The use of **IMGT/DomainGapAlign** and **IMGT/Collier-de-Perles** (for amino acid sequences), **IMGT/V-QUEST** and **IMGT/JunctionAnalysis** (for nucleotide sequences) 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) and whatever the species (e.g. murine and human). Indeed these tools, like the IMGT databases and Web resources, are based on the IMGT-ONTOLOGY concepts of classification (IMGT gene and allele nomenclature approved by HGNC and WHO-IUIS [1,2]), of description (IMGT labels), and of numerotation (IMGT unique numbering for V-DOMAIN [3]). The comparison between V domain sequences include determination of the CDR-IMGT lengths (shown between brackets and separated with dots, e.g. [8.8.13]), percentage of identity between FR-IMGT calculated on 91 amino acids for VH (FR1-: 25, FR2-: 17, FR3-: 38, FR4-IMGT: 11) and 89 for V-KAPPA (FR1-: 26, FR2-: 17, FR3-: 36, FR4-IMGT: 10) and evaluation of the number of IMGT physicochemical classes changes [4,5].

The IMGT unique numbering and 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 unique numbering and tools based on the IMGT-ONTOLOGY concepts bridge the gap between sequences and 3D structures facilitating the analysis of antibody/antigen interactions.

[1] Lefranc and Lefranc. The Immunoglobulin FactsBook, Academic Press (2001)
[2] Lefranc. WHO-IUIS report. Dev Comp Immunol (in press)
[3] Lefranc et al. Dev Comp Immunol 27:55-77 (2003)
[4] Pommié et al. J Mol Recognit 17:17-32 (2004)
[5] Magdelaine-Beuzelin et al. Crit Rev Oncol Hemat 64:210-225 (2007)

IMGT unique numbering and tools for antibody humanization and engineering

Quentin Kaas, François Ehrenmann, Véronique Giudicelli, Patrice Duroux and 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

IMGT Information system® http://imgt.cines.fr

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)

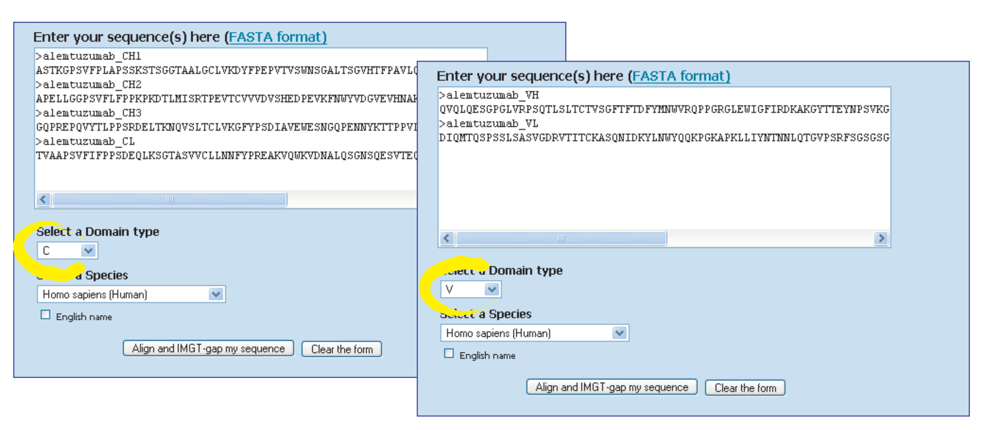
© 2007 F. Ehrenmann and C. Ginestoux
© Copyright 1995-2007 IMGT, the international ImMunoGeneTics information system®

Analysis of antibody amino acid sequences: IMGT/DomainGapAlign IMGT/Collier-de-Perles IMGT/3Dstructure-DB contact analysis

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).



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.

