IMGT/V-QUEST: an algorithm for Immunoglobulin and T cell receptor sequence analysis

Xavier Brochet, Marie-Paule Lefranc, Véronique Giudicelli

IMGT®, the international ImMunoGeneTics information system®, LIGM, Université Montpellier II,
CNRS UPR 1142, IGH 141 rue de la Cardonille, 34396 MONTPELLIER cedex 05, France
Xavier.Brochet@igh.cnrs.fr

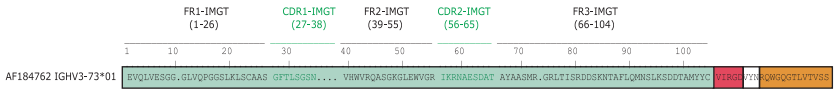
http://imgt.cines.fr

IMGT/V-QUEST (V-QUEry and STandardization) is an integrated alignment tool for the immunoglobulin (IG) and T cell receptor (TR) nucleotide sequences. IMGT/V-QUEST compares germline or rearranged IG or TR variable sequences from users with the IMGT/V-QUEST reference directory sets. The software is based on an original alignment approach which does not accept insertions and deletions. It identifies the closest variable (V), diversity (D) and joining (J) genes and alleles, delimits the framework regions (FR-IMGT) and complementarity determining regions (CDR-IMGT) according to the IMGT unique numbering and localizes the V-GENE mutations in the user sequences.

Objective

User uncharacterized sequence

IMGT/V-QUEST algorithm for sequence analysis



1 Identification of the V-GENE

1/ Identification of the group

Global alignment (without insertion and deletion)

User sequence — Human IG group set

- **Group set** This set contains, for a given species and a given receptor type (IG or TR), a few V-REGION representative sequences with all known different FR-IMGT and CDR-IMGT lengths for each group (IGHV,IGKV,IGLV,TRAV,TRBV, TRGV,TRDV).

The best score identifies the group of user sequence: **IGHV**

Resulting score:

sequence	score
IGHV	1231
IGHV	1228
IGKV	-
IGKV	-
IGLV	-
IGLV	-

The group determines the reference sets to be used in the next alignment and similarity evaluation steps:

*Human gapped IGHV subgroup set for step 1.2
*Human IGHV IMGT reference directory set for step 1.4
*Human IGHJ IMGT reference directory set for step 2
*Human IGHD IMGT reference directory set for step 3

3/ Numbering of the user sequence according to the IMGT unique numbering

Reinsertion of the gaps into the closest IGHV3 sequence according to the IMGT unique numbering

Application of the IMGT unique numbering on user sequence

IMGT unique numbering allows:

- to standardize the description of the V-DOMAINs of IG and TR of all vertebrate species, from fish to human.
- to delimit the framework regions (FR-IMGT) and hypervariable complementarity determining regions (CDR-IMGT), and the lengths of the FR-IMGT and CDR-IMGT
- to always assign the same positions to conserved FR-IMGT amino acids (whatever the antigen receptor (IG or TR), whatever the chain type and/or whatever the species).

Lefranc, M.-P. et al., Dev. Comp. Immunol., 27, 55-77 (2003)

User sequence

The user sequence, numbered according to the IMGT unique numbering, is used in step 1.4

2/ Determination of the V-REGION beginning

Deletion of the gaps from the gapped subgroup set

Gapped subgroup set (Human IGHV)

- **Gapped subgroup set** This set contains, for a given group, representative subgroup V-REGIONs with all known combination of Framework Region (FR) and Complementarity Determining Region (CDR) lengths. These sequences have been numbered according to the IMGT unique numbering.

Global alignment (without insertion and deletion)

User sequence — Gapped subgroup set

Offset

The best alignment determines the start of the V-REGION in the user sequence.

Resulting score:

sequence	score
IGHV3	1235
IGHV1	883
IGHV1	834
IGHV1	768

The best score determines the sequence to be used as a model for the numbering: here **IGHV3**

4/ Identification of the closest V-GENEs and alleles

IMGT/V-QUEST identifies the 5 closest V-GENEs and alleles by similarity evaluation and characterization of the mutations

User sequence with the IMGT unique numbering

IMGT reference directory set (Human IGHV)

- **IMGT reference directory set** This set contains all reference sequences isolated from the functional, open reading frame (ORF) and in frame pseudogene V alleles, that have been gapped according to the IMGT unique numbering.

