Forum

Here, Marie-Paule Lefranc describes the IMGT database, a unique numbering system for immunoglobulins, TcRs, and immunoglobulin-like domains which represents a step forward in analyzing and comparing the structure and evolution of these molecules.

The IMGT Unique Numbering for Immunoglobulins, T-Cell Receptors, and Ig-Like Domains

Marie-Paule Lefranc

IMGT, the international ImMunoGeneTics database (http://imgt.cnusc.fr:8104) is an integrated database, specializing in immunoglobulins (Ig), T-cell receptors (TcR), and major histocompatibility complex (MHC) molecules of all species that was created in 1989 by Marie-Paule Lefranc (Montpellier II University, CNRS, Montpellier, France) [1, 2]. The IMGT unique numbering defined by Marie-Paule Lefranc for immunoglobulins and T-cell receptors has allowed a redefinition of the limits of the variable region framework (FR-IMGT) and complementarity-determining-regions (CDR-IMGT), leading, for the first time, to a standardized description of mutations, allelic polymorphisms, and structural configurations in 3D representation, whatever the antigen receptor, the chain type, or the species. This approach is highly valuable for comparative analysis and evolution studies of the Ig-like domain.

IMGT comprises three databases: LIGM-DB (more than 25,000 sequences of Ig and TcR), MHC/HLA-DB, and PRIMER-DB (the last two are still under development). In order to easily compare sequences of Ig and TcR from all species, a unique numbering has been defined [3]. The IMGT unique numbering relies on the high conservation of the structure of the variable region. This numbering, set up after aligning more than 5000 sequences, takes into account and combines the definition of the framework (FR) and complementarity determining regions (CDR) [4], structural data from X-ray diffraction studies [5], and the characterization of the hypervariable loops [6].

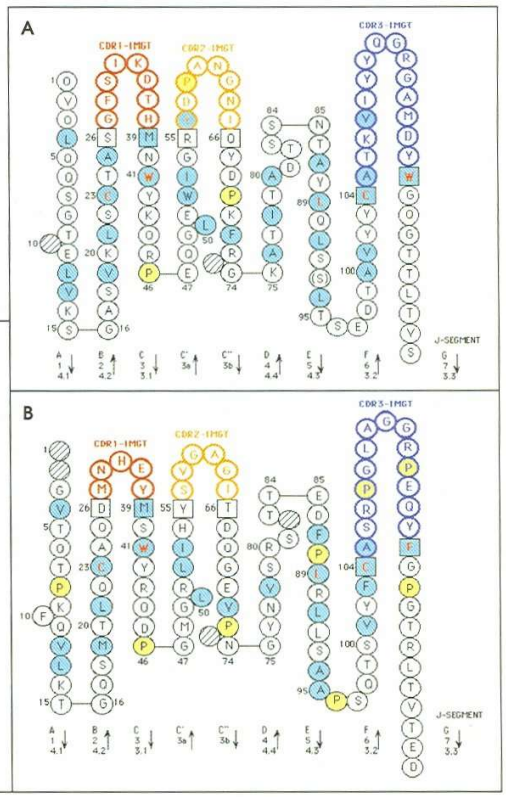

Figure 1. Graphical representation of (**A**) IgH V-D-J region of the mouse antibody E5.2 Fv fragment, and (**B**) TRB V-D-J region of the human TcR αβ A6, with the IMGT numbering, available at the IMGT home page from http://imgt.cnusc.fr:8104. CDR-IMGT lengths are (8.8.15) and (5.6.14), respectively for (**A**) and (**B**). Amino acids are in the one-letter abbrevation. Hydrophobic amino acids (hydropathy index with positive value) and tryptophan residues (W) found at a given position in more than 50% of analyzed Ig and TcR sequences are in blue. All prolines (P) are in yellow. The CDR-IMGT are limited by amino acids shown in squares, which belong to the neighboring FR-IMGT. Cross-hatched circles correspond to missing positions according to the IMGT numbering. Missing positions of the CDRs are not shown in this representation. Arrows indicate the direction of the beta sheets and their different designations in Ig structure. *IMGT modification: IMGT numbering for the graphical representation of the IgH V-D-J region of the mouse antibody E5.2 Fv fragment revealed a conflict between the amino-acid sequence and the atomic coordinate data in the PDB file 1DVF: Leu-Ser-Leu (SEQRES lines) and Leu-Leu-Leu (ATOM lines). The original sequence (8) is Leu-Ser-Ser-Leu (positions 91 to 94 in the IMGT numbering; the serine added at position 93 is shown in parentheses) (8).

