

# Ressources IMGT et IMGT-ONTOLOGY dans le contexte du Web sémantique médical

IMGT, the international ImMunoGeneTics information system®,  
<http://imgt.cines.fr>

Céline Protat<sup>a</sup>, Véronique Giudicelli<sup>a</sup>, Chantal Ginestoux<sup>a</sup>, Valérie Thouvenin-Contet<sup>a</sup>, Denys Chaume<sup>a</sup> et Marie-Paule Lefranc<sup>a,b</sup>

<sup>a</sup>Laboratoire d'ImmunoGénétique Moléculaire - LIGM  
UPR CNRS 1142, Institut de Génétique Humaine  
141, rue de la Cardonille, F-34396 Montpellier Cedex 5  
[lefranc@ligm.igh.cnrs.fr](mailto:lefranc@ligm.igh.cnrs.fr)

<sup>b</sup> Institut Universitaire de France

## Abstract

IMGT, the international ImMunoGeneTics information system® (<http://imgt.cines.fr>) est un système d'information intégré de grande qualité, spécialisé dans les immunoglobulines, les récepteurs T et le complexe majeur d'histocompatibilité de l'homme et des autres vertébrés, créé en 1989, par le Laboratoire d'ImmunoGénétique Moléculaire (LIGM), à l'Université Montpellier II, CNRS, Montpellier, France. IMGT comprend plusieurs bases de données de séquences, de gènes et de structures 3D, des ressources Web de 8000 pages HTML ("IMGT Marie-Paule page") et des outils d'analyse interactifs. Les connaissances IMGT et les concepts d'IMGT-ONTOLOGY sont présentés dans le contexte du Web sémantique médical.

## Mots clés:

IMGT-ONTOLOGY; Ontologie; Immunogénétique; Immunoinformatique; Informatique médicale

## Introduction

IMGT, the international ImMunoGeneTics information system®, <http://imgt.cines.fr> [1], créé en 1989 à Montpellier, est le premier et, à ce jour, le seul système d'information intégré international en Immunogénétique spécialisé dans les immunoglobulines (IG), les récepteurs des lymphocytes T (TR) et les molécules du complexe majeur d'histocompatibilité (CMH) de l'homme et des autres vertébrés. IMGT constitue un outil unique, proposant aux chercheurs et aux cliniciens un accès aux séquences d'ADN et de protéines, aux données structurales et informations génétiques de l'ensemble de ces molécules du système immunitaire, ainsi qu'une expertise détaillée de ces données, et des outils d'analyse interactifs. Les données d'IMGT sont annotées et expertisées selon les règles de la charte scientifique IMGT (*IMGT Scientific chart*), basées sur IMGT-ONTOLOGY. Par la qualité de ses données scientifiques, sa distribution facile et son interface conviviale, IMGT a d'importantes implications en recherche médicale (répertoire dans les maladies autoimmunes et infectieuses, Sida, leucémies, lymphomes, myélomes), en recherche vétérinaire (répertoires des espèces domestiques), pour l'étude de la diversité et de l'évolution des gènes de la réponse immunitaire adaptative, en biotechnologie et ingénierie des anticorps (banques combinatoires, *phage displays*) et pour les approches thérapeutiques (greffes, immunothérapie). IMGT est librement accessible à <http://imgt.cines.fr>.

## 1. Description des ressources IMGT disponibles sur le Web

IMGT comprend **3 bases de données de séquences** (IMGT/LIGM-DB, une base complète des séquences d'IG et TR de l'homme et des autres vertébrés, créée en 1989 par LIGM, Montpellier, sur le Web depuis juillet 1995. En février 2003, IMGT/LIGM-DB contenait 67.000 séquences d'IG et de TR de 105 espèces. IMGT/MHC-DB, des bases de séquences CMH hébergées à EBI, IMGT/PRIMER-DB, une base d'oligonucléotides d'IG et TR), **une base de données de gènes** (IMGT/GENE-DB), **une base de données de structures 3D** (IMGT/3Dstructure-DB) qui fournit l'identification des gènes et allèles IMGT et les Colliers de Perles des IG, TR et CMH dont la structure 3D est connue, créée par LIGM, sur le Web depuis novembre 2001 et qui contenait, en février 2003, 596 fichiers de coordonnées atomiques, **des ressources Web** comprenant 8000 pages HTML ("IMGT Marie-Paule page") et **des outils interactifs** pour les analyses de séquences (IMGT/V-QUEST, IMGT/JunctionAnalysis, IMGT/Allele-Align, IMGT/PhyloGene) et génomes (IMGT/GeneSearch, IMGT/GeneView, IMGT/LocusView) [1].

