Unique database numbering system for immunogenetic analysis

The International ImMunoGeneTics (IMGT) database[1] is an integrated online resource specializing in immunoglobulins (Igs), T-cell receptors (TCRs) and major histocompatibility complex (MHC) molecules of all species. IMGT comprises two databases: LIGM-DB (more than 23500 sequences of Igs and TCRs), and MHC/HLA-DB (in development). In order to compare sequences of Igs and TCRs from all species easily, a unique numbering system has been created for IMGT that relies on the high conservation of the variable (V)-region structure. This numbering, set up after aligning more than 3000 sequences, takes into account and combines the definition of the framework (FR) and complementarity-determining regions (CDRs)[2], structural data from X-ray diffraction studies[3] and the characterization of the hypervariable loops[4].

The unique numbering is used as the output of the IMGT/DNAPLOT alignment tool, which analyses rearranged sequences according to IMGT criteria. IMGT/DNAPLOT, freely available from the IMGT home page (Box 1), was programmed by H-H. Althaus and W. Müller at the Institute for Genetics in Köln, for use with the IMGT-specific dataset provided by V. Giudicelli and M-P. Lefranc at the LIGM Laboratoire d'ImmunoGénétique Moléculaire of Montpellier. In IMGT/DNAPLOT, a rearranged sequence is compared with the set of germline IMGT reference sequences (functional or with an open reading frame). The results are aligned with the rearranged sequence, and show the sequences of the most homologous V-gene, D-segment (for heavy chains) and J-segment alleles. The aligned sequences are displayed according to the IMGT unique numbering system and with the FR-IMGT and CDR-IMGT delimitations. The advantages of this system are listed in Box 2.

By facilitating the comparison between the sequences and the description of alleles and mutations, the IMGT unique numbering system represents a considerable advance in the analysis of the Ig and TCR sequences of all species. Moreover, it gives insight into the structural configuration of the V domain and opens interesting views on the evolution of the sequences of the V-set, since this numbering has been applied with success to all the sequences belonging to the V-set of the Ig superfamily, including nonrearranging sequences in vertebrates (e.g. CD4, CTX) and in invertebrates (e.g. *Drosophila* amalgam and fasciclin II).

Box 1. IMGT home page

All information described in this *Update* article is freely available from IMGT at http://imgt.cnusc.fr:8104

Box 2. Advantages of the unique IMGT numbering system

- Easy comparison between sequences coding the V regions, whatever the antigen receptor (Ig or TCR), the chain type (heavy or light chains for Ig; α, β, γ or δ chains for TCR), or the species
- Conserved amino acids always have the same position (e.g. Cys23, Trp41, Leu89, Cys104); the hydrophobic amino acids of the FR regions are also found in conserved positions (see http://imgt.cnusc.fr:8104/textes/numbering/Hum_TRGV9.html for TCR image and http://imgt.cnusc.fr:8104/textes/numbering/Hum_IGLV6.html for Ig image)
- Permits the redefinition of the limits of the FR and CDR; the FR-IMGT and CDR-IMGT lengths provide crucial information that characterizes V regions belonging to a group, a subgroup and/or a gene
- FR amino acids (and codons) located at the same position in different sequences can be compared without requiring sequence alignments; this also holds for amino acids belonging to CDR-IMGT of the same length
- Facilitates the standardization of the description of V-region mutations and allelic polymorphisms; these mutations and allelic polymorphisms are described by comparison to the 'reference sequences' defined in IMGT

Abbreviations: CDR, complementarity-determining region; FR, framework; Ig, immunoglobulin; TCR, T-cell receptor; V, variable.

IMGT is funded by the European Union's BIOMEDI and BIOTECH programmes, the CNRS (Centre National de la Recherche Scientifique) and the MENRT (Ministère de l'Education Nationale, de la Recherche et de la Technologie). Subventions have been received from Association pour la Recherche sur le Cancer, Association de Recherche sur la Polyarthrite, Fondation pour la Recherche Médicale and the Région Languedoc-Roussillon. M-P. Lefranc is the IMGT initiator and coordinator.

Marie-Paule Lefranc (lefranc@ligm.crbm.cnrs-mop.fr) is at the Laboratoire d'Immunogénétique Moléculaire, LIGM, UMR 5535, CNRS, Université Montpellier II, 1919 route de Mende, 34293 Montpellier Cedex 5, France.

References

1 Giudicelli, V., Chaume, D., Bodmer, J. *et al.* (1997) *Nucleic Acids Res.* 25, 206–211
2 Kabat, E.A., Wu, T.T., Reid-Miller, M., Perry, H.M. and Gottesman, K.S. (1987) *Sequences of Proteins of Immunological Interest*, Public Health Service, NIH
3 Satow, Y., Cohen, G.H., Padlan, E.A. and Davies, D.R. (1986) *J. Mol. Biol.* 190, 593–604
4 Chothia, C. and Lesk, A.M. (1987) *J. Mol. Biol.* 196, 901–917