IMGT®, the international ImMunoGeneTics information system® (http://www.imgt.org) provides a standardized way to compare immunoglobulin sequences and to delimit the FR-IMGT and CDR-IMGT in the process of antibody humanization and engineering, whatever the chain type (heavy and light), whatever the species (e.g. murine and human) [1-3]. IMGT® is based on the **IMGT-ONTOLOGY** concepts of classification (IMGT gene and allele nomenclature approved by HGNC and WHO-IUIS), of description (IMGT labels), and of numerotation (IMGT unique numbering). **IMGT/HighV-QUEST** analyses NGS data (150.000 sequences per batch). **IMGT/DomainGapAlign** analyses amino acid sequences per domain, provides the CDR-IMGT and FR-IMGT delimitation and the evaluation of the number of IMGT physicochemical classes changes. **IMGT/mAb-DB** provides links to **IMGT/3Dstructure-DB** (if 3D structures are known) and to IMGT/2Dstructure-DB and IMGT Colliers de Perles (if sequences are available in INN or in the literature). These new tools and database contribute to the invaluable help brought by IMGT® for the analysis of recombinant antibody sequences and further characterization of specificity or potential immunogenicity.

[1] Lefranc M.-P. Mol. Biotechnol. 40, 101-111 (2008)
[2] Lefranc M.-P. et al. Nucleic Acids Res 37, D1008-1012 (2009)
[3] Ehrenmann F. et al. Nucleic Acids Res 38, D301-307 (2010)

The IMGT unique numbering and the IMGT/HighV-QUEST and IMGT/DomainGapAlign tools provide the delimitations of the FR-IMGT and CDR-IMGT for the analysis of antibody paratope or loop grafting in antibody engineering.

The IMGT/mAb-DB with its links to the other IMGT® databases and tools, based on the IMGT-ONTOLOGY concepts, bridge the gap between sequences and 3D structures facilitating the analysis of antibody/antigen interactions.

IMGT/HighV-QUEST, IMGT/DomainGapAlign and IMGT/mAb-DB: novel IMGT® contribution for antibody engineering and humanization

Véronique Giudicelli, Eltaf Alamyar, François Ehrenmann, Claire Poiron, Chantal Ginestoux, Patrice Duroux, Marie-Paule Lefranc

IMGT®, the international ImMunoGeneTics information system®, LIGM, Université Montpellier 2
CNRS UPR1142, IGH, 141 rue de la Cardonille, 34396 MONTPELLIER cedex 05, France
Marie-Paule.Lefranc@igh.cnrs.fr

Im Muno Gene Tics Information system®
http://www.imgt.org

IMGT® is funded by the Centre National de la Recherche Scientifique (CNRS), the Ministère de l'Enseignement Supérieur et de la Recherche, Université Montpellier 2, Agence Nationale de la Recherche BIOSYS (ANR-06-BYOS-0005-01) and the European ImmunoGrid project (IST-2004-0280069)

IMGT founder and director: Marie-Paule Lefranc (Marie-Paule.Lefranc@igh.cnrs.fr)
Bioinformatics manager: Véronique Giudicelli (Veronique.Giudicelli@igh.cnrs.fr)

Computer manager: Patrice Duroux (Patrice.Duroux@igh.cnrs.fr)
Interface design: Chantal Ginestoux (Chantal.Ginestoux@igh.cnrs.fr)

© 2010 F. Ehrenmann and C. Ginestoux
© Copyright 1995-2010 IMGT®, the international ImMunoGeneTics information system®

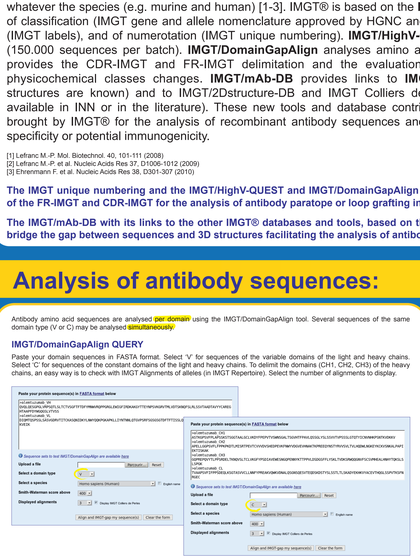
Analysis of antibody sequences:

IMGT/DomainGapAlign NEW

Antibody amino acid sequences are analysed per domain using the IMGT/DomainGapAlign tool. Several sequences of the same domain type (V or C) may be analysed simultaneously.