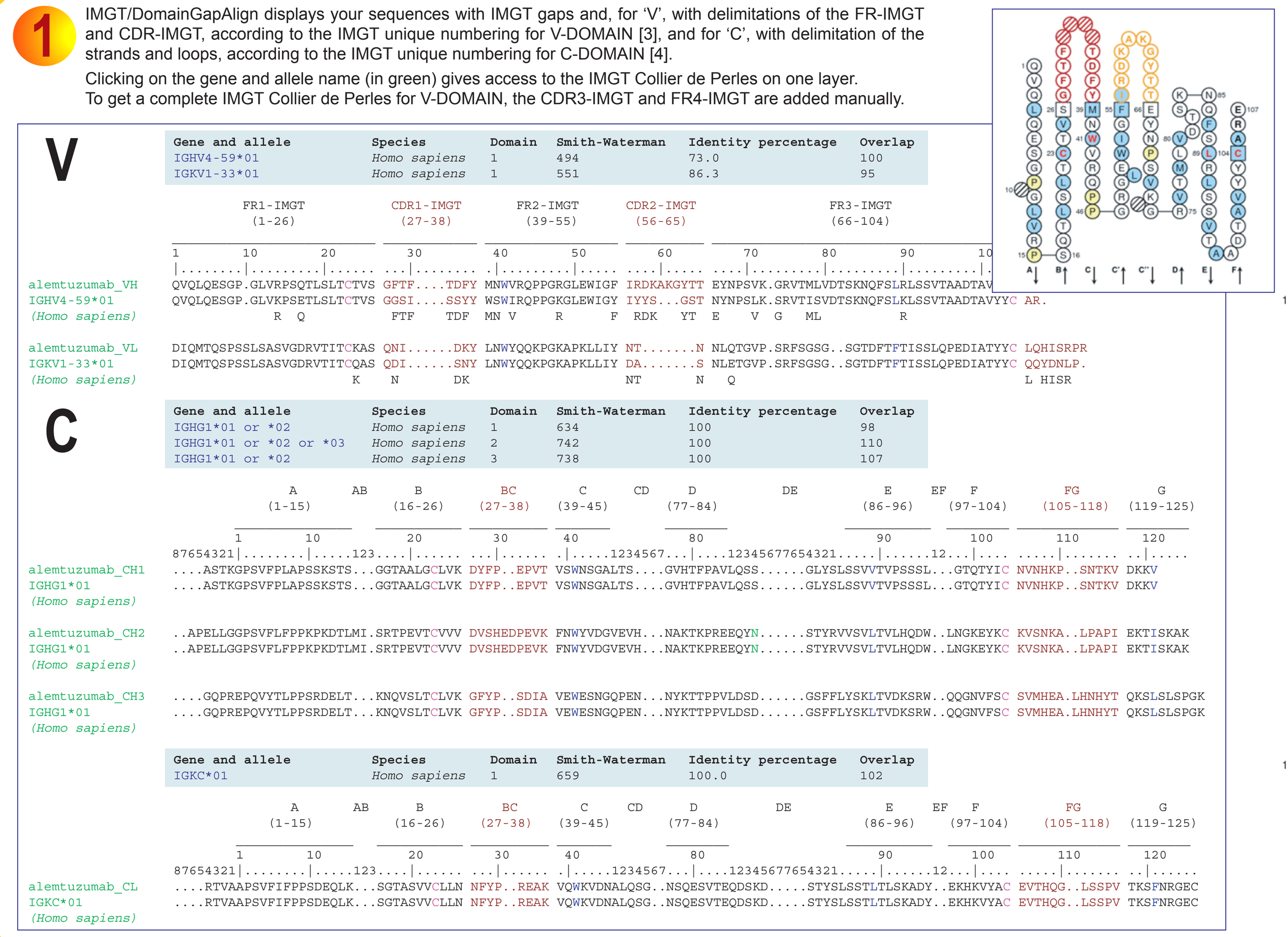
For instance, the V-REGION of the alemtuzumab domains are identified as having 73 % and 86.32 % identity with the *Homo sapiens* IGHV4-59*01 and IGKV1-33*01, respectively.

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.

