

Bioinformatique et ontologies

Marie-Paule Lefranc

Université Montpellier 2
Institut de Génétique Humaine, UPR CNRS 1142

Collège Doctoral, Université Sud de France
22 mai 2013, 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 continue 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.

Mercredi 22 mai 2013:

Marie-Paule Lefranc

Immunogénétique et Immunoinformatique
IMGT-ONTOLOGY axioms et concepts

Jeudi 23 mai 2013:

Véronique Giudicelli

Ontologies et Protégé

Vendredi 24 mai 2013:

Patrice Duroux

Ontologies et Système d'information

Joumana Michaloud

Ontologies et Contrôle de qualité

Mercredi 22 mai 2013

1. IMGT®

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

2. Ontologie

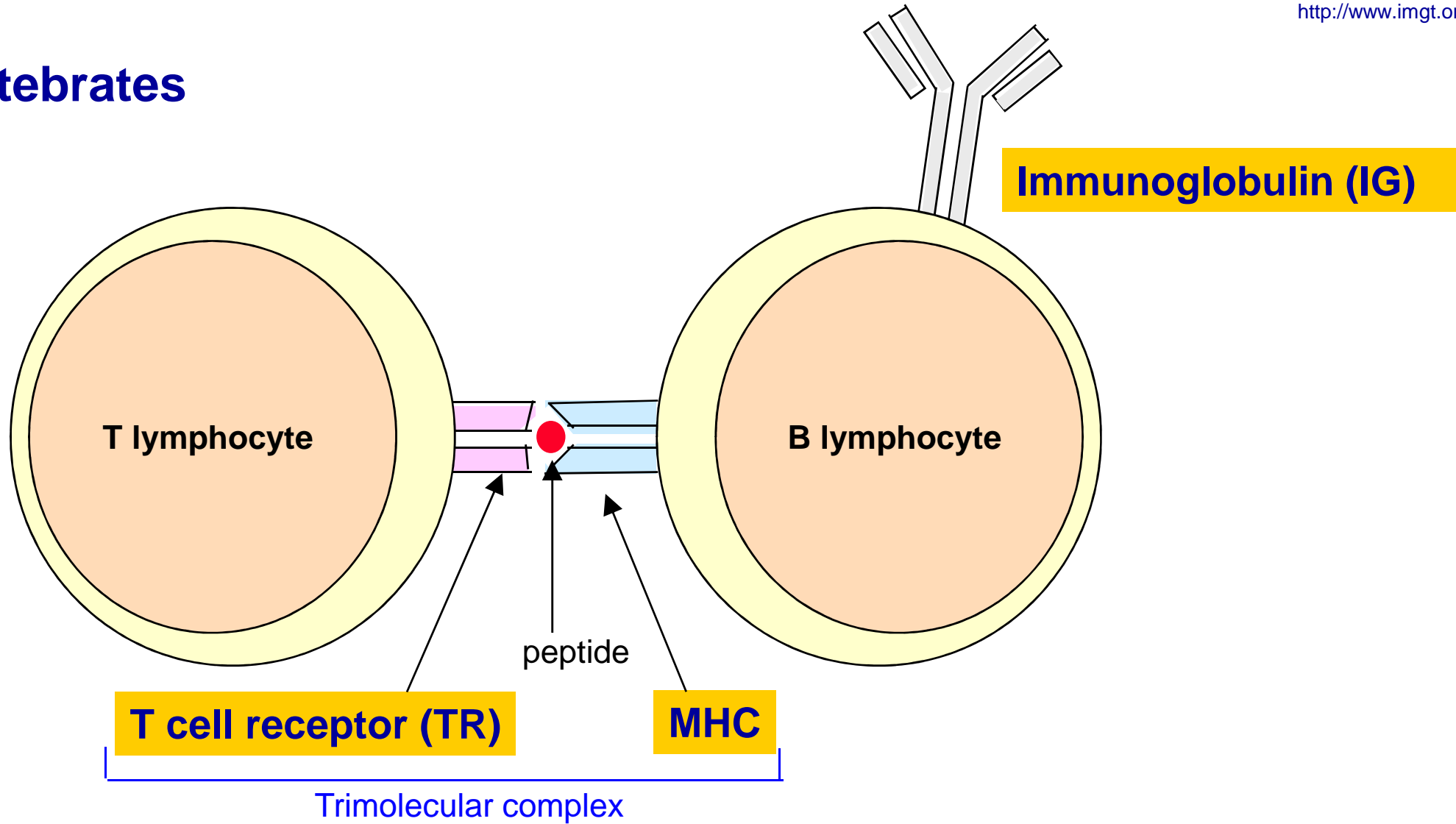
Qu'est-ce qu'une ontologie?

