PERGAMON

Developmental and Comparative Immunology 27 (2003) 55–77

Developmental & Comparative Immunology

www.elsevier.com/locate/devcompimm

IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains

Marie-Paule Lefranc*,[1], Christelle Pommié, Manuel Ruiz, Véronique Giudicelli, Elodie Foulquier, Lisa Truong, Valérie Thouvenin-Contet, Gérard Lefranc

IMGT, Laboratoire d'ImmunoGénétique Moléculaire (LIGM), Université Montpellier II, UPR CNRS 1142, IGH, 141 rue de la Cardonille, 34396 Montpellier Cedex 5, France

Received 2 April 2002; revised 24 May 2002; accepted 29 May 2002

Abstract

IMGT, the international ImMunoGeneTics database (http://imgt.cines.fr) is a high quality integrated information system specializing in immunoglobulins (IG), T cell receptors (TR) and major histocompatibility complex (MHC) of human and other vertebrates. IMGT provides a common access to expertly annotated data on the genome, proteome, genetics and structure of the IG and TR, based on the IMGT Scientific chart and IMGT-ONTOLOGY. The IMGT unique numbering defined for the IG and TR variable regions and domains of all jawed vertebrates has allowed a redefinition of the limits of the framework (FR-IMGT) and complementarity determining regions (CDR-IMGT), leading, for the first time, to a standardized description of mutations, allelic polymorphisms, 2D representations (Colliers de Perles) and 3D structures, whatever the antigen receptor, the chain type, or the species. The IMGT numbering has been extended to the V-like domain and is, therefore, highly valuable for comparative analysis and evolution studies of proteins belonging to the IG superfamily.

Keywords: IMGT; Immunoglobulin; T cell receptor; Variable domain; Immunoglobulin superfamily; Numbering; 3D structure; Colliers de Perles

1. Introduction

IMGT, the international ImMunoGeneTics database (http://imgt.cines.fr) [1] is a high quality integrated information system specializing in immunoglobulins (IG), T cell receptors (TR) and major histocompatibility complex (MHC) molecules of human and other vertebrates [1–3]. IMGT provides a common access to expertly annotated data on the genome, proteome, genetics and structure of the IG and TR based on the IMGT Scientific chart and IMGT-ONTOLOGY [4]. The IMGT standardized description of mutations, allelic polymorphisms, 2D and 3D structure representations, is based on a unique numbering system, which can be applied to any antigen receptor, whatever the chain type or the species. In this paper, we describe the IMGT unique numbering for the IG and TR variable regions and domains of all jawed vertebrates and its extension to the IG superfamily V-like domains.

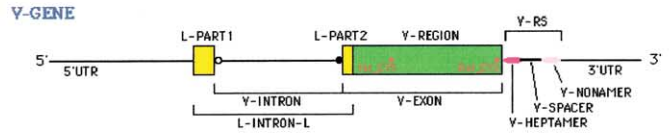
* Corresponding author. Tel.: +33-4-99-61-99-65; fax: +33-4-99-61-99-01.

E-mail address: lefranc@ligm.igh.cnrs.fr (M.-P. Lefranc).

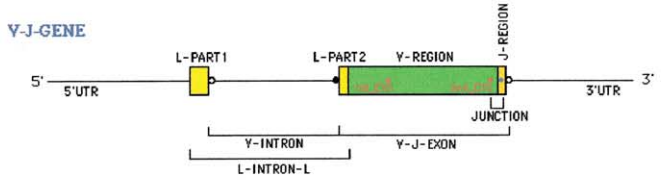
[1] IMGT site: http://imgt.cines.fr

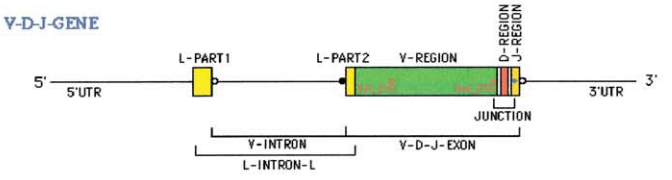
0145-305X/03/$ - see front matter
PII: S0145-305X(02)00039-3

Germline Genomic DNA

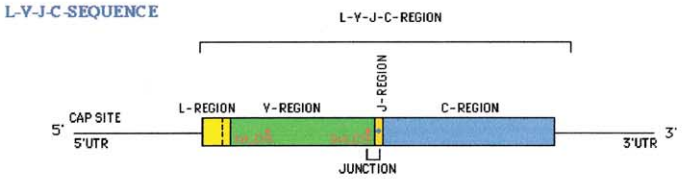


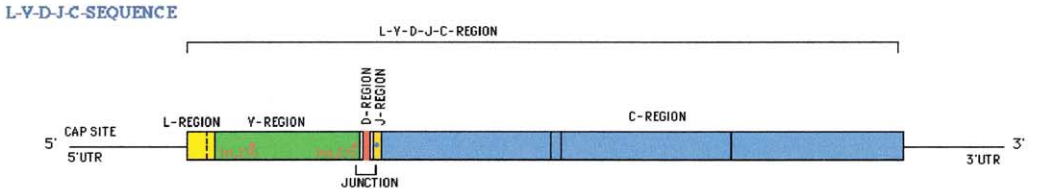
Rearranged Genomic DNA





Rearranged cDNA





Legend

- DONOR-SPLICE
- ACCEPTOR-SPLICE
- J-TRP or J-PHE
- N regions

2. IMGT unique numbering for V-REGION

2.1. Overview

In order to easily compare the variable regions (V-REGION) of IG and TR from all species, a unique numbering has been defined [5,6]. This 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) [7], structural data from X-ray diffraction studies [8], and the characterization of the hypervariable loops [9]. The delimitations of the FR- and CDR-IMGT regions have been defined, based on the longest ones found in the IMGT/LIGM-DB multiple alignments. In a first step, the IMGT unique numbering has been set up for the V-REGION that is up to the V-HEPTAMER of germline genes (including the germline CDR3-IMGT) [10,11], and up to the 2nd-CYS of rearranged V–J and V–D–J genes (Fig. 1). In the IMGT unique numbering, the conserved amino acids always have the same position, for instance Cysteine 23 (1st-CYS), Tryptophan 41 (CONSERVED-TRP), Leucine 89, Cysteine 104 (2nd-CYS) (Table 1). The hydrophobic amino acids of the framework regions are also found in conserved positions. The IMGT unique numbering is used in protein displays [12,13] and in 2D representations designated as Colliers de Perles (Fig. 2) which are available for all the human germline functional and open reading frame (ORF) IG genes [10] and TR genes [11], and annotated genes from other species (mouse, teleostei, etc.) (IMGT repertoire [14–16], http://imgt.cines.fr). Correspondence between the IMGT unique numbering and the other numberings has been established for the IG and TR V-REGIONs [6,10,11] (IMGT Scientific chart, http://imgt.cines.fr). Table 2 shows the correspondence between the IMGT unique numbering and the Kabat numberings.

2.2. FR-IMGT regions

The FR1-IMGT region of IG and TR comprises positions 1–26 (25–26 amino acids, depending on the V-GENE group or subgroup) with 1st-CYS at position 23 (Tables 1 and 2; Fig. 2). The FR2-IMGT region of IG and TR comprises positions 39–55 (16–17 amino acids) with CONSERVED-TRP at position 41. The FR3-IMGT region of IG and TR comprises positions 66–104 (36–39 amino acids, depending on the V-GENE group or subgroup) with a conserved hydrophobic amino acid at position 89 and the 2nd-CYS at position 104.

2.3. CDR-IMGT regions

The CDR1-IMGT region of IG and TR comprises positions 27–38 (5–12 amino acids, depending on the V-GENE group or subgroup; Tables 1 and 2). The longest CDR1-IMGT have 12 amino acids. For shorter CDR1-IMGT, less than 12 amino acids, gaps are created (missing positions, hatched in Collier de Perles (Fig. 2), or not shown in structural data representations). As an example, in a CDR1-IMGT with eight amino acids, positions 27–34 are present, positions 35–38 are missing (Fig. 2). The unique concept of 'CDR1-IMGT' replaces the diverse definitions and delimitations of the 'CDR1', 'H1 loop' and 'L1 loop' found in the literature (Table 3). The CDR2-IMGT region of IG and TR comprises positions 56–65 (0–10 amino acids, depending on the V-GENE group or subgroup; Tables 1 and 2). The longest CDR2-IMGT has 10 amino acids. For shorter CDR2-IMGT, less than 10 amino acids, gaps are created (missing positions, hatched in Collier de Perles, or not shown in structural data representations). As an example, in a CDR2-IMGT with eight amino acids, positions 56–63 are present, positions 64 and 65 are missing (Fig. 2). The unique concept of 'CDR2-IMGT' replaces the diverse definitions and delimitations of the 'CDR2', 'H2 loop' and 'L2 loop' found in the literature (Table 3). The CDR3-IMGT region of germline IG and TR V-GENEs comprises position 105–116 (2–12 amino acids, depending on

Fig. 1. V-REGION representations in germline genomic DNA, rearranged genomic DNA, and rearranged cDNA. IMGT standardized labels are in capital letters and are described in the IMGT Scientific chart (http://imgt.cines.fr).

Table 1
Definition of the FR- and CDR-IMGT

	FR1-IMGT	CDR1-IMGT	FR2-IMGT	CDR2-IMGT	FR3-IMGT	Specific to V-REGION of germline V-GENEs	Specific to V-DOMAIN; for rearranged V–J-GENEs and V–D–J-GENEs	
						Germline CDR3-IMGT (1)	Rearranged CDR3-IMGT (2)	FR4-IMGT (2)
Amino acid numbering	1 → 26 (C 23)	27 → 38	39 → 55 (W 41)	56 → 65	66 → 104 (C 104)	105 → 116	105 → 117 (112.1, 111.1, 112.2, 111.2, etc.)	118 → 129
Number of amino acids	25–26	5–12	16–17	0–10	36–39	2–12	2–13 (more than 13 amino acids)	10–12

(C 23) 1st-CYS, (W 41) CONSERVED-TRP, (C 104) 2nd-CYS, position 118 corresponds to J-PHE or J-TRP as described in the text. IMGT notes: (1) The germline CDR3-IMGT is specific of the V-REGION of germline V-GENEs. It comprises 0, 1 or 2 nucleotide(s) before the V-HEPTAMER. (2) The rearranged CDR3-IMGT and the FR4-IMGT are specific of the V-DOMAIN (V–J-REGION or V–D–J-REGION). They are characteristic of rearranged V–J-GENEs and V–D–J-GENEs, and corresponding cDNAs and proteins. Note that positions of the codons (or amino acids) in the rearranged CDR3-IMGT are not equivalent to those in the germline CDR3-IMGT [10,11], and cannot therefore be directly compared.

the V-GENE group or subgroup) [10–13]. The 3′ end of germline CDR3-IMGT is limited by the V-HEPTAMER. The longest germline CDR3-IMGT have 12 amino acids (Table 1). For shorter germline CDR3-IMGT, positions are missing. As an example, for a germline CDR3-IMGT with two amino acids, only positions 105 and 106 are present [10].

3. IMGT unique numbering for V-DOMAIN

3.1. Overview

Whereas the IMGT unique numbering for V-REGION provides a standardized description of amino acid positions up to the end of the germline CDR3-IMGT, no standardized numbering has so far been defined for the rearranged CDR3-IMGT and the FR4-IMGT of V–J-REGION and V–D–J-REGION of rearranged V–J-GENEs or V–D–J-GENEs, and corresponding cDNAs and proteins. Interestingly, it is the analysis of structural data and sequence comparisons that we were carrying out to apply the IMGT unique numbering for the description of the IG and TR C-DOMAIN (http://imgt.cines.fr), which showed us that the standardized numbering of the FG loop of the C-DOMAIN could be applied to the CDR3-IMGT of rearranged IG and TR sequences. More precisely, the hydrogen bonds between 2nd-CYS 104 and position 119 in the C-DOMAIN correspond structurally to the hydrogen bonds between 2nd-CYS 104 and the Glycine which follows the J-TRP or J-PHE in the J-REGION, and therefore numbered as Glycine 119 (Fig. 3). As a consequence, position of J-TRP and J-PHE is 118. That conservation allowed to extend the IMGT unique numbering to the V-DOMAIN which corresponds to the V–J- and V–D–J-REGION of the IG and TR chains (Fig. 4) [20].

