# IMGT Sequence Profile: A Standardized Visualization for the Immunoglobulin and T Cell Receptor V-REGIONs Applicable to Other Protein Alignments

**Manuel Ruiz** **Marie-Paule Lefranc**
manu@ligm.igh.cnrs.fr lefranc@ligm.igh.cnrs.fr
Laboratoire d'ImmunoGénétique Moléculaire, LIGM, UPR CNRS 1142 IGH,
141 rue de la Cardonille, 34396 Montpellier Cedex 5, France

**Keywords:** sequence profile, protein alignments

## 1 Introduction

Owing to their fundamental role in the antigen recognition by the immune system, Immunoglobulins (Ig) and T cell Receptors (TcR) are extensively studied. The sequence diversity of the Ig and TcR antigen binding sites is located in the variable domains or V-REGIONs, and results from complex and unique molecular mechanisms which include DNA rearrangements, nucleotide deletions and insertions at the rearranged junctions, and somatic hypermutations. IMGT, the international ImMunoGeneTics database (http://imgt.cines.fr:8104) [3, 4, 6, 12] manages the enormous flow of new Ig and TcR sequences and currently contains 35 600 sequences from 81 different vertebrate species. In order to correlate the Ig and TcR sequences with their structure and specificity, our aim was to develop a standardized visualization of the Ig and TcR V-REGIONs taking into account the expertise provided by IMGT.

## 2 Method and Results

### 2.1 Method

Expertised human germline Ig heavy, kappa, lambda (140 sequences) and TcR alpha, beta, gamma and delta (112 sequences) V-REGIONs were described according to the IMGT Scientific chart rules [1, 2, 5-14], and aligned based on structural data using the IMGT unique numbering [5, 7, 8]. This IMGT set represents the first available complete germline repertoire for the human Ig and TcR. From the alignments, the percentage of each amino acid at each sequence position was determined. Histograms were constructed to represent the sequence profiles, with residues colored according to their hydropathy index. This representation, designated as IMGT Sequence Profile, provides a visual means for assessing the contribution of each amino acid, and the percentage of sequences that contain the major component.

### 2.2 Results

IMGT Sequence Profile is a standardized visualization for the Ig and TcR sequences, based on structural data and using the IMGT numbering. This numbering allows, for the first time, a standardization of the Ig and TcR sequence alignments. It permits to easily compare the V-REGION amino acid characteristics between different receptors (Ig or TcR) and between different chains (heavy or light chains for Ig ; alpha, beta, gamma or delta chains for TcR). It permits also to standardize the description and comparison of sites known to be crucial for the conserved intra- or inter-domain packing, and for the hydrogen bonding in the variable domains. IMGT Sequence Profiles, by exploiting this standardization, underline the amino acid composition differences between these groups, visualize structurally important residues in Ig and TcR sequences, and represent the first complete and the most accurate sequence comparison between human germline Ig and TcR variable domains.

## 3 Conclusion

IMGT Sequence Profiles display in a standardized way the amino acid contribution and characteristics in the Ig and TcR variable domains. Sequence profiles are currently established for Ig and TcR expressed sequences, and for sequences from vertebrate species. IMGT Sequence Profiles can be applied to other protein alignments and more particularly to proteins with Ig-like domains.

## References

[1] Barbié, V. and Lefranc, M.-P., IMGT locus in focus – the human immunoglobulin kappa variable (IGKV) genes and joining (IGKJ) segments, *Exp. Clin. Immunogenet.*, 15:171–183, 1998.

[2] Giudicelli, V. and Lefranc, M.-P., Ontology for immunogenetics: IMGT-ONTOLOGY, *Bioinformatics*, in press, 1999.

[3] Giudicelli, V., Chaume, D. and Lefranc, M.-P., IMGT/LIGM-DB : a systematized approach for ImMunoGeneTics database coherence and data distribution improvement, *International Conference on Intelligent Systems ISBM-98*, 59–68, 1998.

[4] Giudicelli, V., Chaume, D., Mennessier, G., Althaus, H.H., Müller, W., Bodmer, J., Malik, A. and Lefranc, M.-P., IMGT, the international ImMunoGeneTics database: a new design for immunogenetics data access, *MEDINFO'1998* (B. Cesnik et al. Eds), IOS Press, Amsterdam, 351–355, 1998.

[5] Lefranc, M.-P., The IMGT unique numbering for immunoglobulins, T-Cell receptors, and Ig-Like domains, *The Immunologist*, 7:132–136, 1999.

[6] Lefranc, M.-P., Giudicelli, V., Ginestoux, C., Bodmer, J., Müller, W., Bontrop, R., Lemaitre, M., Malik, A., Barbié, V. and Chaume, D., IMGT, the international ImMunoGeneTics database, *Nucleic Acid. Res.*, 27:209–212, 1999.

[7] Lefranc, M.-P., IMGT locus in focus : a new section of Experimental and Clinical Immunogenetics, *Exp. Clin. Immunogenet.*, 15:1–7, 1998.

[8] Lefranc, M.-P., Unique database numbering system for immunogenetic analysis, *Immunology Today*, 18:509, 1997.

[9] Martinez, C. and Lefranc, M.-P., IMGT locus in focus – The mouse (Mus musculus) immunoglobulin kappa variable (IGKV) genes and joining (IGKJ) segments, *Exp. Clin. Immunogenet.*, 15:184–193, 1998.

[10] Pallarès, N., Lefebvre, S., Contet, V., Matsuda, F. and Lefranc, M.-P., IMGT locus in focus – the human immunoglobulin heavy variable (IGHV) genes, *Exp. Clin. Immunogenet.*, 16:36–60, 1999.

[11] Pallarès, N., Frippiat, J.-P., Giudicelli, V. and Lefranc, M.-P., IMGT locus in focus – the human immunoglobulin lambda variable (IGLV) genes and joining (IGLJ) segments, *Exp. Clin. Immunogenet.*, 15:8–18, 1998.

[12] Ruiz, M., Giudicelli, V., Ginestoux, C., Stoehr, P., Robinson , J., Bodmer, J., Marsh, S.G., Bontrop, R., Lemaitre, M., Lefranc, G., Chaume, D., and Lefranc, M.-P., IMGT, the international ImMunoGeneTics database, *Nucleic Acids Res.*, 28:219–221, 2000.

[13] Ruiz, M., Pallarès, N., Contet, V., Barbié, V. and Lefranc, M.-P., IMGT locus in focus - the human immunoglobulin heavy diversity (IGHD) and joining (IGHJ) segments, *Exp. Clin. Immunogenet.*, 16:173–184, 1999.

[14] Scaviner, D., Barbié, V., Ruiz, M., and Lefranc, M.-P., IMGT locus in focus – protein displays of the human immunoglobulin heavy, kappa and lambda variable and joining regions, *Exp. Clin. Immunogenet.*, 16:234–240, 1999.