IMGT/DomainGapAlign QUERY

Paste your domain sequences in FASTA format. Select 'V' for sequences of the variable domains of the light and heavy chains. Select 'C' for sequences of the constant domains of the light and heavy chains. To delimit the domains (CH1, CH2, CH3) of the heavy chains, an easy way is to check with IMGT Alignments of alleles (in IMGT Repertoire). Select the number of alignments to display.



IMGT/DomainGapAlign RESULTS

IMGT/DomainGapAlign identifies the closest germline V-REGION (for 'V') and the closest C-DOMAIN (for 'C'). The IMGT gene and allele name of the closest sequence(s) from the IMGT reference directory is (are) provided with a percentage of identity.

For instance, the V-REGION of the alemtuzumab domains are identified as having 73 % and 86.32 % identity with the *Homo sapiens* IGHV4-59*01 and IGKV1-33*01, respectively.

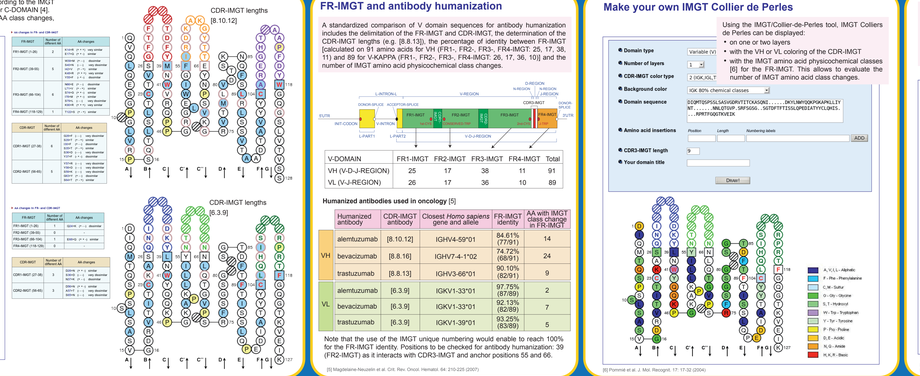
The constant domain of the light chain has 100% identity with *Homo sapiens* IGKC*01. The combined results of CH1 and CH3 of the heavy chain identifies the *Homo sapiens* IGHG1*01 (CH1: K120, CH3: D12, L14) (there is no amino acid difference with IGHG1*02). IGHG1*03 would have been characterized by CH1: R120, CH3: E12, M14.

IMGT nomenclature is based on the CLASSIFICATION axiom and concepts of IMGT-ONTOLOGY [1]. IMGT gene and allele names were approved by the HUGO Nomenclature Committee (HGNC) and by the WHO-IUIS Nomenclature Committee [2]. They are used by Entrez Gene (NCBI), GeneCards, Vega (Wellcome Trust Sanger Institute), Ensembl (EBI)...

[1] Lefranc et al., *In Silico Biology* 5:45-60 (2005)
[2] Lefranc, WHO-IUIS report, Dev. Comp. Immunol. 32:461-463 (2008)

IMGT/DomainGapAlign

IMGT/DomainGapAlign displays your sequences with IMGT gaps and, for 'V', with delimitations of the FR-IMGT and CDR-IMGT, according to the IMGT unique numbering for V-DOMAIN [3], and for 'C', with delimitation of the strands and loops, according to the IMGT unique numbering for C-DOMAIN [4]. Clicking on the 'IMGT Collier de Perles' button gives access to the complete IMGT Collier de Perles with, for V-DOMAIN, the list of AA class changes, compared to the closest germline V and J genes and alleles.



IMGT/Collier-de-Perles

FR-IMGT and antibody humanization

A standardized comparison of V domain sequences for antibody humanization includes the delimitation of the FR-IMGT and CDR-IMGT, the determination of the CDR-IMGT lengths (e.g. [8.8.13]), the percentage of identity between FR-IMGT (calculated on 91 amino acids for VH (FR1-, FR2-, FR3-, FR4-IMGT: 25, 17, 38, 11) and 89 for V-KAPPA (FR1-, FR2-, FR3-, FR4-IMGT: 26, 17, 36, 10)] and the number of IMGT amino acid physicochemical class changes.

