively.

approach refers to the study of the 2D and 3D structures of the IG, TR, MHC, IgSF, MhcSF and RPI, and to the antigen- or ligand-binding characteristics in relationship with the protein functions, polymorphisms and evolution. For each approach, IMGT® provides databases, tools and Web resources (Figure 1 and Table 1). IMGT-Choreography (33), based on the Web service architecture paradigm, has been developed with the goal to enable significant biological and clinical requests involving every part of the IMGT® information system.

IMGT® DATABASES

Gene database

IMGT/GENE-DB (16) is the comprehensive IMGT® genome database. IMGT/GENE-DB is the official repository of all the IG and TR genes and alleles approved by the World Health Organization (WHO)/International Union of Immunological Societies (IUIS) Nomenclature Subcommittee for IG and TR (34,35). In September 2008, IMGT/GENE-DB contained 1911 IG and TR genes from human, mouse and rat and 2909 alleles. Reciprocal links

Downloaded from https://academic.oup.com/nar/article/37/suppl_1/D1006/1014137 by biuupm Bibliotheque Universitaire UPM user on 01 June 2022

Table 1. IMGT® databases, tools and Web resources for genomic, genetic and structural approaches

Approaches	Databases	Tools	Web resources[a]
Genomic	IMGT/GENE-DB (16)	IMGT/GeneView IMGT/LocusView IMGT/CloneSearch IMGT/GeneSearch IMGT/GeneInfo (28) IMGT/GeneFrequency	IMGT Repertoire 'Locus and genes' section: –chromosomal localizations (1,2) –locus representations (1,2) –locus description –gene tables, etc. –potential germline repertoires –lists of genes –correspondence between nomenclatures (1,2)
Genetic	IMGT/LIGM-DB (24) IMGT/PRIMER-DB (25) IMGT/MHC-DB (26)	IMGT/V-QUEST (29) IMGT/JunctionAnalysis (30) IMGT/Allele-Align IMGT/PhyloGene (31) IMGT/DomainDisplay	IMGT Repertoire 'Proteins and alleles' section: –alignments of alleles –protein displays –tables of alleles etc.
Structural	IMGT/3Dstructure-DB (27)	IMGT/DomainGapAlign IMGT/Collier-de-Perles IMGT/DomainSuperimpose IMGT/StructuralQuery (27)	IMGT Repertoire '2D and 3D structures' section: –IMGT Colliers de Perles (2D representations on one layer or two layers) (21,22) –IMGT classes for amino acid characteristics (32) –IMGT Colliers de Perles reference profiles (32) –3D representations

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

exist between IMGT/GENE-DB and the HGNC database (36) and Entrez Gene (17). IMGT-GENE-DB allows a query per gene and allele name. IMGT/GENE-DB interacts dynamically with IMGT/LIGM-DB (24) to download and display human, mouse and rat gene-related sequence data. This is the first example of an interaction between IMGT® databases using the concepts of classification.

Sequence databases

IMGT/LIGM-DB. IMGT/LIGM-DB (24) is the comprehensive IMGT® database of IG and TR nucleotide sequences from human and other vertebrate species, with translation for fully annotated sequences, created in 1989 by LIGM, Montpellier, France, on the Web since July 1995. IMGT/LIGM-DB is the first and the largest IMGT® database. In September 2008, IMGT/LIGM-DB contained 126 667 nucleotide sequences of IG and TR from 223 vertebrate species. The unique source of data for IMGT/LIGM-DB is EMBL-Bank (18) which shares data with the other two generalist databases GenBank (37) and DDBJ (38). IMGT/LIGM-DB sequence data are identified by the EMBL/GenBank/DDBJ accession number. Based on expert analysis, specific detailed annotations are added to IMGT flat files. The Web interface allows searches according to immunogenetic specific criteria and is easy to use without any knowledge in a computing language. Selection is displayed at the top of the resulting sequences page, 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. The IMGT/LIGM-DB annotations (gene and allele name assignment, labels) allow data retrieval not only from IMGT/LIGM-DB, but also from other IMGT® databases. For example, the IMGT/GENE-DB entries provide the IMGT/LIGM-DB accession numbers of the IG and TR cDNA sequences that contain a given V, D, J or C gene. The automatic annotation of rearranged human and mouse cDNA sequences in IMGT/LIGM-DB is performed by IMGT/Automat (39), an internal Java tool that implements IMGT/V-QUEST (29) and IMGT/JunctionAnalysis (30). 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 several Sequence Retrieval System (SRS) sites. IMGT/LIGM-DB can be searched by BLAST or FASTA on different servers (EBI, IGH, Institut Pasteur Paris).

