Bioinformatique et ontologies

Marie-Paule Lefranc

Université Montpellier 2
Institut de Génétique Humaine, UPR CNRS 1142

Module 6334 proposé par la Maison des Ecoles Doctorales
12-14 Mai, 2009, Montpellier

Il y a autant de définitions de la « bioinformatique »
et des « ontologies » qu'il y en a pour la notion de
« gène ».

Au-delà de cette diversité, nous définirons les termes
génériques du domaine.

Nous montrerons comment se construit une ontologie pour
le partage des connaissances.

Nous décrirons pourquoi une ontologie est une nécessité
en bioinformatique,
en particulier lors de l'évolution de systèmes d'information
et lors des modélisations de systèmes biologiques.

A partir d’exemples, nous dégagerons les axiomes et principaux concepts qui ont permis à une ontologie, IMGT-ONTOLOGY, de devenir un paradigme au niveau international.

Nous montrerons que les axiomes d’IMGT-ONTOLOGY qui permettent d’appréhender les connaissances sous différentes facettes sont utilisables pour une représentation multi-échelle (moléculaire, cellulaire, de l’organisme et de population),

et par suite en biologie systémique pour la modélisation de réseaux de régulations, de processus biologiques, de communications entre organes et au sein de population.

- La construction d’une ontologie engendre une dynamique continuelle en recherche fondamentale et en recherche appliquée qui intègre les avancées scientifiques et technologiques du domaine.

- En terme de visibilité internationale et à l’aide d’exemples en recherche clinique (diagnostic des leucémies) et en biotechnologie (ingénierie des anticorps, humanisation des anticorps), nous montrerons comment IMGT-ONTOLOGY permet:
- de gérer efficacement un projet,
- de concilier assurance qualité et créativité,
- et de favoriser une valorisation de la recherche.

Acquis à la fin de la formation:

Cette formation démontrera qu'une approche intégrée des connaissances
d'un système complexe en bioinformatique
est possible à la condition de reposer sur une solide ontologie.

L'étudiant réalisera de plus qu'une ontologie
amène tout naturellement à une assurance qualité
et à une valorisation de la recherche.

Mardi 12 mai 2009:

Marie-Paule Lefranc
IMGT-ONTOLOGY axioms et concepts

Mercredi 13 mai 2009:

Véronique Giudicelli
Ontologies et Protégé

Jeudi 14 mai 2009:

Patrice Duroux
Ontologies et Système d'information

Laetitia Regnier
Ontologies et Contrôle de qualité

http://www.imgt.org

Mardi 12 mai 2009

1. IMGT®

   Quel est le domaine d'expertise d'IMGT®?

2. Ontologie

   Qu'est-ce qu'une ontologie?

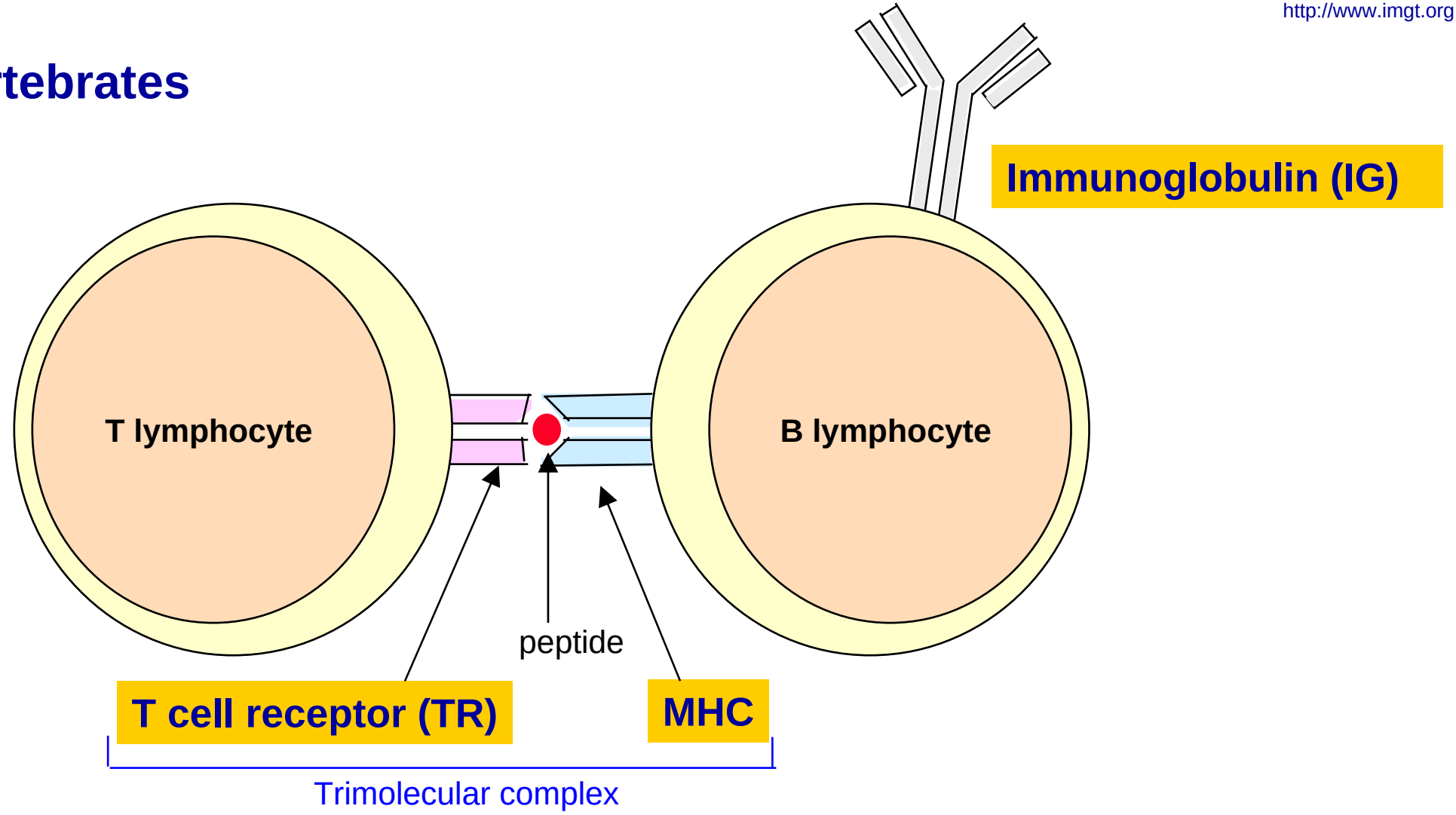
3. IMGT-ONTOLOGY axioms et concepts

4. Exemples d'applications

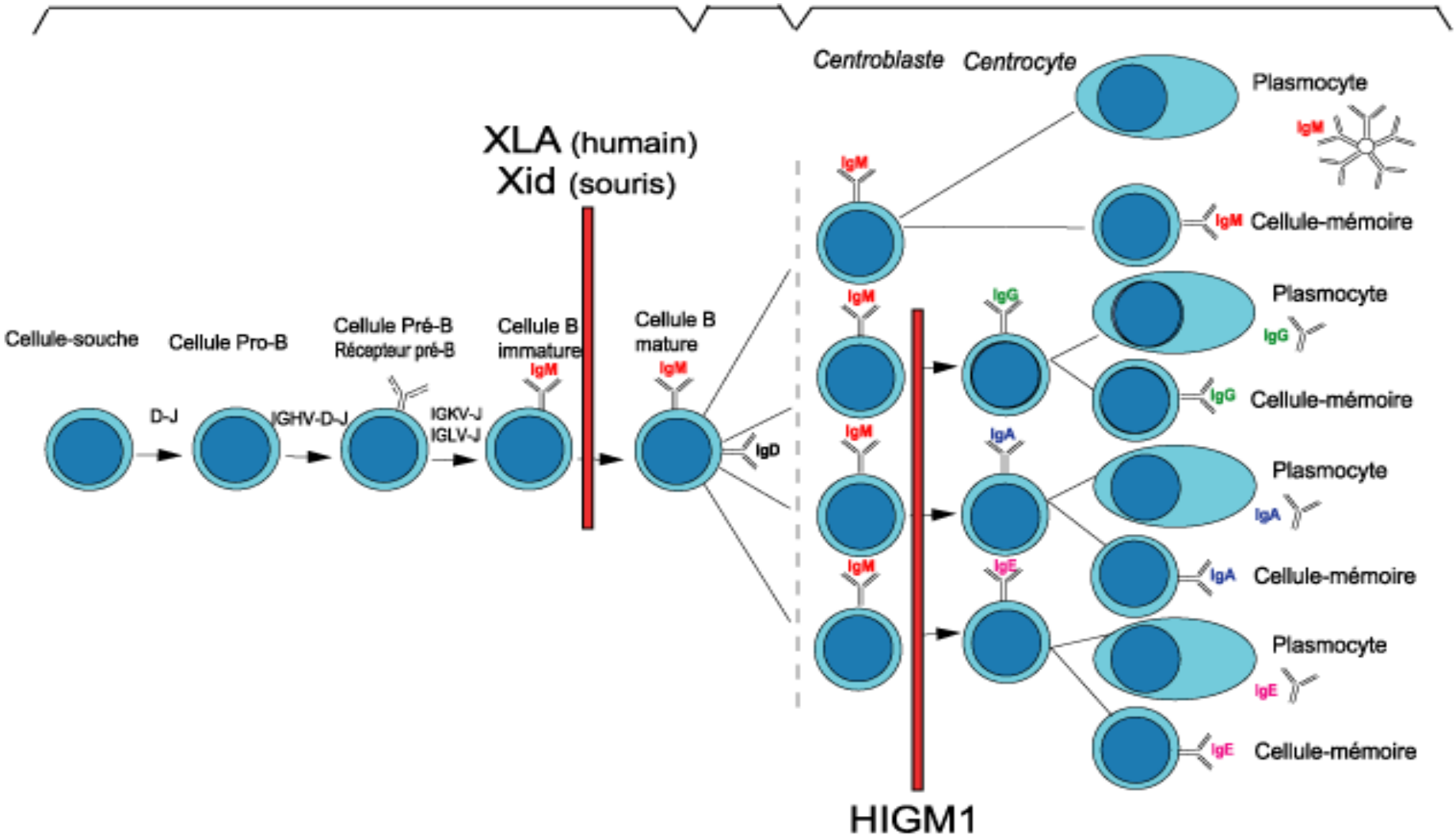
IMGT® domain: the adaptive immune response

http://www.imgt.org

Vertebrates



Bone marrow **Blood** **Lymph nodes, spleen**

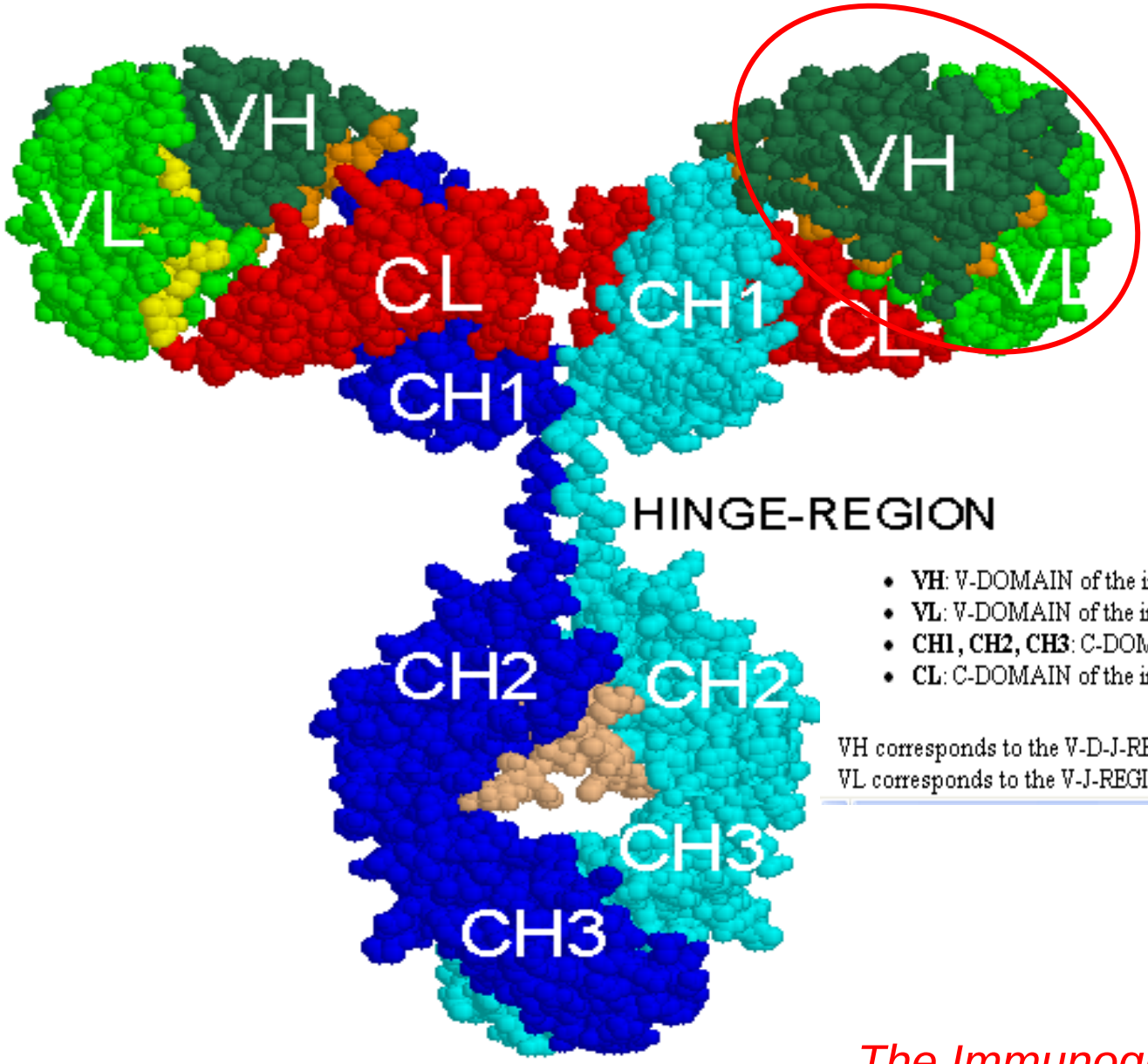


V-D-J and V-J rearrangements **Hypermutations, selection**

http://www.imgt.org

Immunoglobulin or antibody

http://www.imgt.org



IgG

- **VH**: V-DOMAIN of the immunoglobulin heavy chain
- **VL**: V-DOMAIN of the immunoglobulin light chain
- **CH1, CH2, CH3**: C-DOMAIN of the immunoglobulin heavy chain
- **CL**: C-DOMAIN of the immunoglobulin light chain

VH corresponds to the V-D-J-REGION (in green (V), orange (DJ)) of the heavy chain.
VL corresponds to the V-J-REGION (in green (V) and yellow (J)) of the light chain.