3.2. Rearranged CDR3-IMGT

Up to position 104, the IMGT unique numbering for V-DOMAIN follows the IMGT unique numbering for V-REGIONs. The CDR3-IMGT region of rearranged IG and TR V-GENEs and V–D–J-GENEs (and that of corresponding cDNAs and proteins) comprises positions 105–117 (Table 1). These positions correspond to a

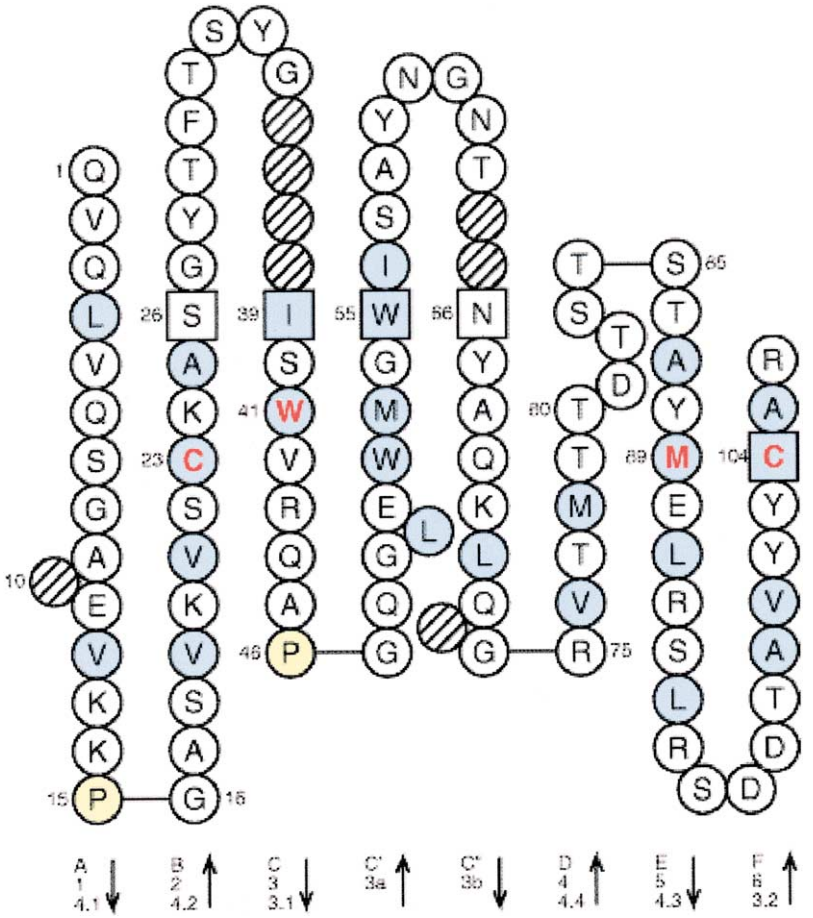


Fig. 2. 2D representation or Collier de Perles of the V-REGION of the human germline IGHV1-18 gene (IMGT/LIGM-DB accession number: M99641). The CDR1- and CDR2-IMGT have a length of eight amino acids (positions 27–34, and 56–63, respectively). Position 10 in the FR1-IMGT and position 73 in the FR3-IMGT are gaps compared to other V-GENE groups or subgroups. Amino acids are shown in the one-letter abbreviation. Position at which hydrophobic amino acids (hydropathy index with positive value: I, V, L, F, C, M, A) and Tryptophan (W) are found in more than 50% of analysed IG and TR sequences are shown in blue. All Proline (P) are shown in yellow. The CDR-IMGT are limited by amino acids shown in squares (anchor positions), which belong to the neighbouring FR-IMGT. Hachured circles or squares correspond to missing positions according to the IMGT unique numbering. Arrows indicate the direction of the beta strands and their different designations in 3D structures (from IMGT repertoire, http://imgt.cines.fr).

rearranged CDR3-IMGT of 13 amino acids (and to a JUNCTION of 15 amino acids, 2nd-CYS 104 and J-TRP or J-PHE 118 being included in the JUNCTION definition). This numbering is convenient to use since 80% of the IMGT/LIGM-DB IG and TR rearranged sequences have a CDR3-IMGT length less than or equal to 13 amino acids (IMGT statistics, October 2001).

For rearranged CDR3-IMGT less than 13 amino acids, gaps are created from the top of the loop, in the following order 111, 112, 110, 113, 109, 114, etc. (Table 4A). For rearranged CDR3-IMGT more than 13 amino acids, additional positions are created, between positions 111 and 112 at the top of the CDR3-IMGT loop, in the following order 112.1, 111.1, 112.2, 111.2, 112.3, 111.3, etc. (Table 4B). The unique concept of CDR3-IMGT replaces the diverse definitions and delimitations of the CDR3, H3 loop found in the literature (Table 3).

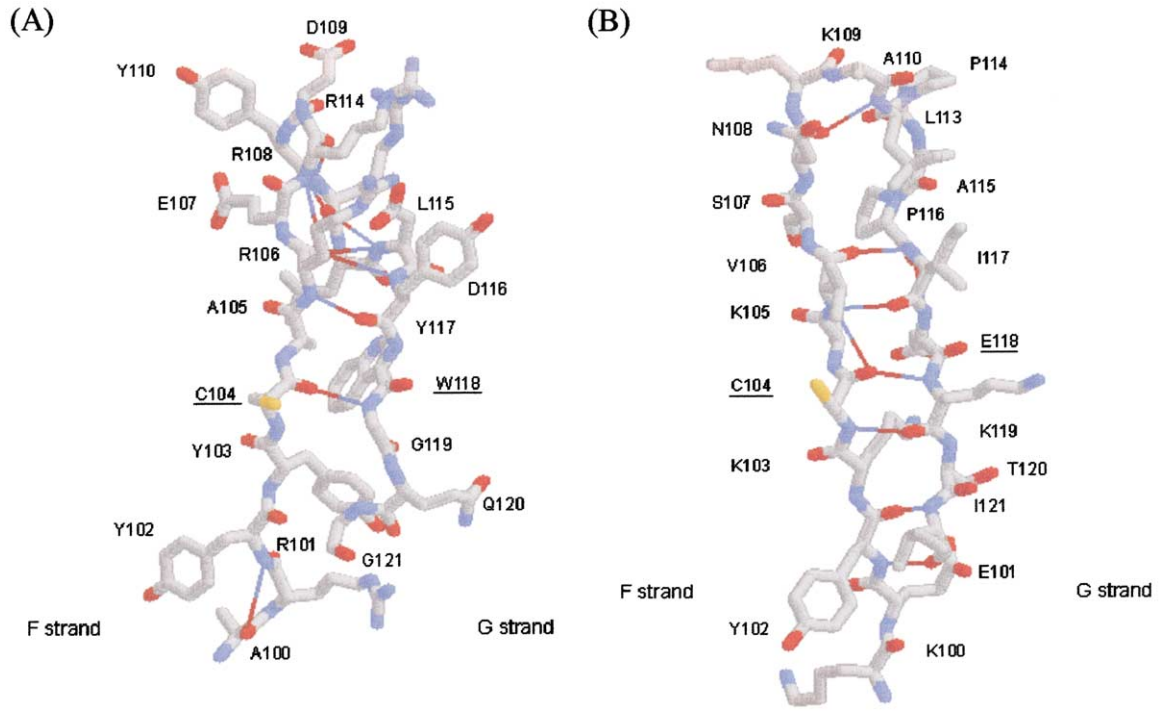


Fig. 3. Hydrogen bounds (A) between the F and G strands of the V-DOMAIN of the heavy chain from mouse D1.3 immunoglobulin Fv (PDB code: 1dvf_B), (B) between the F and G strands of the CH2 (C-DOMAIN) of the human IGHG1 (PDB code: 1mco_H).

3.3. FR4-IMGT

The FR4-IMGT region found in rearranged IG and TR V–J-GENEs and V–D–J-GENEs (and in corresponding cDNAs and proteins) comprises positions 118 (J-PHE or J-TRP 118) to 129 (10–12 amino acids, depending on the J-REGION length). The longest FR4-IMGT have 12 amino acids (IMGT repertoire, http://imgt.cines.fr).

The IMGT unique numbering for V-DOMAIN (i) provides a standardized information on the CDR-IMGT lengths of the V–J-REGION and V–D–J-REGION of all IG and TR chain sequences and 3D structures (Table 5), (ii) a standardized representation of the protein displays and Colliers de Perles of the IG and TR chains [20] (IMGT/3Dstructure-DB, http://imgt.cines.fr). The IMGT unique numbering is also used in more sophisticated representations of the Colliers de Perles on two planes which allow the vizualization of the hydrogen bounds between amino acids belonging to beta strands from the same sheet or from different sheets (Fig. 5).

4. IMGT unique numbering and sequence data analysis

4.1. Allele polymorphisms and somatic hypermutations

The IMGT unique numbering allows a standardized description of mutations, allele polymorphisms and somatic hypermutations of the variable regions and domains. The mutations and allelic polymorphisms are described by comparison to the germline V-, D-, J-REGION from the IMGT reference directory. Based on these criteria, tables of alleles for the germline V, D and J genes have been set up and are available at IMGT, http://imgt.cines.fr. Alignments of alleles are displayed according to the IMGT unique numbering and with the FR- and CDR-IMGT delimitations [10,11]. IMGT genes and alleles are described according to the CLASSIFICATION concept of IMGT-ONTOLOGY [4]. All the human IG and TR genes [10,11] have been approved by the HUGO (HUman Genome Organization) Nomenclature

Table 2
Correspondence between the IMGT unique Lefranc numbering and the Kabat numberings for the IG and TR V-REGIONs

(A) Immunoglobulins

	IGHV: human IGHV6-1					IGKV: human IGKV1-5					IGLV: human IGLV2-23				
FR1-IMGT	**1**	1	cag	GLN	Q	**1**	1	gac	ASP	D	**1**	1	cag	GLN	Q
	2	2	gta	VAL	V	**2**	2	atc	ILE	I	**2**	2	tct	SER	S
	3	3	cag	GLN	Q	**3**	3	cag	GLN	Q	**3**	3	gcc	ALA	A
	4	4	ctg	LEU	L	**4**	4	atg	MET	M	**4**	4	ctg	LEU	L
	5	5	cag	GLN	Q	**5**	5	acc	THR	T	**5**	5	act	THR	T
	6	6	cag	GLN	Q	**6**	6	cag	GLN	Q	**6**	6	cag	GLN	Q
	7	7	tca	SER	S	**7**	7	tct	SER	S	**7**	7	cct	PRO	P
	8	8	ggt	GLY	G	**8**	8	cct	PRO	P	**8**	8	gcc	ALA	A
	9	9	cca	PRO	P	**9**	9	tcc	SER	S	**9**	9	tcc	SER	S
	10		–	–	–	**10**	10	acc	THR	T	**10**	10	–	–	–
	11	10	gga	GLY	G	**11**	11	ctg	LEU	L	**11**	11	gtg	VAL	V
	12	11	ctg	LEU	L	**12**	12	tct	SER	S	**12**	12	tct	SER	S
	13	12	gtg	VAL	V	**13**	13	gca	ALA	A	**13**	13	ggg	GLY	G
	14	13	aag	LYS	K	**14**	14	tct	SER	S	**14**	14	tct	SER	S
	15	14	ccc	PRO	P	**15**	15	gta	VAL	V	**15**	15	cct	PRO	P
	16	15	tcg	SER	S	**16**	16	gga	GLY	G	**16**	16	gga	GLY	G
	17	16	cag	GLN	Q	**17**	17	gac	ASP	D	**17**	17	cag	GLN	Q
	18	17	acc	THR	T	**18**	18	aga	ARG	R	**18**	18	tcg	SER	S
	19	18	ctc	LEU	L	**19**	19	gtc	VAL	V	**19**	19	atc	ILE	I
	20	19	tca	SER	S	**20**	20	acc	THR	T	**20**	20	acc	THR	T
	21	20	ctc	LEU	L	**21**	21	atc	ILE	I	**21**	21	atc	ILE	I
	22	21	acc	THR	T	**22**	22	act	THR	T	**22**	22	tcc	SER	S
	23	22	tgt	**CYS**	**C**	**23**	23	tgc	**CYS**	**C**	**23**	23	tgc	**CYS**	**C**
	24	23	gcc	ALA	A	**24**	24	cgg	ARG	R	**24**	24	act	THR	T
	25	24	atc	ILE	I	**25**	25	gcc	ALA	A	**25**	25	gga	GLY	G
	26	25	tcc	SER	S	**26**	26	agt	SER	S	**26**	26	acc	THR	T
CDR1-IMGT	**27**	26	ggg	GLY	G	**27**	27	cag	GLN	Q	**27**	27	agc	SER	S
	28	27	gac	ASP	D	**28**	* 28	agt	SER	S	**28**	* 27A	agt	SER	S
	29	28	agt	SER	S	**29**	* 29	att	ILE	I	**29**	* 27B	gat	ASP	D
	30	29	gtc	VAL	V	**30**	* 30	agt	SER	S	**30**	* 27C	gtt	VAL	V
	31	30	tct	SER	S	**31**	* 31	agc	SER	S	**31**	* 28	ggg	GLY	G
	32	31	agc	SER	S	**32**	* 32	tgg	TRP	W	**32**	* 29	agt	SER	S
	33	32	aac	ASN	N	**33**	*	–	–	–	**33**	* 30	tat	TYR	Y
	34	33	agt	SER	S	**34**	*	–	–	–	**34**	* 31	aac	ASN	N
	35	* 34	gct	ALA	A	**35**	*	–	–	–	**35**	* 32	ctt	LEU	L
	36	* 35	gct	ALA	A	**36**	*	–	–	–	**36**		–	–	–
	37		–	–	–	**37**	*	–	–	–	**37**		–	–	–
	38		–	–	–	**38**	*	–	–	–	**38**		–	–	–