Other IMGT sequence databases. IMGT/PRIMER-DB (25) is the IMGT® oligonucleotide database on the Web since February 2002. In September 2008, IMGT/PRIMER-DB contained 1864 entries. The database manages standardized information on oligonucleotides (or Primers) and combinations of primers (Sets and Couples) for IG and TR. These primers are useful for combinatorial library constructions, scFv, phage display or microarray technologies. IMGT/PROTEIN-DB, in developement, will contain the translations of the IMGT/LIGM-DB and IMGT/GENE-DB sequences. IMGT/MHC-DB hosted at EBI comprises IMGT/HLA for human MHC (or HLA) and IMGT/MHC-NHP, for MHC of non-human primates (26).

Structure database

IMGT/3Dstructure-DB is the IMGT® 3D structure database, created by LIGM, on the Web since November 2001 (27). IMGT/3Dstructure-DB comprises IG, TR, MHC, IgSF, MhcSF and RPI with known 3D structures. In September 2008, IMGT/3Dstructure-DB contained 1461 atomic coordinate files. These coordinate files extracted from the Protein Data Bank (PDB) (http://www.rcsb.org/pdb/) (40) are renumbered according to the standardized IMGT unique numbering (9–11). The IMGT/3Dstructure-DB cards provide chain details with IMGT

Downloaded from https://academic.oup.com/nar/article/37/suppl_1/D1006/1014137 by biuupm Bibliotheque Universitaire UPM user on 01 June 2022

annotations (receptor, chain and domain description with IMGT labels, assignment of IMGT gene and allele names, domain delimitations and amino acid positions according to the IMGT unique numbering, and IMGT Colliers de Perles on one layer and two layers), contact analysis, downloadable renumbered IMGT/3Dstructure-DB flat files, visualization tools (Jmol and QuickPDB), and external links. IMGT Residue@Position cards provide detailed information on the inter- and intra-domain contacts at each residue position, based on the IMGT unique numbering. The contacts are described per domain (intra- and inter-domain contacts) and annotated in terms of IMGT® labels (chain and domain), positions (IMGT unique numbering), backbone or side-chain implication.

IMGT® TOOLS

Gene tools

The IMGT® gene tools (genomic approach) manage the locus organization and gene location and provide the display of physical maps for the human and mouse IG, TR and MHC loci. They allow to view genes in a locus (IMGT/GeneView, IMGT/LocusView), to search for clones (IMGT/CloneSearch), or to search for genes in a locus (IMGT/GeneSearch, IMGT/GeneInfo) based on IMGT® gene names, functionality or localization on the chromosome. IMGT/GeneFrequency provides a graphical representation of the numbers of cDNA and gDNA IMGT/LIGM-DB sequences containing rearranged IG and TR genes.

Sequence tools

The IMGT® sequence analysis tools (genetic approach) comprise IMGT/V-QUEST (29) for the identification of the V, D and J genes and of their mutations, IMGT/JunctionAnalysis (30) for the analysis of the V-J and V-D-J junctions that confer the antigen receptor specificity, IMGT/Allele-Align for the detection of polymorphisms, IMGT/Phylogene (31) for gene evolution analyses, and IMGT/DomainDisplay for the display of amino acid sequences from the IMGT domain directory. IMGT/V-QUEST (V-QUEry and STandardization) (29) is an integrated software for IG and TR. This tool, which is easy to use, analyses an input of up to fifty IG or TR germline or rearranged variable nucleotide sequences. IMGT/V-QUEST results comprise, for rearranged sequences, the identification of the V, D and J genes and alleles, nucleotide alignments by comparison with the IMGT reference directory, the delimitations of the framework regions (FR-IMGT) and complementarity determining regions (CDR-IMGT) based on the IMGT unique numbering, the protein translation of the input sequences, the result of IMGT/JunctionAnalysis, the description of the mutations and amino acid changes of the V-REGION and the IMGT Collier de Perles representation of the V-DOMAIN.

Structure tools

The IMGT® structure tools bridge the gap between sequences and 3D structures: IMGT/DomainGapAlign analyses amino acid sequences per domain, IMGT/Collier-de-Perles allows to make your own IMGT Collier de Perles, and IMGT/DomainSuperimpose allows to superimpose two domain 3D structures from IMGT/3Dstructure-DB. IMGT/StructuralQuery (27) allows to retrieve the IMGT/3Dstructure-DB entries containing a V-DOMAIN, based on specific structural characteristics of the intramolecular interactions: phi and psi angles, accessible surface area, amino acid type, distance in angstrom between amino acids, and CDR-IMGT lengths.