3. IMGT-ONTOLOGY axioms et concepts

4. Exemples d'applications

IMGT® domain: the adaptive immune response

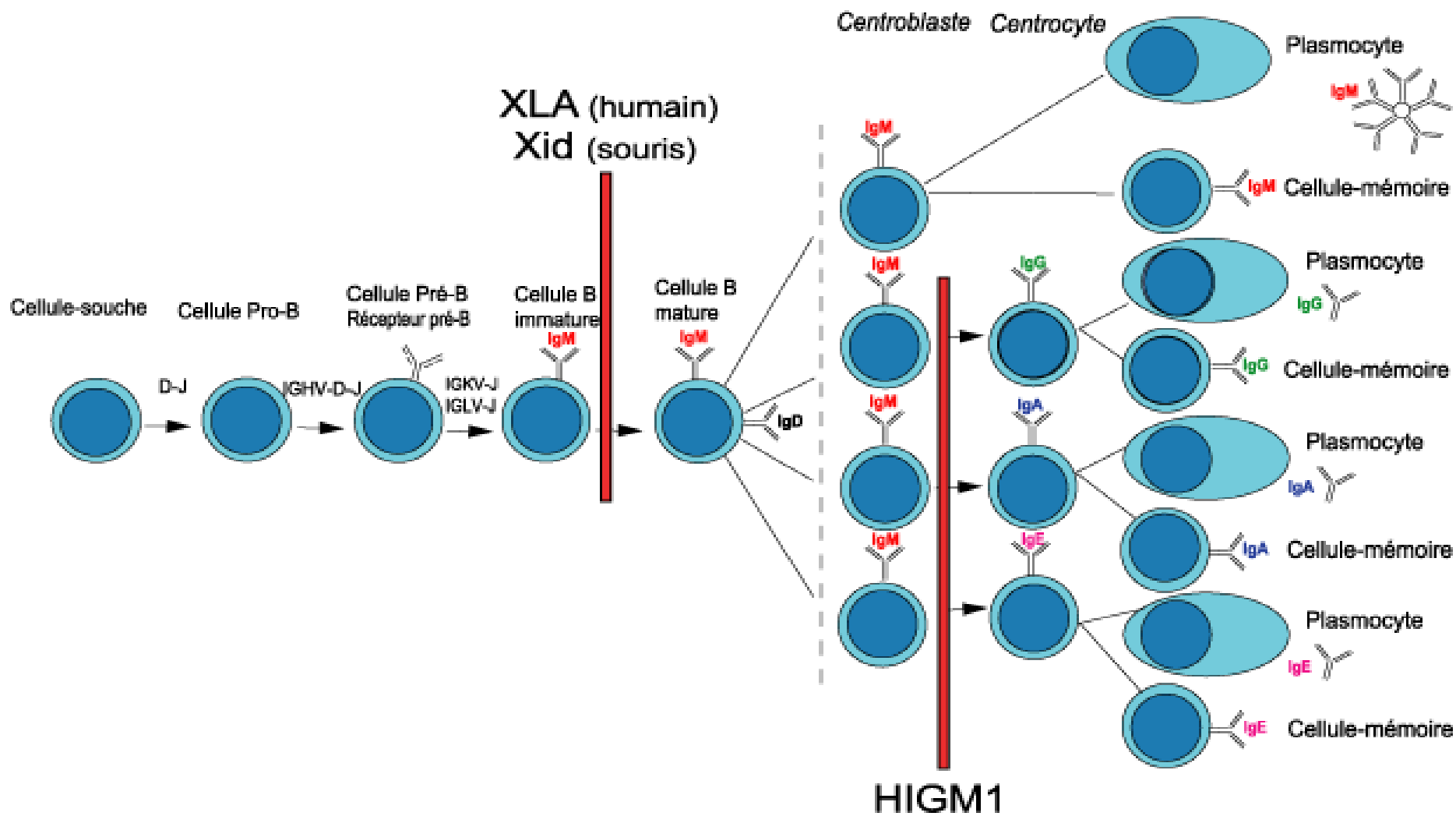
Vertebrates



Bone marrow

Blood

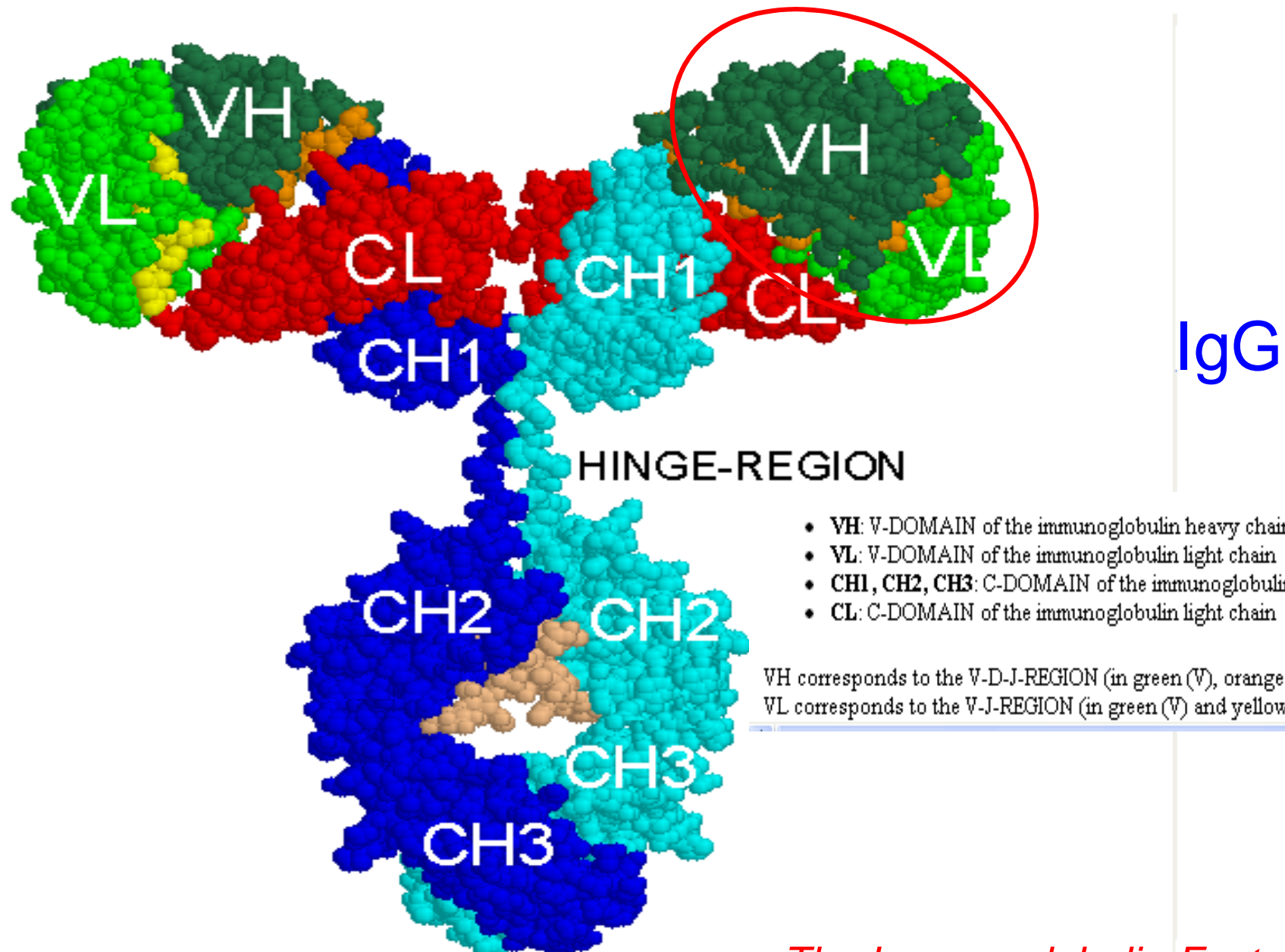
Lymph nodes, spleen



V-D-J and V-J rearrangements

Hypermutations, selection

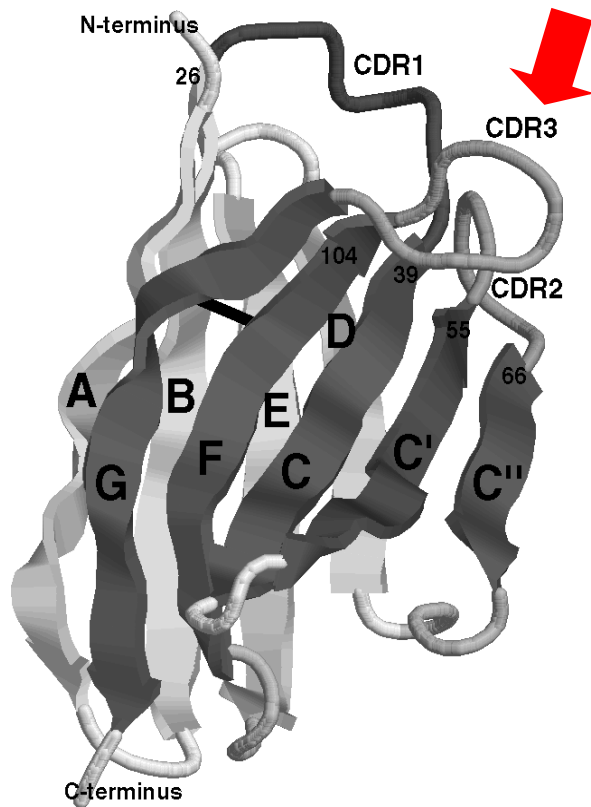
Immunoglobulin or antibody



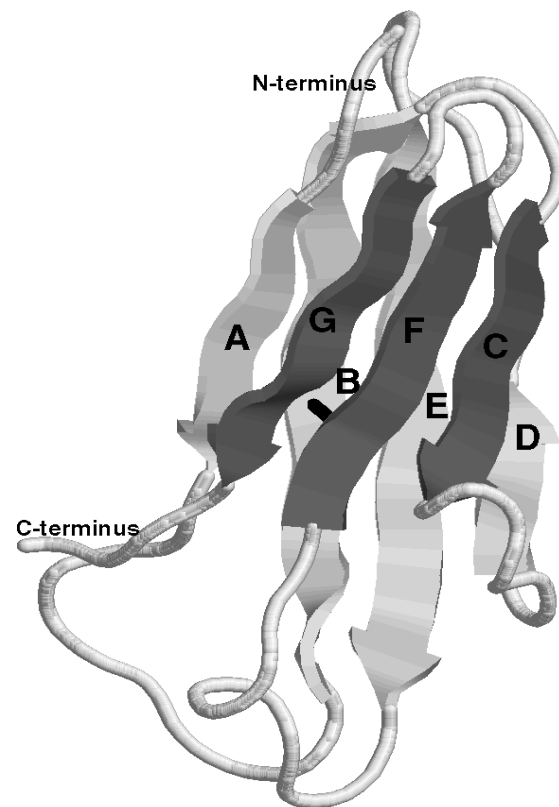
Structural domains

IG and TR

V-DOMAIN



C-DOMAIN

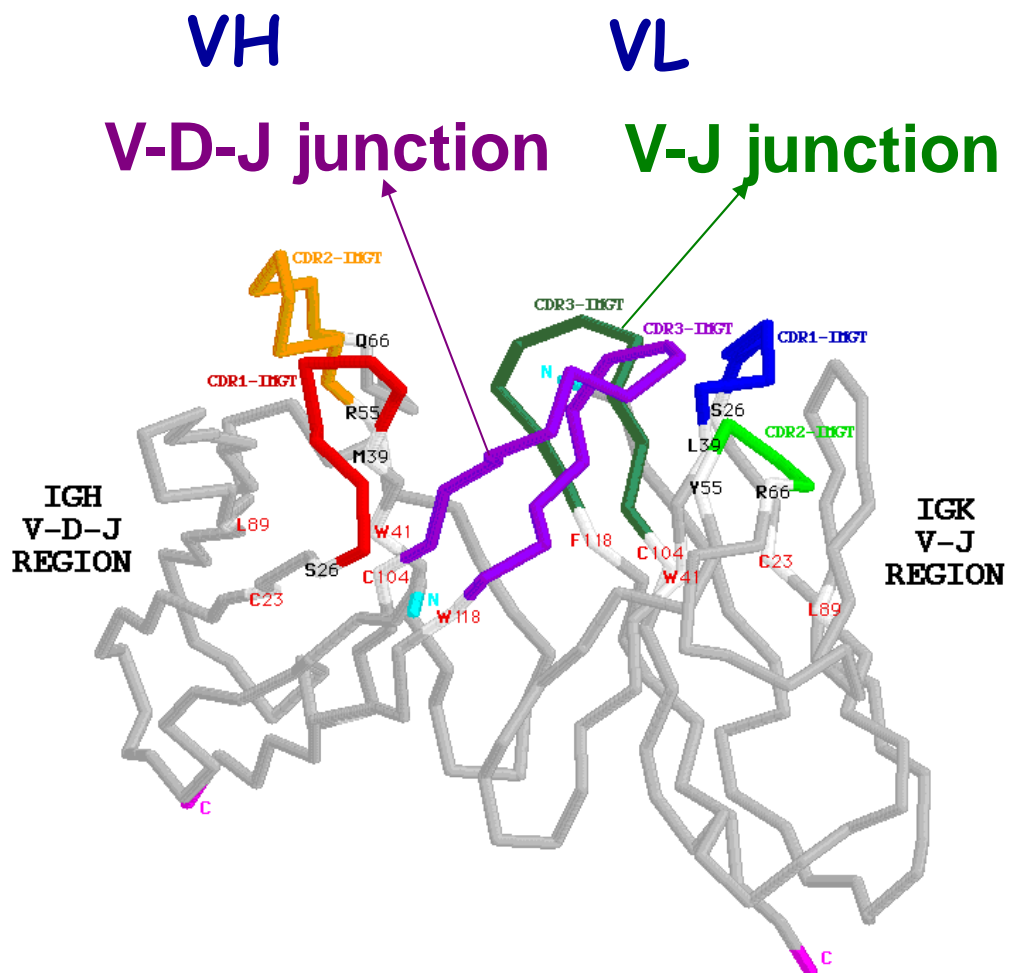


MHC

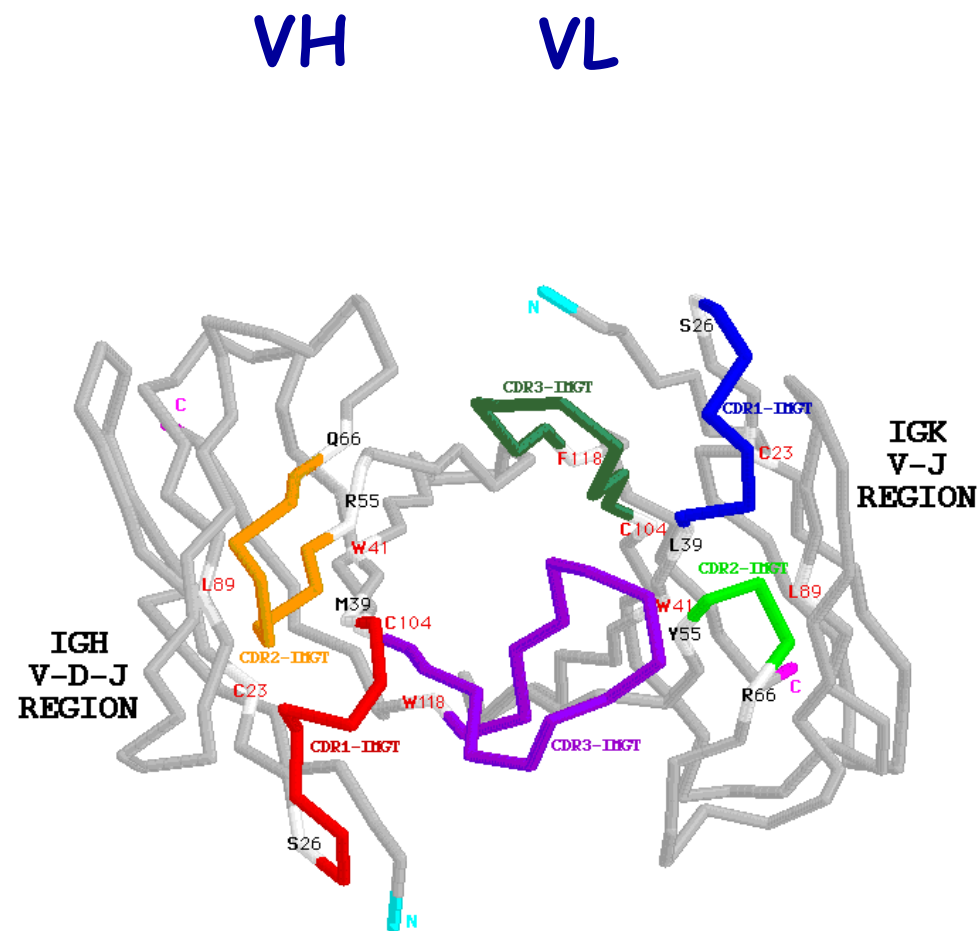
G-DOMAINS



V-DOMAINS: VH and VL



Side view of the V-DOMAINS



View from above the CDRs

Mouse (*Mus musculus*) E5.2Fv

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

V-J junction (104-118)

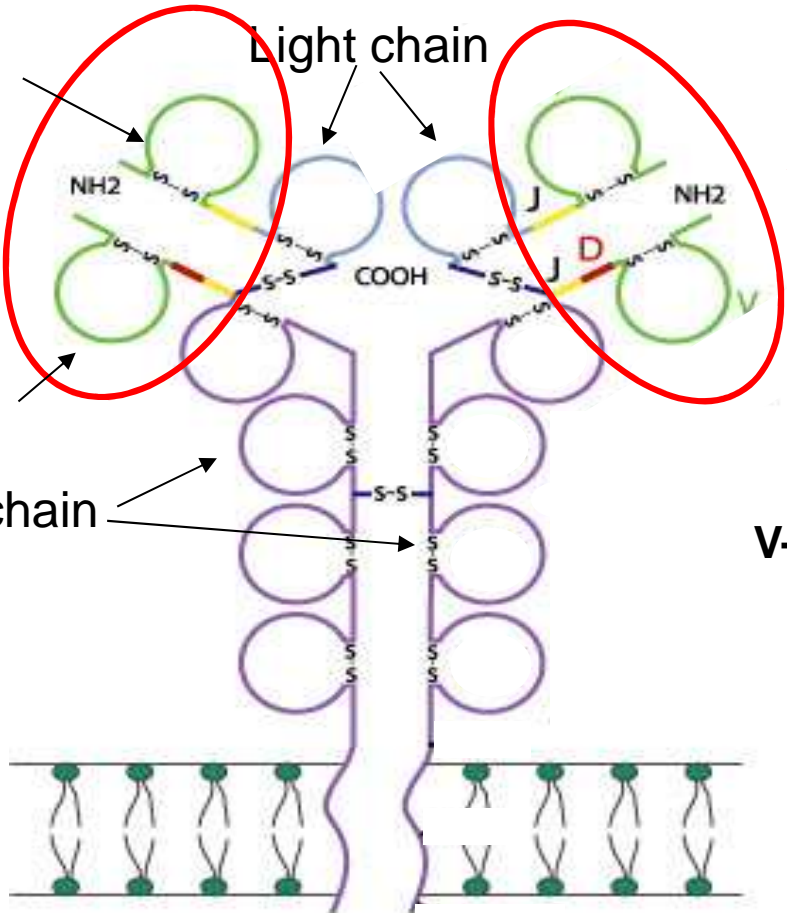
V-D-J junction (104-118)