(*continued on next page*)

Table 2 (*continued*)

(A) Immunoglobulins

	IGHV: human IGHV6-1					IGKV: human IGKV1-5					IGLV: human IGLV2-23				
FR2-IMGT	**39**	* 35A	tgg	TRP	W	**39**	33	ttg	LEU	L	**39**	33	gtc	VAL	V
	40	* 35B	aac	ASN	N	**40**	34	gcc	ALA	A	**40**	34	tcc	SER	S
	41	36	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
	43	38	agg	ARG	R	**43**	37	cag	GLN	Q	**43**	37	caa	GLN	Q
	44	39	cag	GLN	Q	**44**	38	cag	GLN	Q	**44**	38	cag	GLN	Q
	45	40	tcc	SER	S	**45**	39	aaa	LYS	K	**45**	39	cac	HIS	H
	46	41	cca	PRO	P	**46**	40	cca	PRO	P	**46**	40	cca	PRO	P
	47	42	tcg	SER	S	**47**	41	ggg	GLY	G	**47**	41	ggc	GLY	G
	48	43	aga	ARG	R	**48**	42	aaa	LYS	K	**48**	42	aaa	LYS	K
	49	44	ggc	GLY	G	**49**	43	gcc	ALA	A	**49**	43	gcc	ALA	A
	50	45	ctt	LEU	L	**50**	44	cct	PRO	P	**50**	44	ccc	PRO	P
	51	46	gag	GLU	E	**51**	45	aag	LYS	K	**51**	45	aaa	LYS	K
	52	47	tgg	TRP	W	**52**	46	ctc	LEU	L	**52**	46	ctc	LEU	L
	53	48	ctg	LEU	L	**53**	47	ctg	LEU	L	**53**	47	atg	MET	M
	54	49	gga	GLY	G	**54**	48	atc	ILE	I	**54**	48	att	ILE	I
	55	50	agg	ARG	R	**55**	49	tat	TYR	Y	**55**	49	tat	TYR	Y
CDR2-IMGT	**56**	51	aca	THR	T	**56**	50	gat	ASP	D	**56**	50	gag	GLU	E
	57	52	tac	TYR	Y	**57**	51	gcc	ALA	A	**57**	51	ggc	GLY	G
	58	* 52A	tac	TYR	Y	**58**	52	tcc	SER	S	**58**	52	agt	SER	S
	59	* 52B	agg	ARG	R	**59**		–	–	–	**59**		–	–	–
	60	* 53	tcc	SER	S	**60**		–	–	–	**60**		–	–	–
	61	* 54	aag	LYS	K	**61**		–	–	–	**61**		–	–	–
	62	* 55	tgg	TRP	W	**62**		–	–	–	**62**		–	–	–
	63	* 56	tat	TYR	Y	**63**		–	–	–	**63**		–	–	–
	64	* 57	aat	ASN	N	**64**		–	–	–	**64**		–	–	–
	65	*	–	–	–	**65**		–	–	–	**65**		–	–	–
FR3-IMGT	**66**	58	gat	ASP	D	**66**	53	agt	SER	S	**66**	53	aag	LYS	K
	67	59	tat	TYR	Y	**67**	54	ttg	LEU	L	**67**	54	cgg	ARG	R
	68	60	gca	ALA	A	**68**	55	gaa	GLU	E	**68**	55	ccc	PRO	P
	69	61	gta	VAL	V	**69**	56	agt	SER	S	**69**	56	tca	SER	S
	70	62	tct	SER	S	**70**	57	ggg	GLY	G	**70**	57	ggg	GLY	G
	71	63	gtg	VAL	V	**71**	58	gtc	VAL	V	**71**	58	gtt	VAL	V
	72	64	aaa	LYS	K	**72**	59	cca	PRO	P	**72**	59	tct	SER	S
	73		–	–	–	**73**		–	–	–	**73**		–	–	–
	74	65	agt	SER	S	**74**	60	tca	SER	S	**74**	60	aat	ASN	N
	75	66	cga	ARG	R	**75**	61	agg	ARG	R	**75**	61	cgc	ARG	R
	76	67	ata	ILE	I	**76**	62	ttc	PHE	F	**76**	62	ttc	PHE	F
	77	68	acc	THR	T	**77**	63	agc	SER	S	**77**	63	tct	SER	S

Table 2 (*continued*)

(A) Immunoglobulins

	IGHV: human IGHV6-1					IGKV: human IGKV1-5					IGLV: human IGLV2-23				
	78	69	atc	ILE	I	**78**	64	ggc	GLY	G	**78**	64	ggc	GLY	G
	79	70	aac	ASN	N	**79**	65	agt	SER	S	**79**	65	tcc	SER	S
	80	71	cca	PRO	P	**80**	66	gga	GLY	G	**80**	66	aag	LYS	K
	81	72	gac	ASP	D	**81**		–	–	–	**81**		–	–	–
	82	73	aca	THR	T	**82**		–	–	–	**82**		–	–	–
	83	74	tcc	SER	S	**83**	67	tct	SER	S	**83**	67	tct	SER	S
	84	75	aag	LYS	K	**84**	68	ggg	GLY	G	**84**	68	ggc	GLY	G
	85	76	aac	ASN	N	**85**	69	aca	THR	T	**85**	69	aac	ASN	N
	86	77	cag	GLN	Q	**86**	70	gaa	GLU	E	**86**	70	acg	THR	T
	87	78	ttc	PHE	F	**87**	71	ttc	PHE	F	**87**	71	gcc	ALA	A
	88	79	tcc	SER	S	**88**	72	act	THR	T	**88**	72	tcc	SER	S
	89	80	ctg	LEU	L	**89**	73	ctc	LEU	L	**89**	73	ctg	LEU	L
	90	81	cag	GLN	Q	**90**	74	acc	THR	T	**90**	74	aca	THR	T
	91	82	ctg	LEU	L	**91**	75	atc	ILE	I	**91**	75	atc	ILE	I
	92	82A	aac	ASN	N	**92**	76	agc	SER	S	**92**	76	tct	SER	S
	93	82B	tct	SER	S	**93**	77	agc	SER	S	**93**	77	ggg	GLY	G
	94	82C	gtg	VAL	V	**94**	78	ctg	LEU	L	**94**	78	ctc	LEU	L
	95	83	act	THR	T	**95**	79	cag	GLN	Q	**95**	79	cag	GLN	Q
	96	84	ccc	PRO	P	**96**	80	cct	PRO	P	**96**	80	gct	ALA	A
	97	85	gag	GLU	E	**97**	81	gat	ASP	D	**97**	81	gag	GLU	E
	98	86	gac	ASP	D	**98**	82	gat	ASP	D	**98**	82	gac	ASP	D
	99	87	acg	THR	T	**99**	83	ttt	PHE	F	**99**	83	gag	GLU	E
	100	88	gct	ALA	A	**100**	84	gca	ALA	A	**100**	84	gct	ALA	A
	101	89	gtg	VAL	V	**101**	85	act	THR	T	**101**	85	gat	ASP	D
	102	90	tat	TYR	Y	**102**	86	tat	TYR	Y	**102**	86	tat	TYR	Y
	103	91	tac	TYR	Y	**103**	87	tac	TYR	Y	**103**	87	tac	TYR	Y
	104	92	tgt	**CYS**	**C**	**104**	88	tgc	**CYS**	**C**	**104**	88	tgc	**CYS**	**C**
CDR3-IMGT	**105**	93	gca	ALA	A	**105**	89	caa	GLN	Q	**105**	89	tgc	CYS	C
	106	94	aga	ARG	R	**106**	90	cag	GLN	Q	**106**	90	tca	SER	S
	107	95	–	–	–	**107**	91	tat	TYR	Y	**107**	91	tat	TYR	Y
	108					**108**	92	aat	ASN	N	**108**	92	gca	ALA	A
	109					**109**	93	agt	SER	S	**109**	93	ggt	GLY	G
	110					**110**	94	tat	TYR	Y	**110**	94	agt	SER	S
	111					**111**	95	tct	SER	S	**111**	95	agc	SER	S
	112										**112**	95A	act	THR	T
	113										**113**	95B	tta	LEU	L
	114										**114**	95C	–	–	–
	115										**115**	95D	–	–	–

(*continued on next page*)

Table 2 (*continued*)