1. Alignment for V-GENE and allele identification
Closest V-REGIONs (evaluated from the V-REGION first nucleotide to the 2nd-CYS codon)

	Score	Identity
Z27508 IGHV3-73*01	1240	91.49% (269/294 nt)
AB019437 IGHV3-73*02	1231	91.15% (268/294 nt)
X92206 IGHV3-72*01	1024	83.33% (245/294 nt)
X92216 IGHV3-15*01	979	81.63% (240/294 nt)
M99406 IGHV3-15*07	979	81.63% (240/294 nt)

Alignment with FR-IMGT and CDR-IMGT delimitations

2 Identification of the J-GENE

1/ Determination of the J-REGION beginning

IMGT/V-QUEST searches potential J-REGION within 200 nucleotides starting from the 3' end of the V-REGION.
The beginning of the J-REGION is determine with a global alignment (without insertion and deletion) same step as 1.2

User sequence AF184762

IMGT reference directory set (Human IGH) determined in step 1.1

AF184762
V-REGION +1 J-REGION +200
ccatgtattattgtgtgatccggggagatgtttacaaccgacagtggggccaaaccctggtcac

Start of the J-REGION

2/ Identification of the closest J-GENEs and alleles

IMGT/V-QUEST identifies the 5 closest J-GENEs and alleles by similarity evaluation and characterization of the mutations.

User sequence J-REGION determined in step 2.1

IMGT reference directory set (Human IGH) determined in step 1.1

3. Alignment for J-GENE and allele identification
Closest J-REGIONs

	Score	Identity
J00256 IGHJ1*01	179	82.69% (43/52 nt)
X86355 IGHJ4*02	150	79.16% (38/48 nt)
J00256 IGHJ4*01	141	77.08% (37/48 nt)
X86355 IGHJ5*02	138	74.50% (38/51 nt)
M25625 IGHJ4*03	132	75.0% (36/48 nt)

Alignment

3 Identification of the D-GENE

1/ Determination of the D-REGION beginning

IMGT/V-QUEST searches the D-REGION between the 3' end of the V-REGION and the 5' end of the J-REGION.
The beginning of the D-REGION is determined with a global alignment (without insertion and deletion) same step as 1.2

User sequence AF184762

IMGT reference directory set (Human IGHD) determined in step 1.1

AF184762
V-REGION D-REGION J-REGION
ccatgtattattgtgtgatccggggagatgtttacaaccgacagtggggccaaaccctggtcac

Start of the D-REGION

2/ Identification of the closest D-GENEs and alleles

IMGT/V-QUEST identifies the 5 closest D-GENEs and alleles by similarity evaluation and characterization of the mutations.

User sequence D-REGION determined in step 3.1

IMGT reference directory set (Human IGHD) determined in step 1.1

2. Alignment for D-GENE and allele identification
Closest D-REGIONs

	Score	Identity
X13972 IGHD3-10*01	28	72.72% (8/11 nt)
X93615 IGHD3-10*02	19	63.63% (7/11 nt)
J00256 IGHD7-27*01	14	60.0% (6/10 nt)
X97051 IGHD1-16*02	13	62.5% (5/8 nt)
X93614 IGHD3-16*01	10	54.54% (6/11 nt)

Alignment

Conclusion and perspectives

IMGT/V-QUEST provides the detailed analysis of immunoglobulins (IG) and T cell receptors (TR) germline and rearranged sequences which relies on successive global alignments with the IMGT reference directory sets. It also provides a fine analysis of the mutation status of the user sequences. The software works on-line by batches of up to fifty sequences. Combined with the IMGT/JunctionAnalysis, IMGT/V-QUEST is of much value to clinicians and biological scientists studying the repertoire of the immunoglobulins (IG) and T cell receptors (TR) in physiological and pathological immune reponses (autoimmune disease, leukaemias, lymphomas, myelomas, infectious disease...) and for antibody engineering.

Giudicelli, V., Chaume, D. and Lefranc, M.-P., Nucl. Acids Res., 32, W435-440 (2004)

IMGT® is funded by CNRS, MENESR, Université Montpellier II, BIOSTIC-LR2004 Région Languedoc-Roussillon, ACI-IMPBIO IMP82-2004, RNG, ANR (BIOSYS06_135457) and the European ImmunoGrid project (IST-2004-0280069).

© Copyright 1995-2007 IMGT, the international ImMunoGeneTics information system®

©2007 X.Brochet, V. Giudicelli and M.-P. Lefranc