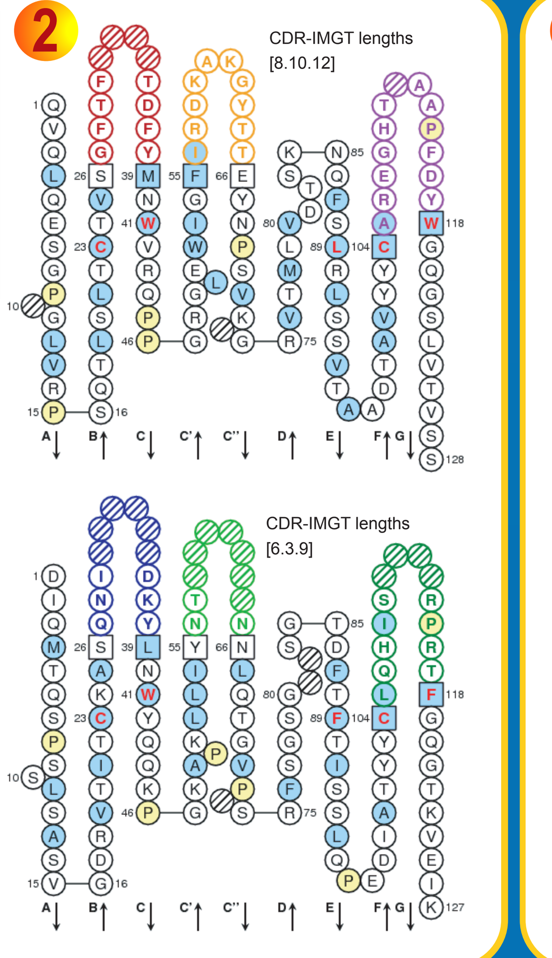
IMGT nomenclature is based on the CLASSIFICATION axiom and concepts of IMGT-ONTOLOGY [1]. IMGT gene and allele names were approved by the HUGO Nomenclature Committee (HGNC) and by the WHO-IUIS Nomenclature Committee [2]. They are used by Entrez Gene (NCBI), GDB, GeneCards, Ensembl (EBI)...

[1] Lefranc et al., *In Silico* Biology, 2005, 5, 45-60, Epub 2005, 5, 0006, 24 Dec 2004.
[2] Lefranc. WHO-IUIS report. Dev Comp Immunol (in press)
[3] Lefranc et al.. Dev Comp Immunol 27:55-77 (2003)
[4] Lefranc et al.. Dev Comp Immunol 29:185-203 (2005)
[5] Pommié et al. J Mol Recognit 17:17-32 (2004)
[6] Kaas et al. Nuc Acids Res 32:D208-D210 (2004)
[7] Magdelaine-Neuzelin et al. Crit Rev Oncol Hematol 64:210-225 (2007)
[8] Giudicelli et al. Nucl Acids Res 32:W435-W440 (2004)
[9] Yousfi Monod et al. Bioinformatics 20:i379-i385 (2004)

1 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 gene and allele name (in green) gives access to the IMGT Collier de Perles on one layer.
To get a complete IMGT Collier de Perles for V-DOMAIN, the CDR3-IMGT and FR4-IMGT are added manually.



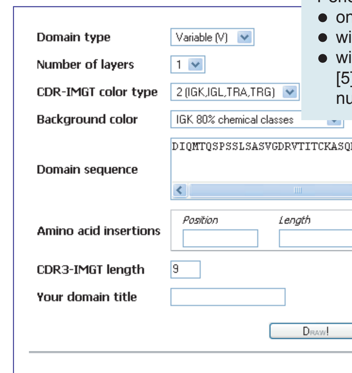
2

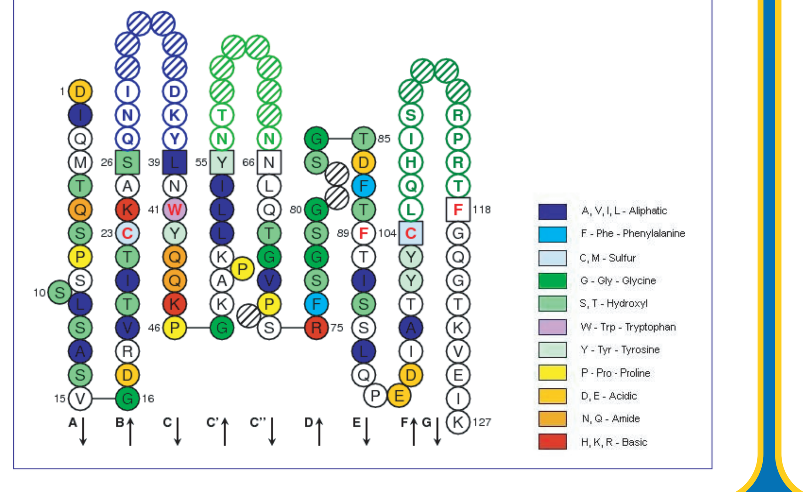


CDR-IMGT lengths
[8.10.12]
CDR-IMGT lengths
[6.3.9]