(B) T cell receptors

	TRBV: human TRBV6-5					TRAV: human TRAV8-6					TRDV: human TRDV2					TRGV: human TRGV3				
FR1-IMGT	**1**	1	aat	ASN	N	**1**	0	agc	ALA	A	**1**	00	gcc	ALA	A	**1**	01	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	ggt	GLY	G	**3**	2	tct	SER	S	**3**	2	gag	GLU	E	**3**	1	aac	ASN	N
	4	4	gtc	VAL	V	**4**	3	gtg	VAL	V	**4**	3	ttg	LEU	L	**4**	2	ttg	LEU	L
	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**	5	cag	GLN	Q	**6**	5	cct	PRO	P	**6**	4	ggg	GLY	G
	7	7	acc	THR	T	**7**	6	ctt	LEU	L	**7**	6	gaa	GLU	E	**7**	5	aga	ARG	R
	8	8	cca	PRO	P	**8**	7	gac	ASP	D	**8**	7	cac	HIS	H	**8**	6	acg	THR	T
	9	9	aaa	LYS	K	**9**	8	agc	SER	S	**9**	8	caa	GLN	Q	**9**	7	aag	LYS	K
	10	10	ttc	PHE	F	**10**	9	caa	GLN	Q	**10**	9	aca	THR	T	**10**	8	tca	SER	S
	11	11	cag	GLN	Q	**11**	10	gtc	VAL	V	**11**	10	gtg	VAL	V	**11**	9	gtc	VAL	V
	12	12	gtc	VAL	V	**12**	11	cct	PRO	P	**12**	11	cct	PRO	P	**12**	10	acc	THR	T
	13	13	ctg	LEU	L	**13**	12	gtc	VAL	V	**13**	12	gtg	VAL	V	**13**	11	agg	ARG	R
	14	14	aag	LYS	K	**14**	13	ttt	PHE	F	**14**	13	tca	SER	S	**14**	12	cag	GLN	Q
	15	15	aca	THR	T	**15**	14	gaa	GLU	E	**15**	14	ata	ILE	I	**15**	13	act	THR	T
	16	16	gga	GLY	G	**16**	15	gaa	GLU	E	**16**	15	ggg	GLY	G	**16**	14	ggg	GLY	G
	17	17	cag	GLN	Q	**17**	16	gcc	ALA	A	**17**	16	gtc	VAL	V	**17**	15	tca	SER	S
	18	18	agc	SER	S	**18**	17	cct	PRO	P	**18**	17	cct	PRO	P	**18**	16	tct	SER	S
	19	19	atg	MET	M	**19**	18	gtg	VAL	V	**19**	18	gcc	ALA	A	**19**	17	gct	ALA	A
	20	20	aca	THR	T	**20**	19	gag	GLU	E	**20**	19	acc	THR	T	**20**	18	gaa	GLU	E
	21	21	ctg	LEU	L	**21**	20	ctg	LEU	L	**21**	20	ctc	LEU	L	**21**	19	atc	ILE	I
	22	22	cag	GLN	Q	**22**	21	agg	ARG	R	**22**	21	agg	ARG	R	**22**	20	act	THR	T
	23	23	tgt	**CYS**	**C**	**23**	22	tgc	**CYS**	**C**	**23**	22	tgc	**CYS**	**C**	**23**	21	tgc	**CYS**	**C**
	24	24	gcc	ALA	A	**24**	23	aac	ASN	N	**24**	23	tcc	SER	S	**24**	22	gat	ASP	D
	25	25	cag	GLN	Q	**25**	24	tac	TYR	Y	**25**	24	atg	MET	M	**25**	23	ctt	LEU	L
	26	26	gat	ASP	D	**26**	25	tca	SER	S	**26**	25	aaa	LYS	K	**26**	24	act	THR	T
CDR1-IMGT	**27**	27	atg	MET	M	**27**	26	tcg	SER	S	**27**	26	gga	GLY	G	**27**	25	gta	VAL	V
	28	28	aac	ASN	N	**28**	27	tct	SER	S	**28**	27	gaa	GLU	E	**28**	26	aca	THR	T
	29	29	cat	HIS	H	**29**	28	gtt	VAL	V	**29**	28	gcg	ALA	A	**29**	27	aat	ASN	N
	30	30	gaa	GLU	E	**30**	29	tca	SER	S	**30**	29	atc	ILE	I	**30**	28	acc	THR	T
	31	* 31	tac	TYR	Y	**31**	30	gtg	VAL	V	**31**	30	ggt	GLY	G	**31**	29	ttc	PHE	F
	32	*	–	–	–	**32**	* 31	tat	TYR	Y	**32**	31	aac	ASN	N	**32**	30	tac	TYR	Y
	33		–	–	–	**33**	*	–	–	–	**33**	32	tac	TYR	Y	**33**	31	–	–	–
	34		–	–	–	**34**	*	–	–	–	**34**	33	tat	TYR	Y	**34**	32	–	–	–
	35		–	–	–	**35**		–	–	–	**35**		–	–	–	**35**		–	–	–
	36		–	–	–	**36**		–	–	–	**36**		–	–	–	**36**		–	–	–
	37		–	–	–	**37**		–	–	–	**37**		–	–	–	**37**		–	–	–
	38		–	–	–	**38**		–	–	–	**38**		–	–	–	**38**		–	–	–

Table 2 (*continued*)

(B) T cell receptors

	TRBV: human TRBV6-5					TRAV: human TRAV8-6					TRDV: human TRDV2					TRGV: human TRGV3				
FR2-IMGT	**39**	32	atg	MET	M	**39**	32	ctc	LEU	L	**39**	34	atc	ILE	I	**39**	33	atc	ILE	I
	40	33	tcc	SER	S	**40**	33	ttc	PHE	F	**40**	34A	aac	ASN	N	**40**	34	cac	HIS	H
	41	34	tgg	**TRP**	**W**	**41**	34	tgg	**TRP**	**W**	**41**	35	tgg	**TRP**	**W**	**41**	35	tgg	**TRP**	**W**
	42	35	tat	TYR	Y	**42**	35	tat	TYR	Y	**42**	36	tac	TYR	Y	**42**	36	tac	TYR	Y
	43	36	cga	ARG	R	**43**	36	gtg	VAL	V	**43**	37	agg	ARG	R	**43**	37	cta	LEU	L
	44	37	caa	GLN	Q	**44**	37	caa	GLN	Q	**44**	38	aag	LYS	K	**44**	38	cac	HIS	H
	45	38	gac	ASP	D	**45**	38	tac	TYR	Y	**45**	39	acc	THR	T	**45**	39	cag	GLN	Q
	46	39	cca	PRO	P	**46**	39	ccc	PRO	P	**46**	40	caa	GLN	Q	**46**	40	gag	GLU	E
	47	40	ggc	GLY	G	**47**	40	aac	ASN	N	**47**	41	ggt	GLY	G	**47**	41	ggg	GLY	G
	48	41	atg	MET	M	**48**	41	caa	GLN	Q	**48**	42	aac	ASN	N	**48**	42	aag	LYS	K
	49	42	ggg	GLY	G	**49**	42	gga	GLY	G	**49**	43	aca	THR	T	**49**	43	gcc	ALA	A
	50	43	ctg	LEU	L	**50**	43	ctc	LEU	L	**50**	44	atc	ILE	I	**50**	44	cca	PRO	P
	51	44	agg	ARG	R	**51**	44	cag	GLN	Q	**51**	45	act	THR	T	**51**	45	cag	GLN	Q
	52	45	ctg	LEU	L	**52**	45	ctt	LEU	L	**52**	46	ttc	PHE	F	**52**	46	cgt	ARG	R
	53	46	att	ILE	I	**53**	46	ctc	LEU	L	**53**	47	ata	ILE	I	**53**	47	ctt	LEU	L
	54	47	cat	HIS	H	**54**	47	ctg	LEU	L	**54**	48	tac	TYR	Y	**54**	48	ctg	LEU	L
	55	48	tac	TYR	Y	**55**	48	aag	LYS	K	**55**	49	cga	ARG	R	**55**	49	tac	TYR	Y
CDR2-IMGT	**56**	49	tca	SER	S	**56**	* 49	tat	TYR	Y	**56**	50	gaa	GLU	E	**56**	50	tat	TYR	Y
	57	50	gtt	VAL	V	**57**	* 50	tta	LEU	L	**57**	51	aag	LYS	K	**57**	51	gac	ASP	D
	58	51	ggt	GLY	G	**58**	* 51	tca	SER	S	**58**	52	gac	ASP	D	**58**	52	gtc	VAL	V
	59	52	gct	ALA	A	**59**	* 52	gga	GLY	G	**59**	*	–	–	–	**59**	53	tcc	SER	S
	60	53	ggt	GLY	G	**60**		–	–	–	**60**		–	–	–	**60**	54	acc	THR	T
	61	54	atc	ILE	I	**61**		–	–	–	**61**		–	–	–	**61**	55	gca	ALA	A
	62		–	–	–	**62**		–	–	–	**62**		–	–	–	**62**	56	agg	ARG	R
	63		–	–	–	**63**		–	–	–	**63**		–	–	–	**63**	* 57	gat	ASP	D
	64		–	–	–	**64**		–	–	–	**64**		–	–	–	**64**		–	–	–
	65		–	–	–	**65**		–	–	–	**65**		–	–	–	**65**		–	–	–
FR3-IMGT	**66**	55	act	THR	T	**66**	* 53	tcc	SER	S	**66**	* 53	atc	ILE	I	**66**	* 58	gtg	VAL	V
	67	56	gac	ASP	D	**67**	* 54	acc	THR	T	**67**	* 54	tat	TYR	Y	**67**	* 59	ttg	LEU	L
	68	57	caa	GLN	Q	**68**	* 55	ctg	LEU	L	**68**	* 55	ggc	GLY	G	**68**	* 60	gaa	GLU	E
	69	58	gga	GLY	G	**69**	* 56	gtt	VAL	V	**69**	* 56	cct	PRO	P	**69**	* 61	tca	SER	S
	70	59	gaa	GLU	E	**70**	* 57	gaa	GLU	E	**70**	* 57	ggt	GLY	G	**70**	* 62	gga	GLY	G
	71	60	gtc	VAL	V	**71**	* 58	agc	SER	S	**71**	* 57A	ttc	PHE	F	**71**	* 63	ctc	LEU	L
	72	61	ccc	PRO	P	**72**	* 59	atc	ILE	I	**72**	* 57B	aaa	LYS	K	**72**	* 64	agt	SER	S
	73	* 62	–	–	–	**73**	*	–	–	–	**73**	* 57C	–	–	–	**73**	* 64A	cca	PRO	P
	74	* 63	aat	ASN	N	**74**	* 60	aac	ASN	N	**74**	58	gac	ASP	D	**74**	* 65	gga	GLY	G
	75	* 64	ggc	GLY	G	**75**	61	ggt	GLY	G	**75**	59	aat	ASN	N	**75**	* 66	aag	LYS	K
	76	65	tac	TYR	Y	**76**	62	ttt	PHE	F	**76**	60	ttc	PHE	F	**76**	* 67	tat	TYR	Y
	77	66	aat	ASN	N	**77**	63	gag	GLU	E	**77**	61	caa	GLN	Q	**77**	* 68	tat	TYR	Y

(*continued on next page*)

Table 2 (*continued*)

(B) T cell receptors

	TRBV: human TRBV6-5					TRAV: human TRAV8-6					TRDV: human TRDV2					TRGV: human TRGV3				
	78	67	gtc	VAL	V	**78**	64	gct	ALA	A	**78**	62	ggt	GLY	G	**78**	* 69	act	THR	T
	79	68	tcc	SER	S	**79**	65	gaa	GLU	E	**79**	63	gac	ASP	D	**79**	* 70	cat	HIS	H
	80	69	aga	ARG	R	**80**	66	ttt	PHE	F	**80**	64	att	ILE	I	**80**	* 71	aca	THR	T
	81	70	tca	SER	S	**81**	67	aac	ASN	N	**81**	65	gat	ASP	D	**81**	* 72	ccc	PRO	P
	82		–	–	–	**82**	68	aag	LYS	K	**82**	66	att	ILE	I	**82**	*	–	–	–
	83	71	acc	THR	T	**83**	69	agt	SER	S	**83**	67	gca	ALA	A	**83**	73	agg	ARG	R
	84	72	aca	THR	T	**84**	70	caa	GLN	Q	**84**	68	aag	LYS	K	**84**	74	agg	ARG	R
	85	73	gag	GLU	E	**85**	71	act	THR	T	**85**	69	aac	ASN	N	**85**	75	tgg	TRP	W
	86	74	gat	ASP	D	**86**	72	tcc	SER	S	**86**	70	ctg	LEU	L	**86**	76	agc	SER	S
	87	75	ttc	PHE	F	**87**	73	ttc	PHE	F	**87**	71	gct	ALA	A	**87**	77	tgg	TRP	W
	88	76	ccg	PRO	P	**88**	74	cac	HIS	H	**88**	72	gta	VAL	V	**88**	78	ata	ILE	I
	89	77	ctc	LEU	L	**89**	75	ttg	LEU	L	**89**	73	ctt	LEU	L	**89**	79	ttg	LEU	L
	90	78	agg	ARG	R	**90**	76	agg	ARG	R	**90**	74	aag	LYS	K	**90**	80	aga	ARG	R
	91	79	ctg	LEU	L	**91**	77	aaa	LYS	K	**91**	75	ata	ILE	I	**91**	81	ctg	LEU	L
	92	80	ctg	LEU	L	**92**	78	ccc	PRO	P	**92**	76	ctt	LEU	L	**92**	82	caa	GLN	Q
	93	81	tcg	SER	S	**93**	79	tca	SER	S	**93**	77	gca	ALA	A	**93**	83	aat	ASN	N
	94	82	gct	ALA	A	**94**	80	gtc	VAL	V	**94**	78	cca	PRO	P	**94**	84	cta	LEU	L
	95	83	gct	ALA	A	**95**	81	cat	HIS	H	**95**	79	tca	SER	S	**95**	85	att	ILE	I
	96	84	ccc	PRO	P	**96**	82	ata	ILE	I	**96**	80	gag	GLU	E	**96**	86	gaa	GLU	E
	97	85	tcc	SER	S	**97**	83	agc	SER	S	**97**	81	aga	ARG	R	**97**	87	aat	ASN	N
	98	86	cag	GLN	Q	**98**	84	gac	ASP	D	**98**	82	gat	ASP	D	**98**	88	gat	ASP	D
	99	87	aca	THR	T	**99**	85	acg	THR	T	**99**	83	gaa	GLU	E	**99**	89	tct	SER	S
	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**	87	gag	GLU	E	**101**	85	tct	SER	S	**101**	91	gtc	VAL	V
	102	90	tac	TYR	Y	**102**	88	tac	TYR	Y	**102**	86	tac	TYR	Y	**102**	92	tat	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**	90	tgt	**CYS**	**C**	**104**	88	tgt	**CYS**	**C**	**104**	94	tgt	**CYS**	**C**
CDR3-IMGT	**105**	93	gcc	ALA	A	**105**	91	gct	ALA	A	**105**	89	gcc	ALA	A	**105**	95	gcc	ALA	A
	106	94	agc	SER	S	**106**	92	gtg	VAL	V	**106**	90	tgt	CYS	C	**106**	96	acc	THR	T
	107	95	agt	SER	S	**107**	93	agt	SER	S	**107**	91	gac	ASP	D	**107**	97	tgg	TRP	W
	108	96	tat	TYR	Y						**108**	92	acc	THR	T	**108**	98	gac	ASP	D
	109	97	–	–	–											**109**	99	agg	ARG	R

For each V-REGION group of immunoglobulins (A) and T cell receptors (B) [4], one germline sequence is shown as an example with, on the left-hand side of each column, the IMGT unique Lefranc numbering (in bold) [6] and the corresponding Kabat numberings [7]. 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 numbering to the other, automatically. 1st-CYS 23, CONSERVED-TRP 41 and 2nd-CYS 104 are in bold.

