

# IMGT/3Dstructure-DB and tools for immunoglobulins (IG) or antibodies, T cell receptors (TR), MHC, IgSF and MhcSF structural data

François EHRENMANN<sup>1</sup> and Marie-Paule LEFRANC<sup>1</sup>

<sup>1</sup> Laboratoire d'ImmunoGénétique Moléculaire, LIGM, Université Montpellier 2, UPR CNRS1142, 141, rue de la Cardonille, 34396, Montpellier, Cedex 5, France  
{Francois.Ehrenmann,Marie-Paule.Lefranc}@igh.cnrs.fr

**Keywords** IMGT, immunoglobulin, antibody, T cell receptor, MHC, IgSF, MhcSF

## 1 Introduction

IMGT/3Dstructure-DB (1,2) is the three-dimensional (3D) structure database of IMGT®, the international ImMunoGenetics information system® (<http://www.imgt.org>) (3), the global reference in immunogenetics and immunoinformatics. Major breakthroughs characterize IMGT® and, therefore, IMGT/3Dstructure-DB: a standardized identification (IMGT keywords), a standardized nomenclature (IMGT gene and allele names), a standardized description (IMGT labels), and a standardized numerotation (IMGT unique numbering). IMGT-ONTOLOGY concepts have been crucial in bridging the gap between sequences and 3D structures (4,5) in IMGT/3Dstructure-DB database and in the IMGT/DomainGapAlign (2) and IMGT/Collier-de-Perles (6,7) tools.

## 2 IMGT/3Dstructure-DB

The IMGT/3Dstructure-DB structural data are extracted from the Protein Data Bank (PDB) and annotated according to the IMGT-ONTOLOGY concepts of classification, using internal tools and IMGT/DomainGapAlign. IMGT/3Dstructure-DB provides the closest genes and alleles that are expressed in the amino acid sequences of the 3D structures, by aligning these sequences with the IMGT domain reference directory. Each entry in the database is detailed in an IMGT/3Dstructure-DB card. Eight tabs are available at the top of each card: 'Chain details', 'Contact analysis', 'Paratope and epitope', '3D visualization Jmol or QuickPDB', 'Renumbered IMGT file', 'IMGT numbering comparison', 'References and links' and 'Printable card'. As an example, the 'Chain details' comprises information, first, on the chain itself (chain ID, chain length, IMGT chain description...), then on each domain, starting from the N-terminal end (IMGT domain description, gene and allele names...). IMGT/Colliers de Perles on two layers, available for the variable (V) and constant (C) type domains, are

displayed with hydrogen bonds. 'Contact analysis' provides contacts between structural units (domains or ligand) and are obtained by a local program written in C in which atoms are considered to be in contact when no water molecule can take place between them. The atom contact types and categories for each amino acid are provided in 'IMGT Residue@Position cards'. 'IMGT pMHC contact sites' graphically represent, in IMGT Colliers de Perles, the MHC amino acid positions that contact the peptide side chains in pMHC complexes, and thus allow comparison of pMHC interactions. These features are part of the information provided by IMGT/3Dstructure-DB. In June 2010, the IMGT/3Dstructure-DB database manages 2242 coordinate files (PDB, INN and Kabat).

## 3 IMGT/Collier-de-Perles

IMGT Colliers de Perles are 2D graphical representations (6,7) available for the V type domain (V-DOMAIN of IG and TR, V-LIKE-DOMAIN of IgSF other than IG and TR), C type domain (C-DOMAIN of IG and TR, C-LIKE-DOMAIN of IgSF other than IG and TR) and groove (G) type domain (G-DOMAIN of MHC, G-LIKE-DOMAIN of MhcSF other than MHC). Any domain represented by an IMGT Collier de Perles is characterized by the length of its strands, loops and turns and, for the G type, by the length of its helix. IMGT Colliers de Perles are generated with the IMGT/Colliers-de-Perles tool which allows the users to draw IMGT Colliers de Perles starting from their own amino acid sequences. Sequences have to be gapped according to the IMGT unique numbering (using for example IMGT/DomainGapAlign). Adjustements are possible manually in the window for unusual features. IMGT/Collier-de-Perles tool can be customized to display the CDR-IMGT according to the IMGT Color menu or to visualize the amino acids according to their hydropathy, volume or IMGT physicochemical classes (8).