The IMGT unique numbering has several advantages:

1. It allows an easy comparison between sequences coding the variable 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.
2. In the IMGT unique numbering, the conserved amino acids always have the same position, for instance cystein 23, tryptophan 41, leucine 89, cystein 104. The hydrophobic amino acids of the framework regions are also found in conserved positions (see Table 1 and Figure 1).
3. This unique numbering has allowed the redefinition of the limits of the FR and CDR. The FR-IMGT and CDR-IMGT lengths themselves become crucial information characterizing the variable regions belonging to a group, a subgroup, and/or a gene (see Appendix).
4. Framework 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.
5. The unique numbering is used as the output of IMGT/DNAPLOT alignment tool which analyzes rearranged sequences according to IMGT criteria. IMGT/DNAPLOT, freely available from the IMGT home page, was developed in collaboration with H.-H. Althaus and W. Müller (Institute for Genetics, Cologne, Germany), for the program, and V. Giudicelli and M.-P. Lefranc (LIGM, Montpellier, France) for the IMGT data set and numbering [7]. In IMGT/DNAPLOT, a rearranged sequence is compared to the set of germline IMGT reference sequences (functional or with an open reading frame). The results show, aligned with the rearranged sequence, the sequences of the most homologous V-gene, D-segment (for heavy chains), and J-segment alleles.
6. The IMGT unique numbering has allowed a standardization of the description of the mutations and allelic polymorphisms of the variable regions. These mutations and allelic polymorphisms are described by comparison to the "reference sequences" defined in IMGT. Allele alignments are displayed according to the IMGT unique numbering and with the FR-IMGT and CDR-IMGT delimitations (see Figure 2).
7. The IMGT unique numbering permits comparison of the CDR-IMGT structural configuration in 3D representations of the Ig and TcR regions (see Figure 3).

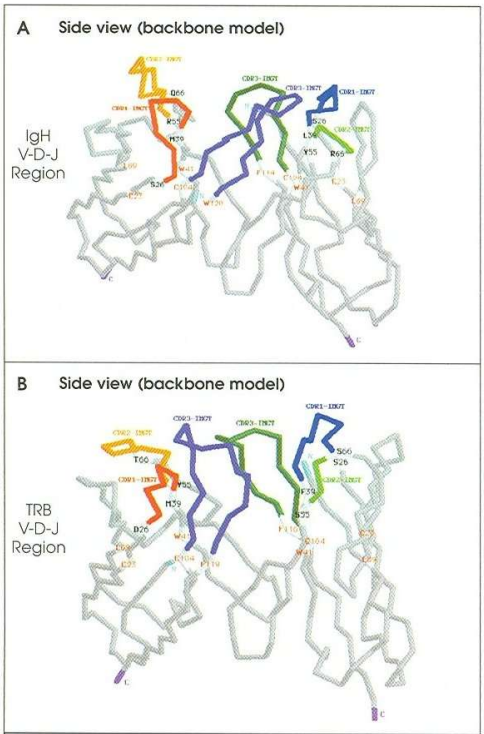

Figure 3. 3D representation of (**A**) mouse antibody E5.2 Fv fragment, and (**B**) human TcR αβ A6 variable regions, with the IMGT numbering, available at the IMGT Marie-Paule page from http://imgt.cnusc.fr:8104. CDR-IMGT regions are colored as follows: IgH and TRB V-D-J region: CDR1-IMGT (red), CDR2-IMGT (orange), CDR3-IMGT (purple), IGK and TRA V-J region: CDR1-IMGT (blue), CDR2-IMGT (green), CDR3-IMGT (green-blue). The conserved C23, W41, and L89, and the FR-IMGT amino acids limiting the CDR-IMGT regions (positions 26 and 39 for the CDR1-IMGT, 55 and 66 for the CDR2-IMGT, 104 and conserved W or F of the J-region for the CDR3-IMGT), are shown in white. These images have been generated with the program RASMOL version 2.6 (9) from the PDB file 1DVF (10), containing the atomic coordinates of the complex between the idiotopic antibody D1.3 Fv fragment and the anti-idiotopic antibody E5.2 Fv fragment (only the E5.2 Fv fragment is displayed here), and from the PDB file 1AO7 (11), containing the atomic coordinates of the complex between human TcR αβ A6, viral peptide Tax, and human MHC molecule HLA-A201 (only the human TcR αβ A6 variable regions are displayed here).

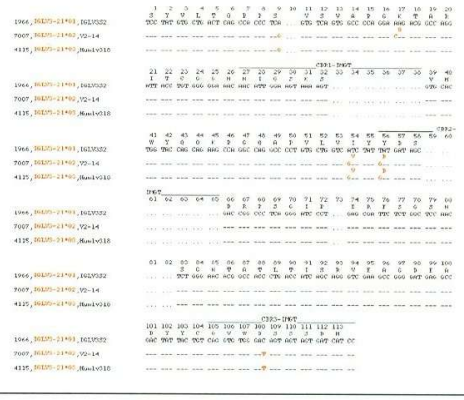

Figure 2. An example of allele alignments: Alignment of the human IGLV3-21 alleles, available at the IMGT home page from http://imgt.cnusc.fr:8104. The aligned sequences are displayed according to the IMGT unique numbering and with the FR-IMGT and CDR-IMGT delimitations. Dots indicate gaps. Dashes indicate similarity.