Immunoglobulin (IG)

T cell receptor (TR)

V-DOMAIN
V-J-REGION

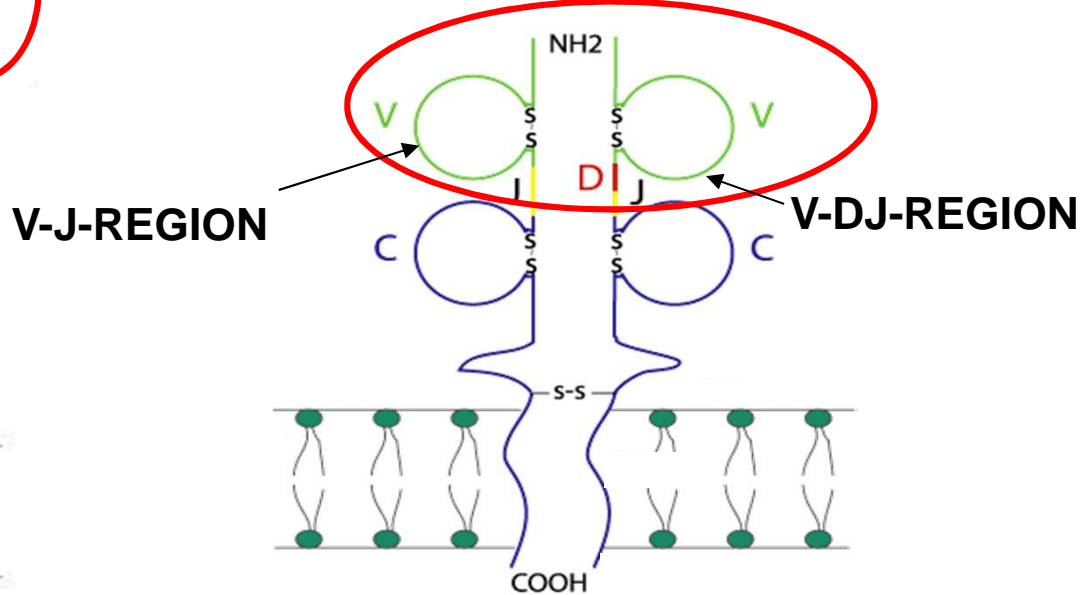
V-DOMAIN
V-D-J-REGION



Membrane IgM

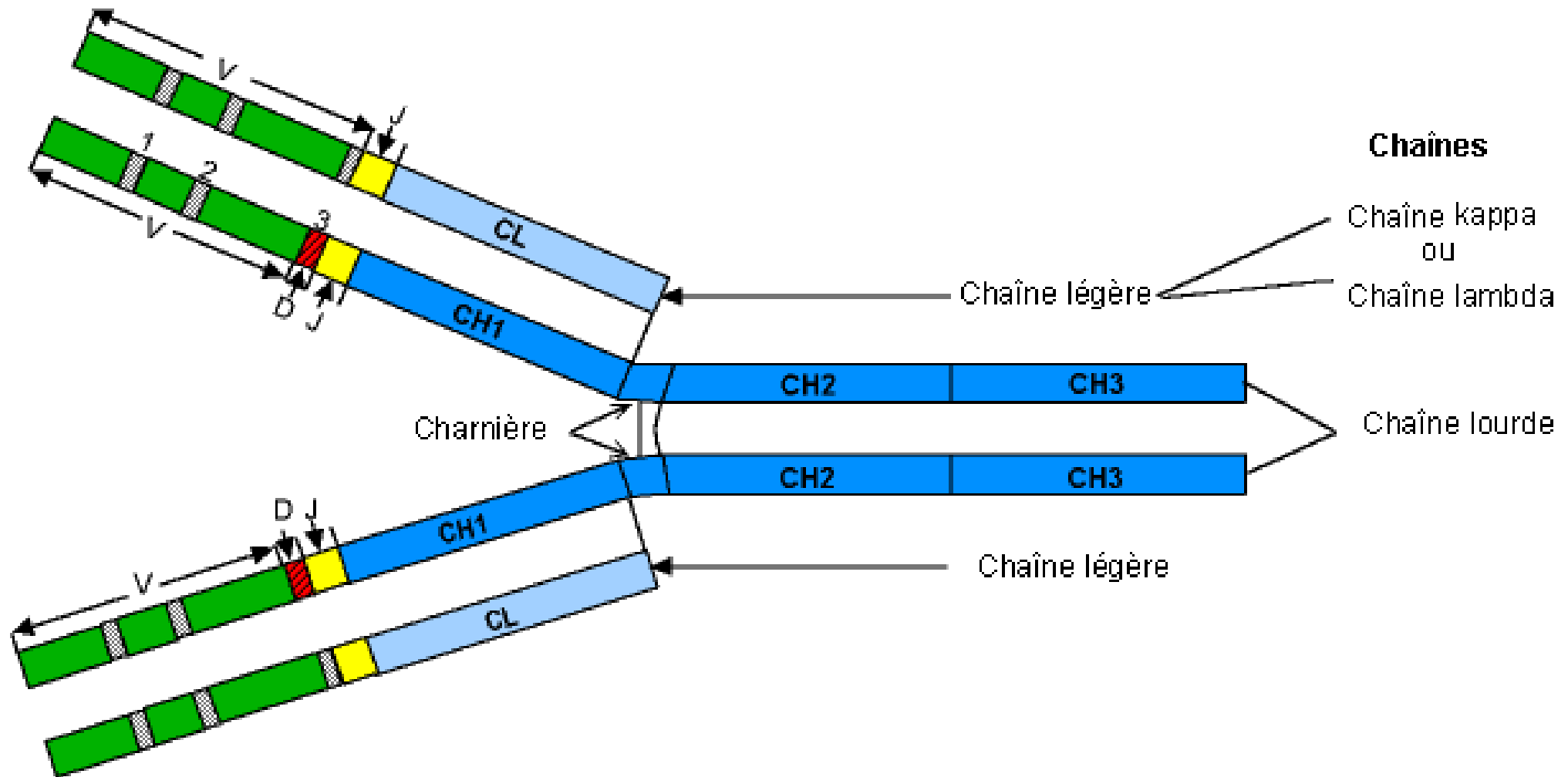
Contribution of the
2 V-DOMAINS
to the antigen binding site

Alpha - Beta
Gamma - Delta



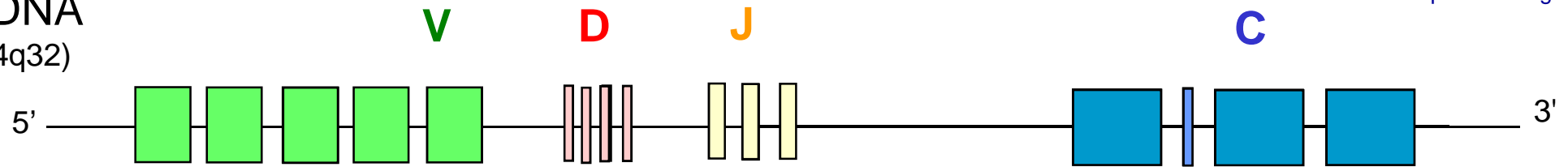
T cell receptor

Immunoglobulin IgG



Immunoglobulin (IG) synthesis

genomic DNA
(IGH Locus 14q32)