The Immunoglobulin FactsBook, 2001

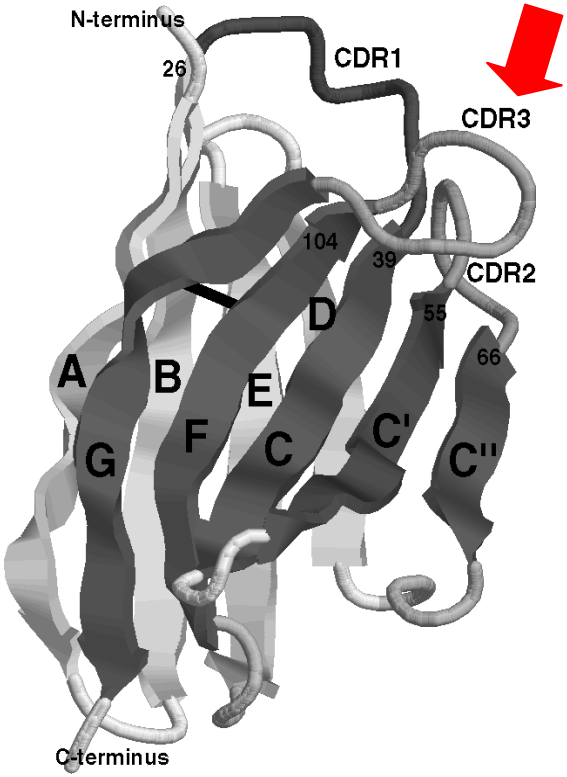
Structural domains

http://www.imgt.org

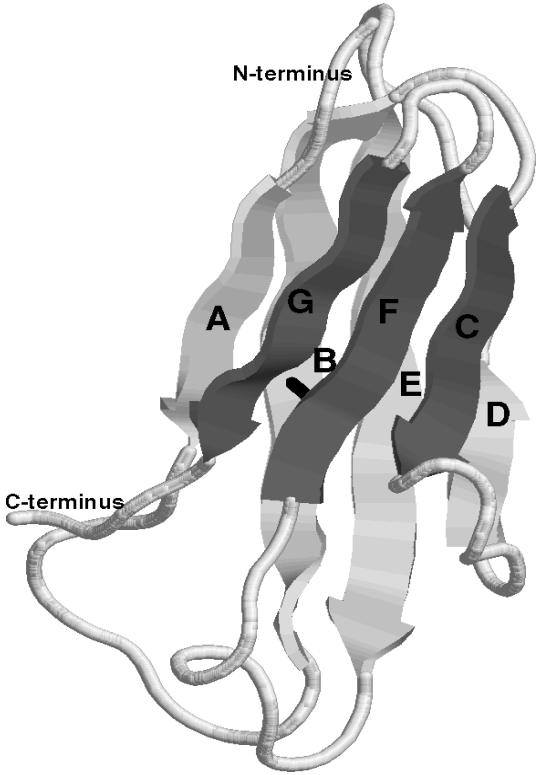
IG and TR

MHC

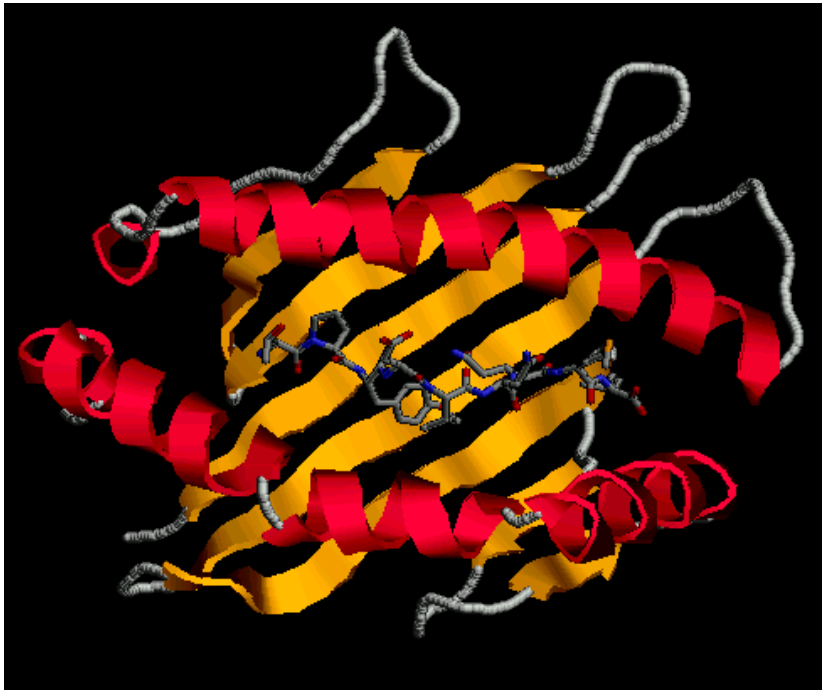
V-DOMAIN



C-DOMAIN



G-DOMAINs

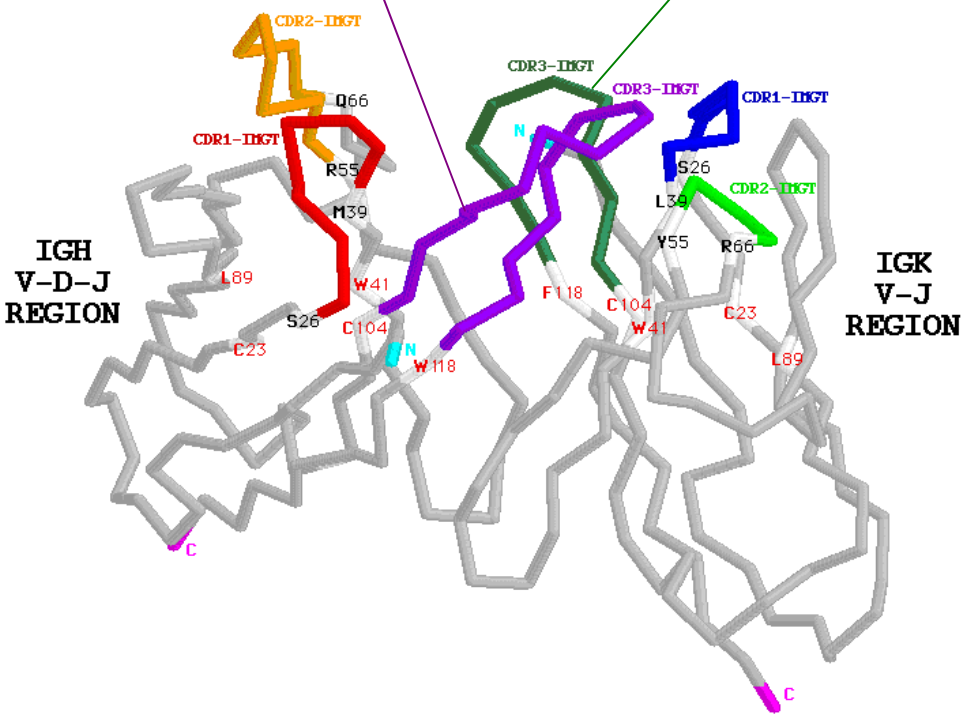


V-DOMAINs: VH and VL

http://www.imgt.org

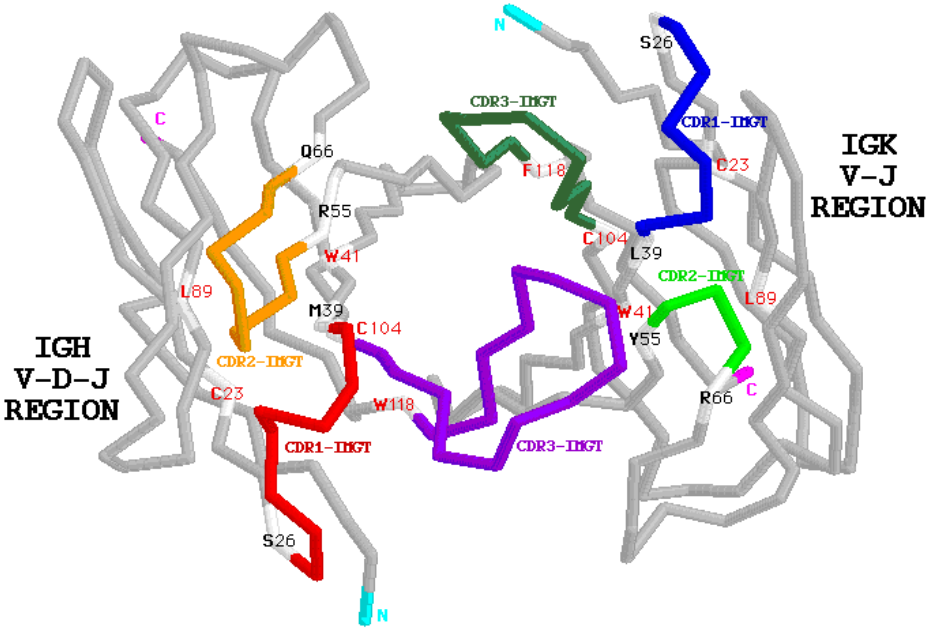
VH VL

V-D-J junction V-J junction



Side view of the V-DOMAINs

VH VL



View from above the CDRs

***Mouse(Mus musculus)* E5.2Fv**

CDR3-IMGT= Complementarity determining region (105-117)

V-J junction (104-118)

V-D-J junction (104-118)

Immunoglobulin (IG)

T cell receptor (TR)

http://www.imgt.org

V-DOMAIN

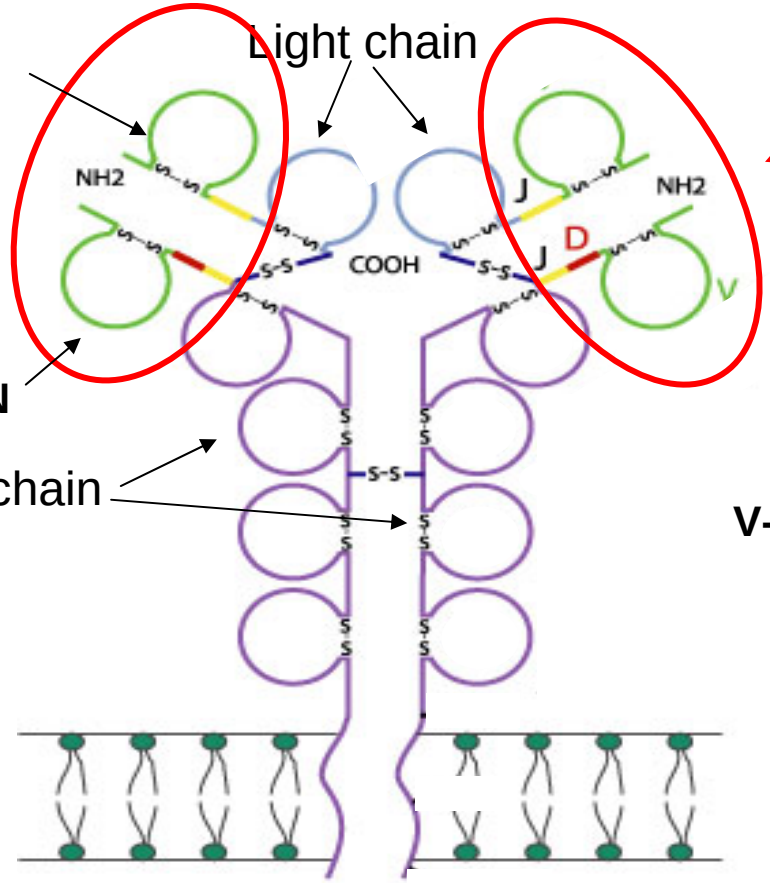
V-J-REGION

Light chain

V-DOMAIN

V-D-J-REGION

Heavy chain



Membrane IgM

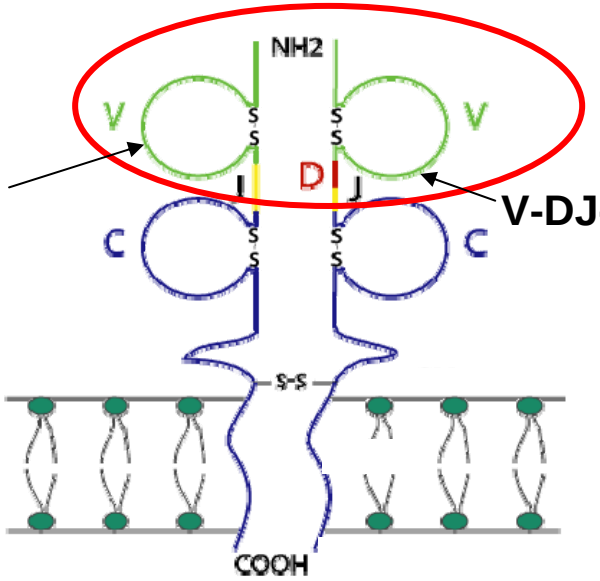
Contribution of the
2 V-DOMAINs
to the antigen binding site