Table 1. Correspondence Between the Unique IMGT Numbering Defined by M.-P. Lefranc (3) and the Kabat Numberings (4)

	IGH		IGK		IGL		TRBV		TRAV		TRDV		TRGV	
	IMGT	Human IGHV6-1	IMGT	Human IGKV1-5	IMGT	Human IGLV2-23	IMGT	Human TRBV6-5	IMGT	Human TRAV1S3	IMGT	Human TRDV2	IMGT	Human TRGV3
	1	1 cag GLN Q	1	1 gac ASP D	1	1 cag GLN Q	1	1 aat ASN N	1	0 agc ALA A	1	00 gcc ALA A	1	01 tct SER S
	2	2 gta VAL V	2	2 atc ILE I	2	2 tct SER S	2	2 gct ALA A	2	1 cag GLN Q	2	1 att ILE I	2	00 tcc SER S
	3	3 cag GLN Q	3	3 cag GLN Q	3	3 gcc ALA A	3	3 ggt GLY G	3	2 tct SER S	3	2 gag GLU E	3	1 aac ASN N
	4	4 ctg LEU L	4	4 atg MET M	4	4 ctg LEU L	4	4 gtc VAL V	4	3 gtg VAL V	4	3 ttg LEU L	4	2 ttg LEU L
	5	5 cag GLN Q	5	5 acc THR T	5	5 act THR T	5	5 act THR T	5	4 acc THR T	5	4 gtg VAL V	5	3 gaa GLU E
	6	6 cag GLN Q	6	6 cag GLN Q	6	6 cag GLN Q	6	6 cag GLN Q	6	5 cag GLN Q	6	5 cct PRO P	6	4 ggg GLY G
	7	7 tca SER S	7	7 tct SER S	7	7 cct PRO P	7	7 acc THR T	7	6 ctt LEU L	7	6 gaa GLU E	7	5 aga ARG R
	8	8 ggt GLY G	8	8 cct PRO P	8	8 gcc ALA A	8	8 cca PRO P	8	7 gac ASP D	8	7 cac HIS H	8	6 acg THR T
	9	9 cca PRO P	9	9 tcc SER S	9	9 tcc SER S	9	9 aaa LYS K	9	8 agc SER S	9	8 caa GLN Q	9	7 aag LYS K
	10	--- --- -	10	10 acc THR T	10	10 --- --- -	10	10 ttc PHE F	10	9 caa GLN Q	10	9 aca THR T	10	8 tca SER S
FR1-	11	10 gga GLY G	11	11 ctg LEU L	11	11 gtg VAL V	11	11 cag GLN Q	11	10 gtg VAL V	11	10 gtg VAL V	11	9 gtc VAL V
IMGT	12	11 ctg LEU L	12	12 tct SER S	12	12 tct SER S	12	12 gtc VAL V	12	11 cct PRO P	12	11 cct PRO P	12	10 acc THR T
	13	12 gtg VAL V	13	13 gca ALA A	13	13 ggg GLY G	13	13 ctg LEU L	13	12 gtc VAL V	13	12 gtg VAL V	13	11 agg ARG R
	14	13 aag LYS K	14	14 tct SER S	14	14 tct SER S	14	14 aag LYS K	14	13 ttt PHE F	14	13 tca SER S	14	12 cag GLN Q
	15	14 ccc PRO P	15	15 gta VAL V	15	15 cct PRO P	15	15 aca THR T	15	14 gaa GLU E	15	14 ata ILE I	15	13 act THR T
	16	15 tcg SER S	16	16 gga GLY G	16	16 gga GLY G	16	16 gga GLY G	16	15 gaa GLU E	16	15 ggg GLY G	16	14 ggg GLY G
	17	16 cag GLN Q	17	17 gac ASP D	17	17 cag GLN Q	17	17 cag GLN Q	17	16 gcc ALA A	17	16 gtc VAL V	17	15 tca SER S
	18	17 acc THR T	18	18 aga ARG R	18	18 tcg SER S	18	18 agc SER S	18	17 cct PRO P	18	17 cct PRO P	18	16 tct SER S
	19	18 ctc LEU L	19	19 gtc VAL V	19	19 atc ILE I	19	19 atg MET M	19	18 gtg VAL V	19	18 gcc ALA A	19	17 ggt ALA A
	20	19 tca SER S	20	20 acc THR T	20	20 acc THR T	20	20 aca THR T	20	19 gag GLU E	20	19 acc THR T	20	18 gaa GLU E
	21	20 ctc LEU L	21	21 atc ILE I	21	21 atc ILE I	21	21 ctg LEU L	21	20 ctg LEU L	21	20 ctc LEU L	21	19 atc ILE I
	22	21 acc THR T	22	22 act THR T	22	22 tcc SER S	22	22 cag GLN Q	22	21 agg ARG R	22	21 agg ARG R	22	20 act THR T
	23	**22 tgt CYS C**	**23**	**23 tgc CYS C**	**23**	**23 tgc CYS C**	**23**	**23 tgt CYS C**	**23**	**22 tgc CYS C**	**23**	**22 tgc CYS C**	**23**	**21 tgc CYS C**
	24	23 gcc ALA A	24	24 cgg ARG R	24	24 act THR T	24	24 gcc ALA A	24	23 aac ASN N	24	23 tcc SER S	24	22 gat ASP D
	25	24 atc ILE I	25	25 gcc ALA A	25	25 gga GLY G	25	25 cag GLN Q	25	24 tac TYR Y	25	24 atg MET M	25	23 ctt LEU L
	26	25 tcc SER S	26	26 agt SER S	26	26 acc THR T	26	26 gat ASP D	26	25 tca SER S	26	25 aaa LYS K	26	24 act THR T
	27	26 ggg GLY G	27	27 cag GLN Q	27	27 agc SER S	27	27 atg MET M	27	26 tcg SER S	27	26 gga GLY G	27	25 gta VAL V
	28	27 gac ASP D	28	*28 agt SER S	28	*27A agt SER S	28	28 aac ASN N	28	27 tct SER S	28	27 gaa GLU E	28	26 aca THR T
	29	28 agt SER S	29	*29 att ILE I	29	*27B gat ASP D	29	29 cat HIS H	29	28 gtt VAL V	29	28 gcg ALA A	29	27 aat ASN N
	30	29 gtc VAL V	30	*30 agt SER S	30	*27C gtt VAL V	30	30 gaa GLU E	30	29 tca SER S	30	29 atc ILE I	30	28 acc THR T
	31	30 tct SER S	31	*31 agc SER S	31	*28 ggg GLY G	31	*31 tac TYR Y	31	30 gtg VAL V	31	30 ggt GLY G	31	29 ttc PHE F
	32	31 agc SER S	32	*32 tgg TRP W	32	*29 agt SER S	32	* --- --- -	32	*31 tat TYR Y	32	31 aac ASN N	32	30 tac TYR Y
CDR1-	33	32 aac ASN N	33	*--- --- -	33	*30 tat TYR Y	33	--- --- -	33	*--- --- -	33	32 tac TYR Y	33	31 --- --- -
IMGT	34	33 agt SER S	34	*--- --- -	34	*31 aac ASN N	34	--- --- -	34	*--- --- -	34	33 tat TYR Y	34	32 --- --- -
	35	*34 gct ALA A	35	*--- --- -	35	*32 ctt LEU L	35	--- --- -	35	--- --- -	35	--- --- -	35	--- --- -
	36	*35 gct ALA A	36	*--- --- -	36	--- --- -	36	--- --- -	36	--- --- -	36	--- --- -	36	--- --- -