## 4 IMGT/DomainGapAlign

IMGT/DomainGapAlign is a tool which aligns the user amino acid sequences with the IMGT domain reference directory, identifies the closest V, C and G domains, creates gaps according to the IMGT unique numbering and highlights differences with the closest reference(s). For an antibody V domain sequence, IMGT/DomainGapAlign identifies the closest germline V-REGION and J-REGION, and provides a delimitation of the strands, framework regions (FR-IMGT) and CDR-IMGT. The IMGT gene and allele name of the closest sequence(s) from the IMGT domain reference directory is provided with a percentage of identity and a Smith-Waterman score. Regions and domains are highlighted using the IMGT Color menu. Several sequences can be analysed simultaneously and users can choose how many alignments to display for each sequence. The IMGT Colliers de Perles are generated from the gapped sequences provided by the IMGT/DomainGapAlign tool.

## 5 IMGT/DomainDisplay

IMGT/DomainDisplay is a tool which provides the display of amino acid sequences from the IMGT domain directory. This directory contains, for the IG and TR, amino acid sequences of the domains encoded by the C genes and the translation of the germline V and joining (J) genes. The identified genes are classified based on the IMGT nomenclature of IG and TR genes and alleles that was approved in 1999 by the Human Genome Organisation (HUGO) Nomenclature Committee (HGNC), entered in IMGT/GENE-DB, and endorsed by the World Health Organization (WHO)-International Union of Immunological Societies (IUIS) Nomenclature Committee. Entrez Gene at the National Center for Biotechnology Information (NCBI), Vertebrate Genome Annotation (Vega) at the Wellcome Trust Sanger Institute, and Ensembl at the European Bioinformatics Institute (EBI) currently use IMGT nomenclature.

## 6 Conclusion

IMGT/3Dstructure-DB and tools are widely used by researchers, particularly for antibody engineering and humanization design (1). Indeed they allow to precisely define and to easily compare amino acid sequences of the FR-IMGT and CDR-IMGT, between the nonhuman (mouse, rat...) and the closest human V domains. A recent analysis performed on humanized antibodies used in

oncology underlines the importance of a correct delimitation of the CDR to be grafted. IMGT/3Dstructure-DB facilitates the identification of potential immunogenic residues at given positions in chimeric or humanized antibodies, including those of the C domains. These therapeutic applications emphasize the importance of the IMGT/3Dstructure-DB standardized approach that bridges the gap between sequences and 3D structures whatever the species. Since 2008, amino acid sequences of monoclonal antibodies (mAb, suffix -mab) and of fusion proteins for immune applications (FPIA, suffix -cept) from the WHO/International Nonproprietary Name (INN) programme have been entered in IMGT®.

## References

- [1] Q. Kaas, M. Ruiz and M.-P. Lefranc, 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, 2004.
- [2] F. Ehrenmann, Q. Kaas and M.-P. Lefranc, IMGT/3Dstructure-DB and IMGT/DomainGapAlign: a database and a tool for immunoglobulins or antibodies, T cell receptors, MHC, IgSF and MhcSF. *Nucleic Acids Res.* 38(Database issue):D301-7, 2010.
- [3] M.-P. Lefranc, V. Giudicelli, C. Ginestoux, J. Jabado-Michaloud, G. Folch, F. Bellahcene, Y. Wu, E. Gemrot, X. Brochet, J. Lane, L. Regnier, F. Ehrenmann, G. Lefranc and P. Duroux, IMGT®, the international ImMunoGeneTics information system®. *Nucleic Acids Res.*, 37,1006-1012, 2009.
- [4] Q. Kaas, F. Ehrenmann and M.-P. Lefranc, IG, TR and IgSF, MHC and MhcSF: what do we learn from the IMGT Colliers de Perles? *Brief. Funct. Genomic Proteomic*, 6:253-264, 2007.
- [5] M.-P. Lefranc, V. Giudicelli, L. Regnier and P. Duroux, IMGT, a system and an ontology that bridge biological and computational spheres in bioinformatics. *Brief Bioinform.*, 9, 263-275, 2008.
- [6] M. Ruiz and M.-P. Lefranc, IMGT gene identification and Colliers de Perles of human immunoglobulins with known 3D structures. *Immunogenetics*, 53, 857-883, 2002.
- [7] Q. Kaas and M.-P. Lefranc, IMGT Colliers de Perles: standardized sequence-structure representations of the IgSF and MhcSF superfamily domains. *Current Bioinformatics*, 2, 21-30, 2007.
- [8] C. Pommier, S. Levadoux, R. Sabatier, G. Lefranc and M.-P. Lefranc, IMGT standardized criteria for statistical analysis of immunoglobulin V-REGION amino acid properties *J. Mol. Recognit.*, 17, 17-32, 2004.