Alpha - Beta

Gamma - Delta

V-J-REGION

V-DJ-REGION

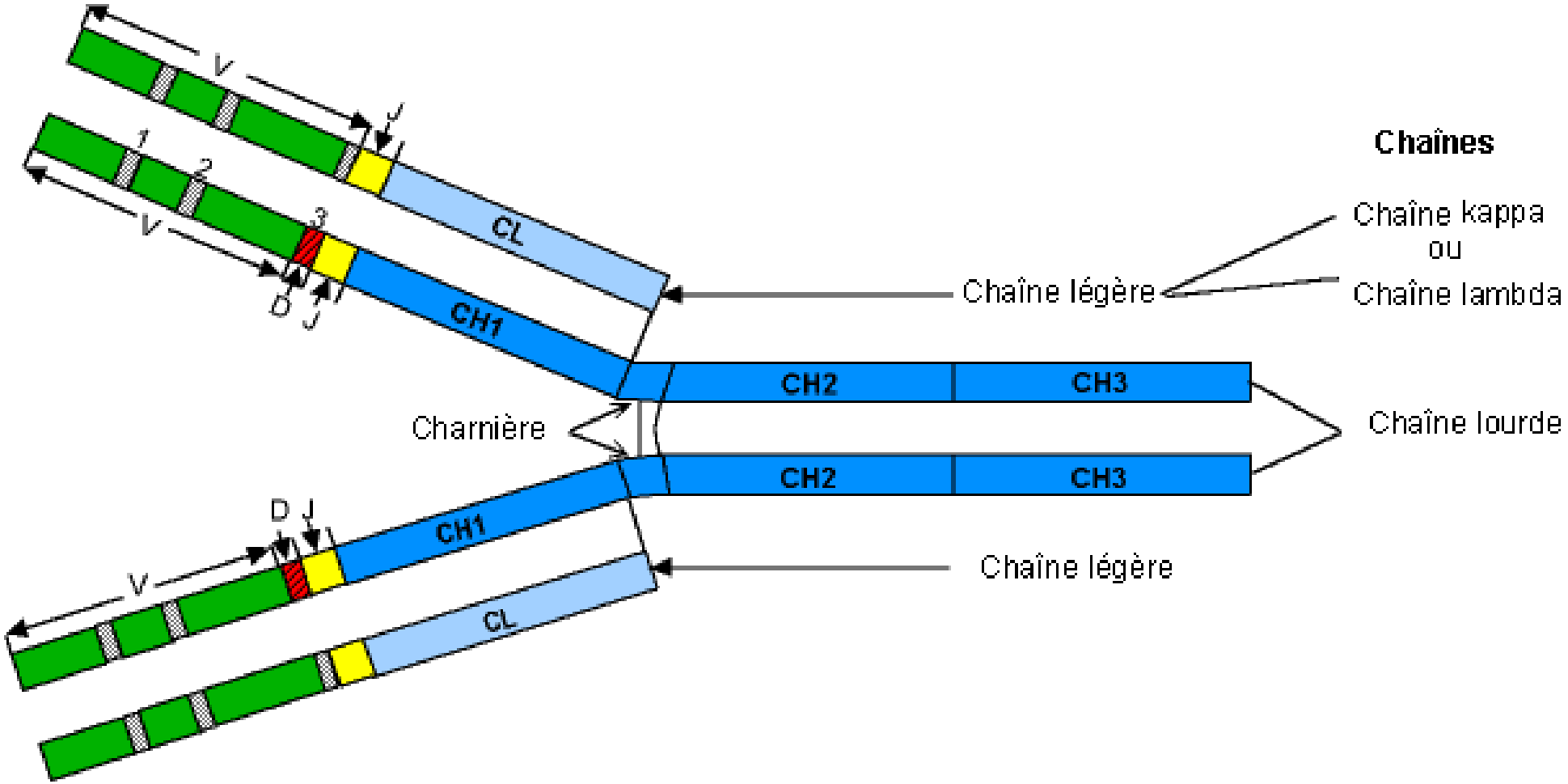


T cell receptor

The Immunoglobulin FactsBook, 2001

Immunoglobulin IgG

http://www.imgt.org

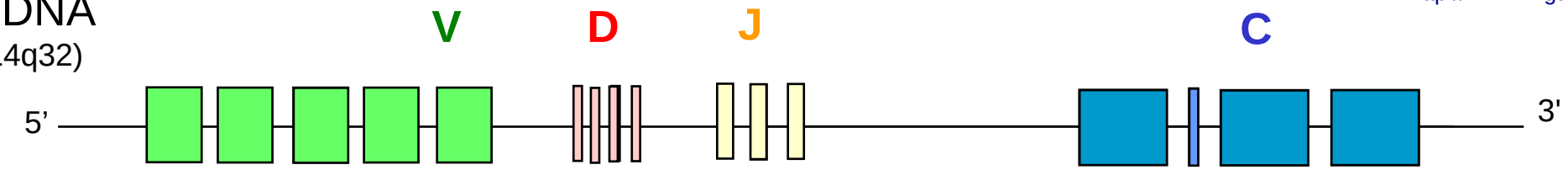


IMGT Repertoire, http://www.imgt.org

Immunoglobulin (IG) synthesis

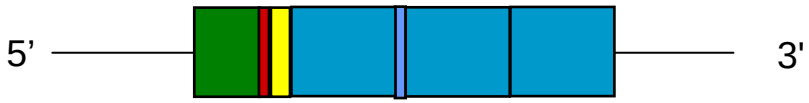
http://www.imgt.org

genomic DNA
(IGH Locus 14q32)

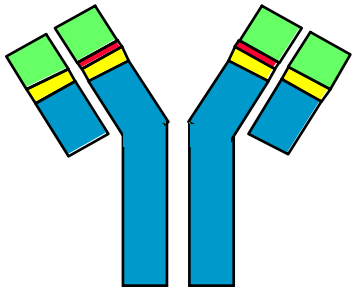


rearranged DNA

mRNA



2×10^{12} different IG per individual



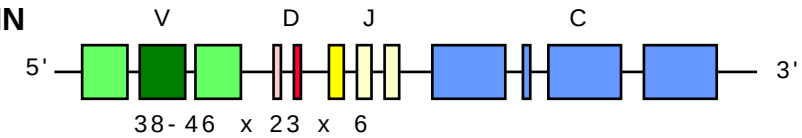
IMGT Repertoire, http://www.imgt.org

Immunoglobulin (IG) synthesis

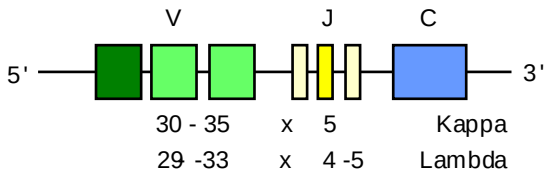
http://www.imgt.org

150
FUNCTIONAL IG GENES

HEAVY CHAIN



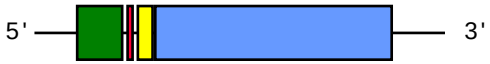
LIGHT CHAIN



6300 **POTENTIAL RECOMBINATIONS**

185 +165 **POTENTIAL RECOMBINATIONS**

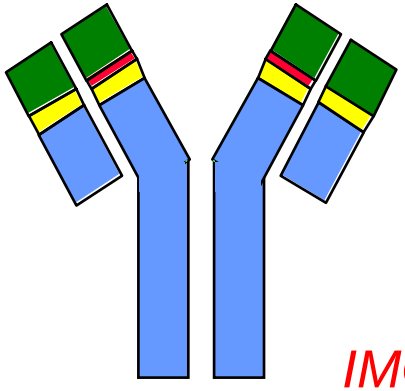
N-DIVERSITY
SOMATIC MUTATIONS
x 1000



ABOUT **6.3×10^6** POSSIBILITIES

ABOUT **3.5×10^5** POSSIBILITIES

2×10^{12}
DIFFERENT ANTIBODIES



IMGT Repertoire, http://www.imgt.org

IMGT® http://www.imgt.org

http://www.imgt.org

IMGT®, the international ImMunoGeneTics information system®

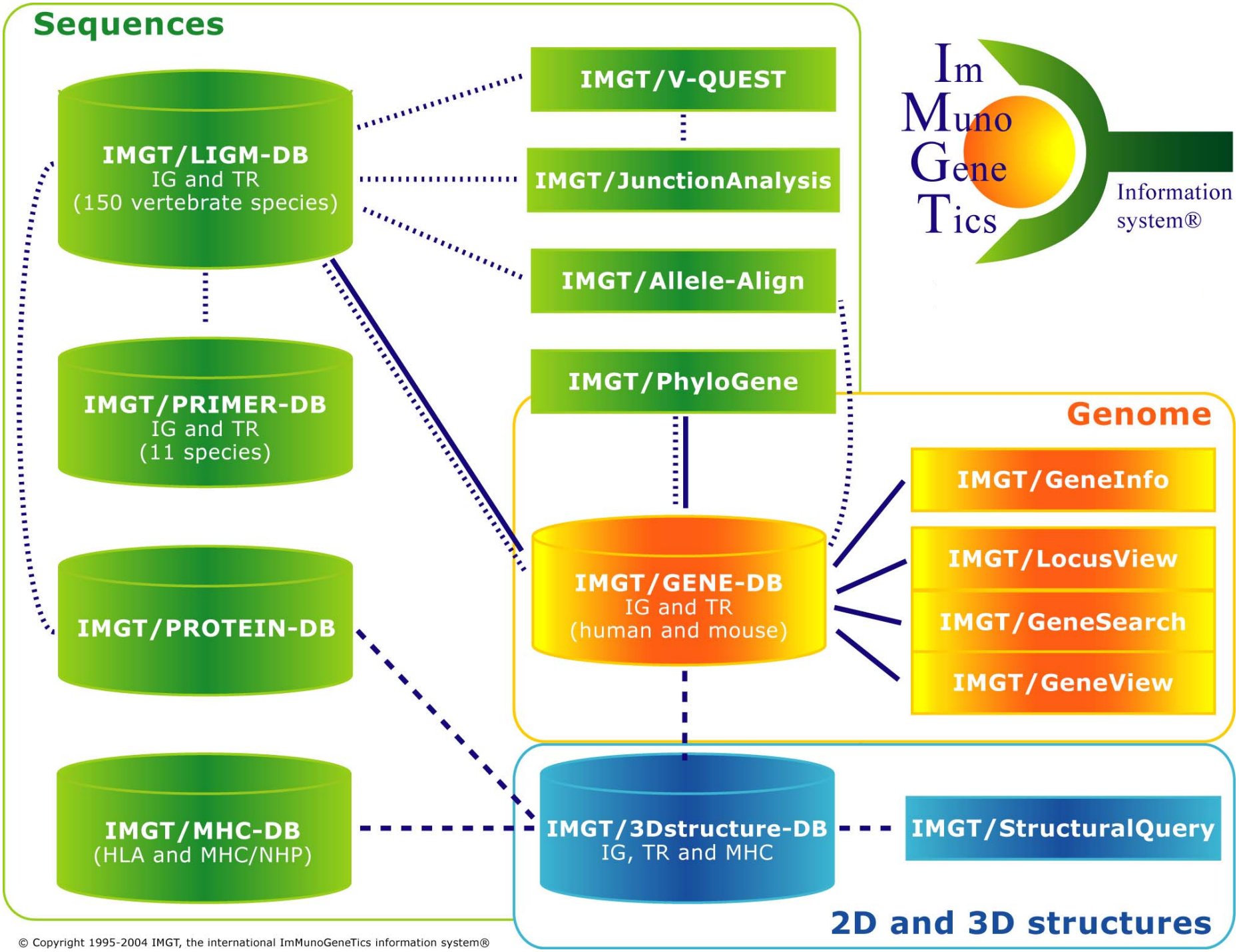
Created in 1989 at Montpellier, France (University Montpellier 2 and CNRS)

IMGT® is the international reference in immunogenetics and immunoinformatics.

IMGT® comprises:

- 6 databases
- 15 on-line tools
- more than 10,000 HTML pages of Web resources.

IMGT® receives 150.000 requests per month.

