Journées Ouvertes de Biologie Informatique et de Mathématiques (JOBIM)
Marseille, France, Juillet 3-6, 2018
Abstracts des Journées, Brun C and Ballester B (eds.), A597, P201, p. 614-617.

IMGT/mAb-DB and IMGT/2Dstructure-DB for IMGT standard definition of an antibody: from receptor to amino acid changes

Mélissa Cambon *† [1], Karima Cherouali [1], Anjana Kushwaha [1], Véronique Giudicelli [1], Patrice Duroux‡ [1], Sofia Kossida§ [1], Marie-Paule Lefranc¶ [1]

[1] IMGT®, the international ImMunoGeneTics information system®, Laboratoire d'ImmunoGénétique Moléculaire (LIGM) – Institut de Génétique Humaine (IGH) CNRS Université de Montpellier UMR 9002 – 141 rue de la Cardonille 34396 Montpellier Cedex 5, France

INTRODUCTION

IMGT®, the international ImMunoGeneTics information system®, http://www.imgt.org [1], is the global reference in immunogenetics and immunoinformatics [2], founded in 1989 by Marie-Paule Lefranc at Montpellier (Université de Montpellier and CNRS). IMGT® is a high-quality integrated knowledge resource specialized in the immunoglobulins (IG) or antibodies, T cell receptors (TR), major histocompatibility (MH) of humans and other vertebrate species, and in the immunoglobulin superfamily (IgSF), MH superfamily (MhSF) and related proteins of the immune system (RPI) of vertebrates and invertebrates.

IMGT® has been built on the IMGT-ONTOLOGY axioms and concepts, which bridged the gap between genes, sequences, and three-dimensional (3D) structures. The concepts include the IMGT® standardized keywords (concepts of identification), IMGT® standardized labels (concepts of description), IMGT® standardized nomenclature (concepts of classification), IMGT unique numbering, and IMGT Colliers de Perles (concepts of numerotation) [2]. IMGT® comprises seven databases, 15,000 pages of web resources, and 17 online tools [1]. Annotated data of IMGT/mAb-DB [3] and IMGT/2Dstructure-DB [4] are being used to generate the IMGT standard definition of an antibody, from receptor to amino acid changes.

METHODOLOGY

Therapeutic proteins found in IMGT/mAb-DB and IMGT/2Dstructure-DB include the IG or antibodies (defined as containing at least one IG variable domain), the fusion protein for immune application (FPIA), the composite protein for clinical application (CPCA) and related protein of the immune system (RPI).

*Speaker
†Corresponding author: melissa.cambon@igh.cnrs.fr
‡Corresponding author: patrice.duroux@igh.cnrs.fr
§Corresponding author: sofia.kossida@igh.cnrs.fr
¶Corresponding author: marie-paule.lefranc@igh.cnrs.fr

Journées Ouvertes de Biologie Informatique et de Mathématiques (JOBIM)
Marseille, France, Juillet 3-6, 2018
Abstracts des Journées, Brun C and Ballester B (eds.), A597, P201, p. 614-617.

IMGT/2Dstructure-DB on-line since 2001 contains 5056 entries of which 3531 are IG with amino acid (AA) sequences from different sources (2821 PDB, 374 INN, 336 Kabat). IMGT/2Dstructure-DB was implemented on the model of IMGT/3Dstructure-DB in order to manage AA sequences of multimeric receptors. Each chain is described in the "Chain details" section which comprises information first on the chain itself, then per domain. Chain and domain annotation includes the IMGT gene and allele names (CLASSIFICATION), region and domain delimitations (DESCRIPTION) and domain AA positions according to the IMGT unique numbering (NUMEROTATION). The closest IMGT® genes and alleles (found expressed in each domain of a chain) and the complementarity determining region (CDR)-IMGT lengths are identified with the integrated IMGT/DomainGapAlign tool [5], which aligns the AA sequences with the IMGT/DomainDisplay AA domain reference sequences [1]. The IMGT reference sequences are acquired by all the upstream work of manual biocuration.

IMGT/mAb-DB, the IMGT database created as an interface for therapeutic proteins, contains 797 entries which comprise 682 IG, 25 FPIA, 44 CPCA and 41 RPI. IMGT/mAb-DB provides the receptor identification in one of the categories (IG, FPIA, CPCA, RPI, and potentially TR and MH if entries become available), links to IMGT/2Dstructure-DB (for entries with AA sequences available) and to IMGT/3Dstructure-DB (for entries with three-dimensional structures available), target name with the HGNC nomenclature (and cross-reference to it), clinical indications, authority decisions and links related to them.

The IMGT standard definition of an antibody can be generated from the IMGT annotated data for the IG entered in both databases, whatever its format (complete IgG, Fab, F(ab')2, scFv...) and whatever its species (*Homo sapiens*, *Mus musculus*, chimeric, humanized...).

RESULTS

Data coming from IMGT/mAb-DB and IMGT/2Dstructure-DB involved in the composition of the sentences of the IMGT standard definition of an antibody includes: specificity, receptor identification, chain identification, positions of domains and disulfide bridges, mutations and the closest IMGT V and J genes and alleles of the amino acid sequences, CDR-IMGT lengths, amino acid changes (polymorphic or engineered).

Currently, the data model (both the Java classes in the implementation of IMGT/mAb-DB and the relational database schema using mapping technology) handles the receptor identification as a linear syntactical construction using terms (basic lexicon) and operators. Parenthesis or brackets grouping (for series-parallel, respectively) are shown with an optional number suffix (including 1 in specific case), and two combinators '-' and '_' for fusion and covalent association between different chains of a receptor. If the substance is a result of a fusion the suffix 'fusion' is added.