V-DOMAIN	FR1-IMGT	FR2-IMGT	FR3-IMGT	FR4-IMGT	Total
VH (V-D-J-REGION)	25	17	38	11	91
VL (V-J-REGION)	26	17	36	10	89

Humanized antibodies used in oncology [5]

	Humanized antibody	CDR-IMGT antibody	Closest *Homo sapiens* gene and allele	FR-IMGT identity	AA with IMGT class change in FR-IMGT
VH	alemtuzumab	[8.10.12]	IGHV4-59*01	84.61% (77/91)	14
	bevacizumab	[8.8.16]	IGHV7-4-1*02	74.72% (68/91)	9
	trastuzumab	[8.8.13]	IGHV3-66*01	90.10% (82/91)	4
VL	alemtuzumab	[6.3.9]	IGKV1-33*01	97.75% (87/89)	2
	bevacizumab	[6.3.9]	IGKV1-33*01	92.13% (82/89)	7
	trastuzumab	[6.3.9]	IGKV1-39*01	93.25% (83/89)	5

Note that the use of the IMGT unique numbering would enable to reach 100% for the FR-IMGT identity. Positions to be checked for antibody humanization: 39 (FR2-IMGT) as it interacts with CDR3-IMGT and anchor positions 55 and 66.

[5] Magdelaine-Beuzelin et al. Crit. Rev. Oncol. Hematol. 64: 210-225 (2007)

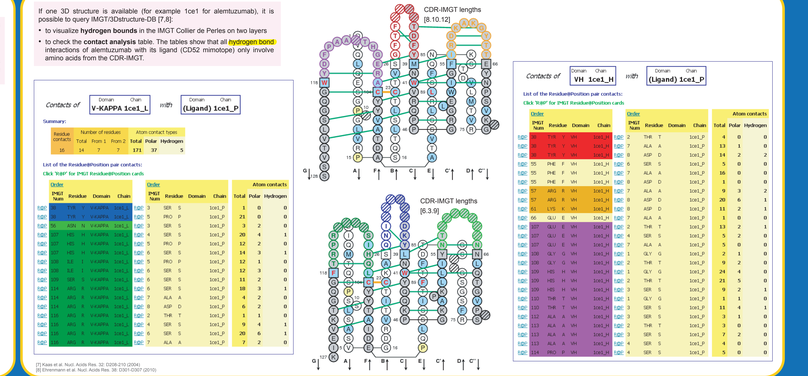
Make your own IMGT Collier de Perles

Using the IMGT/Collier-de-Perles tool, IMGT Colliers de Perles can be displayed:
- on one or two layers
- with the VH or VL coloring of the CDR-IMGT
- with the IMGT amino acid physicochemical classes [6] for the FR-IMGT. This allows to evaluate the number of IMGT amino acid class changes.

[6] Pommié et al. J. Mol. Recogn. 17: 17-32 (2004)

IMGT/3Dstructure-DB contact analysis

If one 3D structure is available (for example 1ce1 for alemtuzumab), it is possible to query IMGT/2Dstructure-DB [7, 8]:
- to visualize **hydrogen bounds** in the IMGT Collier de Perles on two layers
- to check the **contact analysis** table. The tables show that all hydrogen bond interactions of alemtuzumab with its ligand (CD52 mimotope) only involve amino acids from the CDR-IMGT.

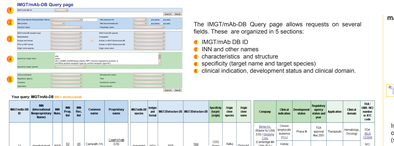


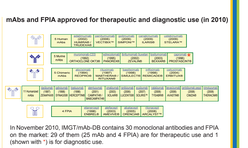
[7] Kaas et al. Nucl. Acids Res. 32: D208-210 (2004)
[8] Ehrenmann et al. Nucl. Acids Res. 38: D301-D307 (2010)

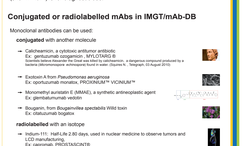
IMGT/mAb-DB NEW

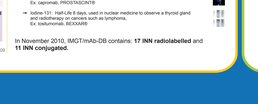
The IMGT® database for therapeutic monoclonal antibodies

IMGT/mAb-DB is the monoclonal antibodies database of IMGT®, the international ImMunoGeneTics information system® (http://www.imgt.org) that is the global reference in immunogenetics and immunoinformatics. IMGT/mAb-DB provides a unique expertised resource on immunoglobulins (IG) or monoclonal antibodies (mAb) with clinical indications, and on fusion proteins for immune applications (FPIA). IMGT/mAb-DB is a relational database using the open source MySQL (http://www.mysql.com) management system database.
Since 2008, amino acid sequences of mAb (suffix -mab) and of FPIA (suffix -cept) from the World Health Organization (WHO)/International Nonproprietary Name (INN) Programme have been entered in IMGT®.