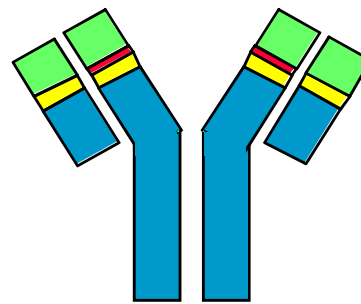


rearranged
DNA



mRNA

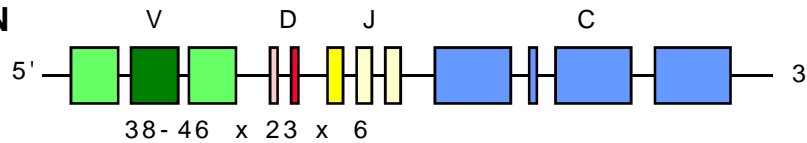
2×10^{12} different IG
per individual



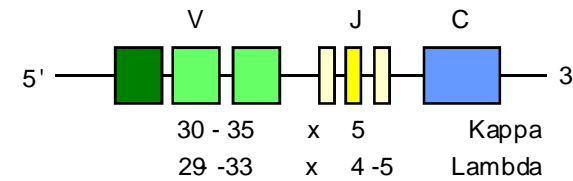
Immunoglobulin (IG) synthesis

150 FUNCTIONAL IG GENES

HEAVY CHAIN



LIGHT CHAIN



6300 POTENTIAL RECOMBINATIONS

185 + 165 POTENTIAL RECOMBINATIONS

N-DIVERSITY
SOMATIC MUTATIONS
x 1000

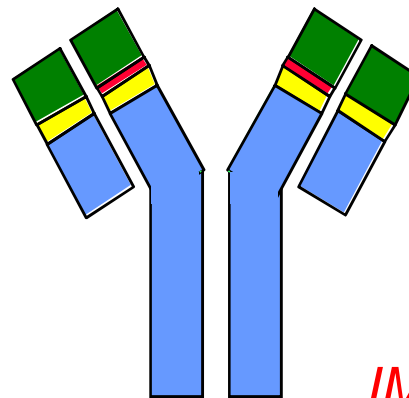


ABOUT 6.3×10^6 POSSIBILITIES

ABOUT 3.5×10^5 POSSIBILITIES

2×10^{12}

DIFFERENT ANTIBODIES



IMGT® <http://www.imgt.org>



IMGT®, the international ImMunoGeneTics information system®

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

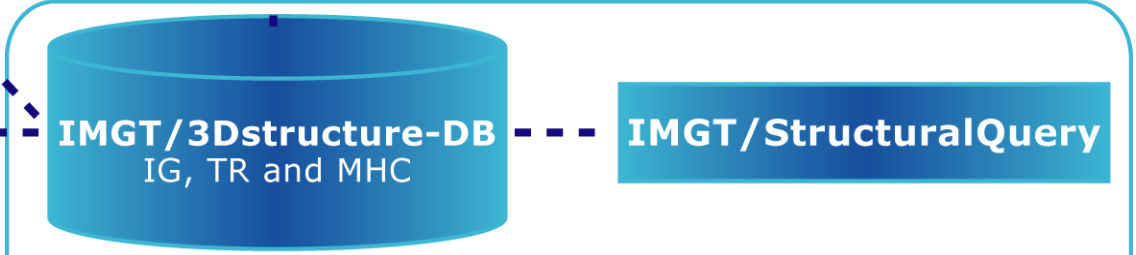
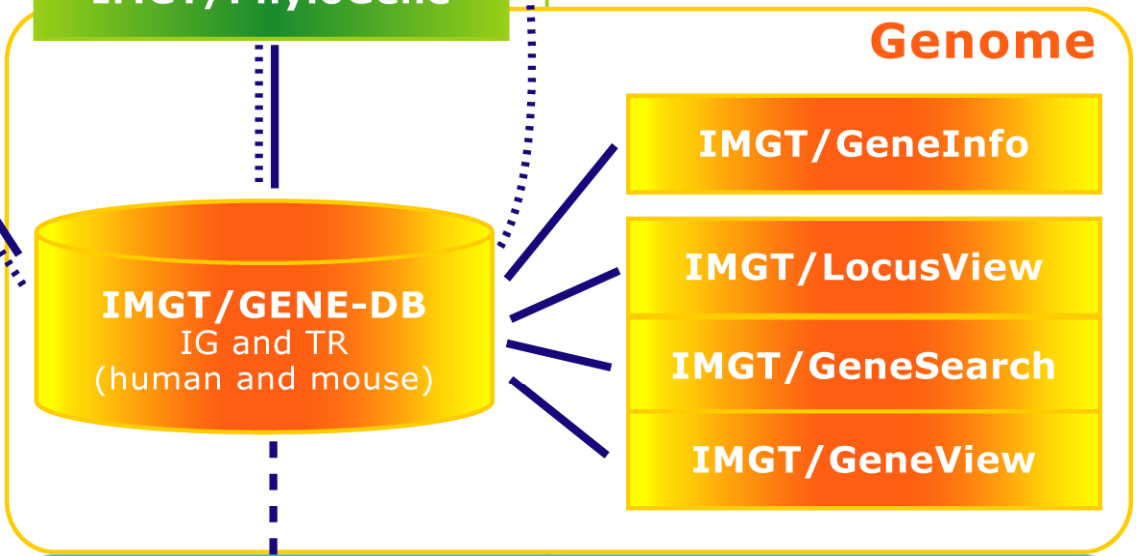
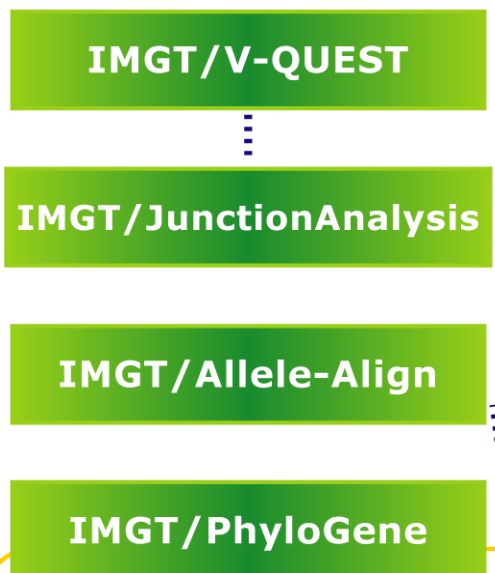
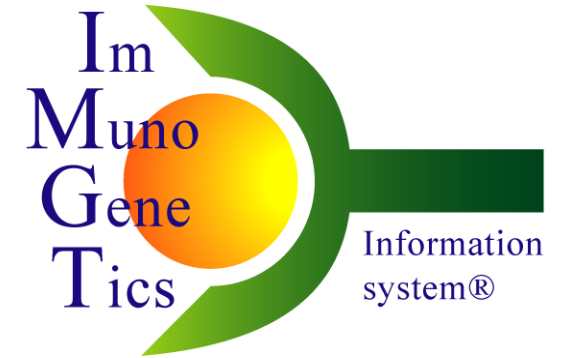
IMGT® is the international reference in immunogenetics and immunoinformatics.

IMGT® comprises:

- 7 databases
- 17 online tools
- more than 15,000 HTML pages of Web resources.

IMGT® receives 150.000 requests per month.

Sequences

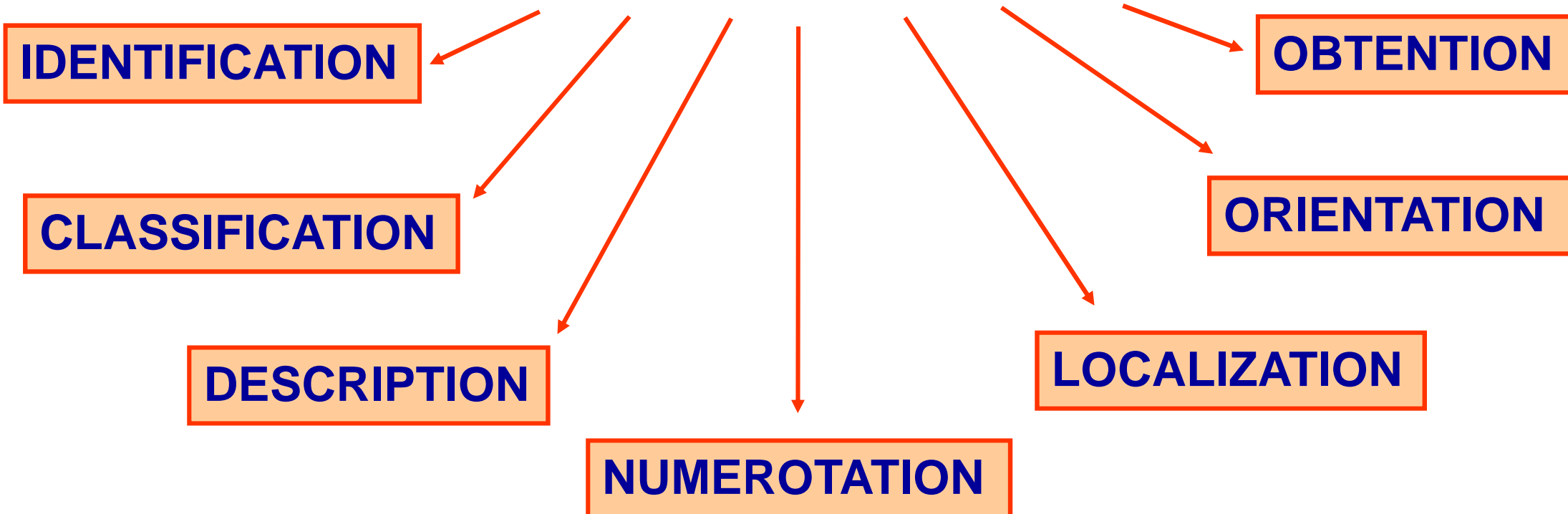


Why and how has IMG T[®] become a paradigm towards Systems Biology?

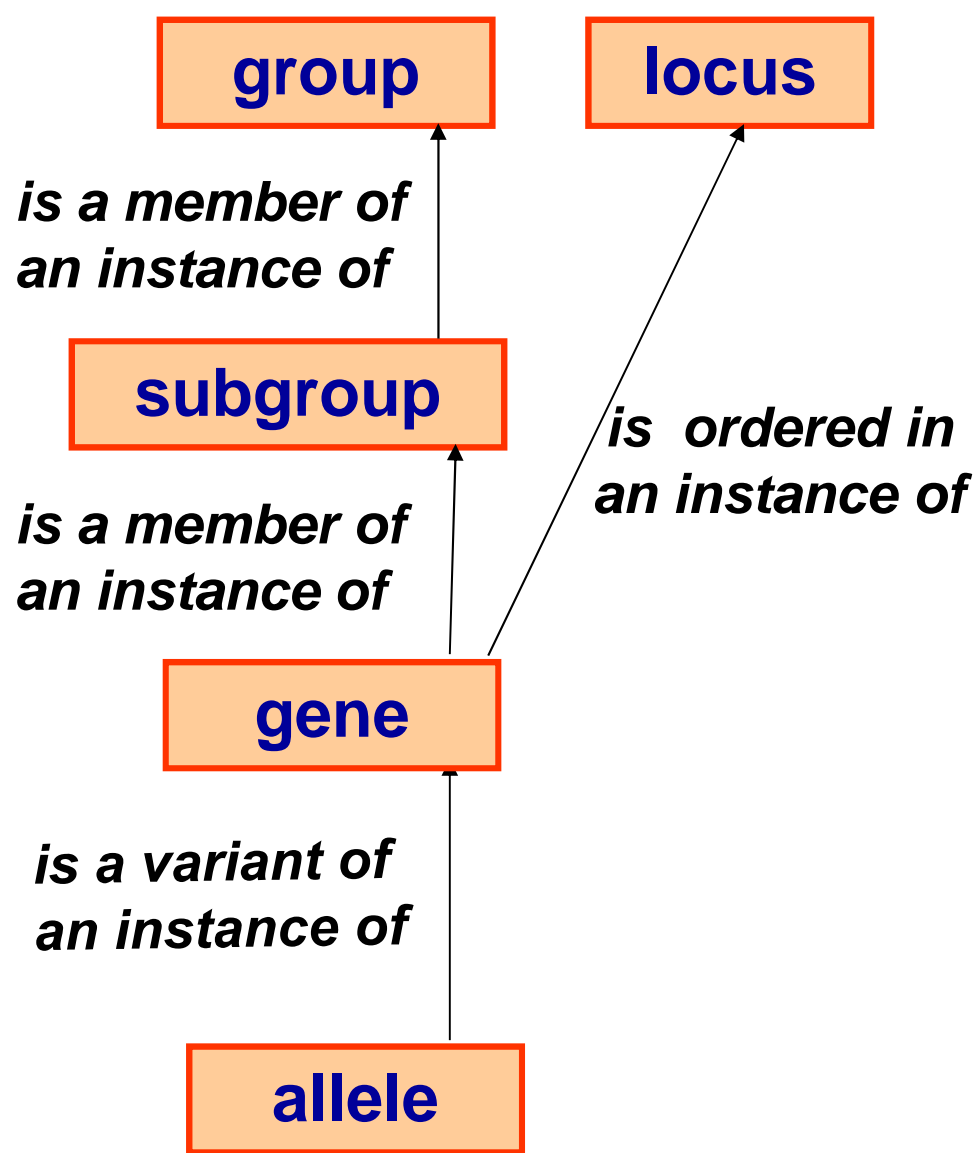
IMGT-ONTOLOGY axioms and concepts

IMGT-ONTOLOGY seven axioms:

