

IMGT-ONTOLOGY for immunogenetics and immunoinformatics information systems

Véronique GIUDICELLI¹ and Marie-Paule LEFRANC¹

¹ Laboratoire d'ImmunoGénétique Moléculaire, LIGM, Université Montpellier 2, UPR CNRS 1142, Institut de Génétique Humaine, IGH, 141 rue de la Cardonille, 34396 Montpellier Cedex 5, Montpellier, France
{Véronique.Giudicelli, Marie-Paule.Lefranc}@igh.cnrs.fr

Abstract *IMGT-ONTOLOGY, the first ontology for immunogenetics and immunoinformatics, manages the immunogenetics knowledge through diverse facets relying on seven axioms and represents a paradigm for the elaboration of integrated ontologies in system biology.*

Keywords IMGT, immunogenetics, immunoinformatics, ontology, information system, system biology.

1 Introduction

IMGT®, the international ImMunoGeneTics information system® (<http://www.imgt.org>), is the reference in immunogenetics and immunoinformatics. IMGT® standardizes and manages the complex immunogenetics data which include the immunoglobulins (IG) or antibodies, the T cell receptors (TR), the major histocompatibility complex (MHC) and the related proteins of the immune system (RPI) which belong to the immunoglobulin superfamily (IgSF) and to the MHC superfamily (MhcSF) [1]. The accuracy and consistency of IMGT® data and the coherence between the different IMGT® components (databases, tools and Web resources) are based on IMGT-ONTOLOGY, the first ontology for immunogenetics and immunoinformatics [2]. IMGT-ONTOLOGY manages the immunogenetics knowledge through diverse facets relying on seven axioms, IDENTIFICATION, CLASSIFICATION, DESCRIPTION, NUMEROTATION, LOCALIZATION, ORIENTATION and OBTENTION, that postulate that objects, processes and relations have to be identified, described, classified, numerotated, localized, orientated, and the way they are obtained, determined. These axioms constitute the Formal IMGT-ONTOLOGY, also designated as IMGT-Kaleidoscope [3].

2 IMGT-ONTOLOGY axioms and concepts

The IDENTIFICATION axiom has generated the concepts of identification which allow to identify any biological objects, processes and relations in IMGT®. They provide the terms and rules that were necessary to define the IMGT standardized keywords

used, in IMGT® databases, for the identification of IG, TR or MHC nucleotide and protein sequences, and structures according to their fundamental biological and immunogenetics characteristics.

The CLASSIFICATION axiom provides the rules that are necessary to classify the IG and TR genes. Indeed, the IG and TR genes belong to highly polymorphic multigene families organized as clusters in several loci in the genome [4,5]: therefore their classification requires a strong knowledge standardization. As a major contribution, the concepts of classification allowed to set up a unique nomenclature of human IG and TR genes, which was approved by the Human Genome Organisation (HUGO) Nomenclature Committee (HGNC) in 1999 and has become the community standard. The IG and TR genes are managed in the IMGT/GENE-DB database [6]. IMGT/GENE-DB database entries are cross-referenced by HGNC database, GenAtlas, Entrez Gene (NCBI) and Vega (Wellcome Trust Sanger Institute).

The DESCRIPTION axiom and related concepts correspond to the standardization of terms and rules which allow to describe the structural and functional characteristic features of the IG, TR and MHC nucleotide and protein sequences, and 3D structures. Description concepts include IMGT standardized labels and the topological relationships that are used for the annotation process. Interestingly, 64 of the IMGT labels have been integrated by Sequence Ontology (<http://www.sequenceontology.org/>).

The NUMEROTATION axiom and the concepts of numerotation determine the principles of a unique numbering for variable, constant and groove domains in IG, TR and MHC sequences and 3D structures. The concept of 'IMGT unique numbering' and its graphical representation, the 'IMGT Collier

de Perles' represent a major breakthrough and are the flashpoint of IMGT® since they allow to bridge the gap between sequences and structures [7,8]. Thanks to both concepts, IMGT® provides a standardized and a fine description of allelic polymorphisms of IG and TR genes and of the somatic mutations in IG sequences. It also provides the rules for the delimitations of the framework and complementarity determining regions of IG and TR allowing the standardization of the contact analysis between residues in 3 D structures.