	37	--- --- -	37	*--- --- -	37	--- --- -	37	--- --- -	37	--- --- -	37	--- --- -	37	--- --- -
	38	--- --- -	38	*--- --- -	38	--- --- -	38	--- --- -	38	--- --- -	38	--- --- -	38	--- --- -
	39	*35A tgg TRP W	39	33 ttg LEU L	39	33 gtc VAL V	39	32 atg MET M	39	32 ctc LEU L	39	34 atc ILE I	39	33 atc ILE I
	40	*35B aac ASN N	40	34 gcc ALA A	40	34 tcc SER S	40	33 tcc SER S	40	33 ttc PHE F	40	34A aac ASN N	40	34 cac HIS H
	41	**36 tgg TRP W**	**41**	**35 tgg TRP W**	**41**	**35 tgg TRP W**	**41**	**34 tgg TRP W**	**41**	**34 tgg TRP W**	**41**	**35 tgg TRP W**	**41**	**35 tgg TRP W**
	42	37 atc ILE I	42	36 tat TYR Y	42	36 tac TYR Y	42	35 tat TYR Y	42	35 tat TYR Y	42	36 tac TYR Y	42	36 tac TYR Y
	43	38 agg ARG R	43	37 cag GLN Q	43	37 caa GLN Q	43	36 cga ARG R	43	36 gtg VAL V	43	37 agg ARG R	43	37 cta LEU L
	44	39 cag GLN Q	44	38 cag GLN Q	44	38 cag GLN Q	44	37 caa GLN Q	44	37 caa GLN Q	44	38 aag LYS K	44	38 cac HIS H
FR2-	45	40 tcc SER S	45	39 aaa LYS K	45	39 cac HIS H	45	38 gac ASP D	45	38 tac TYR Y	45	39 acc THR T	45	39 cag GLN Q
IMGT	46	41 cca PRO P	46	40 cca PRO P	46	40 cca PRO P	46	39 cca PRO P	46	39 ccc PRO P	46	40 caa GLN Q	46	40 gag GLU E
	47	42 tcg SER S	47	41 ggg GLY G	47	41 ggc GLY G	47	40 ggc GLY G	47	40 aac ASN N	47	41 ggt GLY G	47	41 ggg GLY G
	48	43 aga ARG R	48	42 aaa LYS K	48	42 aaa LYS K	48	41 atg MET M	48	41 caa GLN Q	48	42 aac ASN N	48	42 aag LYS K
	49	44 ggc GLY G	49	43 gcc ALA A	49	43 gcc ALA A	49	42 ggg GLY G	49	42 gga GLY G	49	43 aca THR T	49	43 gcc ALA A
	50	45 ctt LEU L	50	44 cct PRO P	50	44 ccc PRO P	50	43 ctg LEU L	50	43 ctc LEU L	50	44 atc ILE I	50	44 cca PRO P
	51	46 gag GLU E	51	45 aag LYS K	51	45 aaa LYS K	51	44 agg ARG R	51	44 cag GLN Q	51	45 act THR T	51	45 cag GLN Q
	52	47 tgg TRP W	52	46 ctc LEU L	52	46 ctc LEU L	52	45 ctg LEU L	52	45 ctt LEU L	52	46 ttc PHE F	52	46 cgt ARG R
	53	48 ctg LEU L	53	47 ctg LEU L	53	47 atg MET M	53	46 att ILE I	53	46 ctc LEU L	53	47 ata ILE I	53	47 ctt LEU L
	54	49 gga GLY G	54	48 atc ILE I	54	48 att ILE I	54	47 cat HIS H	54	47 ctg LEU L	54	48 tac TYR Y	54	48 ctg LEU L
	55	50 agg ARG R	55	49 tat TYR Y	55	49 tat TYR Y	55	48 tac TYR Y	55	48 aag LYS K	55	49 cga ARG R	55	49 tac TYR Y
	56	51 aca THR T	56	50 gat ASP D	56	50 gag GLU E	56	49 tca SER S	56	*49 tat TYR Y	56	50 gaa GLU E	56	50 tat TYR Y
	57	52 tac TYR Y	57	51 gcc ALA A	57	51 ggc GLY G	57	50 gtt VAL V	57	*50 tta LEU L	57	51 aag LYS K	57	51 gac ASP D
	58	*52A tac TYR Y	58	52 tcc SER S	58	52 agt SER S	58	51 ggt GLY G	58	*51 tca SER S	58	52 gac ASP D	58	52 gtc VAL V
	59	*52B agg ARG R	59	--- --- -	59	--- --- -	59	52 ggt ALA A	59	*52 gga GLY G	59	* --- --- -	59	53 tcc SER S
CDR2-	60	*53 tcc SER S	60	--- --- -	60	--- --- -	60	53 ggt GLY G	60	--- --- -	60	--- --- -	60	54 acc THR T
IMGT	61	*54 aag LYS K	61	--- --- -	61	--- --- -	61	54 atc ILE I	61	--- --- -	61	--- --- -	61	55 gca ALA A
	62	*55 tgg TRP W	62	--- --- -	62	--- --- -	62	--- --- -	62	--- --- -	62	--- --- -	62	56 agg ARG R
	63	*56 tat TYR Y	63	--- --- -	63	--- --- -	63	--- --- -	63	--- --- -	63	--- --- -	63	*57 gat ASP D
	64	*57 aat ASN N	64	--- --- -	64	--- --- -	64	--- --- -	64	--- --- -	64	--- --- -	64	--- --- -
	65	* --- --- -	65	--- --- -	65	--- --- -	65	--- --- -	65	--- --- -	65	--- --- -	65	--- --- -
	66	58 gat ASP D	66	53 agt SER S	66	53 aag LYS K	66	55 act THR T	66	*53 tcc SER S	66	*53 atc ILE I	66	*58 gtg VAL V
	67	59 tat TYR Y	67	54 ttg LEU L	67	54 cgg ARG R	67	56 gac ASP D	67	*54 acc THR T	67	*54 tat TYR Y	67	*59 ttg LEU L
	68	60 gca ALA A	68	55 gaa GLU E	68	55 ccc PRO P	68	57 caa GLN Q	68	*55 ctg LEU L	68	*55 ggc GLY G	68	*60 gaa GLU E
	69	61 gta VAL V	69	56 agt SER S	69	56 tca SER S	69	58 gga GLY G	69	*56 gtt VAL V	69	*56 cct PRO P	69	*61 tca SER S
	70	62 tct SER S	70	57 ggg GLY G	70	57 ggg GLY G	70	59 gaa GLU E	70	*57 gaa GLU E	70	*57 ggt GLY G	70	*62 gga GLY G
	71	63 gtg VAL V	71	58 gtc VAL V	71	58 gtt VAL V	71	60 gtc VAL V	71	*58 agc SER S	71	*57A ttc PHE F	71	*63 ctc LEU L
	72	64 aaa LYS K	72	59 cca PRO P	72	59 tct SER S	72	61 ccc PRO P	72	*59 atc ILE I	72	*57B aaa LYS K	72	*64 agt SER S
	73	--- --- -	73	--- --- -	73	--- --- -	73	*62 --- --- -	73	* --- --- -	73	*57C --- --- -	73	*64A cca PRO P
	74	65 agt SER S	74	60 tca SER S	74	60 aat ASN N	74	*63 aat ASN N	74	*60 aac ASN N	74	58 gac ASP D	74	*65 gga GLY G