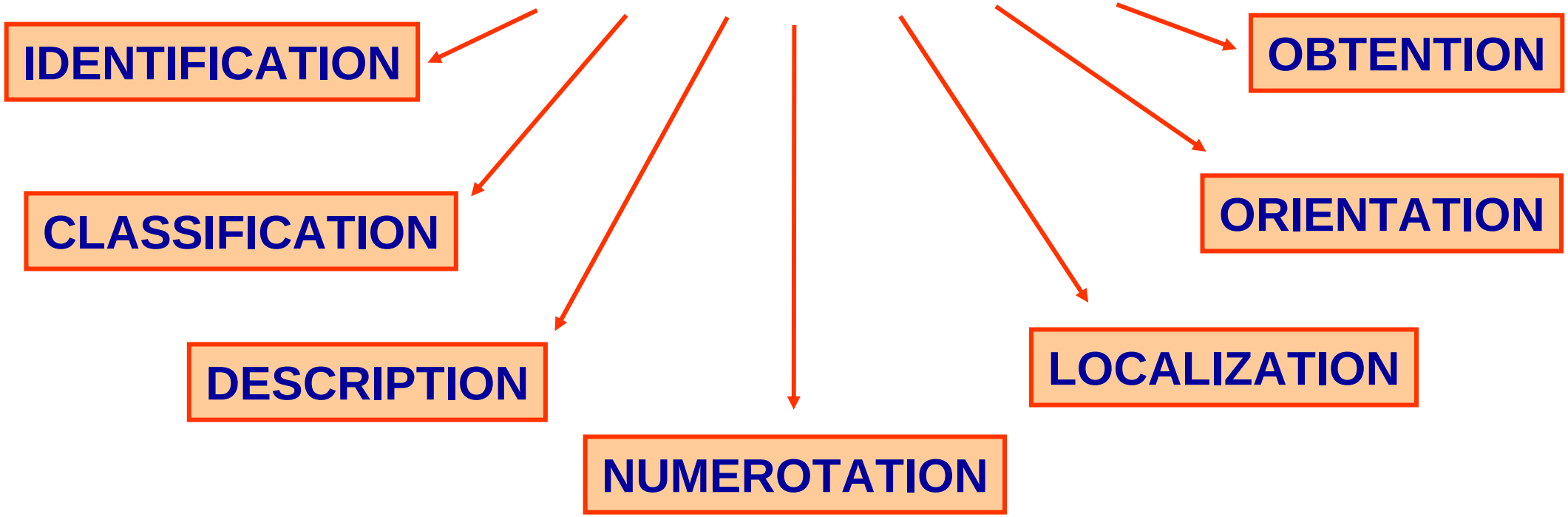


IMGT-ONTOLOGY axioms and concepts

http://www.imgt.org

IMGT-ONTOLOGY seven axioms:

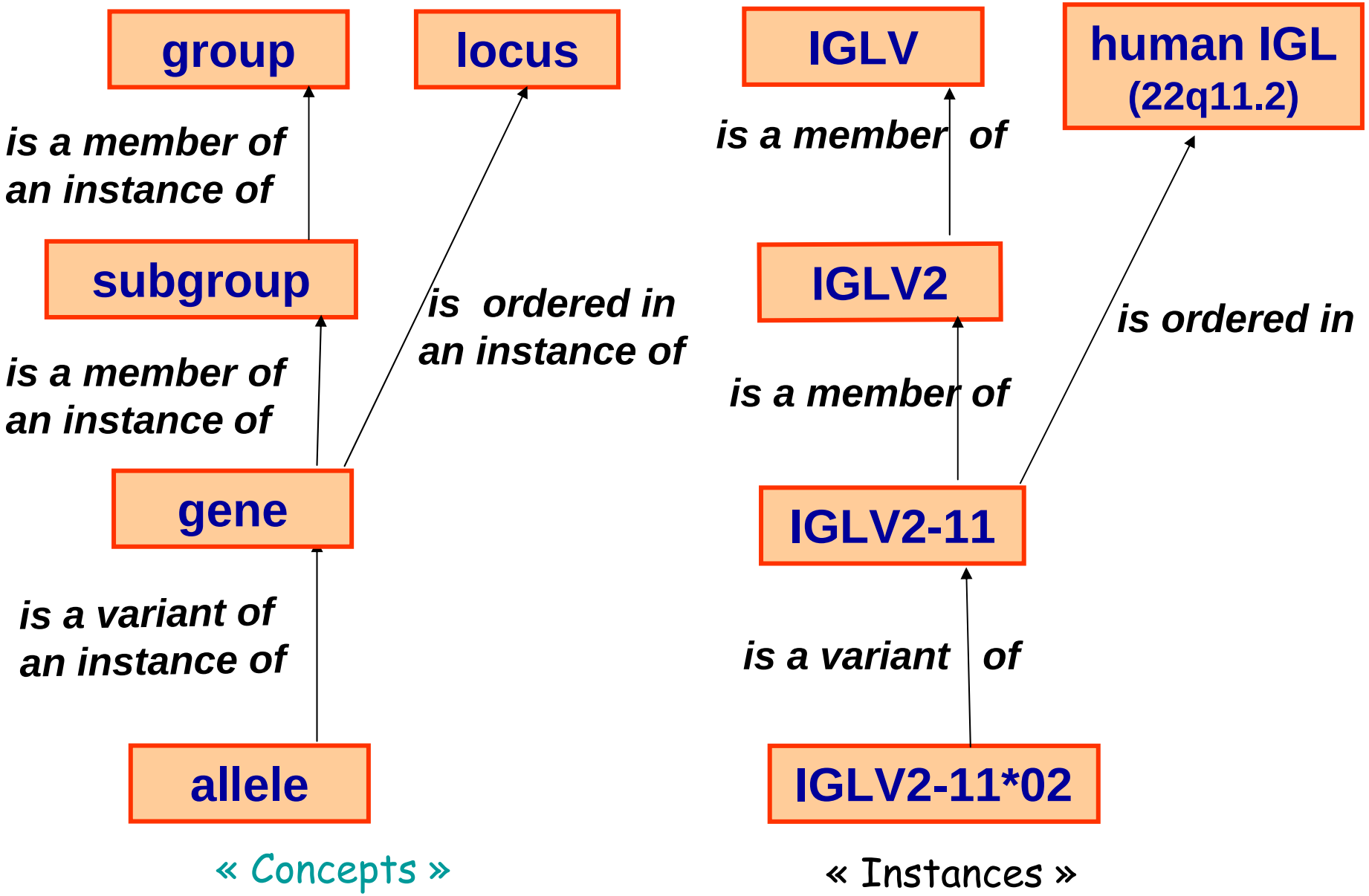
To share, reuse and represent knowledge in Immunogenetics and Life Sciences



Giudicelli and Lefranc, Bioinformatics 1999

CLASSIFICATION axiom

http://www.imgt.org



IMGT Marie-Paule page: IMGT repertoire - Netscape

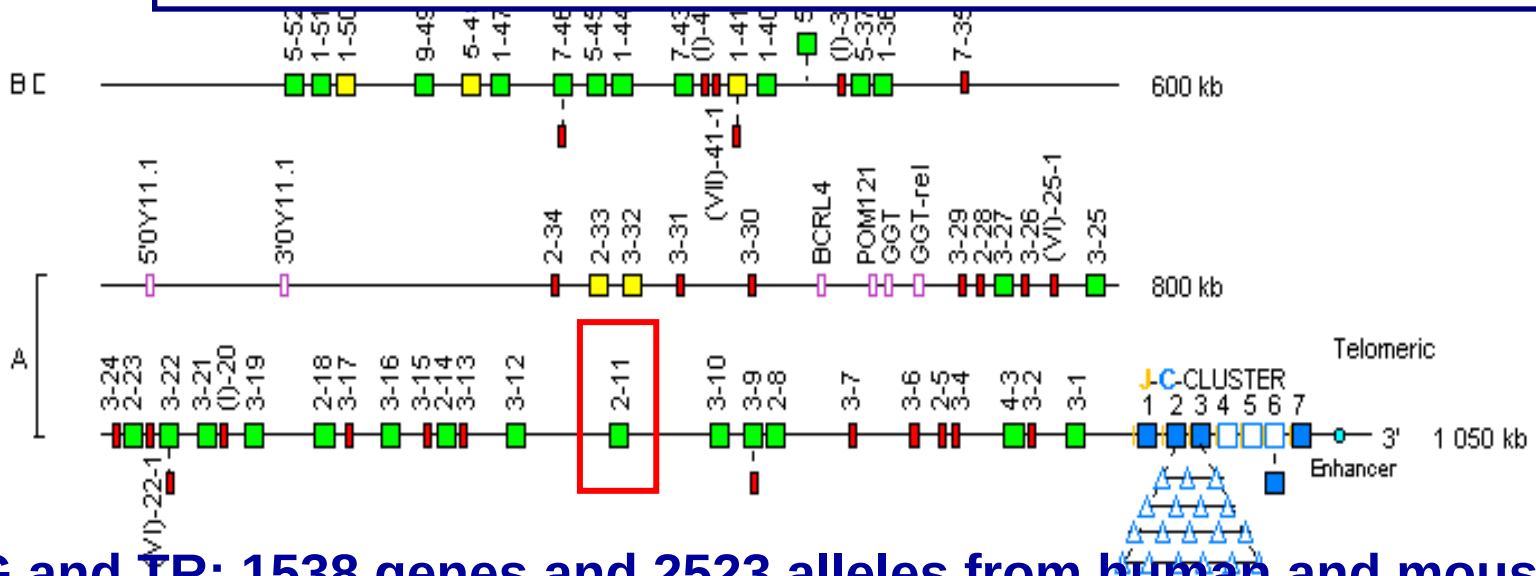
Fichier Edition Afficher Aller Communicator Aide

Locus representation: Human IGL

Human IGL 2

WELCOME !
to IMGT/GENE-DB

THE
INTERNATIONAL
IMMUNOGENETICS
INFORMATION SYSTEM®



IG and TR: 1538 genes and 2523 alleles from human and mouse

Document : chargé

CLASSIFICATION axiom

http://www.imgt.org

- The IMGT-ONTOLOGY main concepts of classification include ‘group’, ‘subgroup’, ‘gene’, ‘allele’.

- They allowed to set up the nomenclature for IG and TR genes (V, D, J, C genes).

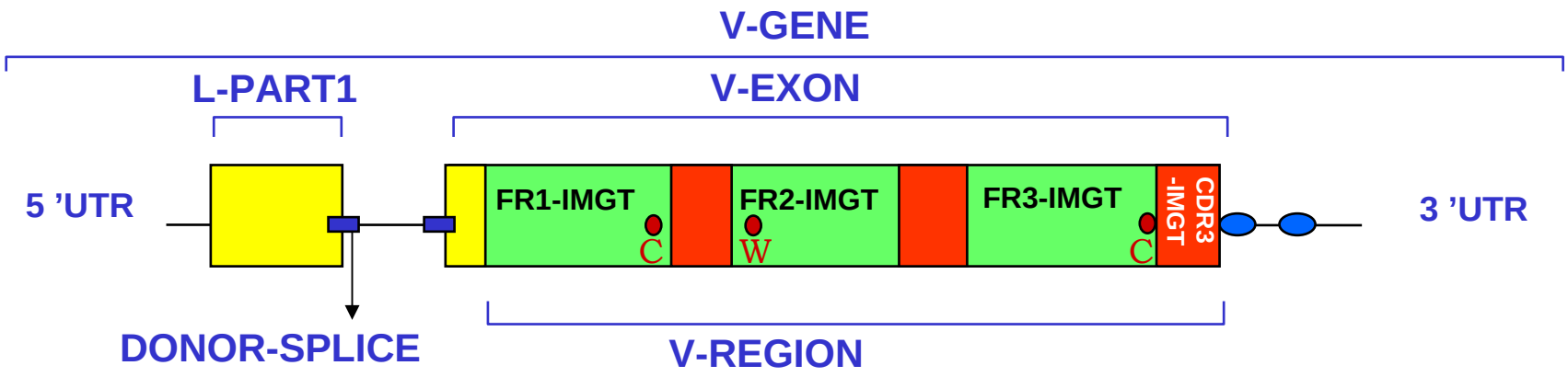
- IMGT gene names were approved by HGNC in 1999 and entered in GDB, LocusLink and Entrez Gene (NCBI).

- IMGT/GENE-DB is the international reference database for IG and TR genes (direct links from Entrez Gene NCBI).

- WHO-IUIS/IMGT 2007 report (*Dev. Comp. Immunol., Immunogenetics*).

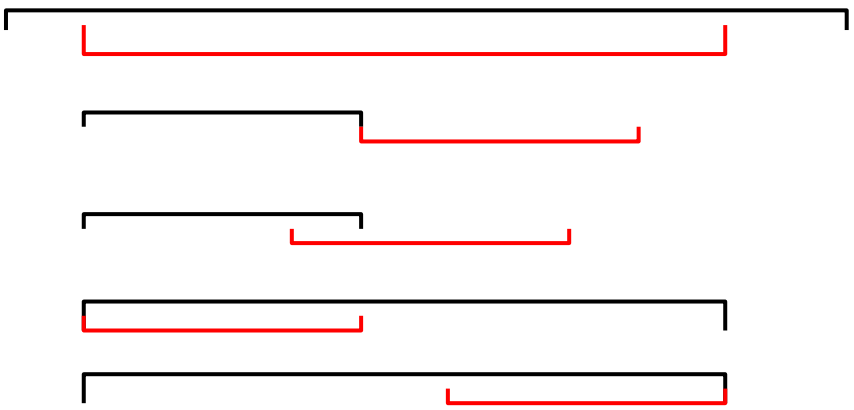
DESCRIPTION axiom

PROTOTYPE for a V-GENE



Label 1	Label 2
V-GENE	V-EXON
FR3-IMGT	CDR3-IMGT
L-PART1	DONOR-SPLICE
V-REGION	FR1-IMGT
V-REGION	CDR3-IMGT

Relations entre Labels

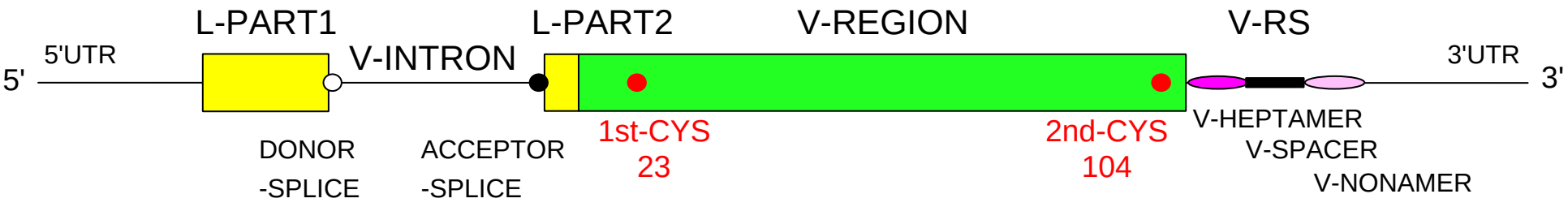


An example of V-GENE

http://imgt.cines.fr

>X62106.0|HSVI2|*Homo sapiens* VI-2 gene for immunoglobulin heavy chain

```
tgagagctcc gttcctcacc atggactgga cctggaggat cctcttcttg gtggcagcag      60
ccacaggtaa gaggctccct agtcccagtg atgagaaaga gattgagtcc agtccaggga     120
gatctcatcc acttctgtgt tctctccaca ggagcccact cccaggtgca gctggtgcag     180
tctggggctg aggtgaagaa gcctggggcc tcagtgaagg tctcctgcaa ggcttctgga     240
tacaccttca ccggctacta tatgcactgg gtgcgacagg cccctggaca agggcttgag     300
tggatgggat ggatcaaccc taacagtggt ggcacaaact atgcacagaa gtttcagggc     360
agggtcacca tgaccaggga cacgtccatc agcacagcct acatggagct gagcaggctg     420
agatctgacg acacggccgt gtattactgt gcgagagaca cagtgtgaaa acccacatcc     480
tgagggtgtc agaaacccaa gggaggaggc ag
```