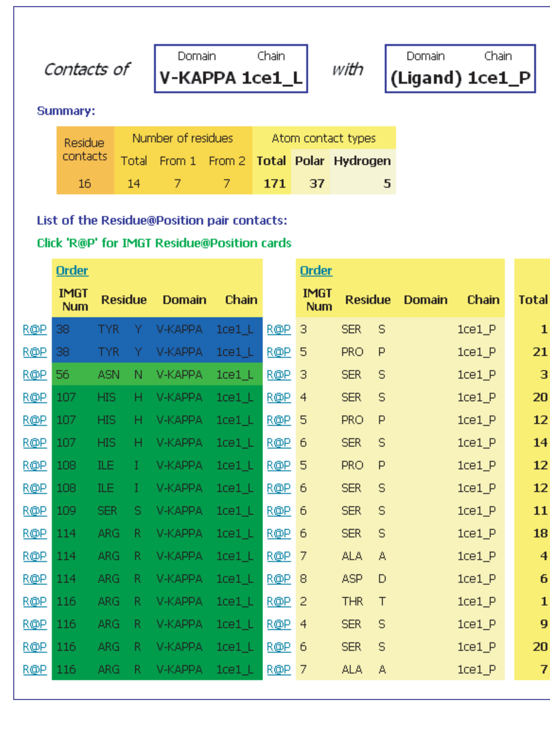
3 Make your own IMGT Collier de Perles

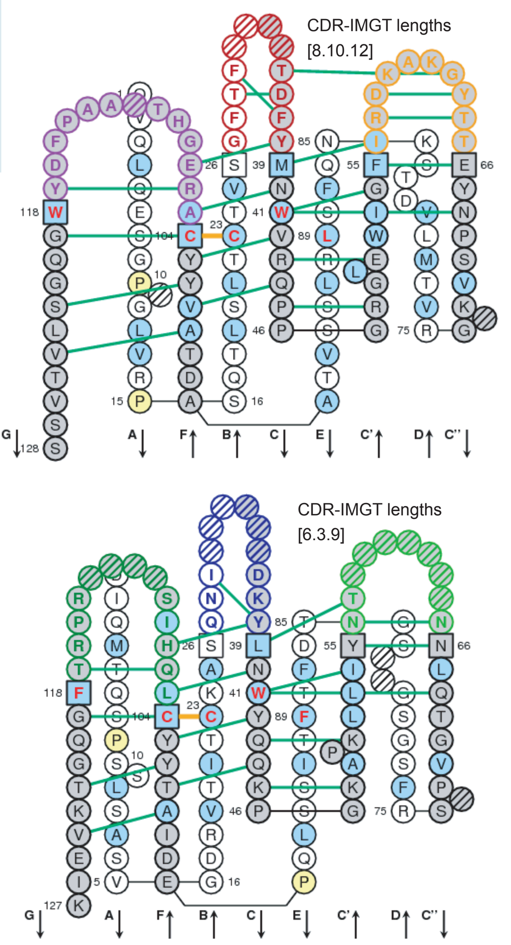
Using the IMGT/Collier-de-Perles tool, IMGT Colliers de Perles can be displayed:
- on one or two layers
- with the VH or VL coloring of the CDR-IMGT
- with the IMGT amino acid physicochemical classes [5] for the FR-IMGT. This allows to evaluate the number of IMGT amino acid class changes.