	75	66 cga ARG R	75	61 agg ARG R	75	61 cgc ARG R	75	*64 ggc GLY G	75	61 ggt GLY G	75	59 aat ASN N	75	*66 aag LYS K
	76	67 atc ILE I	76	62 ttc PHE F	76	62 ttc PHE F	76	65 tac TYR Y	76	62 ttt PHE F	76	60 ttc PHE F	76	*67 tat TYR Y
	77	68 acc THR T	77	63 agc SER S	77	63 tct SER S	77	66 aat ASN N	77	63 gag GLU E	77	61 caa GLN Q	77	*68 tat TYR Y
	78	69 atc ILE I	78	64 ggc GLY G	78	64 ggc GLY G	78	67 gtc VAL V	78	64 gct ALA A	78	62 ggt GLY G	78	*69 act THR T
	79	70 aac ASN N	79	65 agt SER S	79	65 tcc SER S	79	68 tcc SER S	79	65 gaa GLU E	79	63 gac ASP D	79	*70 cat HIS H
	80	71 cca PRO P	80	66 gga GLY G	80	66 aag LYS K	80	69 aga ARG R	80	66 ttt PHE F	80	64 att ILE I	80	*71 aca THR T
	81	72 gac ASP D	81	--- --- -	81	--- --- -	81	70 tca SER S	81	67 aac ASN N	81	65 gat ASP D	81	*72 ccc PRO P
	82	73 aca THR T	82	--- --- -	82	--- --- -	82	--- --- -	82	68 aag LYS K	82	66 att ILE I	82	* --- --- -
FR3-	83	74 tcc SER S	83	67 tct SER S	83	67 tct SER S	83	71 acc THR T	83	69 agt SER S	83	67 gca ALA A	83	73 agg ARG R
IMGT	84	75 aag LYS K	84	68 ggg GLY G	84	68 ggc GLY G	84	72 aca THR T	84	70 caa GLN Q	84	68 aag LYS K	84	74 agg ARG R
	85	76 aac ASN N	85	69 aca THR T	85	69 aac ASN N	85	73 gag GLU E	85	71 act THR T	85	69 aac ASN N	85	75 tgg TRP W
	86	77 cag GLN Q	86	70 gaa GLU E	86	70 acg THR T	86	74 gat ASP D	86	72 tcc SER S	86	70 ctg LEU L	86	76 agc SER S
	87	78 ttc PHE F	87	71 ttc PHE F	87	71 gcc ALA A	87	75 ttc PHE F	87	73 ttc PHE F	87	71 gct ALA A	87	77 tgg TRP W
	88	79 tcc SER S	88	72 act THR T	88	72 tcc SER S	88	76 ccg PRO P	88	74 cac HIS H	88	72 gta VAL V	88	78 ata ILE I
	89	80 ctg LEU L	89	73 ctc LEU L	89	73 ctg LEU L	89	77 ctc LEU L	89	75 ttg LEU L	89	73 ctt LEU L	89	79 ttg LEU L
	90	81 cag GLN Q	90	74 acc THR T	90	74 aca THR T	90	78 agg ARG R	90	76 agg ARG R	90	74 aag LYS K	90	80 aga ARG R
	91	82 ctg LEU L	91	75 atc ILE I	91	75 atc ILE I	91	79 ctg LEU L	91	77 aaa LYS K	91	75 ata ILE I	91	81 ctg LEU L
	92	82A aac ASN N	92	76 agc SER S	92	76 tct SER S	92	80 ctg LEU L	92	78 ccc PRO P	92	76 ctt LEU L	92	82 caa GLN Q
	93	82B tct SER S	93	77 agc SER S	93	77 ggg GLY G	93	81 tcg SER S	93	79 tca SER S	93	77 gca ALA A	93	83 aat ASN N
	94	82C gtg VAL V	94	78 ctg LEU L	94	78 ctc LEU L	94	82 gct ALA A	94	80 gtc VAL V	94	78 cca PRO P	94	84 cta LEU L
	95	83 act THR T	95	79 cag GLN Q	95	79 cag GLN Q	95	83 gct ALA A	95	81 cat HIS H	95	79 tca SER S	95	85 atc ILE I
	96	84 ccc PRO P	96	80 cct PRO P	96	80 gct ALA A	96	84 ccc PRO P	96	82 ata ILE I	96	80 gag GLU E	96	86 gaa GLU E
	97	85 gag GLU E	97	81 gat ASP D	97	81 gag GLU E	97	85 tcc SER S	97	83 agc SER S	97	81 aga ARG R	97	87 aat ASN N
	98	86 gac ASP D	98	82 gat ASP D	98	82 gac ASP D	98	86 cag GLN Q	98	84 gac ASP D	98	82 gat ASP D	98	88 gat ASP D
	99	87 acg THR T	99	83 ttt PHE F	99	83 gag GLU E	99	87 aca THR T	99	85 acg THR T	99	83 gaa GLU E	99	89 tct SER S
	100	88 gct ALA A	100	84 gca ALA A	100	84 gct ALA A	100	88 tct SER S	100	86 gct ALA A	100	84 ggg GLY G	100	90 ggg GLY G
	101	89 gtg VAL V	101	85 act THR T	101	85 gat ASP D	101	89 gtg VAL V	101	87 gag GLU E	101	85 tct SER S	101	91 gtc VAL V
	102	90 tat TYR Y	102	86 tat TYR Y	102	86 tat TYR Y	102	90 tac TYR Y	102	88 tac TYR Y	102	86 tac TYR Y	102	92 tat TYR Y
	103	91 tac TYR Y	103	87 tac TYR Y	103	87 tac TYR Y	103	91 ttc PHE F	103	89 ttc PHE F	103	87 tac TYR Y	103	93 tac TYR Y
	104	**92 tgt CYS C**	**104**	**88 tgc CYS C**	**104**	**88 tgc CYS C**	**104**	**92 tgt CYS C**	**104**	**90 tgt CYS C**	**104**	**88 tgt CYS C**	**104**	**94 tgt CYS C**
	105	93 gca ALA A	105	89 caa GLN Q	105	89 tgc CYS C	105	93 gcc ALA A	105	91 gct ALA A	105	89 gcc ALA A	105	95 gcc ALA A
	106	94 aga ARG R	106	90 cga GLN Q	106	90 tca SER S	106	94 agc SER S	106	92 gtg VAL V	106	90 tgt CYS C	106	96 acc THR T
	107	95 --- --- -	107	91 tat TYR Y	107	91 tat TYR Y	107	95 agt SER S	107	93 agt SER S	107	91 gac ASP D	107	97 tgg TRP W
CDR3-	108		108	92 aat ASN N	108	92 gca ALA A	108	96 tat TYR Y			108	92 acc THR T	108	98 gac ASP D
IMGT	109		109	93 agt SER S	109	93 ggt GLY G	109	97 --- --- -					109	99 agg ARG R
	110		110	94 tat TYR Y	110	94 agt SER S								
	111		111	95 tct SER S	111	95 agc SER S								
					112	95A act THR T								
					113	95B tta LEU L								
					114	95C --- --- -								
					115	95D --- --- -								