The LOCALIZATION axiom postulates that molecules, cells, organs, organisms or populations and their processes and relations have to be localized in time or space. At the molecular level in the field of immunogenetics, the concepts of localization allow to characterize the localization of IG, TR and MHC genes and proteins.

The ORIENTATION axiom defines the rules for the orientation of objects in IMGT®. In the context of genome analysis, it has led to set the 'Genomic orientation' concept (for chromosome, locus and gene) and the 'DNA strand orientation' concept.

The OBTENTION axiom has generated a set of standardized terms that precise, for any object of IMGT®, its origins ('Origin' concept) and the conditions in which the sequences have been obtained ('Methodology' concept).

3 Conclusion

The axioms of IMGT-ONTOLOGY have been essential for the conceptualization of the molecular immunogenetics knowledge and for the creation of IMGT®. All the components of the IMGT® integrated system have been developed, based on standardized concepts and relations, making IMGT® a system and an ontology that bridge biological and computational spheres in bioinformatics [9]. IMGT-ONTOLOGY concepts are available, for the biologists and IMGT® users, in the IMGT Scientific chart [1]. They are being formalized step by step in OWL language with the Protégé ontology editor. A first version has been published on NCBO Bioportal site (<http://bioportal.bioontology.org/>) and is also available from the 'IMGT downloads'. The concepts of IMGT-ONTOLOGY are currently used for the exchange and the sharing of knowledge in: (i) fundamental and medical research (repertoire analysis of the IG antibody sites and of the TR recognition sites in normal and pathological situations such as autoimmune diseases and infectious diseases), (ii) veterinary research, (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 (scFv, phage displays, combinatorial libraries, chimeric, humanized and human antibodies), (vi) diagnostics and (vii) therapeutical approaches (grafts, immunotherapy, vaccinology). IMGT-ONTOLOGY was a key component in the elaboration and setting up of standards of the European ImmunoGrid project (<http://www.immunogrid.org/>) whose aim is to define the essential concepts for modelling of the immune system. IMGT-ONTOLOGY can also be used for multi-scale level approaches at the molecule, cell, tissue, organ, organism or population level, emphasizing the generalization of the application domain. In that way IMGT-ONTOLOGY represents a paradigm for the elaboration of ontologies for immunogenetics and immunoinformatics information systems.

References

- [1] 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.
- [2] V. Giudicelli and M.-P. Lefranc, Ontology for Immunogenetics: IMGT-ONTOLOGY. *Bioinformatics*, 15:1047-1054, 1999.
- [3] P. Duroux, Q. Kaas, X. Brochet, J. Lane, C. Ginestoux, M.-P. Lefranc and V. Giudicelli, IMGT-Kaleidoscope, the Formal IMGT-ONTOLOGY paradigm. *Biochimie*, 90:570-583, 2008.
- [4] M.-P. Lefranc and G. Lefranc, *The Immunoglobulin FactsBook*, Academic Press, London, UK, 2001.
- [5] M.-P. Lefranc and G. Lefranc, *The T Cell Receptor FactsBook*, Academic Press, London, UK, 2001.
- [6] V. Giudicelli, D. Chaume and M.-P. Lefranc, IMGT/GENE-DB: a comprehensive database for human and mouse immunoglobulin and T cell receptor genes. *Nucleic Acids Res.*, 33:256-261, 2005.
- [7] M.-P. Lefranc, C. Pommié, R. Ruiz, V. Giudicelli, E. Foulquier, L. Truong, V. Thouvenin-Contet and G. Lefranc, IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains. *Dev. Comp. Immunol.*, 27:55-77, 2003.
- [8] 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.
- [9] 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.