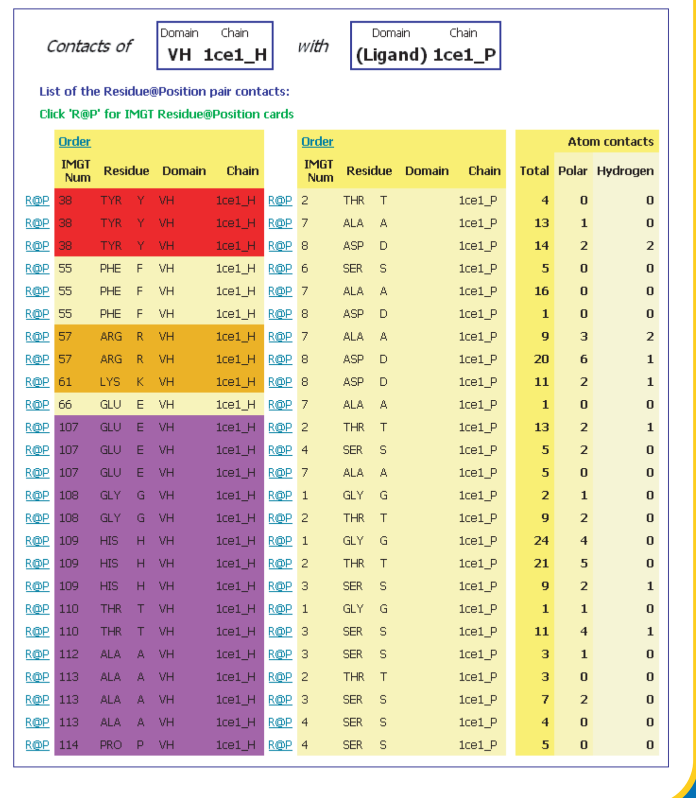


4 If one 3D structure is available (for example 1ce1 for alemtuzumab), it is possible to query IMGT/3Dstructure-DB [6]:
- to visualize **hydrogen bounds** in the IMGT Collier de Perles on two layers
- to check the **contact analysis** table. This table shows that all hydrogen bond interactions of alemtuzumab with its ligand (CD52 mimotope) only involve amino acids from the CDR-IMGT.



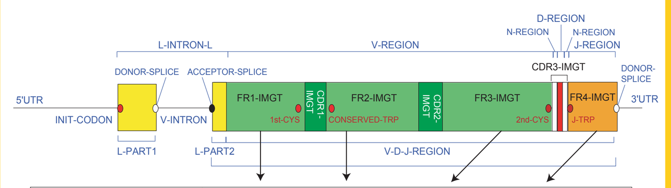


CDR-IMGT lengths
[8.10.12]
CDR-IMGT lengths
[6.3.9]



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 [7]

	Humanized antibody	CDR-IMGT antibody	Closest *Homo sapiens* gene and allele	FR-IMGT identity	AA with IMGT class change
VH	alemtuzumab	[8.10.12]	IGHV4-59*01	84.61% (77/91)	2
	bevacizumab	[8.8.16]	IGHV7-4-1*02	74.72% (68/91)	1
	trastuzumab	[8.8.13]	IGHV3-66*01	90.10% (82/91)	0
VL	alemtuzumab	[6.3.9]	IGKV1-33*01	97.75% (87/89)	1
	bevacizumab	[6.3.9]	IGKV1-33*01	92.13% (82/89)	2
	trastuzumab	[6.3.9]	IGKV1-39*01	93.25% (83/89)	1

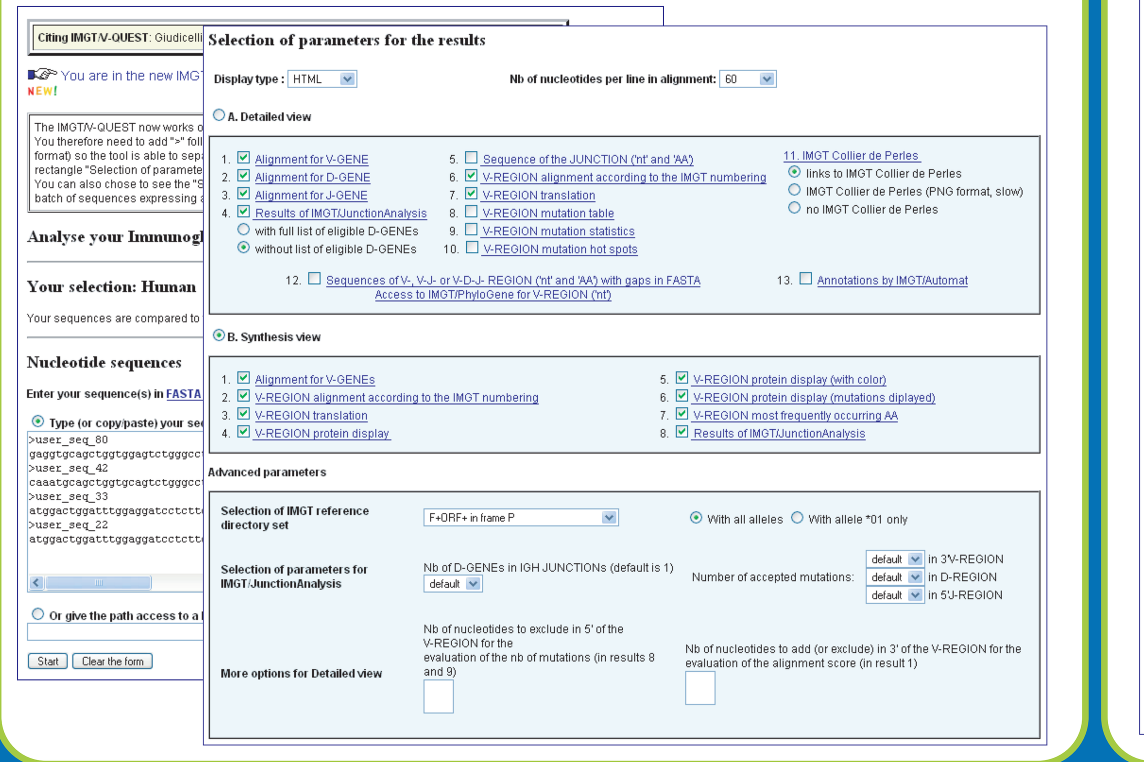
Note that the use of the IMGT unique numbering would enable to reach 100% for the FR-IMGT identity. Positions to be checked for antibody humanization: 39 (FR2-IMGT) as it interacts with CDR3-IMGT and anchor positions 55 and 66.

