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

Mélissa Cambon, Karima Cherouali, Anjana Kushwaha, Véronique Giudicelli, Patrice Duroux, Sofia Kossida and Marie-Paule Lefranc
IMGT®, the international ImMunoGeneTics information system®, Laboratoire d'ImmunoGénétique Moléculaire (LIGM),
Institut de Génétique Humaine (IGH), UMR 9002 CNRS-UM, Université de Montpellier (UM), Montpellier (France)

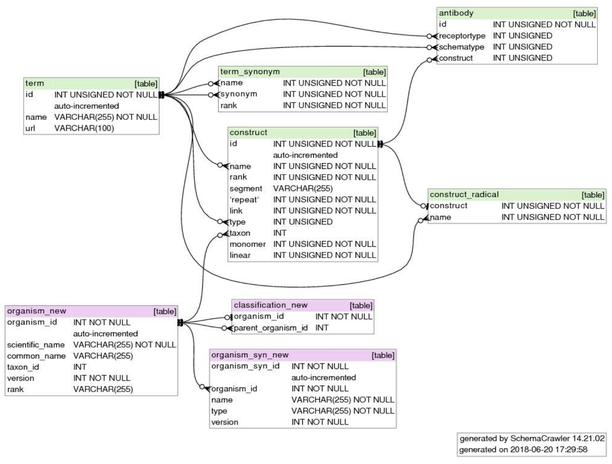
http://www.imgt.org

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 [5] to amino acid changes [6].
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).

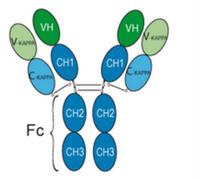
[1] Lefranc M.-P. and al. Nucl. Acids Res. 43:D413-422 (2015) PMID: 25378316 [2] Lefranc M-P. Front. Immunol. 5:22 (2014) PMID: 24600447 [3] Poiron C. et al. Abstract 13, JOBIM Montpellier (2010) [4] Kaas, Q. et al. Nucl. Acids. Res. 32:D208-210 (2004). PMID: 14681396 [5] Ehrenmann F. et al. Nucl. Acids Res., 38:D301-307 (2010). PMID: 19900967 [6] Lefranc M.-P. Lefranc G. Methods Mol. Biol. 2012; 882, 635-680. PMID: 22665258

Data from IMGT/mAb-DB

IMGT/mAb-DB on-line since 2010 is the IMGT® database created as an interface for therapeutic proteins. It contains 807 entries which comprise 690 IG, 25 FPIA, 45 CPCA and 42 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 (3D) structures available), target name with the HGNC nomenclature (and cross-reference to it), clinical indications, authority decisions and links related to them.

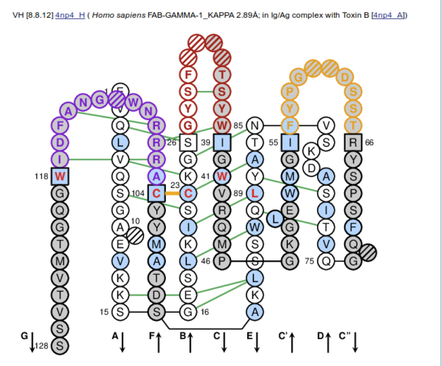


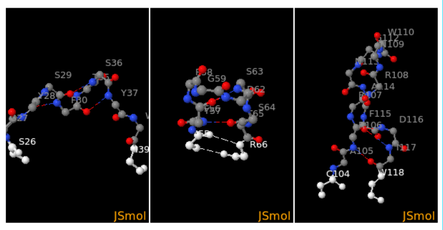
The relational diagram illustrates the part of the data model (both the Java classes in the implementation of IMGT/mAb-DB and the relational database schema using mapping technology) that handles the receptor identification as a linear syntactical construction using terms (basic lexicon) and operators. This enable to formulate all the receptor indentifications present in the database and to face the new ones in a standardized way.