For each V-region group, one germline sequence is shown with, on the left hand side of each column, the IMGT unique numbering (in bold) defined by M.-P. Lefranc (3) and the corresponding numbering according to Kabat and coworkers (4). Positions of missing amino acids (shown with dashes) are reported to the 3' end of the CDR-IMGT. Asterisks indicate positions for which it is not possible to make changes from one one numbering to the other, automatically. Graphical representations of the sequences in the IMGT unique numbering are available from the IMGT homepage, at http://imgt.cnusc.fr:8104. Human IGHV6-1 (X92224), human IGKV1-5 (Z00001), human IGLV2-23 (X14616), human TRBV6-5 (L36092), human TRAV1S3 (X02850), human TRDV2 (X15207), human TRGV3 (M13430). See Figure 1B for human rearranged TRBV6-5 (PDB:1AO7).

By facilitating comparisons between the sequences and the descriptions of alleles and mutations, the IMGT unique numbering represents a big step forward in the analysis of the Ig and TcR sequences of all species. Indeed, it gives insight into the structural configuration of the variable 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 immunoglobulin superfamily, including nonrearranging sequences in vertebrates (CD4, CTX, etc.) and in invertebrates (*Drosophila* amalgam, *Drosophila* fasciclin II, etc.) (graphical representations available at the IMGT home page from http://imgt.cnusc.fr:8104).