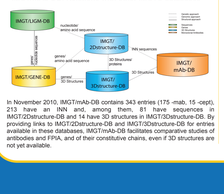






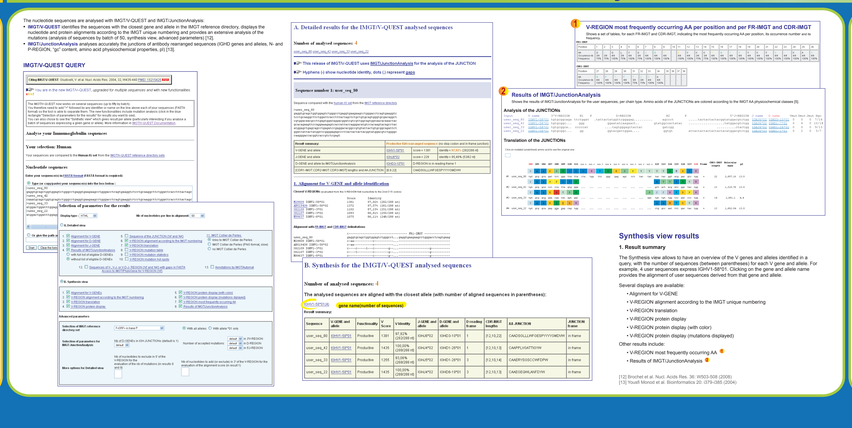


IMGT/mAb-DB: Relations with other IMGT® databases



In November 2010, IMGT/mAb-DB contains 343 entries (175 -mab, 15 -cept), 213 have an INN and, among them, 81 have sequences in IMGT/2Dstructure-DB and 14 have 3D structures in IMGT/3Dstructure-DB. By providing links to IMGT/2Dstructure-DB and IMGT/3Dstructure-DB for entries available in these databases, IMGT/mAb-DB facilitates comparative studies of antibodies and FPIA, and of their constitutive chains, even if 3D structures are not yet available.

IMGT/V-QUEST and IMGT/JunctionAnalysis



IMGT/HighV-QUEST NEW

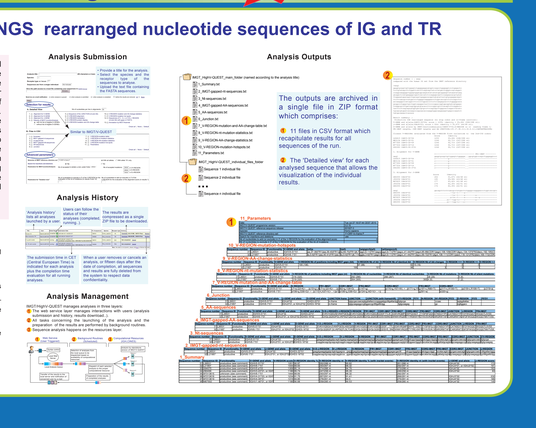
High-Throughput analysis of NGS rearranged nucleotide sequences of IG and TR

IMGT/HighV-QUEST has been developed by IMGT®, the international ImMunoGeneTics Information System® [14], to answer the problematic of the analysis of the antigen receptor data from Next-Generation Sequencing (NGS). The analysis of the expressed repertoires of antigen receptors - immunoglobulins (IG) or antibodies and T cell receptors (TR) - represents a crucial challenge for the study of the adaptive immune response in normal and disease-related situations. IMGT/HighV-QUEST, the high-throughput version of IMGT/V-QUEST, can analyse batches of 150,000 IG and TR rearranged sequences in one run. The tool provides an easy to use and highly customizable interface which is already familiar to IMGT/V-QUEST users. The standardized analysis is based on the IMGT-ONTOLOGY concepts [15,16] and the results are as delivered in IMGT/V-QUEST, including the identification of the variable, diversity and joining genes and alleles, the detailed analysis of the junctions, the analysis of the mutations... The results are formatted in CSV and text in order to facilitate statistical processing.

IMGT/High-V-QUEST is available from the IMGT® Home page (http://www.imgt.org).

IMGT/High-V-QUEST, the high-throughput version of IMGT/V-QUEST can be used for:
- the analysis of repertoire of IG antibody sites and TR recognition sites in normal and pathological situations,
- the identification and characterization of new alleles,
- the qualitative and quantitative analysis of mutations
- the evaluation of the frequency of the variable (V), diversity (D) and joining (J) genes and alleles expressed in the input sequences,
- the analysis (lengths, sequences) of the third complementarity determining region or CDR3-IMGT (which corresponds to the most diverse loop of IG and TR amino acid chains in contact with the antigen).

This work was granted access to the HPC resources of CINES under the allocation 2010-03602 made by GENCI (Grand Equipement National de Calcul Intensif).



[14] Lefranc M.-P. et al., Nucleic Acids Res. 37, 1006-1012 (2009)
[15] Giudicelli, V. and Lefranc, M.-P., Bioinformatics 15, 1047-1054 (1999)
[16] Duroux P. et al., Bioinfomat 90, 570-583 (2008)