IMGT/mAb-DB ID	411
INN (International Nonproprietary Name)	bezlotoxumab
INN number	9608
INN proposed list	107 (2012)
INN recommended list	69 (2013)
Common name	MK-6072, CDB-1, MDX-1388, MK-3415A (combination of actoxumab and bezlotoxumab), formerly CDB 1
Proprietary name	ZINPLAVA™
Species	*Homo sapiens*
IMGT receptor type	IG
Format (legend)	
Receptor identification	IgG1 - kappa
Radiolabelled/ Conjugated/ Fused	
Origin clone species	
Origin clone name	
IMGT/2Dstructure-DB	9608
IMGT/3Dstructure-DB	4np4
Specificity target name [species]	toxin B [*Clostridium difficile*]

Comparing CDR-IMGT 3D

Using 'Domains and sequence alignment' in 'Display results' it is possible to compare all CDR-IMGT of a selection of V domains showing for each of them the IMGT Colliers de Perles [2], the three 3D loops with the anchors at the bottom in the same plane, and a CDR-IMGT amino acid (AA) table, with at the corresponding IMGT numerotation positions, some physical properties and their contacts (inter and intra chains). This may help to compare the characteristics of the same V domain (in terms of their amino acid sequence) in two different complexes: free of target or bound to one (for instance IMGT/mAb-ID 462).





CDR-IMGT amino acid (AA) table

Anchor	CDR1-IMGT 8 AA								Anchor	Anchor	CDR2-IMGT 8 AA								Anchor	Anchor	CDR3-IMGT 12 AA												Anchor
Positions	26	27	28	29	30	35	36	37	38	39	55	56	57	58	60	62	63	64	65	66	104	105	106	107	108	109	110	112	113	114	115	116	117	118
Amino acid	S	G	Y	S	F	T	S	Y	W	I	I	F	Y	P	G	D	S	S	T	R	C	A	R	R	R	N	W	G	N	A	F	D	I	W
Conformation	Strand	Coil	Coil	Turn	Turn	Turn	Turn	Turn	Coil	Strand	Strand	Strand	Strand	Turn	Turn	Turn	Turn	Strand	Strand	Strand	Strand	Strand	Strand	Strand	Turn	Turn	Turn	Turn	Turn	Coil	Strand	Strand	Strand	Strand
Phi	-149.72	90.99	-145.98	-79.18	-52.47	-80.47	-102.59	-132.74	-78.83	-129.94	-130.83	-141.77	-99.10	-68.51	-75.68	-128.21	64.44	-77.73	-127.42	-144.41	-113.96	-135.88	-99.53	-84.03	-73.67	-91.60	175.07	-126.86	71.64	-72.36	-95.97	-80.07	-119.85	-116.45
Psi	151.46	-2.95	171.73	99.27	-40.05	-0.09	-2.85	142.71	148.80	125.88	161.85	133.99	109.91	-23.13	-13.53	-21.64	16.55	137.38	148.40	133.00	139.08	149.86	143.35	144.28	152.55	26.88	120.59	71.41	10.65	153.92	104.80	-34.28	131.84	151.20
ASA	41.2	60.8	29.3	43.9	5.6	68.6	17.2	8.2	0.4	0.2	0.0	16.0	25.5	15.5	34.8	3.8	51.8	5.2	39.1	59.7	0.0	0.0	17.0	0.0	23.7	3.0	32.3	1.8	0.0	1.8	0.0	36.7	39.3	31.3
H bonds	Q3 (2)			S36			S29		R107	F36	R66 (2)	I39	D62 S64		Y37		Y37		I55 (2)		G119	G40 (2)	I117	W38	A114					R108			R106		