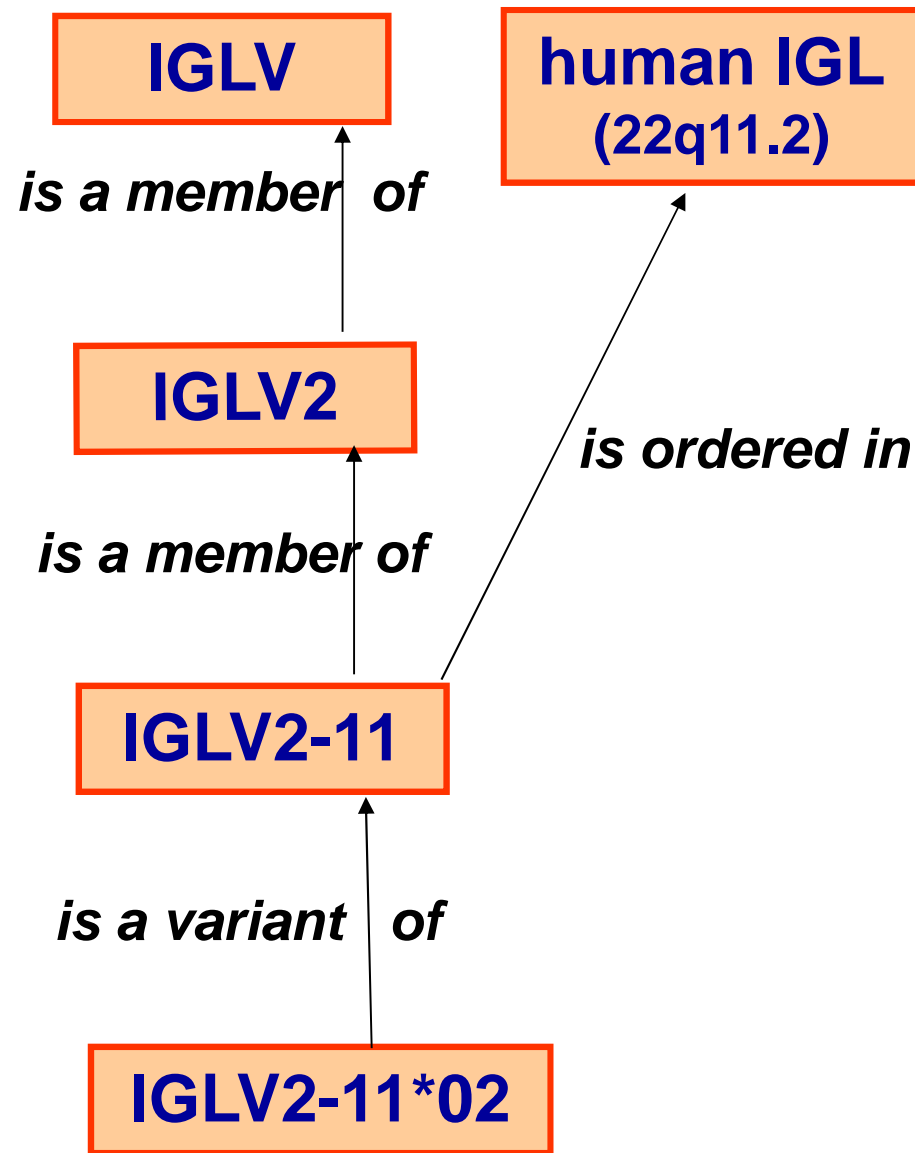
To share, reuse and represent knowledge
in Immunogenetics and Life Sciences



CLASSIFICATION axiom



« Concepts »



« Instances »

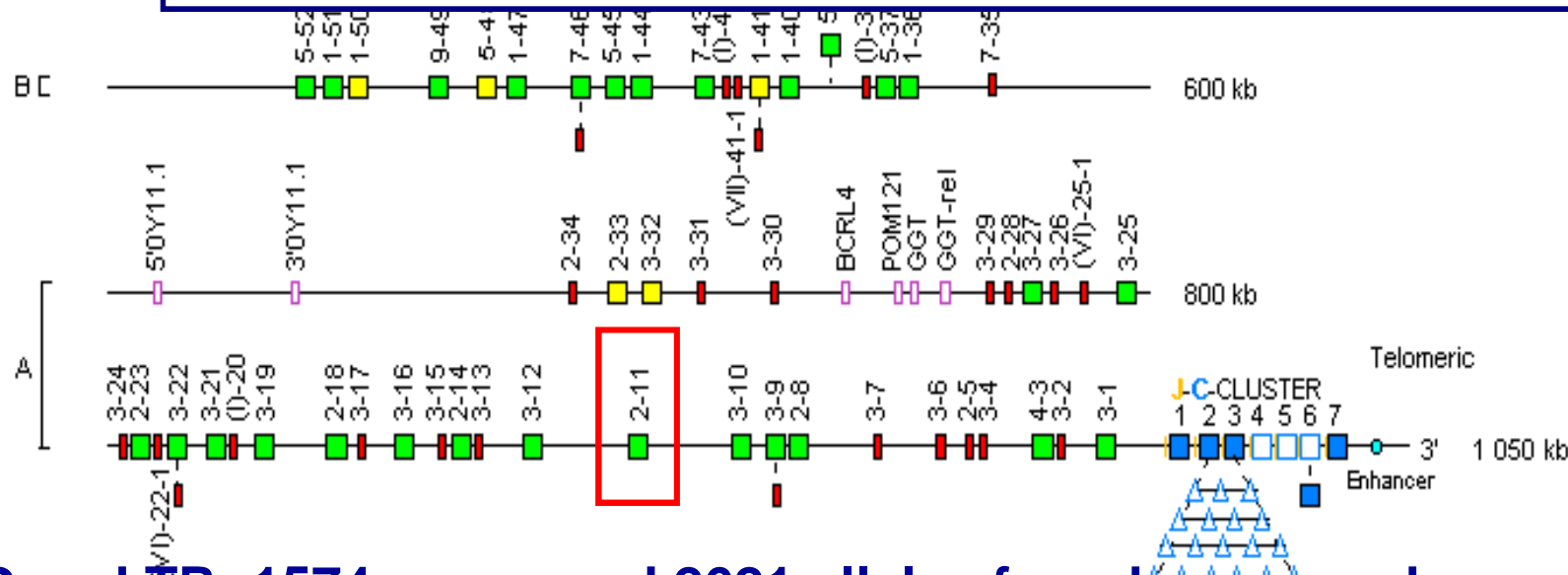


Locus representation: Human IGL

Human IGL 2

**WELCOME !
to IMGT/GENE-DB**

THE
INTERNATIONAL
IMMUNOGENETICS
INFORMATION SYSTEM®



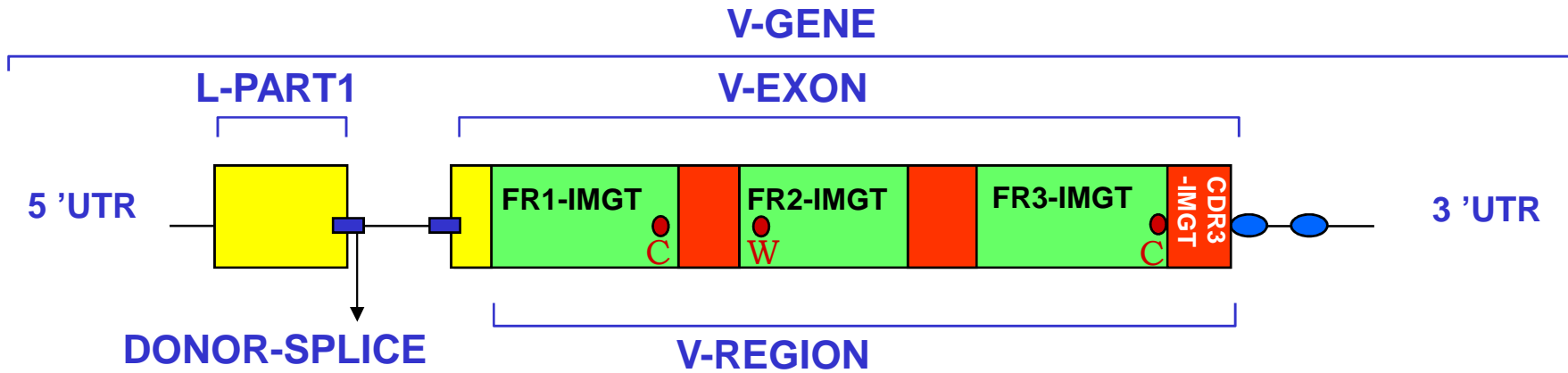
IG and TR: 1574 genes and 2621 alleles from human and mouse

CLASSIFICATION axiom

- 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

Relations entre Labels

V-GENE

V-EXON



FR3-IMGT

CDR3-IMGT



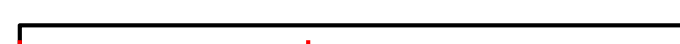
L-PART1

DONOR-SPLICE



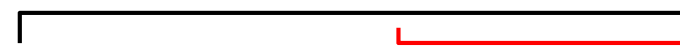
V-REGION

FR1-IMGT



V-REGION

CDR3-IMGT

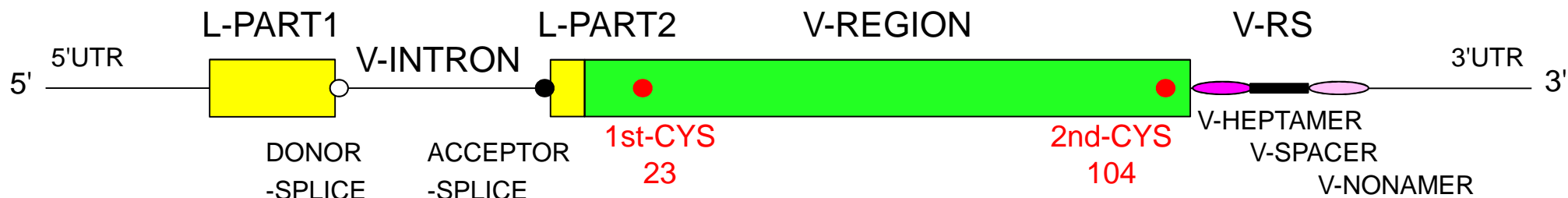


An example of V-GENE

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

```

tgagagctcc gttcctcacc atggactgga cctggaggat cctcttcttg gtggcagcag      60
ccacaggaa gaggctccct agtcccagtg atgagaaaga gattgagtcc agtccagggga     120
gatctcatcc acttctgtgt tctctccaca ggagccact ccaggtgca gctggtgcag      180
tctggggctg aggtgaagaa gcctggggcc tcagtgaagg tctcctgca ggcttctgga      240
tacaccttca ccggctacta tatgcactgg gtgcgacagg ccctggaca agggcttgag      300
tggatgggat ggatcaacc taacagtggg ggcacaaact atgcacagaa gtttcagggc      360
agggtcacca tgaccagggg cacgtccatc agcacagcct acatggagct gagcaggctg      420
agatctgacg acacggccgt gtattactgt gcgagagaca cagtgtgaaa acccacatcc     480
tgagggtgtc agaaacccaa gggaggaggc ag
  
```



IMGT/LIGM-DB



<http://www.igmt.org>

**D
E
S
C
R
I
P
T
I
O
N**

IMGT/LIGM-DB Consultation module v3 - Mozilla Firefox

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

CLASSIFICATION

170,685 sequences from 335 species

IMGT-ONTOLOGY:

277 IMGT labels for sequences

285 IMGT labels for 3D structures

SO (Sequence ontology):