IMGT/LIGM-DB

http://www.imgt.org

IMGT/LIGM-DB Consultation module v3 - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

```
FH   Key                 Location/Qualifiers
FH
FT   L-V-D-J-C-SEQUENCE  <1..375>
FT                       /partial
FT                       /db_xref="taxon:9606"
FT                       /cell_type="B-cell hybridoma 2F7"
FT                       /IMGT_note="automatically annotated with IMGT tools"
FT                       /organism="Homo sapiens"
FT   V-D-J-REGION        1..375
FT                       RLSRAASGFTFSSYGMHWVRQAP
FT                       NSKNTLYLQMNSLRAEDTAVYYC
FT
FT   V-REGION            1..296
FT                       /allele="IGHV3-33*01, putative"
FT                       /gene="IGHV3-33"
FT                       /CDR_length="[8.8.18]"
FT                       /putative_limit="3' side"
FT                       /translation="QVHLVESGGAVFHPGRSLRLSRAASGFTFSSYGMHWVRQAP
FT                       AKGLEWVAVIWYDGSNKYYADSVKGRFTISRDNSKNTLYLQMNSLRAEDTAVYYC
FT                       AK"
FT   FR1-IMGT            1..75
FT                       /AA_IMGT="1 to 26, AA 10 is missing"
FT                       /translation="QVHLVESGGAVFHPGRSLRLSRAAS"
FT   CDR1-IMGT           76..99
FT                       /AA_IMGT="27 to 34"
FT                       /translation="GFTFSSYG"
FT   FR2-IMGT            100..150
FT                       /AA_IMGT="39 to 55"
FT                       /translation="MHWVRQAPAKGLEWVAV"
FT   CONSERVED-TRP       106..108
FT   CDR2-IMGT           151..174
FT                       /AA_IMGT="56 to 63"
FT                       /translation="IWYDGSNK"
FT   FR3-IMGT            175..288
FT                       /AA_IMGT="66 to 104, AA 73 is missing"
FT                       /translation="YYADSVKGRFTISRDNSKNTLYLQMNSLRAEDTAVYYC"
```

Done

DESCRIPTION

CLASSIFICATION

136.575 sequences from 235 species

IMGT-ONTOLOGY:
277 IMGT labels for sequences
285 IMGT labels for 3D structures

SO (Sequence ontology):
67 IMGT labels

DESCRIPTION axiom

http://www.imgt.org

- The IMGT-ONTOLOGY concepts of description comprise the standardized IMGT labels and relations.

- They have allowed to describe the IG, TR and MHC sequences and 3D structures, whatever the receptor type, the chain type, or the species.

- They are particularly useful to describe IG, TR, and MHC and their complexes (IG/antigen, TR/pMHC).

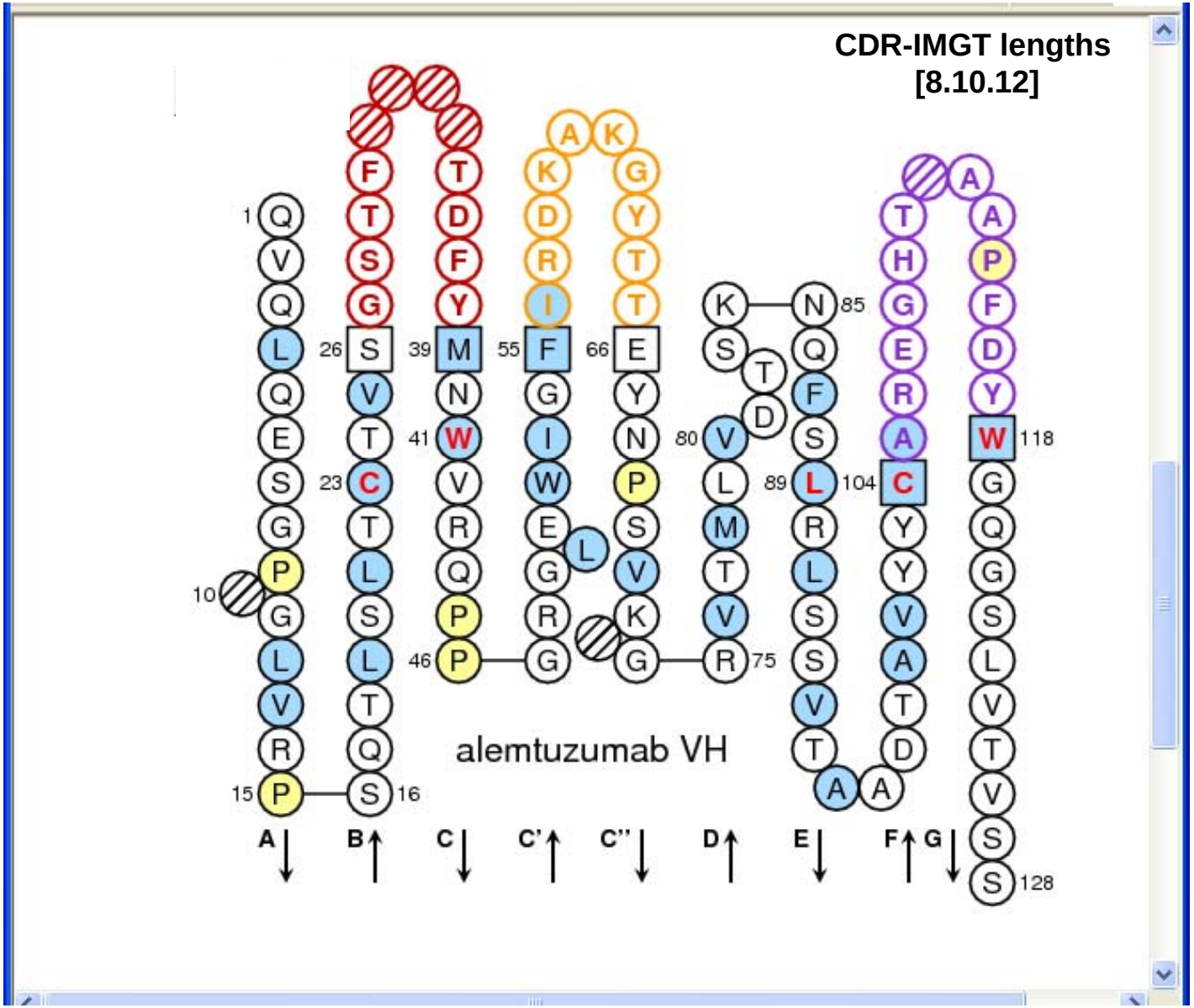
- It is possible to query the IMGT® databases (IMGT/LIGM-DB for sequences, IMGT/3Dstructure-DB for 3D structures) with IMGT labels.

- Sequence Ontology (SO) includes IMGT labels.

NUMEROTATION axiom

http://www.imgt.org

IMGT Collier de Perles

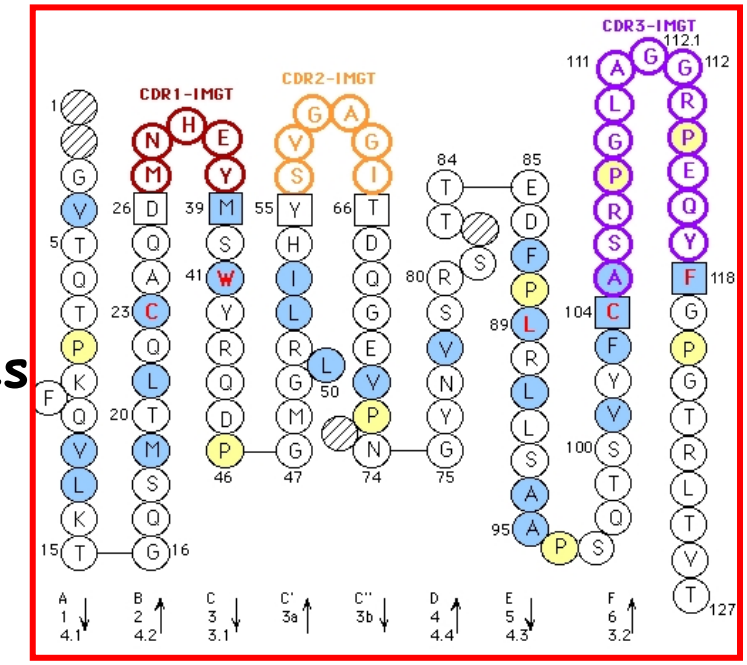


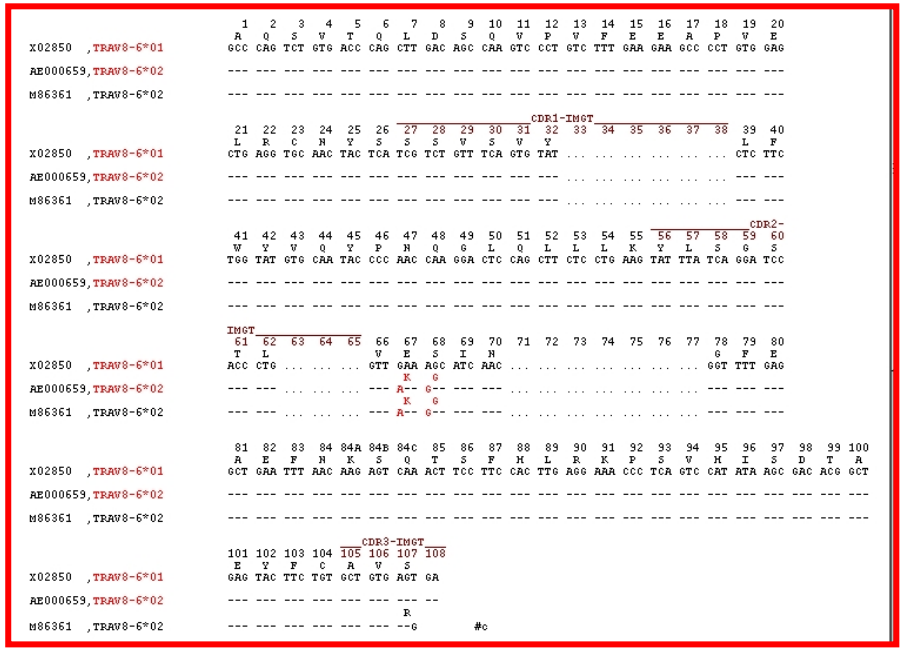
Lefranc et al. Dev. Comp. Immunol. ***27****, 55-77 (2003)*