IMGT standard definition

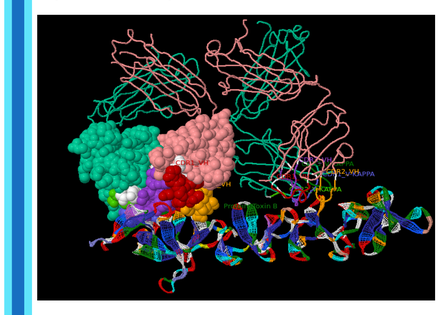
Data coming from IMGT/mAb-DB and IMGT/2Dstructure-2D involved in the composition of the sentences of the IMGT standard definition of an antibody includes: specificity, receptor identification, chain identification, positions of domains and disulfide bridges, mutations and the closest IMGT® V and J genes and alleles of the amino acid sequences, CDR-IMGT lengths, amino acids changes (polymorphic, including allotypes [6], or engineered).

immunoglobulin G1-kappa, anti-[*Clostridium difficile* toxin B)], *Homo sapiens* monoclonal antibody;
gamma1 heavy chain (1-449) [*Homo sapiens* VH (IGHV5-51*01 (94.9%) -(IGHD)-IGHJ3*02 (93.8%) [8.8.12] (1-119) -IGHG1*03, G1m3, nG1m1 (CH1 R120 (216) (120-217), hinge (218-232), CH2 (233-342), CH3 E12 (358), M14 (360) (343-447), CHS (448-449)) (120-449)], (222-215')-disulfide with kappa light chain (1'-215') [*Homo sapiens* V-KAPPA (IGKV3-20*01 (100%) -IGKJ1*01(100%)) [7.3.9] (1'-108') - IGKC*01 Km3 A45.1 (154), V101 (192) (109'-215')];
dimer (228-228'':231-231'')-bisdisulfide

IMGT gene and allele names [2] in black (V,C) or brown (J). Information on allotypes [6] in green. CDR-IMGT lengths [2] in red.

Data from IMGT/3Dstructure-DB

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. The closest IMGT® genes and alleles (identified for 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.



Three-dimensional (3D) representation of two Fab (4np4) of ***bezlotoxumab*** in complex with a target (*Clostridium difficile* toxin B) (Fc are not in the 3D structure). The heavy chain fragment (VH and CH1) and the light chain (V-KAPPA and C-KAPPA) of each Fab are in pink and green, respectively.

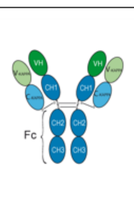
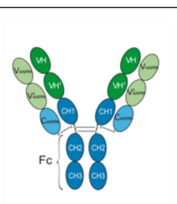
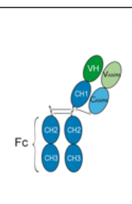
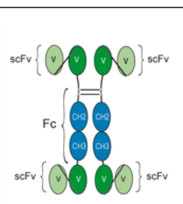
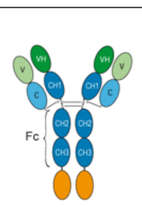
The illustration is given by '3D visualization with Jmol' of IMGT/3Dstructure-DB showing the 'Complex colored by chains and CDRs' with 'spacefill' view for one couple VH and VL and a 'meshribbon' view with 'amino acid type' coloration for the target.

Chain details of **bezlotoxumab Fab**, IG, FAB-GAMMA-1_KAPPA *Homo sapiens* (human) [4np4_H,4np4_L]