Moreover, IMGT/mAb-DB entries are illustrated by a graphical representation which is an "abstract picture" providing visualization of the different categories of the therapeutic substances.

Polymorphic AA changes of the allotypes are described with their position in the IMGT domain and chain and by their position in the sequence. Allotypes are allelic antigenic determinants identified in humans on the immunoglobulin (IG) gamma1, gamma2, gamma3 and alpha2 heavy chains (they are designated as G1m, G2m, G3m and A2m allotypes, respectively), and on the kappa light chain (Km allotypes) [6]. Recently, allotypes regained a lot of attention, owing to the development of therapeutic monoclonal antibodies and their potential immunogenicity. Therapeutic antibodies are most frequently of the IgG1 isotype, and to avoid a potential immunogenicity, the constant region of the gamma1 chains are often engineered to replace the

G1m3 allotype by the "less immunogenic" G1m17 (CH1 R120> K) (G1m17 is more extensively found in different populations) [6]. The description links the allotypes to the IGHG and IGKC genes and alleles, respectively [2].

CONCLUSIONS AND PERSPECTIVES

The IMGT standard definition of an antibody is, per se, a paradigm for any other protein, receptor or ligand, natural, engineered or synthetic. Indeed, antibodies are widely used in clinical applications and for therapeutic purposes owing to their high level of specificity and affinity and to their structure in domains well fitted for antibody engineering and very diverse novel formats.

The main syntactical expressions are necessary and sufficient to provide a standardized IMGT definition schema for any antibody or related receptor type at the receptor, chain, domain and AA level.

They include the species, the IMGT receptor type with an optional complement for radiolabelled or conjugated or fused elements (identification) and a list of chain and domain labels (description) including identity percentage to the closest IMGT gene or allele (classification and numerotation).

Amino acids in the IGHG constant regions of the IG heavy chains are frequently engineered to modify the effector properties of the therapeutic monoclonal antibodies. In order to enrich the description with that information, we recently establish the IMGT engineered variant nomenclature [7] (using the IMGT unique numbering and IMGT chain and domain) for positions of the AA changes involved in antibody-dependent cellular (ADCC), antibody-dependent cellular phagocytosis (ADCP), complement-dependent cytotoxicity (CDC), half-life, reported in the literature. The IMGT engineered variant nomenclature [6] also includes AA changes at positions of interest in antibody engineering: knobs-into-holes AA changes is a rational design strategy, used for heterodimerization of the heavy (H) chains, in the production of bispecific IgG antibodies and controlling half-IG exchange is also a strategy for the generation of bispecific IgG1.

Implementation of Natural Language Processing (NLP) techniques in a Java tool will be considered in order to generate the definition as automatically and accurately as possible.

REFERENCES

Lefranc M-P, Giudicelli V, Duroux P, Jabado-Michaloud J, Folch G, Aouinti S, Carillon E, Duvergey H, Houles A, Paysan-Lafosse T, Hadi-Saljoqi S, Sasorith S, Lefranc G, Kossida S. IMGT®, the international ImMunoGeneTics information system® 25 years on. Nucl. Acids Res., 43(Database issue):D413-22 (2015). PMID: 25378316

Lefranc M-P. Immunoglobulin (IG) and T cell receptor genes (TR): IMGT® and the birth and rise of immunoinformatics. Front Immunol., 5:22 (2014). PMID: 24600447

Poiron C, Wu Y, Ginestoux C, Ehrenmann F, Duroux P and Lefranc M-P. *IMGT/mAb-DB: the IMGT® database for therapeutic monoclonal antibodies. 11èmes Journéees Ouvertes de Biologie, Informatique et Mathématiques (*JOBIM), Montpellier, September 7-9, Abstract 13 (2010).

http://www.sfbi.fr/sites/default/files/jobim/jobim2010/index59385938.html?q=fr/node/55#IMGTmAbDB:_the_IMGT_database_for_therapeutic_monoclonal_antibodies._

Kaas, Q., Ruiz, M. and Lefranc, M.-P. IMGT/3Dstructure-DB and IMGT/StructuralQuery, a database and a tool for immunoglobulin, T cell receptor and MHC structural data. Nucl. Acids. Res., 32:D208-D210 (2004). PMID: 14681396

Ehrenmann F, Kaas Q, Lefranc M-P. IMGT/3Dstructure-DB and IMGT/DomainGapAlign: a database and a tool for immunoglobulins or antibodies, T cell receptors, MHC, IgSF and MhcSF. Nucl. Acids Res., 38:D301-307 (2010). PMID: 19900967

Lefranc M.-P, Lefranc G. Human Gm, Km and Am allotypes and their molecular characterization: a remarkable demonstration of polymorphism. In: B. Tait, F. Christiansen (Eds.), Immunogenetics, chap. 34, Humana Press, Springer, New York, USA. Methods Mol. Biol. 2012; 882, 635-680. PMID: 22665258

Cambon M, Sasorith S, Lefranc M-P. Amino acid positions involved in ADCC, ADCP, CDC, half-life and half-IG exchange. IMGT®, the international ImMunoGenetics information system® http://www.imgt.org.

http://www.imgt.org/IMGTeducation/Tutorials/IGandBcells/_UK/IGproperties/Tableau3.html

Created: 11/12/2012. Version: 12/12/2017

Keywords: IMGT, immunoinformatics, immunogenetics, immunoglobulin, antibody, IMGT/mAb, DB, IMGT/2Dstructure, DB