IMGT® WEB RESOURCES

IMGT Repertoire

The IMGT® Web resources for genomic, genetic and structural approaches are compiled in the sections of the IMGT Repertoire and provide a synthetic view of data managed in the databases and tools.

Genomics Web resources. The IMGT® genomics resources are compiled in the 'Locus and genes' section which includes 'Chromosomal localizations', 'Locus representations', 'Locus description', 'Gene exon/intron organization', 'Gene exon/intron splicing sites', 'Gene tables', 'Potential germline repertoires', lists of IG and TR genes and links between IMGT®, HGNC, Entrez Gene and OMIM, and correspondence between nomenclatures (1,2). The IMGT Repertoire 'Probes and RFLP' section provides data on gene insertion/deletion.

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

Structural Web resources. The IMGT® structural resources are compiled in the '2D and 3D structures' section which includes IMGT Colliers de Perles (21,22), FR-IMGT and CDR-IMGT lengths, amino acid chemical characteristics profiles (32). To appropriately analyse the amino acid resemblances and differences between IG, TR, MHC and RPI chains, eleven IMGT classes were defined for the amino acid 'chemical characteristics' properties and used to set up IMGT Colliers de Perles references profiles. 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 visualization of 3D representations of IG and TR variable domains allows rapid correlation between protein sequences and 3D data.

Other Web resources

In addition to the IMGT Scientific chart and IMGT Repertoire, other major components of the IMGT® Web resources comprise The IMGT Medical page, The IMGT Veterinary page, The IMGT Biotechnology page,

Downloaded from https://academic.oup.com/nar/article/37/suppl_1/D1006/1014137 by biuupm Bibliotheque Universitaire UPM user on 01 June 2022

IMGT Education, IMGT Lexique, IMGT Aide-Mémoire, Tutorials, IMGT Index, and external links (IMGT Blocnotes, The IMGT Immunoinformatics page, Interesting links) and IMGT other accesses (SRS,MRS).

CONCLUSION

Since July 1995, IMGT® has been available on the Web at the IMGT Home page http://www.imgt.org (Montpellier, France). IMGT® has an exceptional response with more than 150 000 requests a month. The information is of much value to clinicians and biological scientists in general. IMGT® databases, tools, and Web resources are extensively queried and used by scientists from both academic and industrial laboratories, who are equally distributed between the United States, Europe and the remaining world. IMGT® is used in very diverse domains: (i) fundamental and medical research (repertoire analysis of the IG antibody recognition sites and of the TR recognition sites in normal and pathological situations such as autoimmune diseases, infectious diseases, AIDS, leukemias, lymphomas and 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 and vaccinology). The creation of dynamic interactions between the IMGT® databases and tools, using Web services and IMGT-ML, and the design of IMGT-Choreography, represent novel and major developments of IMGT®, the international reference in immunogenetics and immunoinformatics.

CITING IMGT

Users are requested to cite this article and quote the IMGT home page URL, http://www.imgt.org.

ACKNOWLEDGEMENTS

We thank all the IMGT® users from academic and industrial laboratories and the clinicians and scientists from the European Research Initiative on CLL who help promoting standardization. IMGT® has received the National Bioinformatics Platform RIO label since the RIO creation in 2001 (CNRS, INSERM, CEA, INRA) and the National Bioinformatics Platform IBiSA label since the IBiSA creation in 2007. IMGT® is an Institutional Academic Member of the International Medical Informatics Association. IMGT® is a registered mark of the Centre National de la Recherche Scientifique.

FUNDING

The BIOMED1 (BIOCT930038), Biotechnology BIOTECH2 (BIO4CT960037) and 5th PCRDT Quality of Life and Management of Living Resources programmes (QLG2-2000-01287) programmes of the European Union (EU). IMGT® is currently supported by the CNRS, the Ministère de l'Enseignement Supérieur et de la Recherche (MESR) (Université Montpellier 2 Plan Pluri-Formation, Institut Universitaire de France), Réseau National des Génopoles, the Région Languedoc-Roussillon, the Agence Nationale de la recherche ANR (BIOSYS06_135457, FLAVORES), and the EU ImmunoGrid (IST-028069). Funding for open access charge: CNRS.

Conflict of interest statement. None declared.