## 2. IMGT-ONTOLOGY et charte scientifique IMGT

Les données d'IMGT sont annotées et expertisées selon les règles et les concepts d'IMGT-ONTOLOGY, première ontologie dans le domaine de l'immunogénétique et de l'immunoinformatique [2,3]. IMGT dispose ainsi d'une sémantique spécifique bien définie et de la description standardisée des relations qui lient les différents concepts [4,5]. IMGT-ONTOLOGY repose sur cinq concepts: IDENTIFICATION, DESCRIPTION, CLASSIFICATION, NUMEROTATION, OBTENTION. Le concept IDENTIFICATION regroupe les connaissances qui permettent aux utilisateurs de sélectionner efficacement les séquences IG et TR sur le principe de la recherche documentaire. Le concept DESCRIPTION regroupe le vocabulaire et l'ensemble des règles qui permettent de décrire l'état, l'organisation et la composition des séquences des IG et TR. Le concept CLASSIFICATION regroupe les connaissances dont on dispose pour classer les gènes des IG et TR [6,7]. Ce concept a permis d'établir une nomenclature standardisée des gènes IG et TR [6,7] qui a été approuvée en 1999 par le HUMAN Genome Organization (HUGO) Nomenclature Committee (HGNC) et, par suite, une interopérabilité entre IMGT et Genome DataBase GDB (Canada), LocusLink au NCBI (Etats-Unis) et GeneCards (Israël). Le concept NUMEROTATION regroupe les connaissances et les règles qui permettent de positionner les connaissances dans un espace à une ou plusieurs dimensions. Ce concept a permis la mise en place de la numérotation unique IMGT pour toutes les séquences, données génétiques et structurales des IG, TR et CMH, et ceci, quels que soient le type de chaîne et l'espèce. Le concept OBTENTION regroupe les concepts qui permettent de sélectionner les séquences selon leur origine ou les pathologies auxquelles ils sont associés. Les concepts d'IMGT-ONTOLOGY sont à la base des règles de *IMGT Scientific chart* accessibles à <http://imgt.cines.fr>. IMGT-ONTOLOGY évolue par l'ajout de nouveaux concepts en fonction des avancées de la biotechnologie, de l'immunoinformatique et de la recherche médicale.

IMGT-ONTOLOGY a été décrite avec les schémas XML: IMGT-ML. Ces schémas codifient les contraintes référentielles et de domaines des concepts. Ils peuvent alors être utilisés par les traitements informatiques pour contrôler la cohérence du système d'information et pour distribuer les données dans un format XML conforme à IMGT-ONTOLOGY et pour favoriser l'interopérabilité avec d'autres ontologies.

Les outils d'aide à l'annotation s'appuient sur les métadonnées d'IMGT-ONTOLOGY et permettent à IMGT d'être une source de connaissance unique en immunogénétique et immunoinformatique. De janvier 2000 à décembre 2002, le serveur WWW d'IMGT à Montpellier a été consulté par plus de 180.000 sites différents, avec une moyenne de 120.000 requêtes par mois.

### Remerciements

Nous remercions très vivement toute l'équipe IMGT qui contribue avec compétence et enthousiasme au développement d'IMGT. IMGT est financée par le programme 5<sup>ème</sup> PCRDT (QLG2-2000-01287) de l'Union Européenne, le CNRS, le Ministère de l'Éducation Nationale et de la Recherche.

### Références

- [1] Lefranc M.-P. IMGT, the international ImMunoGeneTics database®, <http://imgt.cines.fr>. *Nucl Acids Res*, 2003: 31, 307-310.
- [2] Giudicelli V. Conception d'une ontologie en immunogénétique et développement d'un module de cohérence pour le contrôle de qualité de IMGT/LIGM-DB. *Thèse de doctorat*, Université Montpellier II. 1998
- [3] Giudicelli V., Lefranc M.-P. Ontology for Immunogenetics: IMGT-ONTOLOGY. *Bioinformatics*, 1999: 12, 1047-1054.
- [4] Giudicelli V., Chaume D., Lefranc M.-P. *IMGT/LIGM-DB: A systematized approach for ImMunoGeneTics database coherence and data distribution improvement*. Proceedings Sixth International Conference on Intelligent Systems for Molecular Biology ISBM-98, AAAI Press, Menlo Park, CA, 1998: pp. 59-68.
- [5] Giudicelli V., Chaume D., Mennessier G., Althaus H.-H., Müller W., Bodmer J., Malik A., Lefranc M.-P. *IMGT, the international ImMunoGeneTics database: a new Design for Immunogenetics Data Access*. MEDINFO' 1998 (B. Cesnik et al. Eds), IOS Press, Amsterdam, 1998: pp. 351-355.
- [6] Lefranc M.-P., Lefranc G. *The Immunoglobulin FactsBook*. Academic Press, London, UK, 2001: 458 pages.
- [7] Lefranc M.-P., Lefranc G. *The T cell receptor FactsBook*. Academic Press, London, UK, 2001: 398 pages.