Chain ID	**4np4_H**	
Chain length	222	
IMGT chain description	VH-CH1 = VH (1-119) [D1] + CH1 (120-217) [D2]	
Chain sequence	[VH (1-119) [D1] EVQLVQSGAEVKKGESLKISCKGSGYSFTSYWIGWVRQMPGKGLEWMGIFYPGDSDTRYSPSFQGQVTISADKSVNTAYLQWSSLKASD][CH1 (120-211) [D2] TAMYYCARRRNWGNAFDIWGQGTMVTVSSASTKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSSG] LYSLSSVVTVPSSSLGTQTYICNVNHKPSNTKVDKRVEPKSC *Sequence in FASTA format* *Sequence in IMGT format*	
V-DOMAIN	IMGT domain description	VH (1-119) [D1]
	IMGT gene and allele name	IGHV5-51*01 (94.9%)(human), IGHV5-51*03 (94.9%)(human) *Alignment details*
	IMGT gene and allele name	IGHJ3*02 (93.8%)(human) *Alignment details*
	2D representation	IMGT Collier de Perles or IMGT Collier de Perles on 2 layers
	Contact analysis	Domain contacts (overview)
	CDR-IMGT lengths	[8.8.12]
	Sheet composition	[A' B D E] [A" C C' C" F G]
	[CDR1] [CDR2] EVQLVQSGA.EVKKGESLKISCKGSGYSF....TSYWIGWVRQMPGKGLEWMGIFYPG..DSSTRYSPSFQ.GQVTISADKSVNTAYLQ [CDR3] WSSLKASDTAMYYCARRRNW.GNAFDIWGQGTMVTVSS *IMGT/DomainGapAlign results*	
V-DOMAIN	IMGT domain description	V-KAPPA (1-108) [D1]
	IMGT gene and allele name	IGKV3-20*01 (100%)(human) *Alignment details*
	IMGT gene and allele name	IGKJ1*01 (100%)(human) *Alignment details*
	2D representation	IMGT Collier de Perles or IMGT Collier de Perles on 2 layers
	Contact analysis	Domain contacts (overview)
	CDR-IMGT lengths	[7.3.9]
	Sheet composition	[A' B D E] [A" C C' C" F G]
	[CDR1] [CDR2] EIVLTQSPGTLSLSPGERATLSCRASQSVS.....SSYLAWYQQKPGQAPRLLIYGA.......SSRATGIP.DRFSGSG..SGTDFTLT [CDR3] ISRLEPEDFAVYYCQQYGS....STWIFGQGTKVEIK *IMGT/DomainGapAlign results*	

Diversity of formats

The IMGT® standard definition of an antibody can be generated from the IMGT® annoted 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...). In the format description, parentheses or brackets grouping (for series-parallel, respectively) are shown with an optional number suffix (including 1 in specific cases), and two combinators '-' and '_' for fusion and covalent association between different chains of receptor.

Graph									
Identification (IMGT/mAb-DB)	IgG1 - kappa	Fab - G1 - kappa	F(ab')2 - G1 - nd	VH - VL - VH' - VL	VH - VH' - VH	VH -VH' -H-Gamma1 - VL- VL' -C-kappa	Fab - G1 - kappa - [Fc - G1]	[scFv]2 - Fc - [scFv]2	IgG1 - nd - Protein
Description (IMGT/2Dstructure-DB)	IG-GAMMA-1_KAPPA	(VH-CH1)_L-KAPPA	None	SCFV-KAPPA-HEAVY_SCFV-LAMBDA-HEAVY	(VH-VH'-VH)	VH -VH' -H-Gamma1 - VL- VL' -C-kappa	(L-KAPPA_H-GAMMA-1_GAMMA-1-CH2-CH3)	None	None

Conclusion & Perspectives

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. Amino acid changes in the IGHG constant regions of the IG heavy chains, either polymorphic (including allotypes) [6], or engineered to modify the effector properties (ADCC, ADCP, CDC, half-life) or to obtain bispecific (knobs-into-holes) therapeutic monoclonal antibodies, are described using the IMGT unique numbering for C domain (Cambon M., Sasorith S. and Lefranc M.-P., http://www.imgt.org/IMGTeducation/Tutorials/IGandBcells/_UK/IGproperties/Tableau3.html).
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.
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.

IMGT® founder and executive director emeritus: Marie-Paule Lefranc (Marie-Paule.Lefranc@igh.cnrs.fr)
IMGT® director: Sofia Kossida (Sofia.Kossida@igh.cnrs.fr)
Bioinformatics manager: Véronique Giudicelli (Veronique.Giudicelli@igh.cnrs.fr)
Computer manager: Patrice Duroux (Patrice.Duroux@igh.cnrs.fr)