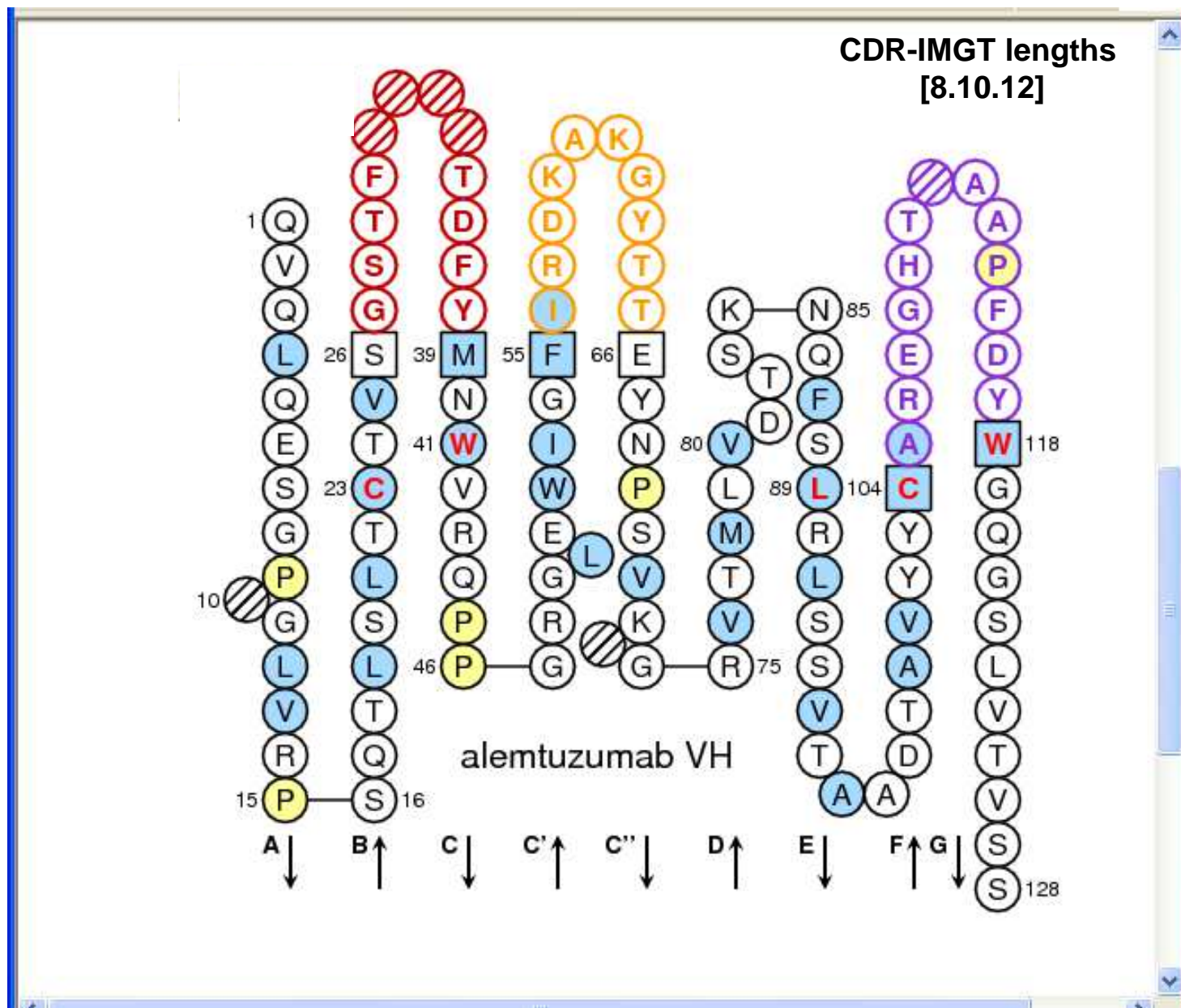
67 IMGT labels

DESCRIPTION axiom

- The IMGT-ONTOLOGY concepts of description comprise the standardized **IMGT labels** and **relations**.
- They have allowed to describe the IG, TR and MH 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/pMH**).
- 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

IMGT
Collier
de
Perles

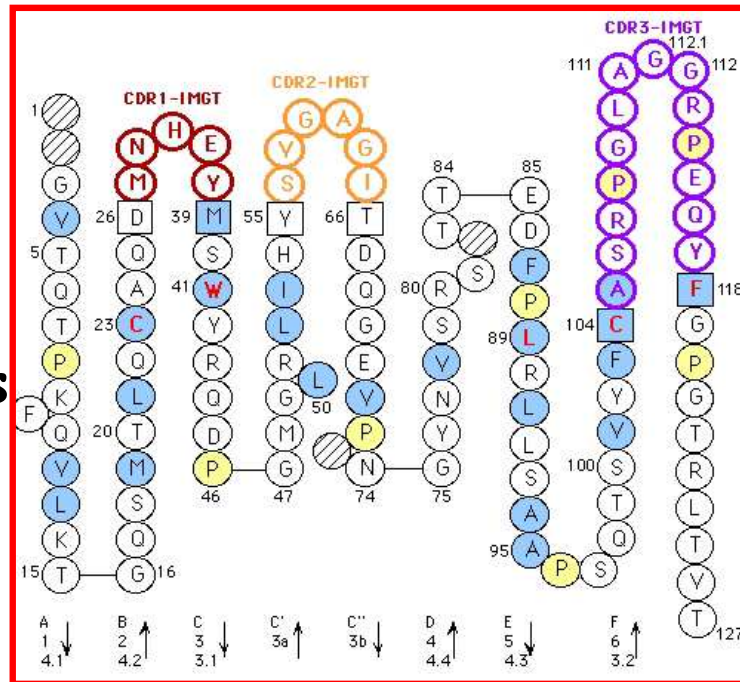


IMGT Web resources: 15 000 pages HTML



<http://www.imgt.org>

IMGT
Collier
de Perles

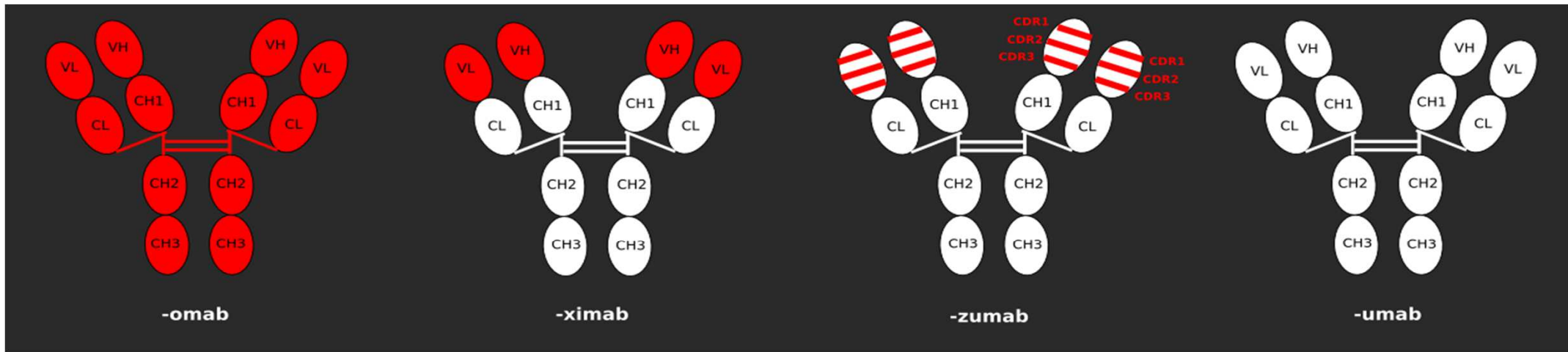


Sequence	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20																									
X02850 ,TRAV8-6*01	A	Q	S	V	T	Q	L	D	S	Q	V	F	F	E	A	P	V	E	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40							
AE000659 ,TRAV8-6*02	GCC	CAG	TCT	GTG	ACC	CAG	CTT	GAC	AGC	CAR	GTC	CCT	GTC	TTT	GAA	GAA	GCC	CCT	GTG	GAG	L	R	C	N	V	S	S	S	V	S	V	Y	Y	Y	Y	Y	Y	Y	L	F					
M86361 ,TRAV8-6*02	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	CTG	AGG	TGC	AAC	TAC	TCA	TGG	TCT	GTT	TCA	GTG	TAT	---	---	---	---	---	---	---	CTC	TTC				
X02850 ,TRAV8-6*01	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60				
AE000659 ,TRAV8-6*02	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	TGG	TAT	GTG	CAR	TAC	CCC	ARC	CAR	GGA	GTC	CAG	CTT	CTC	CTG	ARG	TAT	TTA	TCA	GGA	TCC				
M86361 ,TRAV8-6*02	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---			
X02850 ,TRAV8-6*01	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80				
AE000659 ,TRAV8-6*02	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	T	L	---	---	---	V	E	S	I	N	---	---	---	---	---	---	---	---	---	G	F	E		
M86361 ,TRAV8-6*02	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	ACC	CTG	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	
X02850 ,TRAV8-6*01	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	
AE000659 ,TRAV8-6*02	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	
M86361 ,TRAV8-6*02	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	
X02850 ,TRAV8-6*01	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	81	82	83	84	84A	84C	84C	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	
AE000659 ,TRAV8-6*02	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	A	E	F	N	K	S	Q	T	S	F	H	L	R	K	P	S	V	H	I	S	D	T	A	
M86361 ,TRAV8-6*02	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	GCT	GAA	TTT	ARG	ARG	AGT	CAR	ACT	TCC	TTC	CAC	TTG	AGG	AAA	CCC	TCA	GTG	CAT	ATA	AGC	GAC	ACG	GCT	
X02850 ,TRAV8-6*01	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
AE000659 ,TRAV8-6*02	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
M86361 ,TRAV8-6*02	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
X02850 ,TRAV8-6*01	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	101	102	103	104	105	106	107	108																
AE000659 ,TRAV8-6*02	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	E	V	F	C	A	V	S																	
M86361 ,TRAV8-6*02	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	GAG	TAC	TTC	TGT	GCT	GTG	AGT	GA																
X02850 ,TRAV8-6*01	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
AE000659 ,TRAV8-6*02	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
M86361 ,TRAV8-6*02	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

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.PSEVTAVEGAIQINCTYQ	TSGFYG.....	LSWYQQHDGGAPTFLSY	NALDG.....	LEETG.....	RFSSFLSRSDSYGYLLQLQMKDSASYFC
AE000658, TRAV1-2	GQNIHQ.PTEMTATEGAIQINCTYQ	TSGFNG.....	LFWYQQHAGEAPTFLSY	NVLDTG.....	LEELG.....	RFSSFLSRSKGYSYLLKELQMKDSASYLC
AE000658, TRAV2	KDQVFQ.PSTVASSEGAVVEIFCNHS	VSNAYM.....	FFWYLLHPPGAPRLLVK	GSK.....	PSQQG.....	RYNMTYER.FSSLLILQVREADAAVYYC
AE000658, TRAV3	AQSVAQPEDQVNVVAEGNPLTVKCTYS	VSGNPY.....	LFWYVQYPNRLGQLLLK	YITGDNL...	VKGSY.....	GFEAEFNKTSFHLKPKSALVSDSALYFC
AE000658, TRAV4	LAKTTQ.PISMDSYEGQEVNITCSHN	NIATNDY.....	ITWYQQPPSQGPRFIIQ	GYKT.....	KVTNE.....	VASLFIAPADRKSSTLSLPRVSLSDTAVYYC
AE000659, TRAV5	GEDVEQS.LFLSVREGDSSVINCTYT	DSSSTY.....	LYWYKQEPGAGLQLLTY	IFSNMD.....	MKQDQ.....	RLTVLLNKKDKKLSLRADTQTGDSAIYFC
AE000659, TRAV6	SQKIEQNSEALNIQEGKTAITLTCNYT	NYSPAY.....	LQWYRQDPGPRGPFVLLI	IRENEK.....	EKRKE.....	RLKVTFTDITLKQSLFHITASQPADSATYLC
AE000659, TRAV7	ENQVEHSPHFLGPPQGDVMSMCTYS	VSRFNN.....	LQWYRQNTGMGPKHLLS	MYSAGY...	EKQKG.....	RLNATLLK.NGSSLYITAVQPEDSATYFC
AE000659, TRAV8-1	AQSVSQHNHHVILSEAASLELGCNYS	YGGTVN.....	LFWYVQYPGQHLQLLLK	YFSGDPL...	VKGIK.....	GFEAEFIKSKFSFNLKPKSVQWSDTAEYFC
AE000659, TRAV8-2	AQSVTQLDHSVSVSEGTPLVLLRCNYS	SSYSPS.....	LFWYVQHPNKGQLLLLK	YTSAAITL...	VKGIN.....	GFEAEFKKSETSFHLTKPSAHMSDAAEYFC
AE000659, TRAV8-3	AQSVTQPDIIHITVSEGASLELRCNYS	YGATPY.....	LFWYVQSPGQGLQLLLK	YFSGDITL...	VQGIG.....	GFEAEFKRSQSSFNLRKPSVHWSDAEYFC
AE000659, TRAV8-4	AQSVTQLGSHVSVSEGALVLLRCNYS	SSVPPY.....	LFWYVQYPNQGQLLLLK	YTSAAITL...	VKGIN.....	GFEAEFKKSETSFHLTKPSAHMSDAEYFC
X02850, TRAV8-6	AQSVTQLDSQVPVFEAPVELRCNYS	SSVSVY.....	LFWYVQYPNQGQLLLLK	YLSGSTL...	VESIN.....	GFEAEFNKTSFHLRKP SVHISDTAEYFC
AE000660, TRAV8-7	TQSVTQLDGHITVSEAPLELKCNYT	YSGVPS.....	LFWYVQYSSQSLQLLLK	DLTEATQ...	VKGIR.....	GFEAEFKKSETSFYLRKPSHVSDAEYFC
AE000659, TRAV9-1	QDSWVQTEGQWLPSFGDSLIVMCSYE	TTQVPS.....	LFWYVQYRGGEGPOLHLK	AMKAMD...	KGRNK.....	GFEAMYRKETTSPHLEKSDVQESDAVYFC



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)	

This table is from 2008. Using IMGT/mAb-DB: 1) What are the **withdrawn** mAbs?