REFERENCES

1. Lefranc,M.-P. and Lefranc,G. (2001) *The Immunoglobulin FactsBook*, Academic Press, London, UK, pp. 1–458.
2. Lefranc,M.-P. and Lefranc,G. (2001) *The T cell receptor FactsBook*, Academic Press, London, UK, pp. 1–398.
3. Lefranc,M.-P., Giudicelli,V., Kaas,Q., Duprat,E., Jabado-Michaloud,J., Scaviner,D., Ginestoux,C., Clément,O., Chaume,D. and Lefranc,G. (2005) IMGT, the international ImMunoGeneTics information system. *Nucleic Acids Res.*, **33**, D593–D597.
4. Giudicelli,V. and Lefranc,M.-P. (1999) Ontology for Immunogenetics: the IMGT-ONTOLOGY. *Bioinformatics*, **12**, 1047–1054.
5. Lefranc,M.-P. (2000) Nomenclature of the human immunoglobulin genes. In Coligan,J.E., Bierer,B.E., Margulies,D.E., Shevach,E.M. and Strober,W. (eds), *Current Protocols in Immunology*, John Wiley & Sons, Inc., NJ, Hoboken, pp. A.1P.1–A.1P.37.
6. Lefranc,M.-P. (2000) Nomenclature of the human T cell receptor genes. In Coligan,J.E., Bierer,B.E., Margulies,D.E., Shevach,E.M. and Strober,W. (eds), *Current Protocols in Immunology*, John Wiley & Sons, Inc., NJ, Hoboken, pp. A.1O.1–A.1O.23.
7. Lefranc,M.-P. (1997) Unique database numbering system for immunogenetic analysis. *Immunol. Today*, **18**, 509.
8. Lefranc,M.-P. (1999) The IMGT unique numbering for Immunoglobulins, T cell receptors and Ig-like domains. *The Immunologist*, **7**, 132–136.
9. 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.
10. Lefranc,M.-P., Pommié,C., Kaas,Q., Duprat,E., Bosc,N., Guiraudou,D., Jean,C., Ruiz,M., Da Piedade,I., Rouard,M. *et al.* (2005) IMGT unique numbering for immunoglobulin and T cell receptor constant domains and Ig superfamily C-like domains. *Dev. Comp. Immunol.*, **29**, 185–203.
11. Lefranc,M.-P., Duprat,E., Kaas,Q., Tranne,M., Thiriot,A. and Lefranc,G. (2005) IMGT unique numbering for MHC groove G-DOMAIN and MHC superfamily (MhcSF) G-LIKE-DOMAIN. *Dev. Comp. Immunol.*, **29**, 917–938.
12. Duroux,P., Kaas,Q., Brochet,X., Lane,J., Ginestoux,C., Lefranc,M.-P. and Giudicelli,V. (2008) IMGT-Kaleidoscope, the Formal IMGT-ONTOLOGY paradigm. *Biochimie*, **90**, 570–583.
13. Noy,N.F., Crubezy,M., Fergerson,R.W., Knublauch,H., Tu,S.W., Vendetti,J. and Musen,M.A. (2003) Protege-2000: an open-source ontology-development and knowledge-acquisition environment. *AMIA Annu. Symp. Proc.*, **2003**, 953.
14. Eilbeck,K and Lewis,S.E (2004) Sequence Ontology Annotation Guide. *Com. Funct. Genomics*, **5**, 642–647.
15. Wain,H.M., Bruford,E.A., Lovering,R.C., Lush,M.J., Wright,M.W. and Povey,S. (2002) Guidelines for human gene nomenclature. *Genomics*, **79**, 464–470.

Downloaded from https://academic.oup.com/nar/article/37/suppl_1/D1006/1014137 by biuupm Bibliotheque Universitaire UPM user on 01 June 2022