Table 3
Correspondence between the IMGT unique numbering and the different numberings for CDR of different lengths. The CDR-IMGT lengths based on the standardized IMGT unique Lefranc numbering [6, 10, 11] replace the diverse definitions of the CDR, H and L loops and canonical structures found in the literature [9,17–19]

CDR1-IMGT

IMGT unique numbering [6]	FR1-IMGT C 23	24	25	26	[27	28	29	30	31	32	33	34	CDR1-IMGT 35	36	37	38]	FR2-IMGT 39	40	W 41
													*	*			*	*	
IGHV [17]	22	23	24	25	26	27	28	29	30	[31	32	33	34	35			35A	35B]	36
	22	23	24	25	26	27	28	29	30	[31	32	33	34	35			35A	-]	36
	22	23	24	25	26	27	28	29	30	[31	32	33	34	35			-	-]	36
												*	*	*	*				
IGHV [19]	22	23	24	25	[26	27	28	29	30	31	31a	31b	32]	33			34	35	36
[19]	22	23	24	25	[26	27	28	29	30	31	31a	-	32]	33			34	35	36
[9, 18, 19]	22	23	24	25	[26	27	28	29	30	31	-	-	32]	33			34	35	36

IMGT unique numbering [6]	FR1-IMGT C 23	24	25	26	[27	28	29	30	CDR1-IMGT 31	32	33	34	35	36	37	38]	FR2-IMGT 39	40	W 41
						*	*	*	*	*	*	*	*	*	*	*			
IGKV [17]	23	[24	25	26	27	27A	27B	27C	27D	27E	27F	28	29	30	31	32	33	34]	35
	23	[24	25	26	27	27A	27B	27C	27D	27E	-	28	29	30	31	32	33	34]	35
	23	[24	25	26	27	27A	27B	27C	27D	-	-	28	29	30	31	32	33	34]	35
	23	[24	25	26	27	27A	-	-	-	-	-	28	29	30	31	32	33	34]	35
	23	[24	25	26	27	-	-	-	-	-	-	28	29	30	31	32	33	34]	35
	23	[24	25	26	27	-	-	-	-	-	-	-	29	30	31	32	33	34]	35
									*	*	*	*	*	*	*	*			
IGKV [9, 18]	23	24	25	[26	27	28	29	30	31	31a	31b	31c	31d	31e	31f	32]	33	34	35
	23	24	25	[26	27	28	29	30	31	31a	-	31c	31d	31e	31f	32]	33	34	35
	23	24	25	[26	27	28	29	30	31	-	-	-	-	-	-	32]	33	34	35
	23	24	25	[26	27	28	29	30	-	-	-	-	-	-	-	32]	33	34	35
									*	*	*	*	*	*	*	*			
IGKV [19]	23	24	25	[26	27	28	29	30	30a	30b	30c	30d	30e	30f	31	32]	33	34	35
	23	24	25	[26	27	28	29	30	30a	30b	30c	-	30e	30f	31	32]	33	34	35
	23	24	25	[26	27	28	29	30	30a	30b	30c	30d	-	-	31	32]	33	34	35
	23	24	25	[26	27	28	29	30	30a	-	-	-	-	-	31	32]	33	34	35
	23	24	25	[26	27	28	29	30	-	-	-	-	-	-	31	32]	33	34	35
	23	24	25	[26	27	28	29	30	-	-	-	-	-	-	-	32]	33	34	35

IMGT unique numbering [6]	FR1-IMGT C 23	24	25	26	[27	28	29	30	CDR1-IMGT 31	32	33	34	35	36	37	38]	FR2-IMGT 39	40	W 41
						*	*	*	*	*	*	*	*	*	*	*			
IGLV [17]	23	[24	25	26	27	27A	27B	27C	27D	27E	27F	28	29	30	31	32	33	34]	35
	23	[24	25	26	27	27A	27B	-	-	-	-	28	29	30	31	32	33	34]	35
	23	[24	25	26	27	27A	-	-	-	-	-	28	29	30	31	32	33	34]	35
	23	[24	25	26	27	-	-	-	-	-	-	28	29	30	31	32	33	34]	35
						*	*	*	*	*	*				*	*			
IGLV [9]	23	24	25	[26	27	28	29	30	30a	30b	30c				31	32]	33	34	35
	23	24	25	[26	27	28	29	30	30a	30b	-				31	32]	33	34	35
IGLV [19]	23	24	[25	26	27	28	29	30	30a	30b	30c				31	32]	33	34	35
	23	24	[25	26	27	28	29	30	30a	30b	-				31	32]	33	34	35
	23	24	[25	26	27	-	-	30	30a	30b	30c				31	32]	33	34	35

Number of amino acids

CDR length[17]	CDR1-IMGT[6]
7	**10**
6	**9**
5	**8**

H1"canonical structure" type	H1 length[19]	CDR1-IMGT[6]	Example of proteins (see table 5)
3	9	**10**	
2	8	**9**	
1	7	**8**	Kol, Newm, McPC603, J539

CDR length[17]	CDR1-IMGT[6]
17	**12**
16	**11**
15	**10**
12	**7**
11	**6**
10	**5**

L1"canonical structure" type	L1 length[19]	CDR1-IMGT[6]	Example of proteins (see table 5)
3	13	**12**	McPC603
4	12	**11**	4-4-20
2	7	**6**	Rei
1	6	**5**	J539
3	13	**12**	McPC603
4	12	**11**	4-4-20
5	11	**10**	
6	8	**7**	
2	7	**6**	Rei
1	6	**5**	J539

CDR length[17]	CDR1-IMGT[6]
14	**9**
13	**8**
12	**7**
11	**6**

L1"canonical structure" type	L1 length[9]	CDR1-IMGT[6]	Example of proteins (see table 5)
	10	**9**	Newm
	9	**8**	Kol, Rhe
	L1 length[19]		
2,3	11	**9**	Newm
1	10	**8**	Kol, Rhe
4	9	**7**	Hil

(*continued on next page*)

CDR2-IMGT

IMGT unique-numbering [6]	FR2-IMGT			CDR2-IMGT								FR3-IMGT									CDR length[17]	CDR2 IMGT [6]	
	55	[56	57	58	59	60	61	62	63	64	65]	66	67	68	69	70	71	72	73	74			
				*	*	*	*	*	*	*	*												
IGHV [17]	[50	51	52	52A	52B	52C	53	54	55	56	57	58	59	60	61	62	63	64	-	65]		19	**10**
	[50	51	52	52A	52B	-	53	54	55	56	57	58	59	60	61	62	63	64	-	65]		18	**9**
	[50	51	52	52A	-	-	53	54	55	56	57	58	59	60	61	62	63	64	-	65]		17	**8**
	[50	51	52	-	-	-	53	54	55	56	57	58	59	60	61	62	63	64	-	65]		16	**7**

																					H2 "canonical structure" type	H2 length[19]	CDR2 IMGT [6]	Example of proteins (see table 5)
				*	*	*	*	*	*	*	*													
IGHV [9, 18, 19]	50	51	52	[52a	52b	52c	53	54	55]	56	57	58	59	60	61	62	63	64	-	65	4	6	**10**	McPC603, 4-4-20
	50	51	52	[52a	52b	-	53	54	55]	56	57	58	59	60	61	62	63	64	-	65	-	5	**9**	
	50	51	52	[52a	-	-	53	54	55]	56	57	58	59	60	61	62	63	64	-	65	2,3	4	**8**	Kol, J539
	50	51	52	[-	-	-	53	54	55]	56	57	58	59	60	61	62	63	64	-	65	**1**	3	**7**	Newm

IMGT unique numbering [6]	FR2-IMGT			CDR2-IMGT								FR3-IMGT									CDR length[17]	CDR2-IMGT [6]	
	55	[56	57	58	59	60	61	62	63	64	65]	66	67	68	69	70	71	72	73	74			
IGKV [17]	49	[50	51	52								53	54	55	56]	57	58	59	-	60		7	**3**

																					L2 "canonical structure" type	L2 length[19]	CDR2-IMGT [6]	Example of proteins (see table 5)
IGKV [9, 18, 19]	49	[50	51	52]								53	54	55	56	57	58	59	-	60	1	3	**3**	Rei, McPC603, J539

IMGT unique numbering [6]	FR2-IMGT			CDR2-IMGT								FR3-IMGT									CDR length[17]	CDR2-IMGT [6]	
	55	[56	57	58	59	60	61	62	63	64	65]	66	67	68	69	70	71	72	73	74			
IGLV [17]	49	[50	51	52								53	54	55	56]	57	58	59	-	60		7	**3**

																					L2 "canonical structure" type	L2 length[19]	CDR2-IMGT [6]	Example of proteins (see table 5)
IGLV [9, 18, 19]	49	[50	51	52]								53	54	55	56	57	58	59	-	60	1	3	**3**	Kol, Rhe

REARRANGED CDR3-IMGT

IMGT unique numbering [6]	C						CDR3-IMGT														W	CDR length[17]	CDR3-IMGT [6]	H3 "canonical structure" type	H3 length[19]	CDR3-IMGT [6]	Example of proteins (see table 5)
	104	[105	106	107	108	109	110	111	111.1	111.2	111.3	111.4	112.4	112.3	112.2	112.1	112	113	114	115	116	117]	118						
								*	*	*	*	*	*	*	*	*	*	*	*	*								
IGHV [17]	92	93	94	[95	96	97	98	99	100	100A	100B	100C	100D	100E	100F	100G	100H	100I	100J	100K	101	102]	103		19	**21**			
	92	93	94	[95	96	97	98	99	100	100A	100B	100C	100D	100E	100F	100G	100H	100I	-	-	101	102]	103		17	**19**			
	92	93	94	[95	96	97	98	99	100	100A	100B	100C	-	-	-	-	-	-	-	-	101	102]	103		11	**13**			
	92	93	94	[95	96	97	98	99	100	100A	-	-	-	-	-	-	-	-	-	-	101	102]	103		9	**11**			
								*	*	*	*	*	*	*	*	*	*	*	*	*								
IGHV [9]	92	93	94	95	[96	97	98	99	100	100a	100b	100c	100d	100e	100f	100g	100h	100i	100j	100k	101]	102	103		17	**21**			
	92	93	94	95	[96	97	98	99	100	100a	100b	100c	100d	100e	100f	100g	100h	100i	-	-	101]	102	103		15	**19**	Kol		
	92	93	94	95	[96	97	98	99	100	100a	100b	100c	-	-	-	-	-	-	-	-	101]	102	103		9	**13**	McPC603		
	92	93	94	95	[96	97	98	99	100	100a	-	-	-	-	-	-	-	-	-	-	101]	102	103		7	**11**	Newm		

IMGT unique numbering [6]	C						CDR3-IMGT									W	CDR length [17]	CDR3-IMGT [6]	L3 "canonical structure" type	L3 length[19]	CDR3-IMGT [6]	Example of proteins (see table 5)
	104	[105	106	107	108	109	110	111	111.1	112.1	112	113	114	115	116	117]	118					
							*	*	*	*	*	*	*	*								
IGKV [17]	88	[89	90	91	92	93	94	95	95A	95B	95C	95D	95E	95F	96	97]	98	15	**15**			
	88	[89	90	91	92	93	94	95	-	-	-	-	-	-	96	97]	98	9	**9**			
							*	*														
IGKV [9,19]	88	89	90	[91	92	93	94	95	-	-	-	-	-	-	96]	97	98			1,2	6	**9**	Rei, McPC603, J539