IMGT Web resources: 10 000 pages HTML

http://www.imgt.org

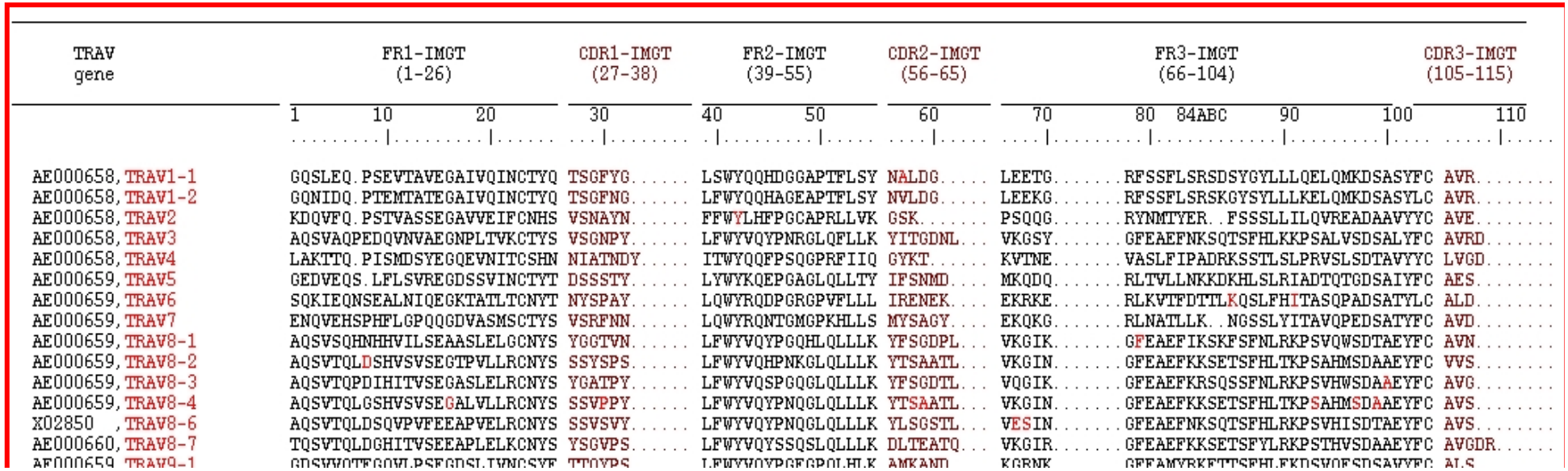
IMGT Collier de Perles



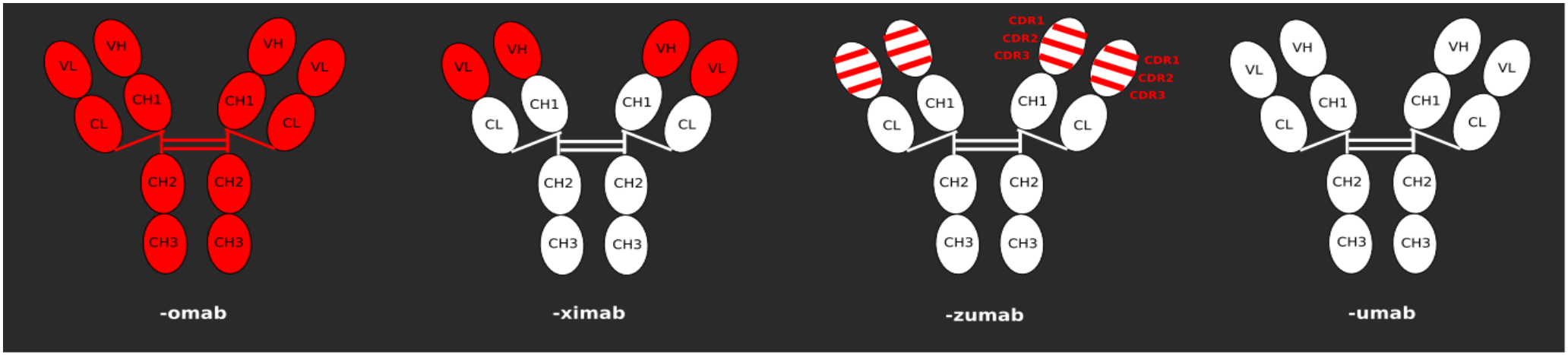


IMGT Alignment of alleles

IMGT Protein Display



TRAV gene	FR1-IMGT (1-26)	CDR1-IMGT (27-38)	FR2-IMGT (39-55)	CDR2-IMGT (56-65)	FR3-IMGT (66-104)	CDR3-IMGT (105-115)
AE000658,TRAV1-1	GQSLEQ.PSEVTAVEGAIVQINCTYQ	TSGFYG.....	LSWYQQHDGGAPTFLSY	NALDG.....	LEETG.....RFSSFLSRSDSYGYLLLQELQMKDSASYFC	AVR.....
AE000658,TRAV1-2	GQNIDQ.PTEMTATEGAIVQINCTYQ	TSGFNG.....	LFWYQQHAGEAPTFLSY	NVLDG.....	LEEKG.....RFSSFLSRSKGYSYLLLKELQMKDSASYLC	AVR.....
AE000658,TRAV2	KDQVFQ.PSTVASSEGAVVEIFCNHS	VSNAYN.....	FFWYLHFPGCAPRLLVK	GSK.....	PSQQG.....RYNMTYER..FSSSLLILQVREADAAVYYC	AVE.....
AE000658,TRAV3	AQSVAQPEDQVNVAEGNPLTVKCTYS	VSGNPY.....	LFWYVQYPNRGLQFLLK	YITGDNL...	VKGSY.....GFEAEFNKSQTSFHLKKPSALVSDSALYFC	AVRD.....
AE000658,TRAV4	LAKTTQ.PISMDSYEGQEVNITCSHN	NIATNDY.....	ITWYQQFPSQGPRFIIQ	GYKT.....	KVTNE.....VASLFIPADRKSSTLSLPRVSLSDTAVYYC	LVGD.....
AE000659,TRAV5	GEDVEQS.LFLSVREGDSSVINCTYT	DSSSTY.....	LYWYKQEPGAGLQLLTY	IFSNMD....	MKQDQ.....RLTVLLNKKDKHLSLRIADTQTGDSAIYFC	AES.....
AE000659,TRAV6	SQKIEQNSEALNIQEGKTATLTCNYT	NYSPAY.....	LQWYRQDPGRGPVFLLL	IRENEK....	EKRKE.....RLKVTFDTTLKQSLFHITASQPADSATYLC	ALD.....
AE000659,TRAV7	ENQVEHSPHFLGPQQGDVASMSCTYS	VSRFNN.....	LQWYRQNTGMGPKHLLS	MYSAGY....	EKQKG.....RLNATLLK..NGSSLYITAVQPEDSATYFC	AVD.....
AE000659,TRAV8-1	AQSVSQHNHHVILSEAASLELGCNYS	YGGTVN.....	LFWYVQYPGQHLQLLLK	YFSGDPL...	VKGIK.....GFEAEFIKSKFSFNLRKPSVQWSDTAEYFC	AVN.....
AE000659,TRAV8-2	AQSVTQLDSHVSVSEGTPVLLRCNYS	SSYSPS.....	LFWYVQHPNKGLQLLLK	YTSAATL...	VKGIN.....GFEAEFKKSETSFHLTKPSAHMSDAAEYFC	VVS.....
AE000659,TRAV8-3	AQSVTQPDIHITVSEGASLELRCNYS	YGATPY.....	LFWYVQSPGQGLQLLLK	YFSGDTL...	VQGIK.....GFEAEFKRSQSSFNLRKPSVHWSDAAEYFC	AVG.....
AE000659,TRAV8-4	AQSVTQLGSHVSVSEGALVLLRCNYS	SSVPPY.....	LFWYVQYPNQGLQLLLK	YTSAATL...	VKGIN.....GFEAEFKKSETSFHLTKPSAHMSDAAEYFC	AVS.....
X02850 ,TRAV8-6	AQSVTQLDSQVPVFEEAPVELRCNYS	SSVSVY.....	LFWYVQYPNQGLQLLLK	YLSGSTL...	VESIN.....GFEAEFNKSQTSFHLRKPSVHISDTAEYFC	AVS.....
AE000660,TRAV8-7	TQSVTQLDGHITVSEEAPLELKCNYS	YSGVPS.....	LFWYVQYSSQSLQLLLK	DLTEATQ...	VKGIR.....GFEAEFKKSETSFYLRKPSTHVSDAAEYFC	AVGDR.....
AE000659,TRAV9-1	GDSVWQTFGQWLPSEGDSLIWNCSYE	TTQYPS.....	LFWYVQYPGEGPQLHLK	AMKAND....	KGRNK.....GFEAMYRKETTSFHLEKDSVQFSDSAVYFC	ALS.....



Immunogenicity

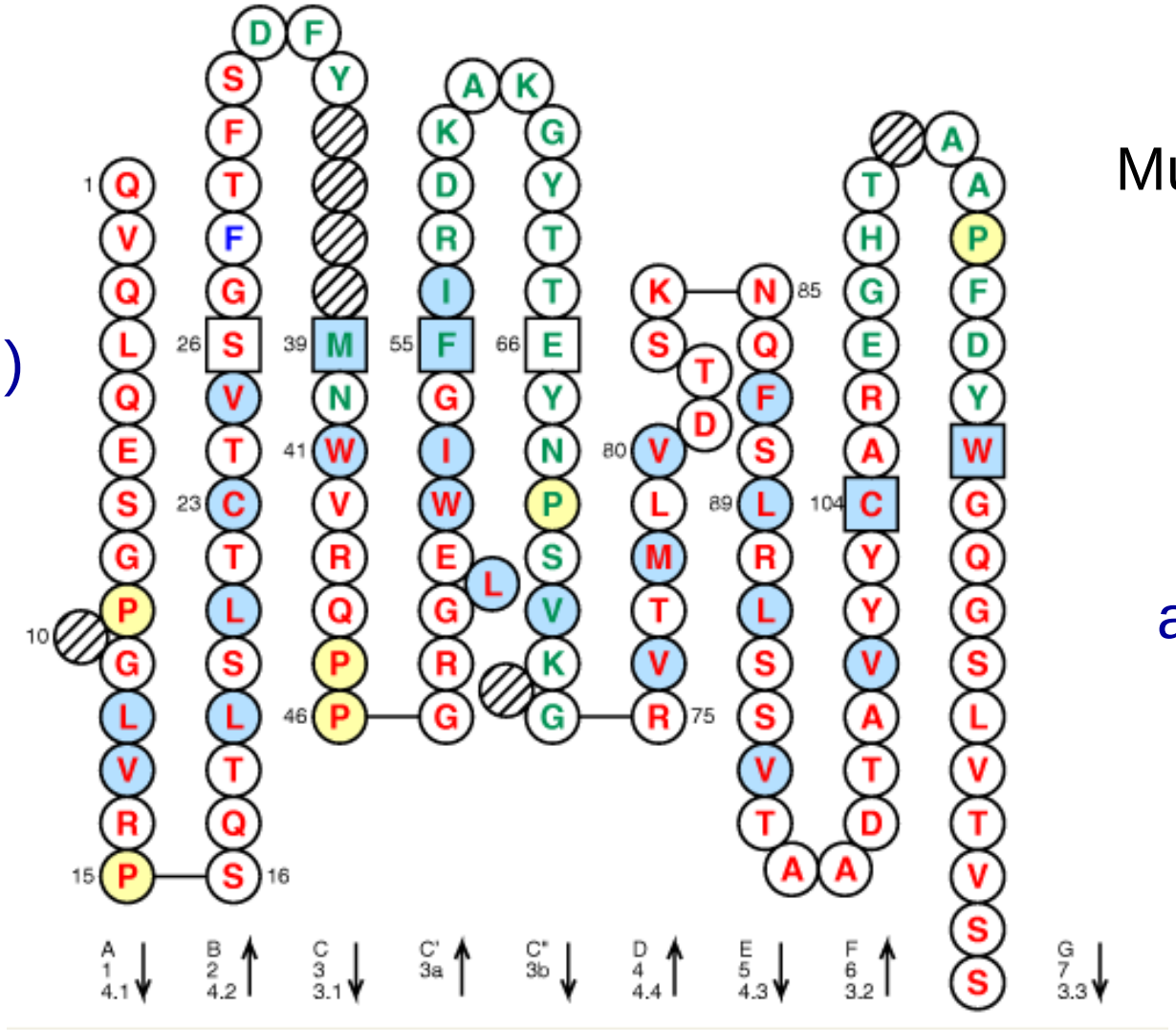
-omab	-ximab	-zumab	-umab
muromonab (1986)	abciximab (1994)	daclizumab (1997)	adalimumab (2002)
edrecolomab (1995)	rituximab (1997)	palivizumab (1998)	panitumumab (2006)
ibritumomab tiuxetan (2002)	basiliximab (1998)	trastuzumab (1998)	
tositumomab (2003)	infliximab (1998)	gemtuzumab ozogamicin (2000)	
	cetuximab (2004)	alemtuzumab (2001)	
		efalizumab (2003)	
		omalizumab (2003)	
		bevacizumab (2004)	
		natalizumab (2004)	
		nimotuzumab (2004)	
		ranibizumab (2006)	
		eculizumab (2007)	
		certolizumab pegol (2008)	

Humanized CAMPATH-1H mutant 1

http://www.imgt.org

VH domain
(V-D-J-REGION)

[8.10.12]



Mutant 1: S28>**F**

Mutant 2:
alemtuzumab
S31>**T**

human

rat

NUMEROTATION axiom

http://www.imgt.org

- The IMGT-ONTOLOGY concepts of numerotation include IMGT unique numbering and IMGT Collier de Perles for V-DOMAIN (IG and TR).

- They have been extended to the C-DOMAIN (IG and TR) and G-DOMAIN (MHC).

- They have allowed to bridge the gap between sequences and 3D structures in IMGT/3Dstructure-DB.

- They are used for mutations, polymorphisms, CDR-IMGT lengths, contact analysis, potential immunogenicity evaluation and paratope definition.

- WHO-INN programme requires the CDR-IMGT lengths for antibody.

Examples of IMGT® tools based on the IMGT-ONTOLOGY concepts

http://www.imgt.org

IMGT/JunctionAnalysis
IMGT/V-QUEST
IMGT/3Dstructure-DB

Immunoglobulin V-D-J generation of sequence diversity

http://www.imgt.org

3'V-REGION | N-REGION | D-REGION | N-REGION | 5'J-REGION

tgtgcgaaa ga tacc agcatattgtg gtggtgactgctat tcc gatt acaactggttcg actcc**tgg**

JUNCTION

C A P Y R G D T Y D Y S W

tgt gcg cca tac cgg ggt gac act tat gat tac tcc tgg

IMGT/JunctionAnalysis: analysis of the IG and TR junctions

http://www.imgt.org

IMGT/JunctionAnalysis Results

Locus IGH
Species Homo sapiens
IMGTrepertoire link Locus representation

Maximum number of mutations :
V-REGION : 2; D-REGION : 4; J-REGION : 2
Deletion criterium : Using patterns
Best D gene choice for a same score : Less mutations

Description of the JUNCTIONs

Click on mutated (underscored) nucleotid to see the original one: c

	Input	V name	V-REGION	P	N1	D-REGION	N2	P	J-REGION	J name	D name	Vmut	Dmut	Jmut	Ngc
#1	Z70256	IGHV2-26*01	tgtgtacg.....		tgttgt	.....gcagcgcctggtac	ccaaatatc		...actttgaccactgg	IGHJ4*02	IGHD6-13*01	1	2	1	5/15
#2	Z70257	IGHV3-7*02	tgtgcgag.		ggatgg	.....cagctcttatgcc	cgccc		ctactggtacttcgatctctgg	IGHJ2*01	IGHD2-2*01	0	2	0	9/11
#3	Z70606	IGHV4-31*03	tgtgcgagag.		c	.gactacg.....	cact		..atgcttttgatgtctgg	IGHJ3*01	IGHD4-17*01	0	0	0	3/5
#4	Z70608	IGHV4-39*05	tgtgc.		cagagta	.....acgatttttggagtggttatt.....	ccccgggggga		..atgcttttgatatctgg	IGHJ3*02	IGHD3-3*01	0	0	0	12/17
#5	Z70610	IGHV4-34*09	tgtgcgagag.		tcgggag	.....cgatttttggagtggttatt.....	cccga	ca	tgatgcttttgatatctgg	IGHJ3*02	IGHD3-3*01	0	0	0	9/12
#6	Z70611	IGHV4-59*01	tgtgcgaga..		ca	.....tggtaactataa.	tgccggcgttg		...actggttcgacccctgg	IGHJ5*02	IGHD3-9*01	0	2	0	9/13
#7	Z70613	IGHV4-59*01	tgtgcgagag.			.....cagcagctggtac	ctccct		....ctttgactactgg	IGHJ4*02	IGHD6-13*01	0	0	0	4/6
#8	Z70614	IGHV4-59*01	tgtgcgaga..		cactataa	.....ttcggggacttat.....	cccctc		.....gactactgg	IGHJ4*02	IGHD3-16*01	0	2	0	7/14
#9	Z70615	IGHV4-59*01	tgtgcgagag.		ggctg	gtaaagaggg.....	tttcggaa		.tactggtacttcgatctctgg	IGHJ2*01	IGHD5-24*01	0	2	0	7/13
#10	Z70616	IGHV4-34*01	tgtgcgagag.		cgg	.....gtttggg.....	ttccc		...actggttcgacccctgg	IGHJ5*02	IGHD3-16*01	0	0	0	6/8
#11	Z70620	IGHV4-30-4*01	tgtgcgagaga		cc	.....ggggcgggatggtt.....	cgg		.gatgcttttgatatctgg	IGHJ3*02	IGHD3-16*01	1	4	0	5/5
#12	Z70621	IGHV4-39*01	tgtgcgagaca		ccacgatttatgg	.....ttcggggagtt.....	tgaccccc		.....ttgactactgg	IGHJ4*02	IGHD3-16*01	0	1	0	12/21
#13	Z70622	IGHV4-39*06	tgtgcgagaga	t	tgccccgctcctgccaaaat	gtattactatggttcgggga.....	tatgtacg		.....tttgactactgg	IGHJ4*03	IGHD3-10*01	0	0	0	15/28