16. 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.
17. Maglott,D., Ostell,J., Pruitt,K.D. and Tatusova,T. (2007) Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res.*, **35**, D26–D31.
18. Cochrane,G., Akhtar,R., Aldebert,P., Althorpe,N., Baldwin,A., Bates,K., Bhattacharyya,S., Bonfield,J., Bower,L., Browne,P. *et al.* (2008) Priorities for nucleotide trace, sequence and annotation data capture at the Ensembl Trace Archive and the EMBL Nucleotide Sequence Database. *Nucleic Acids Res.*, **36**, D5–D12.
19. Wilming,L.G., Gilbert,J.G.R., Howe,K., Trevanion,S., Hubbard,T. and Harrow,J.L. (2008) The vertebrate genome annotation (Vega) database. *Nucleic Acids Res.*, **36**, D753–D760.
20. Bult,C.J., Eppig,J.T., Kadin,J.A., Richardson,J.E. and Blake,J.A. and the Mouse Genome Database Group (2008) The Mouse Genome Database (MGD): mouse biology and model systems. *Nucleic Acids Res.*, **36**, D724–D728.
21. 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.
22. Kaas,Q., Ehrenmann,F. and Lefranc,M.-P. (2007) IG, TR, MHC, IgSf and MhcSF: what do we learn from the IMGT Colliers de Perles? Brief. *Funct. Genomic Proteomic*, **6**, 253–264.
23. Lefranc,M.-P., Giudicelli,V., Regnier,L. and Duroux,P. (2008) IMGT, a system and an ontology that bridge biological and computational spheres in bioinformatics. *Brief. Bioinform.*, **9**, 263–275.
24. Giudicelli,V., Duroux,P., Ginestoux,C., Folch,G., Jabado-Michaloud,J., Chaume,D. and Lefranc,M.-P. (2006) IMGT/LIGM-DB, the IMGT® comprehensive database of immunoglobulin and T cell receptor nucleotide sequences. *Nucleic Acids Res.*, **34**, D781–D784.
25. Folch,G., Bertrand,J., Lemaitre,M. and Lefranc,M.-P. (2004) IMGT/PRIMER-DB. The Molecular Biology Database Collection: 2004 update. *Nucleic Acids Res.*, **32**, 3–22.
26. Robinson,J., Waller,M.J., Parham,P., de Groot,N., Bontrop,R., Kennedy,L.J., Stoehr,P. and Marsh,S.G. (2003) IMGT/HLA and IMGT/MHC sequence databases for the study of the major histocompatibility complex. *Nucleic Acids Res.*, **31**, 311–314.
27. 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.
28. Baum,T.P., Hierle,V., Pasqual,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.
29. Brochet,X., Lefranc,M.-P. and Giudicelli,V. (2008) IMGT/V-QUEST: the highly customized and integrated system for IG and TR standardized V-J and V-D-J sequence analysis. *Nucleic Acids Res*, **36**, W503–W508.
30. Yousfi Monod,M., Giudicelli,V., Chaume,D. and Lefranc,M.-P. (2004) IMGT/JunctionAnalysis: the first tool for the analysis of the immunoglobulin and T cell receptor complex V-J and V-D-J JUNCTIONs. *Bioinformatics*, **20**, i379–i385.
31. 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.
32. 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.
33. Lefranc,M.-P., Clément,O., Kaas,Q., Duprat,E., Chastellan,P., Coelho,I., Combres,K., Ginestoux,C., Giudicelli,V., Chaume,D. *et al.* (2004) IMGT-Choreography for Immunogenetics and Immunoinformatics. *In Silico Biol.*, **5**, 45–60.
34. Lefranc,M.-P. (2008) WHO-IUIS Nomenclature Subcommittee for Immunoglobulins and T cell receptors report Immunoglobulins and T cell receptors report August 2007, 13th International Congress of Immunology, Rio de Janeiro, Brazil. *Dev. Comp. Immunol.*, **32**, 461–463.
35. Lefranc,M.-P. (2007) WHO-IUIS Nomenclature Subcommittee for Immunoglobulins and T cell receptors report. *Immunogenetics*, **59**, 899–902.
36. Bruford,E.A., Lush,M.J., Wright,M.W., Sneddon,T.P., Povey,S. and Birney,E. (2008) The HGNC Database in 2008: a resource for the human genome. *Nucleic Acids Res.*, **36**, D445–D448.
37. Benson,D.A., Karsch-Mizrachi,I., Lipman,D.J., Ostell,J. and Wheeler,D.L. (2008) GenBank. *Nucleic Acids Res.*, **36**, D25–D30.
38. Sugawara,H., Ogasawara,O., Okubo,K., Gojobori,T. and Tateno,Y. (2008) DDBJ with new system and face. *Nucleic Acids Res.*, **36**, D22–D24.
39. Giudicelli,V., Chaume,D., Jabado-Michaloud,J. and Lefranc,M.-P. (2005) Immunogenetics sequence annotation: the strategy of IMGT based on IMGT-ONTOLOGY. *Stud. Health Technol. Inform.*, **116**, 3–8.
40. Henrick,K., Feng,Z., Bluhm,W.F., Dimitropoulos,D., Doreleijers,J.F., Dutta,S., Flippen-Anderson,J.L., Ionides,J., Kamada,C., Krissinel,E. *et al.* (2008) Remediation of the protein data bank archive. *Nucleic Acids Res.*, **36**, D426–D433.

Downloaded from https://academic.oup.com/nar/article/37/suppl_1/D1006/1014137 by biuupm Bibliotheque Universitaire UPM user on 01 June 2022