IMGT unique numbering [6]	C						CDR3-IMGT										W	CDR length [17]	CDR3-IMGT [6]	L3 "canonical structure" type	L3 length[19]	CDR3-IMGT [6]	Example of proteins (see table 5)
	104	[105	106	107	108	109	110	111	111.1	112.1	112	113	114	115	116	117]	118						
							*	*	*	*	*	*	*	*									
IGLV [17]	88	[89	90	91	92	93	94	95	95A	95B	95C	95D	95E	95F	96	97]	98	15	**15**				
	88	[89	90	91	92	93	94	95	95A	95B	-	-	-	-	96	97]	98	11	**11**				
	88	[89	90	91	92	93	94	95	-	-	-	-	-	-	96	97]	98	9	**9**				
							*	*	*	*													
IGLV [9]	88	89	90	[91	92	93	93a	93b	94	95					96]	97	98			2	8	**11**	Kol, Rhe
	88	89	90	[91	92	93	-	-	94	95					96]	97	98			1	6	**9**	Newm
							*	*	*	*													
IGLV [19]	88	89	90	[91	92	93	94	95	95a	95b					96]	97	98			2	8	**11**	Kol, Rhe
	88	89	90	[91	92	93	94	95	-	-					96]	97	98			1	6	**9**	Newm

The correspondence between the IMGT unique numbering [6] and the different numberings [9,17–19], is shown for the CDR1-, CDR2-IMGT and rearranged CDR3-IMGT of the IGHV, IGKV and IGLV genes. Brackets indicate the limits of the CDR in each numbering. Only CDR lengths mentioned in Refs. [9,17–19] are shown. A complete description of the CDR-IMGT lengths is available in IMGT (http://imgt.cines.fr). The identification of the IMGT V-GENE and alleles, V-DOMAIN CDR-IMGT lengths and PDB code of the proteins quoted in this table are given in Table 5. The protein display and Colliers de Perles of the human IG proteins (underlined in right column) are available at IMGT (http://imgt.cines.fr) in *IMGT Repertoire > 2D and 3D structures > Proteins with known 3D structures* and in the IMGT/3D structure-DB database [20].

Table 4
Gaps and additional positions for rearranged CDR3-IMGT

CDR3-IMGT lengths													
(A) Gaps for rearranged CDR3-IMGT less than 13 amino acids													
13	105	106	107	108	109	110	111	112	113	114	115	116	117
12	105	106	107	108	109	110	–	112	113	114	115	116	117
11	105	106	107	108	109	110	–	–	113	114	115	116	117
10	105	106	107	108	109	–	–	–	113	114	115	116	117
9	105	106	107	108	109	–	–	–	–	114	115	116	117
8	105	106	107	108	–	–	–	–	–	114	115	116	117
7	105	106	107	108	–	–	–	–	–	–	115	116	117
6	105	106	107	–	–	–	–	–	–	–	115	116	117
5	105	106	107	–	–	–	–	–	–	–	–	116	117
(B) Additional positions for rearranged CDR3-IMGT more than 13 amino acids													
21	**111**	111.1	111.2	111.3	111.4	112.4	112.3	112.2	112.1	**112**			
20	**111**	111.1	111.2	111.3	–	112.4	112.3	112.2	112.1	**112**			
19	**111**	111.1	111.2	111.3	–	–	112.3	112.2	112.1	**112**			
18	**111**	111.1	111.2	–	–	–	112.3	112.2	112.1	**112**			
17	**111**	111.1	111.2	–	–	–	–	112.2	112.1	**112**			
16	**111**	111.1	–	–	–	–	–	112.2	112.1	**112**			
15	**111**	111.1	–	–	–	–	–	–	112.1	**112**			
14	**111**	–	–	–	–	–	–	–	112.1	**112**			

For rearranged CDR3-IMGT more than 13 amino acids, additional positions are created between positions 111 and 112 (in bold). In a given sequence set with CDR3-IMGT more than 13 amino acids, gaps are created based on the longest CDR3-IMGT in the set. As an example, gaps are shown by comparison to a 21 amino acid long CDR3-IMGT.

Committee in 1999, and links have been made to GDB, Toronto (Canada), LocusLink, NCBI USA and GeneCards.

4.2. IMGT/V-QUEST and IMGT/JunctionAnalysis tools

The unique numbering is used as the output of IMGT/V-QUEST and IMGT/JunctionAnalysis integrated sequence alignment tools which analyse IG and TR variable (germline or rearranged) sequences and junctions according to IMGT criteria (http://imgt.cines.fr). In IMGT/V-QUEST, a variable rearranged sequence is compared to the appropriate sets of V-, D- and J-REGION alleles from the IMGT reference directory which comprise all the functional or ORF germline IMGT reference sequences [10,11]. The results show, aligned with the input sequence, the sequences of the most homologous V- and J-REGION alleles. The aligned V-REGION sequences and the Collier de Perles are displayed according to the IMGT unique numbering and with the FR- and CDR-IMGT delimitations. IMGT/JunctionAnalysis analyses the IG and TR junction of rearranged genes and cDNAs and identifies, if appropriate, the D-REGION alleles (for IG heavy chains and for TR beta and delta chains). IMGT/JunctionAnalyses provides the numbering of the JUNCTION according to the IMGT unique numbering. Based on the longest CDR3-IMGT in the set of analysed sequences, gaps or additional positions are added for CDR3-IMGT less or more than 13 amino acids, respectively (Table 4).

5. IMGT unique numbering and CDR-IMGT comparison

5.1. CDR-IMGT length definition

The CDR-IMGT length (number of codons or amino acids, that is number of occupied positions) is a crucial and original concept of IMGT-ONTOLOGY [4]. The CDR-IMGT lengths characterize the IG and TR V-REGIONs and V-DOMAINs. The length of the three CDR-IMGT, in number of codons (or amino acids), is shown into brackets and separated by dots.

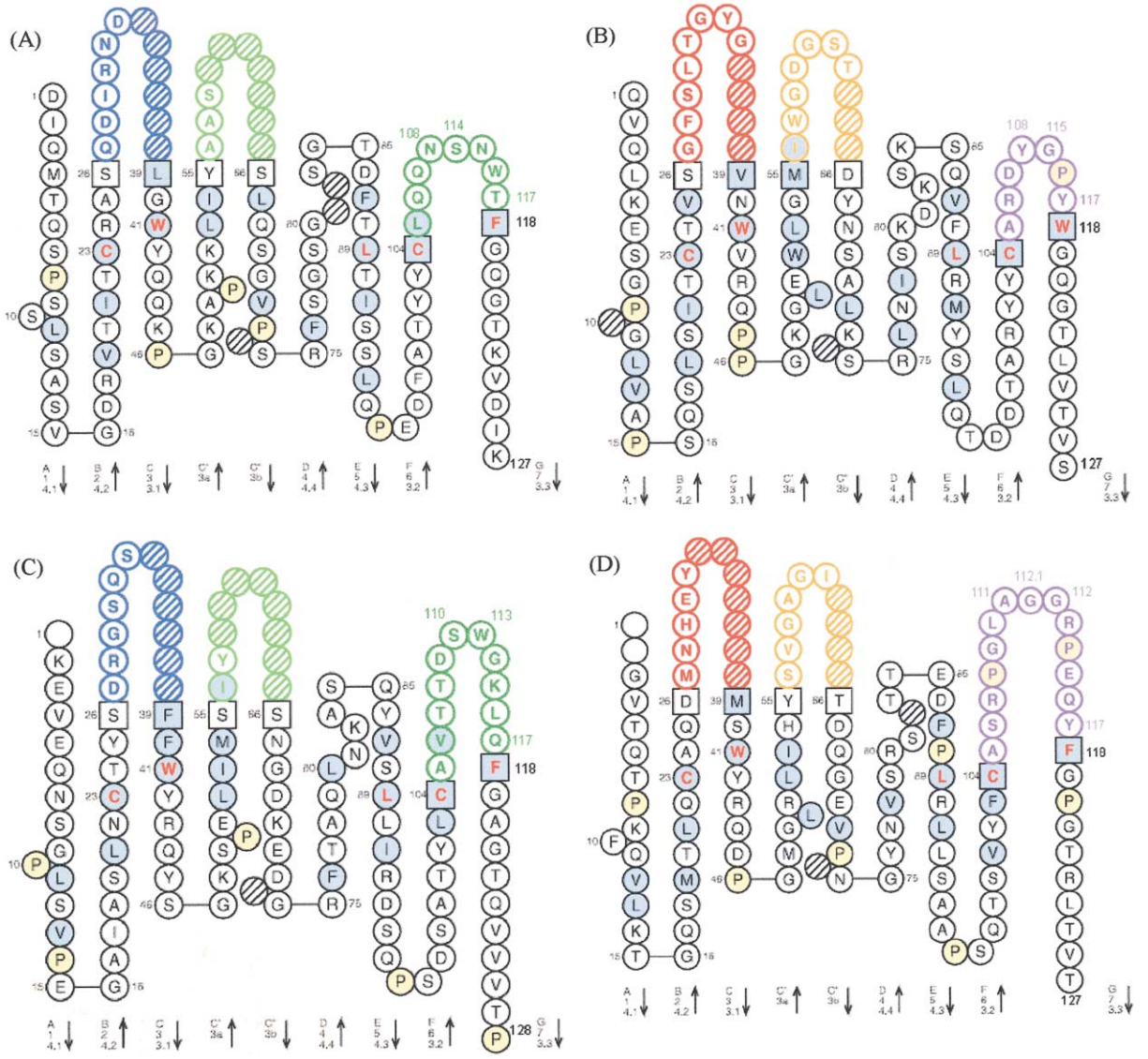
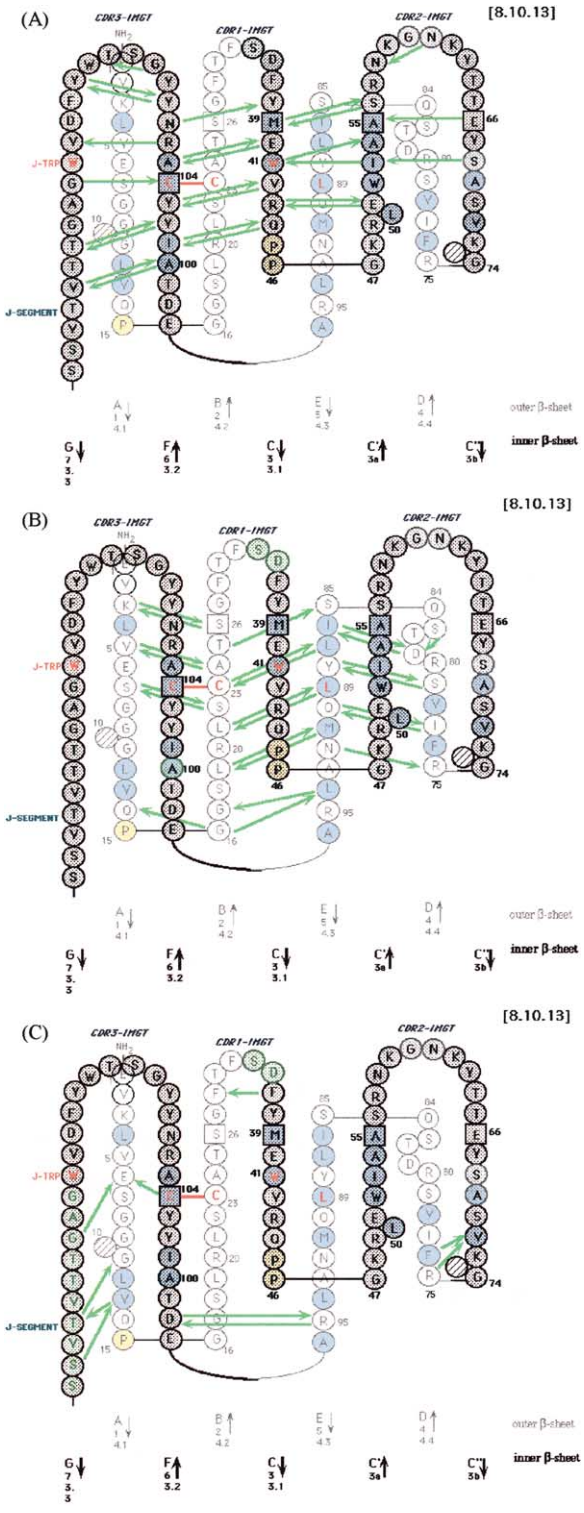


Fig. 4. Colliers de Perles of V-DOMAINs (A) Immunoglobulin IGK V-DOMAIN (V–J-REGION) (Human Mez [6.3.8], PDB code: 1dql_L), (B) Immunoglobulin IGH V-DOMAIN (V–D–J-REGION) (Human M3C65 [8.7.7], PDB code: 1dl7_H), (C) T cell receptor TRA V-DOMAIN (V–J-REGION) (Human A6 [6.2.11], PDB code: 1qrn_D), and (D) T cell receptor TRB V-DOMAIN (V–D–J-REGION) (Human A6 [5.6.14], PDB code: 1qrn_E). For (A) and (C), CDR1-IMGT in blue, CDR2-IMGT in green and CDR3-IMGT in greenblue. For (B) and (D), CDR1-IMGT in red, CDR2-IMGT in orange and CDR3-IMGT in purple.