References

1. Giudicelli V, Chaume D, Bodmer J, Müller W, Busin C, Marsh S, Bontrop R, Lemaitre M, Malik A, Lefranc MP. *Nucleic Acids Res* 1997; 25:206–211
2. Lefranc MP, Giudicelli V, Busin C, Bodmer J, Müller W, Bontrop R, Lemaitre M, Malik A, Chaume D. *Nucleic Acids Res* 1998; 26:297–303
3. Lefranc MP. *Immunol Today* 1997; 18:509
4. Public Health Service (Ed.), *Sequences of proteins of immunological interest*. Washington, DC: NIH, 1987
5. Satow Y, Cohen GH, Padlan EA, Davies DR. *J Mol Biol* 1986; 190:593–604
6. Chothia C, Lesk AM. *J Mol Biol* 1987; 196:901–917
7. Giudicelli V, Chaume D, Mennessier G, Althaus HH, Müller W, Bodmer J, Malik A and Lefranc MP. In Cesnik B. et al. (Eds.) *MEDINFO 1998*. Amsterdam: IOS-Press, 1998, pp. 351–355
8. Goldbaum FA, Fields BA, Cauerhff A, Ysern X, Houdusse A, Eisele JL, Poljak RJ, Mariuzza RA. *J Mol Biol* 1994; 241: 739–743
9. Sayle R, Milner-White EJ. *TIBS* 1995; 20:374–376
10. Braden BC, Fields BA, Ysern X, Dall'Acqua W, Goldbaum FA, Poljak RJ, Mariuzza RA. *J Mol Biol* 1996; 22, 264:137–151
11. Graboczi DN, Ghosh P, Utz U, Fan QR, Biddison WE, Wiley DC. *Nature* 1996; 384:134–141

Appendix: Definition of the FR-IMGT and CDR-IMGT Regions

Overview

FR1-IMGT	1–26 (with 1st_CYS 23)
CDR1-IMGT	from 27 (to 38 for the longest one)
FR2-IMGT	39–55 (with CONSERVED_TRP 41)
CDR2-IMGT	from 56 (to 65 for the longest one)
FR3-IMGT	66–104 (with 2nd_CYS 104)
CDR3-IMGT	
– for germline V genes:	from 105 (to 115 for the longest one)
– for rearranged V-D-J genes:	from 105 to J-TRP or J-PHE not included

	FR1-IMGT	CDR1-IMGT	FR2-IMGT	CDR2-IMGT	FR3-IMGT	CDR3-IMGT (for germline V gene)*
Amino acid numbering	1 ÷ 26 (C 23)	27 ÷ 38	39 ÷ 55 (W 41)	56 ÷ 65	66 ÷ 104 (C 104)	105 ÷ 115
Number of amino acids	25–26 aa	5–12 aa	16–17 aa	0–10 aa	36–39 aa	2–11 aa

* + 0, 1, or 2 nucleotide(s) before the heptamer.

For nucleotide position:
- Multiply amino acid number by 3, for third nucleotide in codon.
- Multiply amino acid number by 3, minus 1 for second nucleotide in codon.
- Multiply amino acid number by 3, minus 2 for first nucleotide in codon.