Terminé

The eleven IMGT amino acid classes according to the physicochemical properties

http://www.imgt.org

'Volume' classes		'Hydropathy' classes									
	in Å^3	Hydrophobic			Neutral			Hydrophilic			
Very large	189-228	F		W		Y					
Large	162-174	I	L	M				K	R		
Medium	138-154	V					H			E	Q
Small	108-117			C	P	T				D	N
Very small	60-90	A			G	S					
		Aliphatic		Sulfur		Hydroxyl	Basic			Acidic	Amide
						Uncharged	Charged				Uncharged
		Nonpolar				Polar					

Pommié et al. J. Mol Recognit. ***17****, 17-32, 2004*

IMGT/JunctionAnalysis: analysis of the IG and TR junctions

http://www.imgt.org

JUNCTION alignments with translation and IMGT AA classes

Click on mutated (underlined) amino acid to see the original one:

		104	105	106	107	108	109	110	111	111.1	111.2	111.3	112.3	112.2	112.1	112	113	114	115	116	117	118
		C	S	P	G	G	S	A	Y						Y	H	E	H	F	Q	Q	W
#1	AY393054	tgt	agt	ccc	ggg	ggt	agt	gct	tat	...	...	...	...	...	tac	cac	gaa	cac	ttc	cag	cag	tgg
		C	V	K	P	T	D	D	D	G				H	R	A	E	Y	F	Q	Y	W
#2	AY393055	tgt	gtg	aaa	ccc	acg	gat	gat	gat	ggc	...	...	...	cac	cgg	gct	gaa	tac	ttc	cag	tac	tgg
		C	S	P	G	G	S	A	Y						Y	H	E	D	F	Q	Q	W
#3	AY393058	tgt	agt	ccc	ggg	ggt	agc	gct	tat	...	...	...	...	...	tac	cac	gaa	gac	ttc	cag	cag	tgg
		C	S	P	G	G	S	A	Y						Y	H	E	H	F	Q	Q	W
#4	AY393072	tgt	agt	ccc	ggg	ggt	agt	gct	tat	...	...	...	...	...	tac	cac	gaa	cac	ttc	cag	cag	tgg
		C	A	R	Q	N	P	P	E	Y	S	G	A	Y	H	D	G	W	F	D	P	W
#5	AY393088	tgt	gcg	aga	caa	aac	ccc	ccc	gag	tat	agt	ggc	gca	tat	cat	gat	ggg	tgg	ttc	gac	ccc	tgg
		C	A	R	E	M	L	Y	G	S	G	G	Y	Y	P	P	D	A	F	E	L	W
#6	AY393089	tgt	gcg	aga	gag	atg	ctc	tat	ggt	tcg	ggg	ggt	tat	tac	ccc	cct	gat	gca	ttt	gag	ctc	tgg
		C	A	R	Q	N	P	P	E	Y	S	G	A	Y	H	D	G	W	F	D	P	W
#7	AY393091	tgt	gcg	aga	cag	aat	ccc	ccc	gag	tat	agt	ggc	gca	tat	cat	gat	ggg	tgg	ttc	gac	ccc	tgg
		C	A	R	E	M	L	Y	G	S	G	G	Y	Y	P	P	D	A	F	E	V	W
#8	AY393092	tgt	gcg	aga	gag	atg	ctc	tat	ggt	tcg	ggg	ggt	tat	tac	ccc	cct	gat	gca	ttt	gag	gtc	tgg
		C	A	R	Q	N	P	P	E	Y	S	G	A	Y	H	D	G	W	F	D	P	W
#9	AY393094	tgt	gcg	aga	cag	aac	ccc	ccc	gag	tat	agt	ggc	gca	tat	cat	gat	ggg	tgg	ttc	gac	ccc	tgg

Yousfi Monod et al. Bioinformatics 20, i379-i385 (2004)

Pommié et al. J. Mol Recognit. 17, 17-32 (2004)

IMGT/V-QUEST: analysis of IG and TR sequences

WELCOME !

to the IMGT/V-QUEST Search page

THE
INTERNATIONAL
IMMUNOGENETICS
INFORMATION SYSTEM®

Im Muno Gene Tics Information system®

http://imgt.cines.fr

Citing IMGT/V-QUEST: Giudicelli, V. et al. Nucl. Acids Res. 2004, 32, W435-440 PMID: 15215425 PDF

☞ You are in the new IMGT/V-QUEST, *upgraded for multiple sequences and with new functionalities.* NEW!

Analyse your Immunoglobulin nucleotide sequences

- Human
- Mouse
- Chondrichthyes
- Teleostei
  - Atlantic cod
  - Channel catfish
  - Rainbow trout
- Sheep

Analyse your T cell Receptor nucleotide sequences

- Human
- Mouse
- Non-human primates

http://www.imgt.org

Analyse your Immunoglobulin sequences

Your selection: Human

Your sequences are compared to the **Human IG set** from the IMGT/V-QUEST reference directory sets

Nucleotide sequences

Enter your sequence(s) in FASTA format (FASTA format is required):

Type (or copy/paste) your sequence(s) into the box below :

```
>AY393054
gctgggttttccttgttgctattttaaaaggtgtccaatgtgaggtgcagctggtggagtctgggggaggcttggtacagccagggcgg
tccctgagactctcctgtgcagcttctggattgacctttggtgattactttatgagctggttccgccaggctccagggaagggactgga
gtgggtaggtttcattaagagcgaaacttatggtgggacaacagaatacgccgcgtctgtgaaaggcagattcatcatctcgagagatg
attccaaaagcatcgcctatttgcaaatgaacagcctggaaaccgaggacacagccatatattactgtagtcccgggggtagtgcttat
taccacgaacacttccagcagtggggcccgggcaccatggtcaccgtctcctcagcctccaccaagggcccatcggtcttccccctggc
accctcctccaagagcacctctgggggcacagcggccctgggctgcctggtcaaggactacttcccc
>AY393055
gctgggttttccttgttgctatttcaaaaggtgtccagtgtgaggtgcagctggtggagactggaggaggcttgatccagcctgggggg
tccctgagactctcctgtgcagcctctgggttcaccgtcagtagcaactacatgagctgggtccgccaggctccagggaaggggctgga
```

Or give the path access to a local file containing your sequence(s) in FASTA format (FASTA format is required):

Parcourir...

Start | Clear the form

Analysis by batches of up to 50 sequences in a single run

http://www.imgt.org

Internet

Giudicelli V. et al.

IMGT/3Dstructure-DB: analysis of the 3D structures

THANK YOU

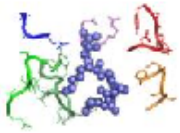
for using IMGT/3Dstructure-DB

THE
INTERNATIONAL
IMMUNOGENETICS
INFORMATION SYSTEM®

Im Muno Gene Tics Information system®

http://imgt.cines.fr

IMGT/3Dstructure-DB card for : 1ce1



Entry code [ ] Search

Query page

IMGT protein name	IMGT receptor type	IMGT receptor description	Ligand(s)	Species	CC	Chain ID
CAMPATH-1H, alemtuzumab, MABCAMPATH®	IG	FAB-GAMMA-1_KAPPA		Humanized	1	[1ce1_H 1ce1_L]
		Peptide	CD52 (synthetic peptide)	Synthetic	1	[1ce1_P]

Experimental technique X-ray diffraction Resolution (in angstrom) 1.90 PDB release date 25-JUN-99

Epitope and Chain details | **Contact analysis** | 3D visualization Jmol or QuickPDB | Renumbered IMGT file | IMGT numbering comparison | References and links | Printable card

Contact analysis:

IMGT/3Dstructure-DB Domain pair contacts (overview) **of 1ce1**

Atom contact types
☐ Non covalent ☐ Covalent

Atom contact categories
☐ (BB) Backbone/backbone

Kaas Q. et al.

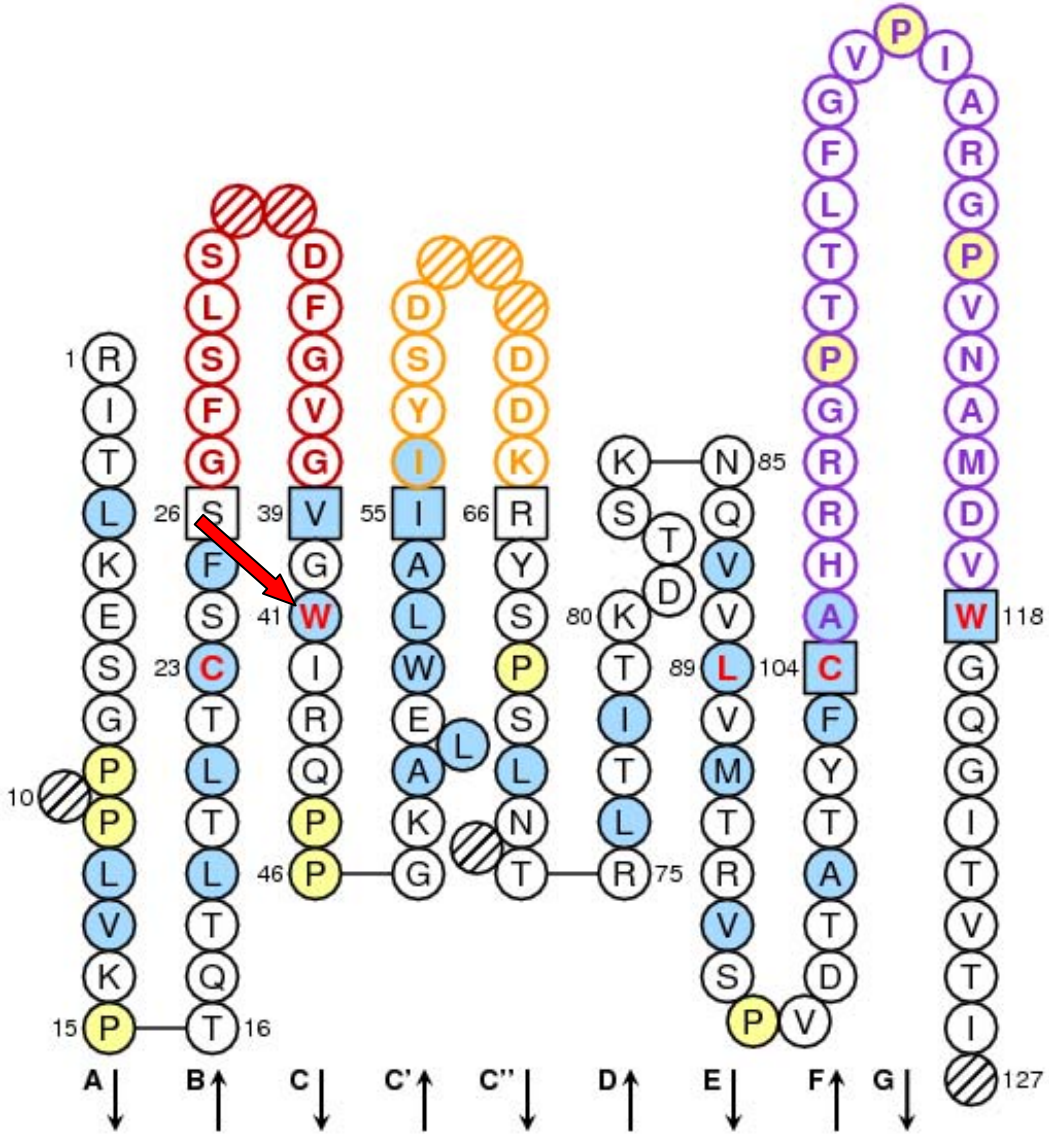
Access to atomic pair contacts in IMGT/3Dstructure-DB

http://www.imgt.org

IMGT Collier de Perles : *Homo sapiens* (Human) IGHV V-DOMAIN from 2F5 (1u8k_B)

CDR-IMGT lengths [10.7.24]

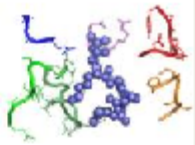
Click on residue
in IMGT Collier de Perles
(or in amino acid sequence)



Atomic pair contacts in IMGT/3Dstructure-DB

http://www.imgt.org

IMGT Residue@Position card



Residue@Position: 41 - TRP (W) - VH - 1u8k_B

General information:

PDB file numbering	36
IMGT file numbering	41
Residue full name	Tryptophan
Formula	C11 H12 N2 O2

IMGT LocalStructure@Position

Secondary structure	Extended conformation
Phi (in degrees)	-122.64
Psi (in degrees)	137.12
ASA (in square angstrom)	0.0

Pair contacts:

Atom contact types

- [x] Non covalent
- [x] Polar
- [x] Hydrogen bond
- [x] Non polar
- [x] Covalent
- [ ] Disulfide

Check all
Uncheck all

Atom contact categories

- [ ] (BB) Backbone/backbone
- [ ] (SS) Side chain/side chain
- [ ] (BS) Backbone/side chain
- [ ] (SB) Side chain/backbone