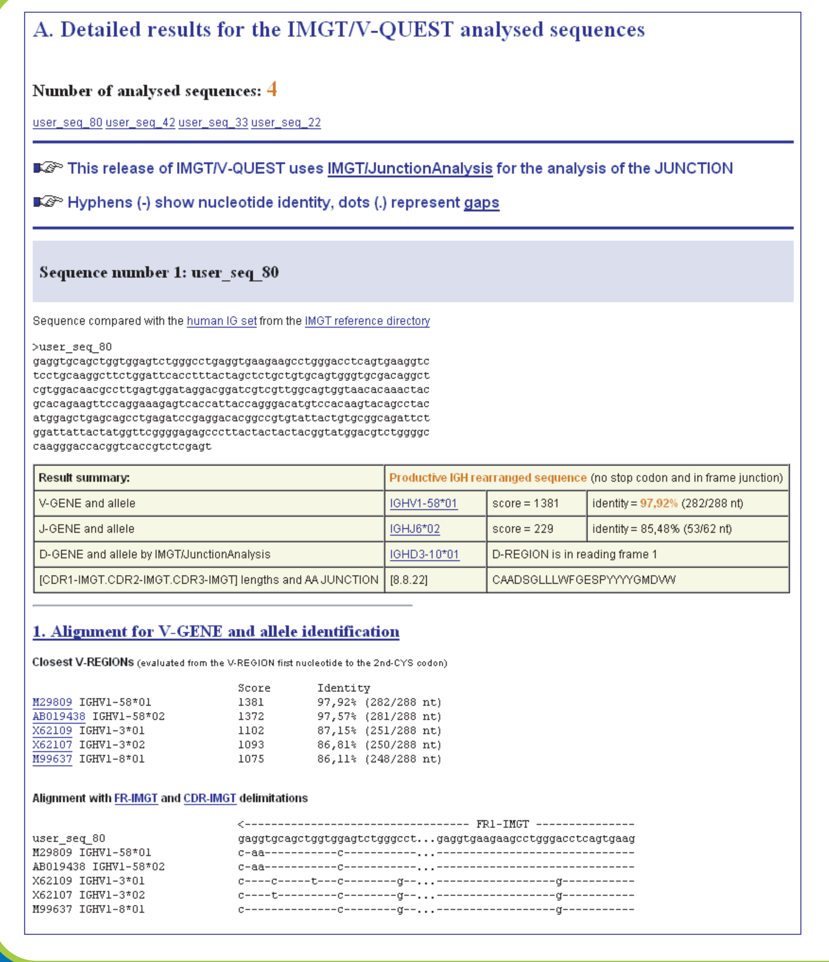
Analysis of antibody nucleotide sequences: IMGT/V-QUEST IMGT/JunctionAnalysis