For example: C 23 tgt 67–69; W41 tgg 121–123

Amino acids are shown by capital letters (one-letter code). Nucleotides are shown by small letters.

FR-IMGT Regions

The FR1-IMGT region of Ig and TcR comprises amino acids 1–26 (25–26 amino acids, depending on the variable gene group or subgroup). 1st_CYS is at position 23. The FR2-IMGT region of Ig and TcR comprises amino acids 39–55 (16–17 amino acids). CONSERVED_TRP is at position 41. The FR3-IMGT region of Ig and TcR comprises amino acids 66–104, and comprises 36–39 amino acids depending on the variable gene group or subgroup. 2nd_CYS is at position 104.

CDR-IMGT Regions and Loops

The limits of the CDR1-IMGT region (for Ig and TcR) and those of the corresponding loop H1 (for IGHV genes) or loop L1 (for IGKV and IGLV genes) are identical. The CDR1-IMGT are limited by amino acids 26 and 39, which belong to the FR1-IMGT and FR2-IMGT regions, respectively. The CDR1-IMGT regions encompass amino acids 27 to 38, and comprise from 5 to 12 amino acids depending on the variable gene group or subgroup. This applies to both germline and rearranged genes. The limits of the CDR2-IMGT region (for Ig and TcR) and those of the corresponding loop H2 (for IGHV genes) or loop L2 (for IGKV and IGLV genes) are identical. CDR2-IMGT are limited by amino acids 55 and 66, which belong to the FR2-IMGT and FR3-IMGT regions, respectively. The CDR2-IMGT regions encompass amino acids 56 to 65, and comprise 0 to 10 amino acids, depending on the variable gene group or subgroup. This applies to both germline and rearranged genes. The CDR3-IMGT region of germline V genes starts with amino acid 105, and comprises 2 to 11 amino acids depending on the variable gene group or subgroup. The 3' end of CDR3-IMGT is limited by the heptamer recombination signal. The CDR3-IMGT region of rearranged Ig and TcR V-D-J genes and the corresponding loop H3 (for rearranged IGHV genes) or loop L3 (for rearranged IGKV and IGLV genes) extend from amino acid 105 to J-TRP or J-PHE not included.

Numbering and Sequence Comparison of the CDR-IMGT

Numbering of the CDR-IMGT is based on the longest ones found in the multiple alignments. The longest CDR-IMGT are:

CDR1-IMGT	Position 27 to 38 (12 amino acids)
CDR2-IMGT	Position 56 to 65 (10 amino acids)
CDR3-IMGT	
– for germline V genes	Position 105 to 115 (11 amino acids)
– for rearranged V-J genes or V-D-J genes	Position 105 to position of the amino acid preceding J-TRP or J-PHE

Numbers are missing in shorter CDR-IMGT. Examples are:

- In a CDR1-IMGT with 5 amino acids, positions 27–31 are present, positions 32–38 are missing.
- In a CDR2-IMGT with 3 amino acids, positions 56–58 are present, positions 59–65 are missing.
- For a germline V gene: In a CDR3-IMGT with 2 amino acids, only positions 105 and 106 are present.
- For a rearranged V-J gene or V-D-J gene: CDR3-IMGT extends from position 105 to position of the amino acid preceding J-TRP or J-PHE.

CDR-IMGT are defined by their length (number of amino acids). For a given germline V gene or V gene subgroup, the length of the three CDR-IMGT may be shown, if necessary, into brackets after the gene name or gene subgroup, and separated by dots. For example, IGKV1-6 (6.3.7) means that in the germline IGKV1-6 gene, the CDR1-IMGT, CDR2-IMGT, and CDR3-IMGT regions are 6, 3, and 7 amino acids long, respectively. Sequence comparisons of CDR-IMGT of the same length can be done directly using the numbering. For example, the codon or amino acid at position 26 can be compared to all codons or amino acids at the same position for all CDR1-IMGT of the same length. Sequence comparisons of CDR-IMGT of different lengths need alignment tools.

Acknowledgments. This paper is dedicated to the memory of Sophie Creuzet and Alan F. Williams. I am grateful to Véronique Giudicelli, Denys Chaume, Gérard Lefranc, Valérie Barbié, Chantal Busin, Géraldine Folch, Sophie Lefebvre, Christèle Martinez, Gérard Mennessier, Nathalie Pallarès, Gaëlle Rousseaux, Manuel Ruiz and Dominique Scaviner for helpful discussions. I thank Nathalie Pallarès, Christian Ranaivo-Norbert for preparing the table, Manuel Ruiz for the figures, Nathalie Serre-Berrière for secretarial work, and Valérie Contet for the preparation of the manuscript. IMGT is funded by the European Union's BIOTECH Programme, the CNRS (Centre National de la Recherche Scientifique), and the MENRT (Ministère de l'Education Nationale, de la Recherche et de la Technologie).

Marie-Paule Lefranc is with the Laboratoire d'ImmunoGénétique Moléculaire, LIGM, Institut de Génétique Humaine UPR CNRS 1142, 141 rue de la Cardonille, F-34396 Montpellier Cédex 5, France (tel. +33 4 9961-9965, fax +33 4 9961-9901, e-mail lefranc@ligm.igh.cnrs.fr).