2) What are the novel **approved** mAbs?

Humanized CAMPATH-1H mutant 1

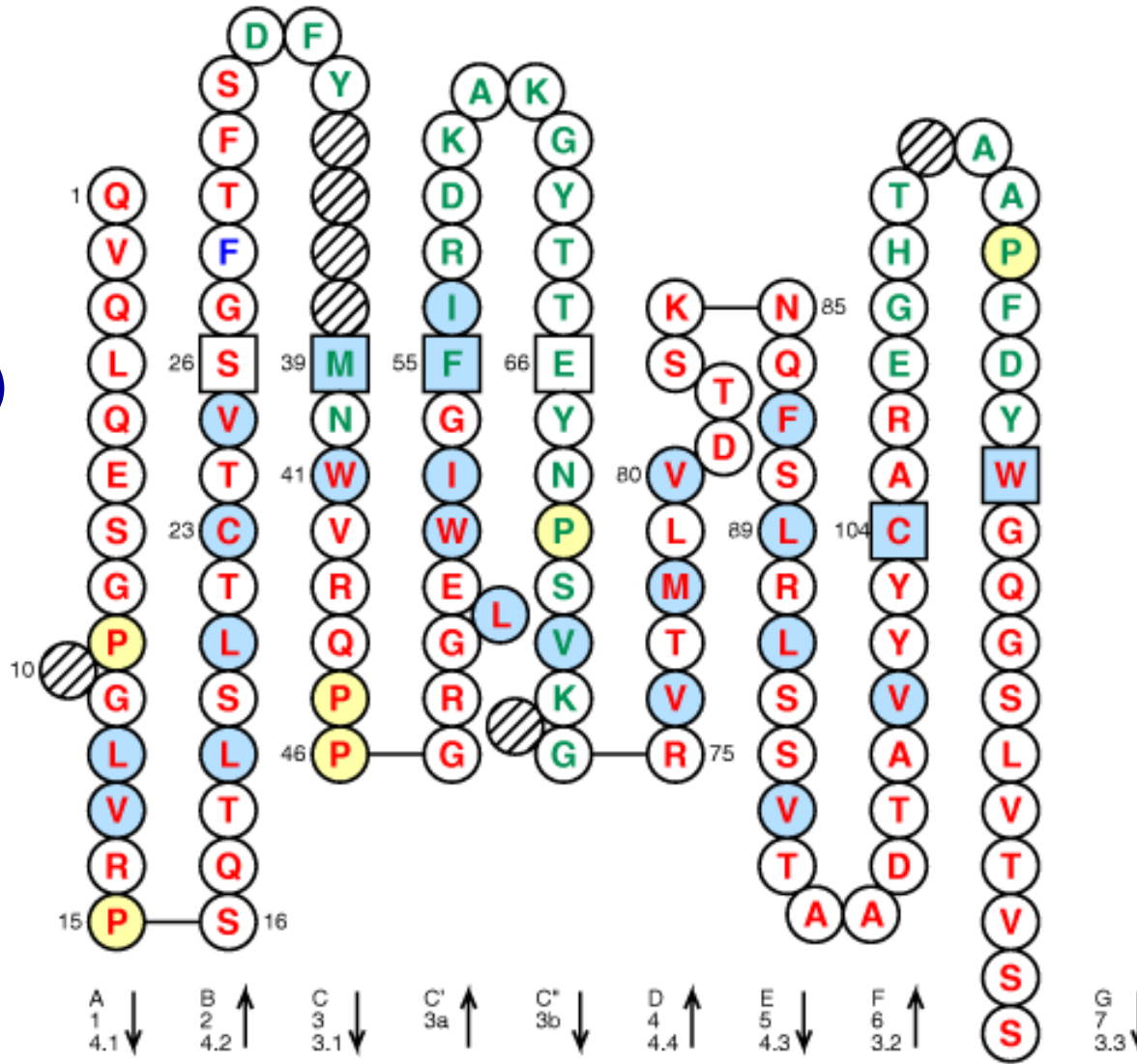
Mutant 1: **S28>F**

Mutant 2:
alemtuzumab
S31>T

VH domain
(V-D-J-REGION)

[8.10.12]

■ human
■ rat



NUMEROTATION axiom

- 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 (MH).
- 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.

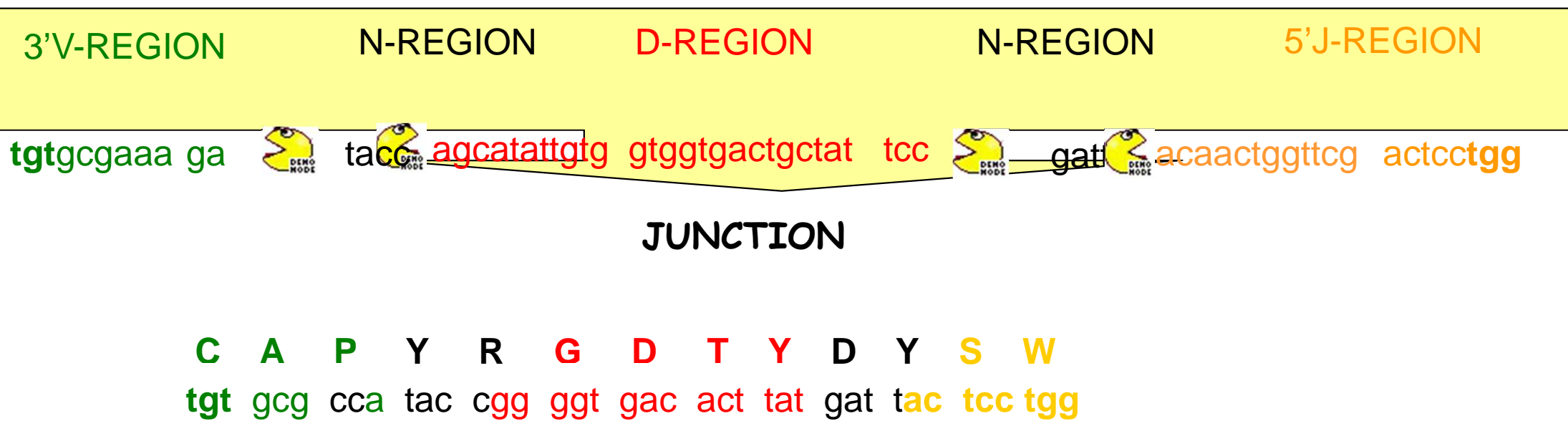
Examples of IMGT® tools based on the IMGT-ONTOLOGY concepts



<http://www.imgt.org>

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

Immunoglobulin V-D-J generation of sequence diversity



IMGT/JunctionAnalysis: analysis of the IG and TR junctions



<http://www.imgt.org>

IMGT/JunctionAnalysis Results

Locus IGH
Species Homo sapiens
IMGT repertoire 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		tgttgtgcagcgcctggtac	ccaaatac		...actttgaccactgg	IGHJ4*02	IGHD6-13*01	1	2	1	5/15
#2	Z70257	IGHV3-7*02		ggatggcagctcttatgcc	cgccc		ctactggactctgatctctgg	IGHJ2*01	IGHD2-2*01	0	2	0	9/11
#3	Z70606	IGHV4-31*03		c	.gactacg.....	cact		..atgcttttgatgctctgg	IGHJ3*01	IGHD4-17*01	0	0	0	3/5
#4	Z70608	IGHV4-39*05		cagagtaacgattttggagtggttatt.....	ccccgggga		..atgctttgatctctgg	IGHJ3*02	IGHD3-3*01	0	0	0	12/17
#5	Z70610	IGHV4-34*09		tcgggagcgattttggagtggttatt.....	cccgga	ca	tgatgctttgatctctgg	IGHJ3*02	IGHD3-3*01	0	0	0	9/12
#6	Z70611	IGHV4-59*01		catggtactataa.	tgccggcgttg		...actggttcgaccctgg	IGHJ5*02	IGHD3-9*01	0	2	0	9/13
#7	Z70613	IGHV4-59*01		cagcagctggtac	ctccct		...cttgactactgg	IGHJ4*02	IGHD6-13*01	0	0	0	4/6
#8	Z70614	IGHV4-59*01		cactataattcggggacttat.....	cccctc	gactactgg	IGHJ4*02	IGHD3-16*01	0	2	0	7/14
#9	Z70615	IGHV4-59*01		ggctg	gtaaagagg.....	tttcggaa		.tactggactctgatctctgg	IGHJ2*01	IGHD5-24*01	0	2	0	7/13
#10	Z70616	IGHV4-34*01		cgggtttggg.....	ttccc		...actggttcgaccctgg	IGHJ5*02	IGHD3-16*01	0	0	0	6/8
#11	Z70620	IGHV4-30-4*01		ccggggcgggatggtt.....	cgg		.gatgctttgatctctgg	IGHJ3*02	IGHD3-16*01	1	4	0	5/5
#12	Z70621	IGHV4-39*01		ccacgatttatgttcggggagtt.....	tgaccccc	ttgactactgg	IGHJ4*02	IGHD3-16*01	0	1	0	12/21
#13	Z70622	IGHV4-39*06	t	tgccccgctcctgccaaaat	gtattactatggttcgggga.....	tatgtacg	ttgactactgg	IGHJ4*03	IGHD3-10*01	0	0	0	15/28

The eleven IMGT amino acid classes according to the physicochemical properties

'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	
Small	108-117		C	P	T		D	
Very small	60-90	A		G	S		N	
		Aliphatic		Sulfur	Hydroxyl	Basic	Acidic	Amide
		Nonpolar			Uncharged	Charged	Uncharged	Polar