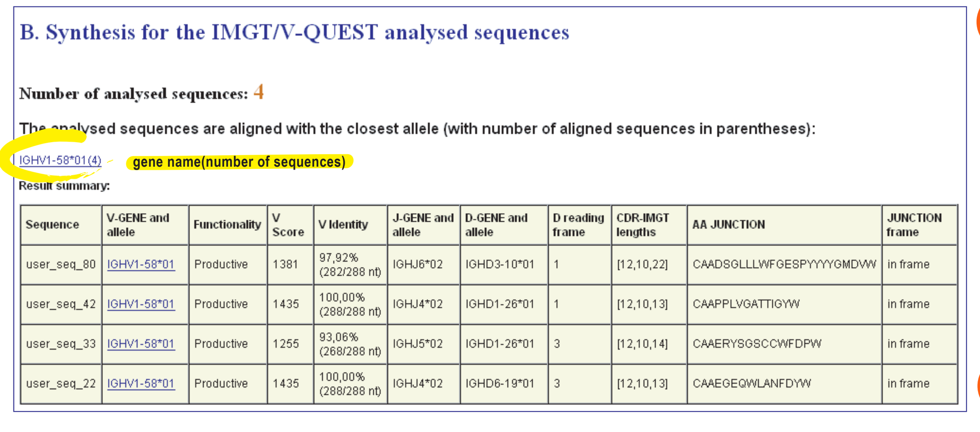
The nucleotide sequences are analysed with IMGT/V-QUEST and IMGT/JunctionAnalysis:
- **IMGT/V-QUEST** identifies the sequences with the closest gene and allele in the IMGT reference directory, displays the nucleotide and protein alignments according to the IMGT unique numbering and provides an extensive analysis of the mutations [8].
- **IMGT/JunctionAnalysis** analyses accurately the junctions of antibody rearranged sequences (IGHD genes and alleles, N- and P-REGION, "gc" content, amino acid physicochemical properties, pI) [9].

New functionalities include:
- analyis of sequences by batch of 50
- synthesis view
- advanced parameters

IMGT/V-QUEST QUERY







Synthesis view results

1. Result summary

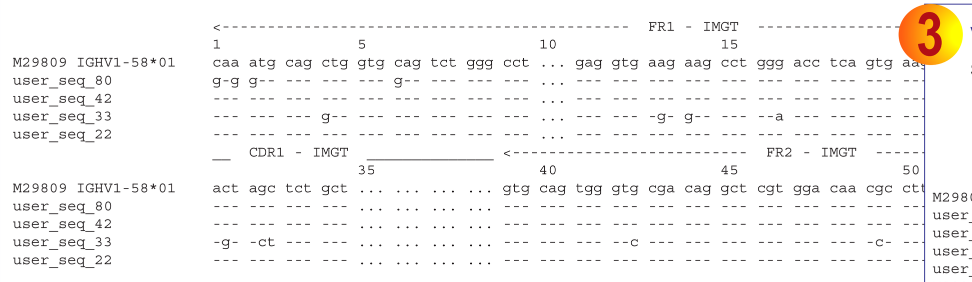
The Synthesis view allows to have an overview of the V genes and alleles identified in a query, with the number of sequences (between parentheses) for each V gene and allele. For example, 4 user sequences express IGHV1-58*01. Clicking on the gene and allele name provides the alignment of user sequences derived from that gene and allele.

Several displays are available:
- Alignment for V-GENE
- V-REGION alignment according to the IMGT unique numbering 1
- V-REGION translation 2
- V-REGION protein display 3
- V-REGION protein display (with color) 4
- V-REGION protein display (mutations displayed)

Other results include:
- V-REGION most frequently occurring AA 5
- Results of IMGT/JunctionAnalysis 6

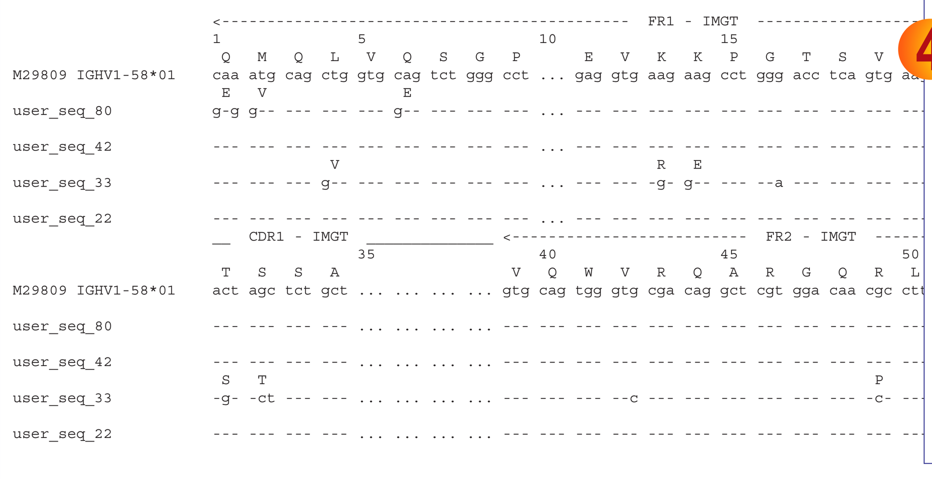
1 V-REGION alignment according to the IMGT unique numbering

Shows the user nucleotide sequences with the IMGT unique numbering and with the FR-IMGT and CDR-IMGT delimitations.