Check all
Uncheck all

Show

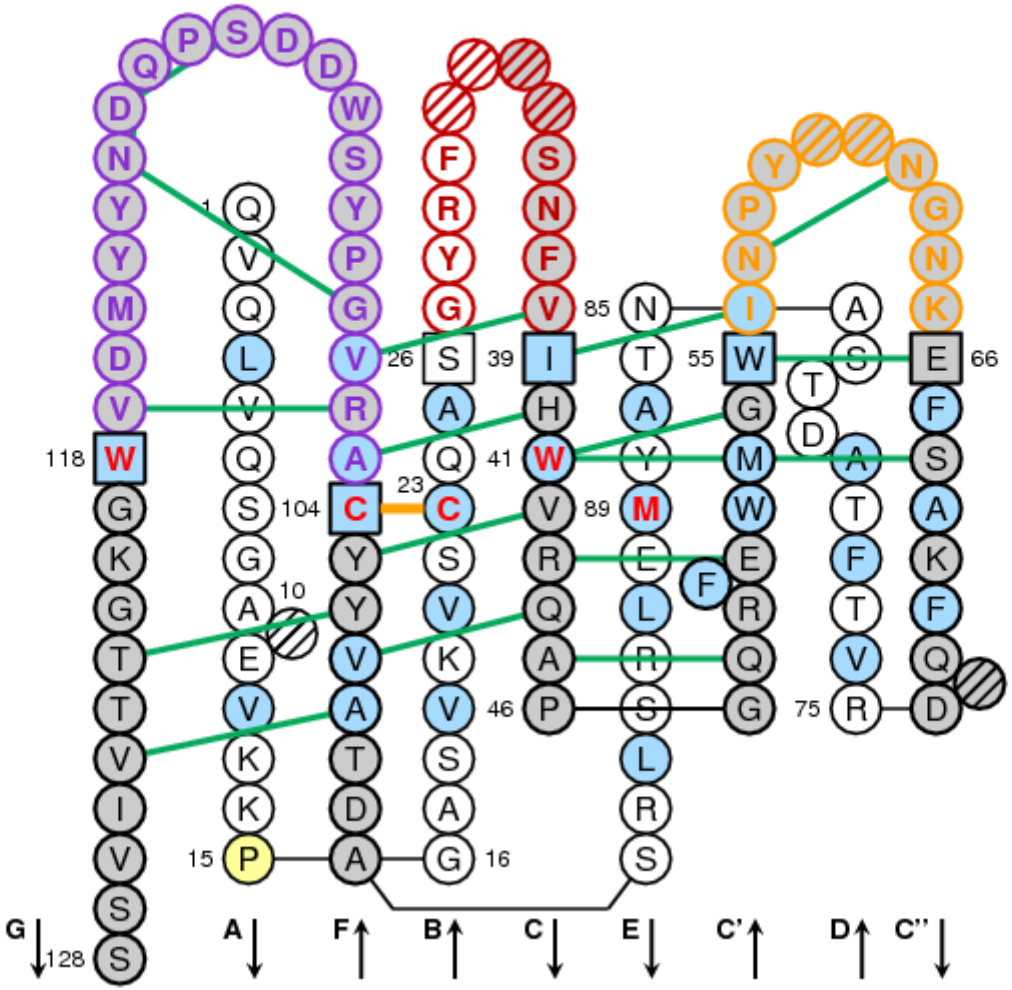
IMGT Num	Residue		Domain	Chain	Atom contacts	Non Covalent	Polar	Hydrogen Bond	Non Polar
6	GLU	E	VH	1u8k_B	6	6	0	0	6
21	LEU	L	VH	1u8k_B	17	17	0	0	17
22	THR	T	VH	1u8k_B	8	8	0	0	8
23	CYS	C	VH	1u8k_B	10	10	0	0	10
39	VAL	V	VH	1u8k_B	2	2	1	0	1
43	ARG	R	VH	1u8k_B	2	2	1	0	1

Hydrogen bonds (IMGT Collier de Perles on 2 layers)

http://www.imgt.org

IMGT Collier de Perles : *Homo sapiens* (Human) IGHV V-DOMAIN from b12 (1hzh_H)

CDR-IMGT lengths [8.8.20]



Contacts VH-(Ligand), V-KAPPA-(Ligand)

http://www.imgt.org

IMGT molecule name	IMGT description	Chain ID	IMGT chain description	Domain number	IMGT domain description
CAMPATH-1H, alemtuzumab, MABCAMPATH®	FAB-GAMMA-1_KAPPA	1ce1_H	VH-CH1	[D1]	VH
				[D2]	CH1
		1ce1_L	L-KAPPA	[D1]	V-KAPPA
				[D2]	C-KAPPA
CD52 (synthetic peptide)	Peptide	1ce1_P	Peptide		

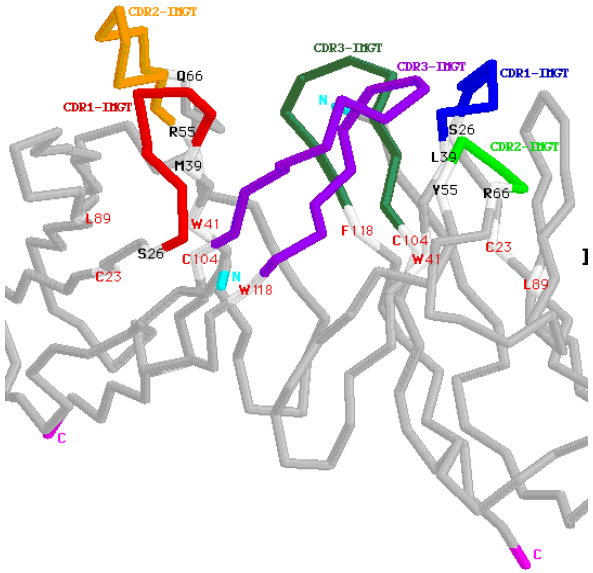
	Unit 1		Unit 2		Residue contacts	Number of residues			Atom contact types		
	Domain	Chain	Domain	Chain		Total	From 1	From 2	Total	Polar	Hydrogen
DomPair	VH	1ce1_H	CH1	1ce1_H	19	17	8	9	125	9	1
DomPair			V-KAPPA	1ce1_L	63	45	24	21	532	61	6
DomPair			(Ligand)	1ce1_P	25	19	12	7	216	40	9
DomPair	CH1	1ce1_H	VH	1ce1_H	19	17	9	8	125	9	1
DomPair			C-KAPPA	1ce1_L	68	58	28	30	498	40	6
DomPair	V-KAPPA	1ce1_L	VH	1ce1_H	63	45	21	24	532	61	6
DomPair			C-KAPPA	1ce1_L	18	18	8	10	137	19	2
DomPair			(Ligand)	1ce1_P	16	14	7	7	171	37	5
DomPair	C-KAPPA	1ce1_L	CH1	1ce1_H	68	58	30	28	498	40	6
DomPair			V-KAPPA	1ce1_L	18	18	10	8	137	19	2

Kaas Q. et al.

Contacts VH-(Ligand)

http://www.imgt.org

	IMGT Num	Residue		Domain	Chain		IMGT Num	Residue		Domain	Chain	Total	Polar	Hydrogen
R@P	38	TYR	Y	VH	1ce1_H	R@P	2	THR	T		1ce1_P	4	0	0
R@P	38	TYR	Y	VH	1ce1_H	R@P	7	ALA	A		1ce1_P	13	1	0
R@P	38	TYR	Y	VH	1ce1_H	R@P	8	ASP	D		1ce1_P	14	2	2
R@P	55	PHE	F	VH	1ce1_H	R@P	6	SER	S		1ce1_P	5	0	0
R@P	55	PHE	F	VH	1ce1_H	R@P	7	ALA	A		1ce1_P	16	0	0
R@P	55	PHE	F	VH	1ce1_H	R@P	8	ASP	D		1ce1_P	1	0	0
R@P	57	ARG	R	VH	1ce1_H	R@P	7	ALA	A		1ce1_P	9	3	2
R@P	57	ARG	R	VH	1ce1_H	R@P	8	ASP	D		1ce1_P	20	6	1
R@P	61	LYS	K	VH	1ce1_H	R@P	8	ASP	D		1ce1_P	11	2	1
R@P	66	GLU	E	VH	1ce1_H	R@P	7	ALA	A		1ce1_P	1	0	0
R@P	107	GLU	E	VH	1ce1_H	R@P	2	THR	T		1ce1_P	13	2	1
R@P	107	GLU	E	VH	1ce1_H	R@P	4	SER	S		1ce1_P	5	2	0
R@P	107	GLU	E	VH	1ce1_H	R@P	7	ALA	A		1ce1_P	5	0	0
R@P	108	GLY	G	VH	1ce1_H	R@P	1	GLY	G		1ce1_P	2	1	0
R@P	108	GLY	G	VH	1ce1_H	R@P	2	THR	T		1ce1_P	9	2	0
R@P	109	HIS	H	VH	1ce1_H	R@P	1	GLY	G		1ce1_P	24	4	0
R@P	109	HIS	H	VH	1ce1_H	R@P	2	THR	T		1ce1_P	21	5	0
R@P	109	HIS	H	VH	1ce1_H	R@P	3	SER	S		1ce1_P	9	2	1
R@P	110	THR	T	VH	1ce1_H	R@P	1	GLY	G		1ce1_P	1	1	0
R@P	110	THR	T	VH	1ce1_H	R@P	3	SER	S		1ce1_P	11	4	1
R@P	112	ALA	A	VH	1ce1_H	R@P	3	SER	S		1ce1_P	3	1	0
R@P	113	ALA	A	VH	1ce1_H	R@P	2	THR	T		1ce1_P	3	0	0
R@P	113	ALA	A	VH	1ce1_H	R@P	3	SER	S		1ce1_P	7	2	0
R@P	113	ALA	A	VH	1ce1_H	R@P	4	SER	S		1ce1_P	4	0	0
R@P	114	PRO	P	VH	1ce1_H	R@P	4	SER	S		1ce1_P	5	0	0



Kaas Q. et al

http://www.imgt.org

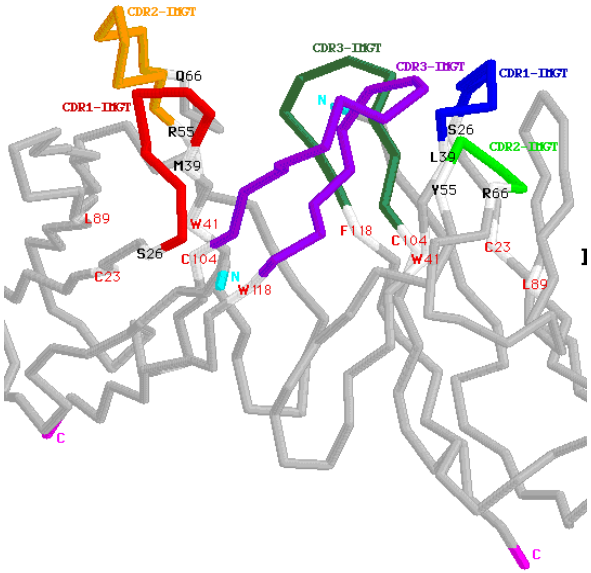
Summary:

Residue contacts	Number of residues			Atom contact types		
	Total	From 1	From 2	Total	Polar	Hydrogen
16	14	7	7	171	37	5

List of the Residue@Position pair contacts:

Click 'R@P' for IMGT Residue@Position cards

	Order						Order					Atom contacts		
	IMGT Num	Residue		Domain	Chain		IMGT Num	Residue		Domain	Chain	Total	Polar	Hydrogen
R@P	38	TYR	Y	V-KAPPA	1ce1_L	R@P	3	SER	S		1ce1_P	1	0	0
R@P	38	TYR	Y	V-KAPPA	1ce1_L	R@P	5	PRO	P		1ce1_P	21	0	0
R@P	56	ASN	N	V-KAPPA	1ce1_L	R@P	3	SER	S		1ce1_P	3	2	0
R@P	107	HIS	H	V-KAPPA	1ce1_L	R@P	4	SER	S		1ce1_P	20	4	1
R@P	107	HIS	H	V-KAPPA	1ce1_L	R@P	5	PRO	P		1ce1_P	12	2	0
R@P	107	HIS	H	V-KAPPA	1ce1_L	R@P	6	SER	S		1ce1_P	14	3	1
R@P	108	ILE	I	V-KAPPA	1ce1_L	R@P	5	PRO	P		1ce1_P	12	1	0
R@P	108	ILE	I	V-KAPPA	1ce1_L	R@P	6	SER	S		1ce1_P	12	3	0
R@P	109	SER	S	V-KAPPA	1ce1_L	R@P	6	SER	S		1ce1_P	11	2	0
R@P	114	ARG	R	V-KAPPA	1ce1_L	R@P	6	SER	S		1ce1_P	18	3	1
R@P	114	ARG	R	V-KAPPA	1ce1_L	R@P	7	ALA	A		1ce1_P	4	2	0
R@P	114	ARG	R	V-KAPPA	1ce1_L	R@P	8	ASP	D		1ce1_P	6	2	0
R@P	116	ARG	R	V-KAPPA	1ce1_L	R@P	2	THR	T		1ce1_P	1	1	0
R@P	116	ARG	R	V-KAPPA	1ce1_L	R@P	4	SER	S		1ce1_P	9	4	1
R@P	116	ARG	R	V-KAPPA	1ce1_L	R@P	6	SER	S		1ce1_P	20	6	1
R@P	116	ARG	R	V-KAPPA	1ce1_L	R@P	7	ALA	A		1ce1_P	7	2	0



Kaas Q. et al

CONCLUSIONS and PERSPECTIVES

http://www.imgt.org

1. The IMGT-ONTOLOGY axioms and concepts: CLASSIFICATION (nomenclature), DESCRIPTION (labels), NUMEROTATION (IMGT unique numbering, IMGT Colliers de Perles)…are acknowledged as the international standards in immunogenetics and immunoinformatics.

2. The WHO-INN programme requires the CDR-IMGT lengths.

3. American and European companies (Centocor Johnson and Johnson USA, Merck USA,..) have adopted the IMGT® tools for antibody engineering and antibody humanization.

4. The IMGT-ONTOLOGY axioms are used for a multiscale and systemic approach (system immunobiology). Concepts are currently described at the cell level (EU ImmunoGrid IST projet).

Many thanks to the IMGT® team at Montpellier, France