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
#1 AY393054	C	S	P	G	G	S	<u>A</u>	Y						Y	<u>H</u>	E	<u>H</u>	F	Q	Q	W
	tgt	agt	ccc	ggg	ggt	agt	<u>gct</u>	tat	tac	<u>cac</u>	gaa	<u>cac</u>	ttc	cag	<u>cag</u>	tgg
#2 AY393055	C	<u>V</u>	K	P	T	D	D	D	G				<u>H</u>	R	A	E	Y	F	Q	<u>Y</u>	W
	tgt	<u>gtg</u>	aaa	ccc	acg	gat	gat	gat	ggc	<u>cac</u>	cgg	gct	gaa	tac	ttc	cag	<u>tac</u>	tgg
#3 AY393058	C	S	P	G	G	S	<u>A</u>	Y						Y	<u>H</u>	E	<u>D</u>	F	Q	Q	W
	tgt	agt	ccc	ggg	ggt	agc	<u>gct</u>	tat	tac	<u>cac</u>	gaa	<u>gac</u>	ttc	cag	<u>cag</u>	tgg
#4 AY393072	C	S	P	G	G	S	<u>A</u>	Y						Y	<u>H</u>	E	<u>H</u>	F	Q	Q	W
	tgt	agt	ccc	ggg	ggt	agt	<u>gct</u>	tat	tac	<u>cac</u>	gaa	<u>cac</u>	ttc	cag	<u>cag</u>	tgg
#5 AY393088	C	A	R	Q	N	P	P	E	Y	S	G	A	Y	<u>H</u>	<u>D</u>	G	W	F	D	P	W
	tgt	gcg	aga	caa	aac	ccc	ccc	gag	tat	agt	ggc	gca	tat	<u>cat</u>	<u>gat</u>	ggg	tgg	ttc	gac	ccc	tgg
#6 AY393089	C	A	R	E	M	L	Y	G	S	G	<u>G</u>	Y	Y	P	P	D	A	F	<u>E</u>	<u>L</u>	W
	tgt	gcg	aga	gag	atg	ctc	tat	ggt	tcg	ggg	<u>ggt</u>	tat	tac	ccc	cct	gat	gca	ttt	<u>gag</u>	<u>ctc</u>	tgg
#7 AY393091	C	A	R	Q	N	P	P	E	Y	S	G	A	Y	<u>H</u>	<u>D</u>	G	W	F	D	P	W
	tgt	gcg	aga	cag	aat	ccc	ccc	gag	tat	agt	ggc	gca	tat	<u>cat</u>	<u>gat</u>	ggg	tgg	ttc	gac	ccc	tgg
#8 AY393092	C	A	R	E	M	L	Y	G	S	G	<u>G</u>	Y	Y	P	P	D	A	F	<u>E</u>	<u>V</u>	W
	tgt	gcg	aga	gag	atg	ctc	tat	ggt	tcg	ggg	<u>ggt</u>	tat	tac	ccc	cct	gat	<u>gca</u>	ttt	<u>gag</u>	<u>gtc</u>	tgg
#9 AY393094	C	A	R	Q	N	P	P	E	Y	S	G	A	Y	<u>H</u>	<u>D</u>	G	W	F	D	P	W
	tgt	gcg	aga	cag	aac	ccc	ccc	gag	tat	agt	ggc	gca	tat	<u>cat</u>	<u>gat</u>	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®



<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](#)

Analyse your Immunoglobulin sequences

Your selection: Human

Your sequences are compared to the **Human IG set** from the [IMGTV-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
gctgggttttcccttgttgctatTTTAAAAGGTGTCCAATGTGAGGTGCAGCTGGTGGAGTCTGGGGGAGGCTTGGTACAGCCAGGGCGG
tccctgagactctcctgtgcagcttctggattgaccttgggtgattactttatgagctgggtccgccaggctccaggggaagggactgga
gtgggtaggtttcattaagagcgaaacttatggtgggacaacagaatacgcgcgctctgtgaaaggcagatcatcatctcgagagatg
attccaaaagcatcgctatTTGCAAATGAACAGCCTGGAAACCGAGGACACAGCCATATATTACTGTAGTCCCAGGGGGTAGTGCTTAT
taccacgaacacttccagcagtggggcccgggaccatggteaccgtctcctcagcctccaccaagggcccatcggctcttccccctggc
acctctccaagagcacctctgggggcacagcggccctgggctgcctggtcaaggactacttcccc
>AY393055
gctgggttttcccttgttgctatTTCAAAGGTGTCCAGTGTGAGGTGCAGCTGGTGGAGACTGGAGGAGGCTTGATCCAGCCTGGGGGG
tccctgagactctcctgtgcagcttctgggttaccgtcagtagcaactacatgagctgggtccgccaggctccaggggaaggggctgga
```

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

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

<http://www.imgt.org>

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

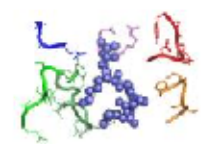
THANK YOU
for using IMGT/3Dstructure-DB

THE
INTERNATIONAL
IMMUNOGENETICS
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, alemruzumab , 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:

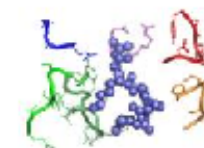
IMGT/3Dstructure-DB Domain pair contacts (overview) of 1ce1

- Atom contact types Atom contact categories
- Non covalent
 - Covalent
 - (BB) Backbone/backbone

Kaas Q. et al.

Atomic pair contacts in IMGT/3Dstructure-DB

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

- Non covalent
- Polar
- Hydrogen bond
- Non polar

- 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

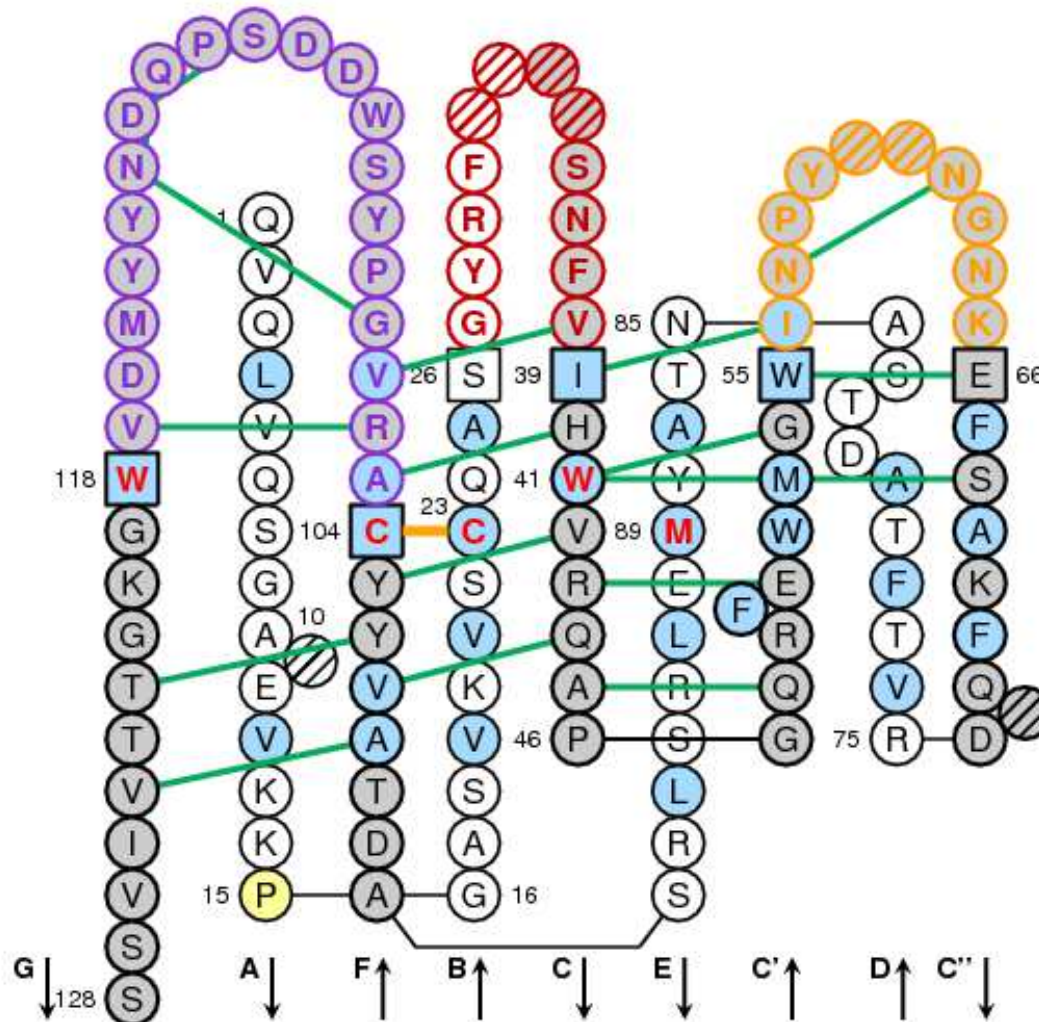
Show

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

Hydrogen bonds (IMGT Collier de Perles on 2 layers)

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

CDR-IMGT lengths [8.8.20]



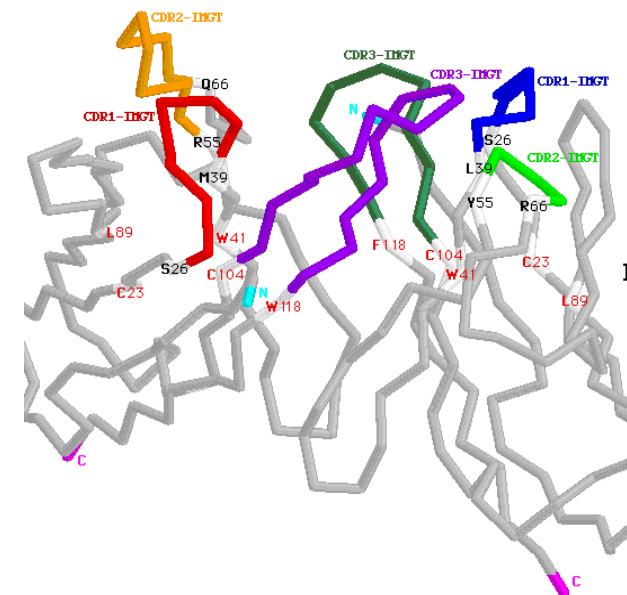
Contacts VH-(Ligand), V-KAPPA-(Ligand)

IMGT molecule name	IMGT description	Chain ID	IMGT chain description	Domain number	IMGT domain description
CAMPATH-1H, alemtezumab , 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

Contacts VH-(Ligand)

	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



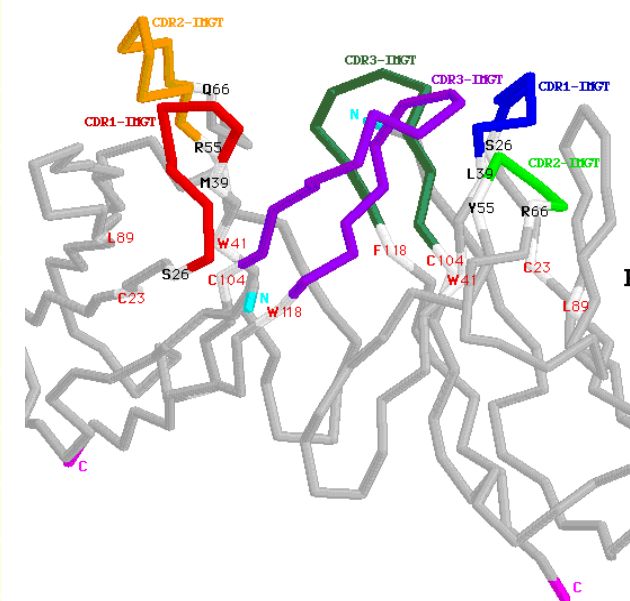
Summary:

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		
IMG T Num	Residue	Domain	Chain		IMG T 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



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 (Janssen 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**).



The IMGT® team, Montpellier, France