Actes des Journées Ouvertes Biologie, Informatique et Mathématiques, JOBIM 2007, Marseille, France (10-12 Juillet 2007), Brun C. and Didier G. (eds), pp.329-330

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

Xavier Brochet[1], Marie-Paule Lefranc[1, 2] and Véronique Giudicelli[1]

[1]IMGT®, the international ImMunoGeneTics information system®, Laboratoire d'ImmunoGénétique Moléculaire, LIGM, Institut de Génétique Humaine IGH, UPR CNRS 1142, 141 rue de la Cardonille, 34396 Montpellier Cedex 5, France.
{Xavier.Brochet, Veronique.Giudicelli}@igh.cnrs.fr
[2]Institut Universitaire de France, Bd St Michel, Paris, France
Marie-Paule.Lefranc@igh.cnrs.fr

Abstract: *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. IMGT/V-QUEST includes IMGT/JunctionAnalysis for the analysis of the JUNCTION(s).*

Keywords: IMGT, Immunogenetics, Immunoinformatics, IMGT unique numbering, sequence alignment.

1 Introduction

The molecular synthesis of the immunoglobulins (IG) and T cell receptors (TR) [1,2] is particularly complex and unique since it generates an extraordinary diversity of the IG and TR repertoires (10^{12} antibodies and 10^{12} TR per individual) which results from several mechanisms at the DNA level: the combinatorial diversity of the variable (V), diversity (D) and joining (J) genes, the N-diversity and, for IG, the somatic hypermutations. One major challenge for IMGT is to provide users with a detailed and accurate characterization of these sequences according to the IMGT Scientific chart rules, based on the IMGT-ONTOLOGY concepts [3]. This objective is achieved with IMGT/V-QUEST [4] (for "V-QUEry and STandardization") an integrated on-line software program which analyses IG and TR germline or rearranged nucleotide sequences. IMGT/V-QUEST identifies the V, D and J genes and alleles by alignment with the germline IG and TR sequences of the IMGT reference directory sets. It

delimits the framework regions (FR-IMGT) and complementarity determining regions (CDR-IMGT) according to the IMGT unique numbering [5]. Moreover, this approach allows to characterize in detail the mutation status of the sequences.

2 Alignment Algorithm

IMGT/V-QUEST performs different alignments with the IMGT reference directory sets, for each identification step of V, J, and D gene and allele. The algorithm alignment is a global pairwise alignment which does not accept insertions and deletions. Indeed neither insertion or deletion steps are necessary during the alignment process since the gaps are already included in the IMGT reference sequences. These gaps are according to the IMGT unique numbering [5] and allow firstly to keep conserved amino acid codons at the same positions in the V region, and secondly to delimit the FR-IMGT and the CDR-IMGT. The IMGT reference directory sets contain the sequences of all known human and mouse IG and TR genes and alleles that may rearrange. Once the first best alignment is obtained, gaps of the identified closest sequence are entered at the same positions into the input V region sequence. Following this numbering step, IMGT/V-QUEST compares the input sequence with the IMGT directory set which has the highest similarity score and identifies successively the closest germline V, J, D genes and alleles. The alignment algorithm is part of the general algorithm of IMGT/V-QUEST that proceeds through a single run of 5 main sequential steps : 1) identification of the V group of the input sequence, 2) search of the closest V genes and alleles, 3) characterization of the mutations in the V region 4) search of the closest J genes and alleles, 5) search of the closest D genes and alleles.

3 Conclusion and Perspectives

IMGT/V-QUEST provides the detailed analysis of IG and 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 [6], IMGT/V-QUEST is of much value to clinicians and biological scientists studying the repertoire of the immunoglobulins and T cell receptors in physiological and pathological immune responses (autoimmune diseases, leukaemias, lymphomas, myelomas, infectious diseases, etc.) and for antibody engineering.

References

[1] M-P. Lefranc, G. Lefranc, The Immunoglobulin FactsBook. Academic Press, London, UK. 458 pages. 2001.

[2] M-P. Lefranc, G. Lefranc, The T cell receptor FactsBook. Academic Press, London, UK. 398 pages. 2001.

[3] V. Giudicelli, M-P Lefranc, Ontology for immunogenetics: IMGT-ONTOLOGY. Bioinformatics 15, 1047-1054, 1999.

[4] V. Giudicelli, D. Chaume, M-P. Lefranc, IMGT/V-QUEST, an integrated software program for immunoglobulin and T cell receptor V-J and V-D-J rearrangement analysis. Nucleic Acids Res., 33:W435-W440, 2004.

[5] M.-P. Lefranc, C. Pommié , M. Ruiz , V. Giudicelli, E. Foulquier, L. Truong , V. Thouvenin-Contet. and G. Lefranc, IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains. Dev. Comp. Immunol., 27: 55-77, 2003.

[6] M. Yousfi Monod, V. Giudicelli, D. Chaume, M.-P. Lefranc, IMGT/JunctionAnalysis: the first tool for the analysis of the immunoglobulin and T cell receptor complex V-J and V-D-J JUNCTIONs. Bioinformatics. 20: i379-i385, 2004.