Example: human germline IGHV1-18 [8.8.2] means that in the human germline IGHV1-18 gene, the CDR1-, CDR2- and CDR3-IMGT have a length of 8, 8 and 2 codons, respectively (Fig. 2). All CDR-IMGT lengths of the human [10,11] and mouse germline genes are available in *IMGT Repertoire > Table of FR and CDR lengths*.

Human TR A6 V-BETA [5.6.14] means that in the human TR A6 V beta domain, the CDR1-, CDR2- and CDR3-IMGT regions have a length of 5, 6 and 14 amino acids, respectively (Fig. 4). The CDR-IMGT lengths of all the human IG chains with known 3D structures are described in IMGT/3Dstructure-DB (http://imgt.cines.fr) [20] and in *IMGT Repertoire > 2*



D and 3D structures > Proteins with known 3D structures.

5.2. Codon or amino acid identifiant

Any codon or amino acid in a CDR-IMGT is characterized by its own IMGT unique numbering and the length (in number of codons or amino acids shown between brackets) of the CDR-IMGT to which it belongs. This identifiant is in itself sufficient to assign a codon or an amino acid in a given CDR-IMGT and to localize it precisely in the CDR-IMGT (positions 27, 56 and 105 being the first position of CDR1-, CDR2- and CDR3-IMGT, respectively). As an example in Fig. 4D, M27[5] would indicate that a methionine is the first amino acid in a five amino acid CDR1-IMGT, G59[6] that a glycine is the fourth amino acid in a six amino acid CDR2-IMGT and A111[14] that an alanine is the seventh amino acid in a 14 amino acid CDR3-IMGT. This also applies for the corresponding codons.

5.3. CDR-IMGT sequence and structural data comparison

Practically, sequence and structural data comparisons of CDR-IMGT of same length can be done directly using the IMGT unique numbering. For example, the codon or amino acid at position 28 can be compared to all codons or amino acids at the same position for all CDR1-IMGT of the same length. This also holds for statistical analysis, position per position.

For multiple sequence alignment and structural data comparisons and statistical analysis, position per position, of CDR-IMGT of different lengths, the codon (or amino acid) identifiants allow a flexibility in

Fig. 5. Colliers de Perles (in space) of the mouse (*Mus musculus*) IGH V-DOMAIN from McPC603 (PDB code: 2mcp_H) with hydrogen bonds (A) between the GFCC′C″ strands of the inner beta sheet, (B) between the ABED strands of the outer beta sheet and (C) between the inner and outer beta sheets. Hydrogen bonds between the strands are represented by green lines. The Colliers de Perles show, on the forefront, the inner GFCC′C″ strands located at the interface with the light chain V-DOMAIN (not shown) and, on the back, the ABED strands located on the outer surface of the heavy chain V-DOMAIN.

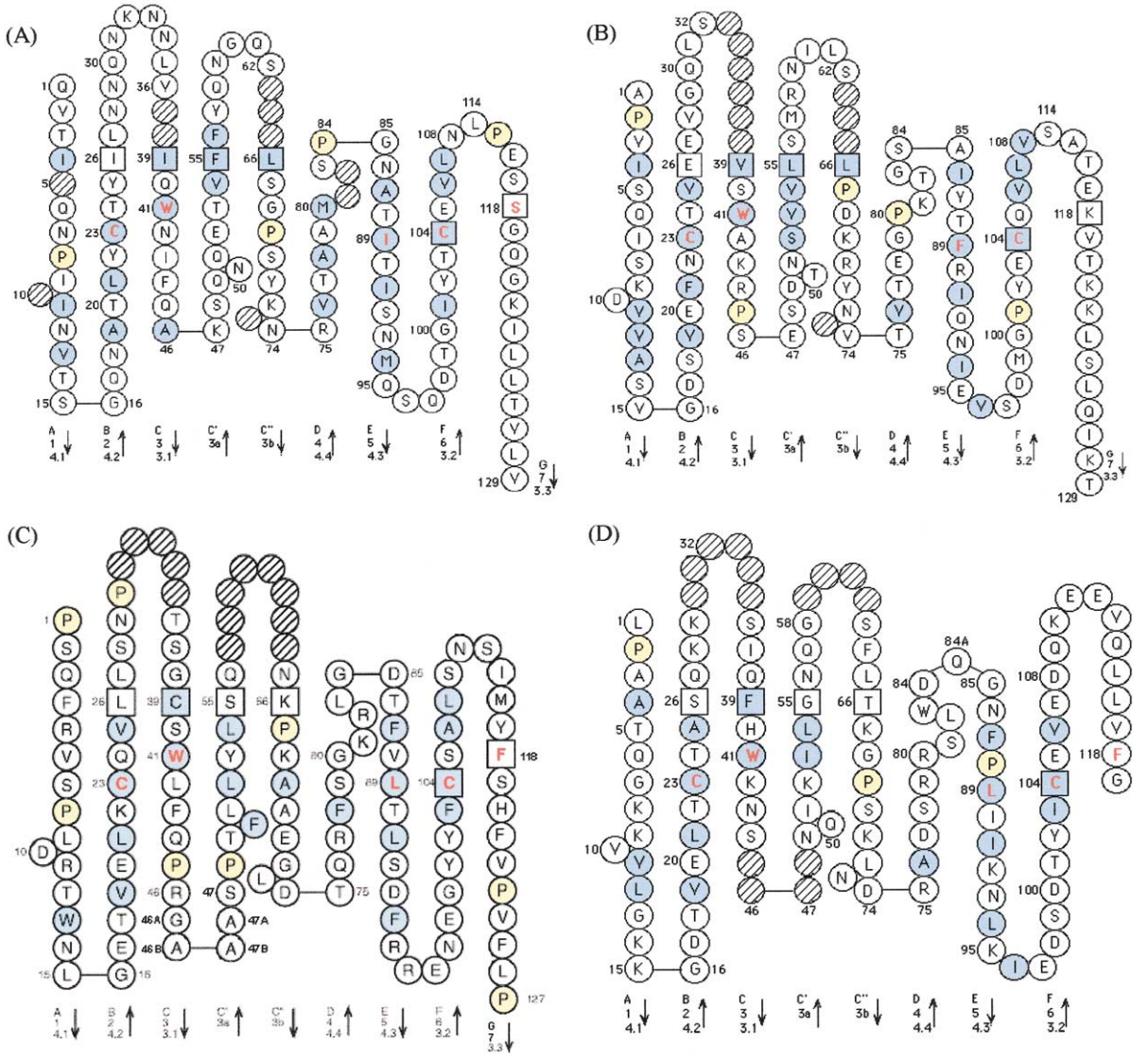


Fig. 6. Colliers de Perles of V-LIKE-DOMAINs (A) CTX [10.7.8] (*Xenopus laevis*) (EMBL/GenBank/DDBJ accession number: U43330), (B) Amalgam domain 1 [6.7.8] (*Drosophila melanogaster*) (EMBL/GenBank/DDBJ accession number: M23561), (C) CD8A [7.2.9] (*Homo sapiens*) (M27161) (PDB code: 1cd8), and (D) CD4 domain 1 [6.6.13] (*H. sapiens*) (U47924) (PDB code: 1cdj). The CTX and CD4 domain 1 V-LIKE-DOMAINs are encoded by two exons instead of being encoded by a single V-EXON (first exon: positions 1–53 for CTX (U43394), positions 1–67 for CD4 (U47924); second exon: positions 54–123 for CTX (U43393), positions 68–119 for CD4 (U47924)). The amalgam V-LIKE-DOMAIN (domain 1) is encoded as part of a unique exon which also encodes two C-LIKE-DOMAINs (domains 2 and 3). The CD8A V-LIKE-DOMAIN (up to position 127) is encoded by a nonrearranging exon. Four additional positions are found in CD8A between 46 and 47. In contrast, four positions (positions 45–48) are missing in CD4 domain 1. Note that gaps can be moved either to the end of CDR-IMGT (A and B) or at the top of the loops (C and D), or they can be omitted as in the CDR3-IMGT (A–D).

the data analysis and results representation. Indeed, if the codon (or amino acid) positions in the IMGT unique numbering and the CDR-IMGT lengths are known, CDR-IMGT gaps can be moved between the anchor positions (26 and 39 for CDR1-IMGT, 55 and 66 for CDR2-IMGT, 104 and 118 for CDR3-IMGT, respectively), depending on the length of the longest CDR-IMGT in the set of analysed sequences and on

Table 5
Identification of the IMGT V-GENE and alleles, V-DOMAIN CDR-IMGT lengths and PDB code of the proteins quoted in Table 3

Species	IG heavy chain				IG light chain (kappa or lambda)		
	Protein name	IGHV gene and allele	CDR-IMGT lengths	PDB code	IGKV or IGLV gene and allele	CDR-IMGT lengths	PDB code
Human	Newm	IGHV4-59 * 04	[8.7.11]	7fab_H	IGLV1-40 * 01	[9.3.9]	7fab_L
	Kol	IGHV3-33 * 01 or IGHV3-33 * 04	[8.8.19]	2fb4_H; 2ig2_H	IGLV1-44 * 01	[8.3.11]	2fb4_L; 2ig2_L
	Rhe				IGLV1-36 * 01	[8.3.11]	2rhe_–
	Hil	IGHV3-33 * 01 or IGHV3-33 * 04	[8.8.14]	8fab_B; 8fab_D	IGLV3-25 * 02	[6.3.9]	8fab_A; 8fab_C
	Rei				IGKV1-33 * 01	[6.3.9]	1rei_A; 1rei_B
Mouse (*Mus musculus*)	McPC603	IGHV7S1 * 01	[8.10.13]	1mcp_H; 2mcp_H	IGKV8-28 * 01	[12.3.9]	1mcp_L; 2mcp_L; 2imm_–
	J539	IGHV4S1 * 01	[8.8.11]	2fbj_H	IGKV4-86 * 01	[5.3.9]	2fbj_L
	4-4-20	IGHV6S1 * 01	[8.10.9]	4fab_H; 1flr_H	IGKV1-110 * 01	[11.3.9]	4fab_L; 1flr_L

the type of analysis performed or results provided. For example, if the longest CDR1-IMGT or CDR2-IMGT is less than 12 or 10 codon or amino acid long, respectively, the resulting gaps can be ignored or moved next to both anchor positions (with an equal number of gaps on both sides of the codons (or amino acids) if even number, or one more gap on the right side of the codons (or amino acids) if odd number). Moreover, if the sequences have CDR-IMGT of different lengths, the corresponding gaps can be moved at the middle of the CDR-IMGT (or top of the loop) (with an equal number of codons or (amino acids) on both sides if the CDR-IMGT length is an even number, or with one more codon (or amino acid) in the left part if it is an odd number; Fig. 6). In all cases, the codon (or amino acid) identifiant provides sufficient information to correctly identify and compare any codon (or amino acid) for any type of CDR-IMGT sequence and structural data comparison, while allowing a flexibility in data analysis and representation.