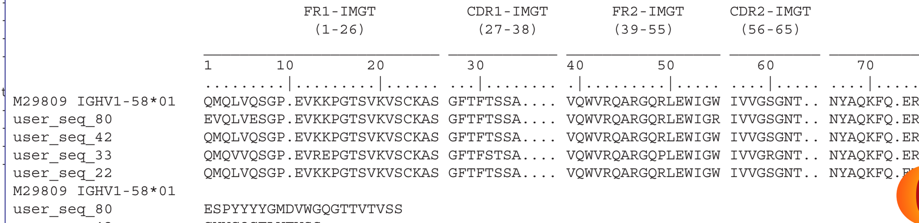
2 V-REGION translation

Shows the user nucleotide sequence and deduced amino acid translation, with the IMGT unique numbering and with the FR-IMGT and CDR-IMGT delimitations.



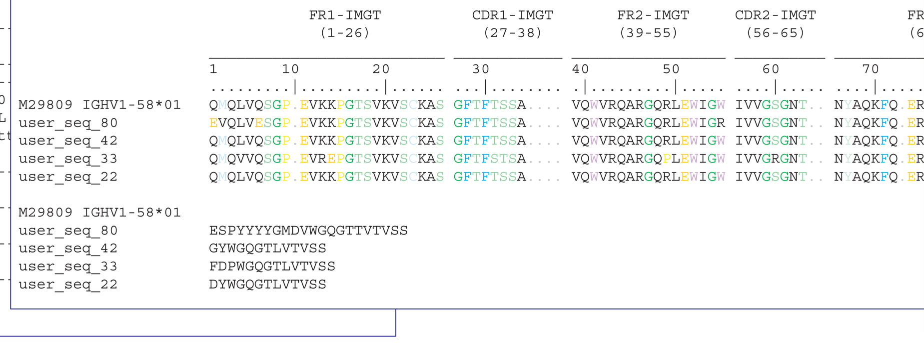
3 V-REGION protein display

Shows the user amino acid sequences aligned with the closest V-REGION allele.



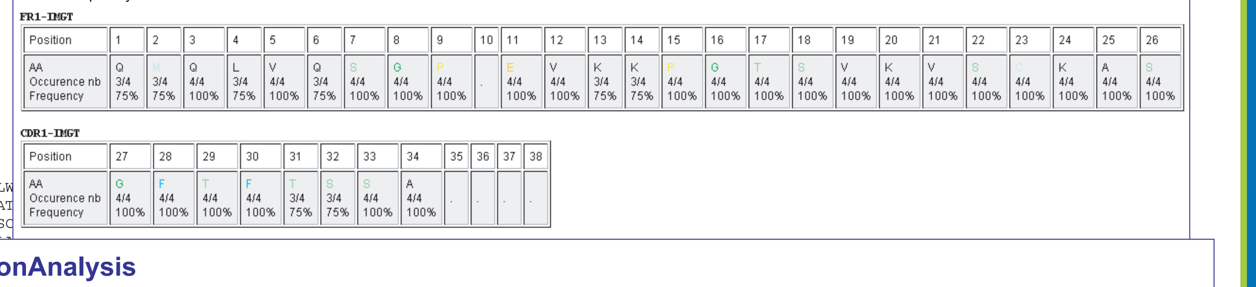
4 V-REGION protein display with colored AA according to the IMGT AA classes

Shows the user amino acid sequences aligned with the closest V-REGION allele, and with colored AA according to the IMGT AA physicochemical classes [5].



5 V-REGION most frequently occurring AA per position and per FR-IMGT and CDR-IMGT

Shows a set of tables, for each FR-IMGT and CDR-IMGT, indicating the most frequently occurring AA per position, its occurrence number and its frequency.



6 Results of IMGT/JunctionAnalysis

Shows the results of IMGT/JunctionAnalysis for the user sequences, per chain type. Amino acids of the JUNCTIONs are colored according to the IMGT AA physicochemical classes [5].