6. IMGT unique numbering, data coherence and structure prediction

As soon as the first Collier de Perles was set up on the web site in December 1997, the enormous potential of the IMGT unique numbering as a means to control data coherence was obvious. Indeed, this Collier de Perles of the mouse antibody E5.2Fv (PDB code: 1dvf) showed that an amino acid was missing in both the ATOM lines and the SEQRES lines of the PDB file, which was confirmed by the original publication (*IMGT Repertoire > IMGT 2D and 3D representations (examples) > Mouse E5.2Fv*). For new sequences, for which there are no 3D structures available, the Colliers de Perles allow to precisely delimit the FR- and CDR-IMGT and considerably help in predicting the structures of these molecules, particularly the antigen binding sites. Thus, the Colliers de Perles for the V-REGION of the human TRGV9 and TRDV2 were entirely confirmed once the 3D structure of a gamma–delta TR became available [21]. This has been extended from IG and TR V-REGIONs and V-DOMAINs from human [23–31] to those of mouse [32,33] and other vertebrates [34,35].

7. IMGT unique numbering for V-LIKE-DOMAIN

The IMGT unique numbering gives insight in the structural configuration of the variable domain but also opens interesting views on the evolution of the sequences of the V-set [22]. These sequences possess at least one V-LIKE-DOMAIN, that is a domain of similar structure as a V-DOMAIN, but found in chains other than IG or TR and include nonrearranging sequences in vertebrates (CD4, CD8, *Xenopus* CTX, etc.) and in invertebrates (*Drosophila* Amalgam, *Drosophila* Fasciclin II, etc.). The IMGT unique numbering has been applied with success to all the sequences belonging to the V-set of the IG superfamily [5,6] and examples are shown in Fig. 6.

In the absence of available 3D structures, the V-LIKE-DOMAIN Colliers de Perles are particularly useful to compare domains of very diverse families. Thus, the V-LIKE-DOMAIN description applies to all the extracellular domains of the bony fish novel immune-type receptors (NITR) described in *Spheroides nephelus* (Southern pufferfish) [36], *Danio rerio* (zebrafish) [37] and *Ictalurus punctatus* (channel catfish) [38]. The IMGT unique numbering allows to show that the distinction between V and V/C2 [36–38] becomes unnecessary. It allows, by comparison with genomic sequences, to delimit the V-LIKE-DOMAIN which is, in the NITR molecules, encoded by an exon. Moreover, it allows to determine the LOOP-IMGT lengths of the BC, C′C″ and FG loops of the V-LIKE-DOMAINs (equivalent of the IG and TR CDR1-, CDR2- and CDR3-IMGT, respectively). For example, the LOOP-IMGT lengths of *I. punctatus* NITR1 domains 1 and 2 are [6.6.9] and [10.4.9], respectively, and those of the NITR6 (AF397459) single domain are [7.7.9]. Colliers de Perles of these V-LIKE-DOMAINs are available at http://imgt.cines.fr (IMGT Repertoire > Immunoglobulin superfamily).

8. Conclusion

The IMGT unique numbering has many advantages. It allows an easy comparison between sequences coding the variable regions and domains, whatever the antigen receptor (IG or TR), the chain type (heavy or light chains for IG; alpha, beta, gamma or delta chains for TR), or the species. This unique numbering has allowed the redefinition of the limits of the FR and CDR. The FR- and CDR-IMGT lengths (number of codons or amino acids, that is number of occupied positions) become crucial information characterizing the variable regions and domains, and the corresponding genes, cDNAs and proteins [10–13]. The CDR lengths based on the standardized IMGT unique numbering [6] replace the diverse definitions and delimitations of the CDR, H and L loops and canonical structures found in the literature [9,17–19]. The IMGT unique numbering has allowed standardized analysis and representations of nucleotide and amino acid sequences (in *IMGT Repertoire*, Tables of FR and CDR lengths, Tables of alleles, Alignments of alleles, Protein displays, Colliers de Perles, 3D structures). The IMGT unique numbering provides standardized results for the IMGT/V-QUEST and IMGT/JunctionAnalysis tools and for queries of the IMGT/LIGM-DB and IMGT/3Dstructures-DB databases.

Moreover, the IMGT unique numbering has been extended to the V-like domain and represents, therefore, a major step forward in analysing and comparing the structure and evolution of the molecules belonging to the IG superfamily.

Acknowledgements

We are grateful to Nathalie Bosc, Géraldine Folch, Chantal Ginestoux, Christèle Jean, Dominique Scaviner and Denys Chaume for helpful discussion. We thank Nora Bonnet-Saidali for help in the preparation of the manuscript. IMGT is funded by the European Union's 5th PCRDT (QLG2-2000-01287) program, the Centre National de la Recherche Scientifique (CNRS), the Ministère de la Recherche et de l'Education Nationale. Subventions have been received from Association pour la Recherche sur le Cancer (ARC) and the Région Languedoc-Roussillon.

References

[1] Lefranc M-P. IMGT, the international ImMunoGeneTics database. Nucl Acids Res 2001;29:207–9.

[2] Lefranc M-P, Giudicelli V, Busin C, Bodmer J, Müller W, Bontrop R, Lemaitre M, Malik A, Chaume D. IMGT, the international ImMunoGeneTics database. Nucl Acids Res 1998;26:297–303.

[3] Lefranc M-P, Giudicelli V, Ginestoux C, Bodmer J, Müller W, Bontrop R, Lemaître M, Malik A, Barbié V, Chaume D. IMGT, the international ImMunoGeneTics database. Nucl Acids Res 1999;27:209–12.

[4] Giudicelli V, Lefranc M-P. Ontology for immunogenetics: IMGT-ONTOLOGY. Bioinformatics 1999;12:1047–54.

[5] Lefranc M-P. Unique database numbering system for immunogenetic analysis. Immunol Today 1997;18:509.

[6] Lefranc M-P. The IMGT unique numbering for immunoglobulins, T cell receptors and Ig-like domains. The Immunologist 1999;7:132–6.

[7] Kabat EA, Wu TT, Reid-Miller M, Perry HM, Gottesman KS. Sequences of proteins of immunological interest. Washington, DC: National Institute of Health; 1987. 804 pp., Publication no. 165-462.

[8] Satow Y, Cohen GH, Padlan EA, Davies DR. Phosphocholine binding immunoglobulin Fab McPC603. An X-ray diffraction study at 2.7 Å. J Mol Biol 1986;190:593–604.

[9] Chothia C, Lesk AM. Canonical structures for the hypervariable regions of immunoglobulins. J Mol Biol 1987;196:901–17.

[10] Lefranc M-P, Lefranc G. The immunoglobulin FactsBook. London: Academic Press; 2001. 458 pp., ISBN: 012441351X.

[11] Lefranc M-P, Lefranc G. The T cell receptor FactsBook. London: Academic Press; 2001. 398 pp., ISBN: 0124413528.

[12] Scaviner D, Barbié V, Ruiz M, Lefranc M-P. Protein displays of the human immunoglobulin heavy, kappa and lambda variable and joining regions. Exp Clin Immunogenet 1999;16:234–40.

[13] Folch G, Scaviner D, Contet V, Lefranc M-P. Protein displays of the human T cell receptor alpha, beta, gamma and delta variable and joining regions. Exp Clin Immunogenet 2000;17:205–15.

[14] Ruiz M, Giudicelli V, Ginestoux C, Stoehr P, Robinson J, Bodmer J, Marsh S, Bontrop R, Lemaitre M, Lefranc G, Chaume D, Lefranc M-P. IMGT, the international ImMunoGeneTics database. Nucl Acid Res 2000;28:219–21.

[15] Lefranc M-P. IMGT ImMunoGeneTics Database. Int BIOforum 2000;4:98–100.

[16] Lefranc M-P. IMGT, the international ImMunoGeneTics database: a high-quality information system for comparative immunogenetics and immunology. Dev Comp Immunol 2002;26:697–705.

[17] Kabat EA, Wu TT, Perry HM, Gottesman KS, Foeller C. Sequences of proteins of immunological interest. Washington, DC: National Institute of Health; 1991. Publication no. 91-3242.

[18] Chothia C, Lesk AM, Tramontano A, Levitt M, Smith-Gill SJ, Air G, Sheriff S, Padlan EA, Davies D, Tulip WR, et al. Conformation of immunoglobulin hypervariable regions. Nature 1989;342:877–83.

[19] Al-Lazikani B, Lesk AM, Chothia C. Standard conformations for the canonical structures of immunoglobulins. J Mol Biol 1997;273:927–48.

[20] Ruiz M, Lefranc M-P. IMGT gene identification and Colliers de Perles of human immunoglobulin with known 3D structures. Immunogenetics DOI 10.1007/s00251-001-0408-6. Immunogenetics 2002;53:857–83.

[21] Allison TJ, Winter CC, Fournié JJ, Bonneville M, Garboczi DN. Structure of a human gamma delta T-cell antigen receptor. Nature 2001;411:820–4.

[22] Williams AF, Barclay AN. The immunoglobulin superfamily—domains for cell surface recognition. Annu Rev Immunol 1988;6:381–405.

[23] Giudicelli V, Chaume D, Bodmer J, Müller W, Busin C, Marsh S, Bontrop R, Lemaitre M, Malik A, Lefranc M-P. IMGT, the international ImMunoGeneTics database. Nucl Acids Res 1997;25:206–11.

[24] Lefranc M-P. Nomenclature of the human immunoglobulin heavy (IGH) genes. Exp Clin Immunogenet 2001;18:100–16.

[25] Lefranc M-P. Nomenclature of the human immunoglobulin kappa (IGK) genes. Exp Clin Immunogenet 2001;18:161–74.

[26] Lefranc M-P. Nomenclature of the human immunoglobulin lambda (IGL) genes. Exp Clin Immunogenet 2001;18:242–54.

[27] Pallarès N, Frippiat JP, Giudicelli V, Lefranc M-P. The human immunoglobulin lambda variable (IGLV) genes and joining (IGLJ) segments. Exp Clin Immunogenet 1998;15:8–18.

[28] Barbié V, Lefranc M-P. The human immunoglobulin kappa variable (IGKV) genes and joining (IGKJ) segments. Exp Clin Immunogenet 1998;15:171–83.

[29] Pallarès N, Lefebvre S, Contet V, Matsuda F, Lefranc M-P. The human immunoglobulin heavy variable (IGHV) genes. Exp Clin Immunogenet 1999;16:36–60.

[30] Folch G, Lefranc M-P. The human T cell receptor beta variable (TRBV) genes. Exp Clin Immunogenet 2000;17:42–54.

[31] Scaviner D, Lefranc M-P. The human T cell receptor alpha variable (TRAV) genes. Exp Clin Immunogenet 2000;17:83–96.

[32] Bosc N, Lefranc M-P. The mouse (*Mus musculus*) T cell receptor beta variable (TRBV), diversity (TRBD), and joining (TRBJ) genes. Exp Clin Immunogenet 2000;17:216–28.

[33] Bosc N, Contet V, Lefranc M-P. The mouse (*Mus musculus*) T cell receptor delta variable (TRDV), diversity (TRDD), and joining (TRDJ) genes. Exp Clin Immunogenet 2001;18:51–8.

[34] Artero S, Lefranc M-P. The Teleostei immunoglobulin heavy IGH genes. Exp Clin Immunogenet 2000;17:148–61.

[35] Artero S, Lefranc M-P. The Teleostei immunoglobulin light IGL1 and IGL2 V, J and C genes. Exp Clin Immunogenet 2000;17:162–72.

[36] Strong SJ, Mueller MG, Litman RT, Hawke NA, Haire RN, Miracle AL, Rast JP, Amemiya CT, Litman GW. A novel multigene family encodes diversified variable regions. Proc Natl Acad Sci USA 1999;96:15080–5.

[37] Yoder JA, Mueller MG, Wei S, Corliss BC, Prather DM, Willis T, Litman RT, Djeu JY, Litman GW. Immune-type receptor genes in zebrafish share genetic and functional properties with genes encoded by the mammalian leukocyte receptor cluster. Proc Natl Acad Sci USA 2001;98:6771–6.

[38] Hawke NA, Yoder JA, Haire RN, Mueller MG, Litman RT, Miracle AL, Stuge T, Shen L, Miller N, Litman GW. Extraordinary variation in a diversified family of immune-type receptor genes. Proc Natl Acad Sci USA 2001;98: 13832–7.