

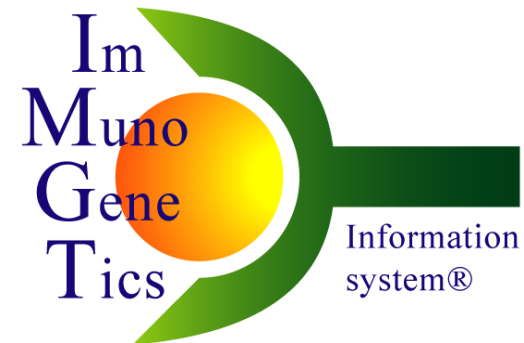
Computational resources for immunoinformatics (IMGT)

Marie-Paule Lefranc

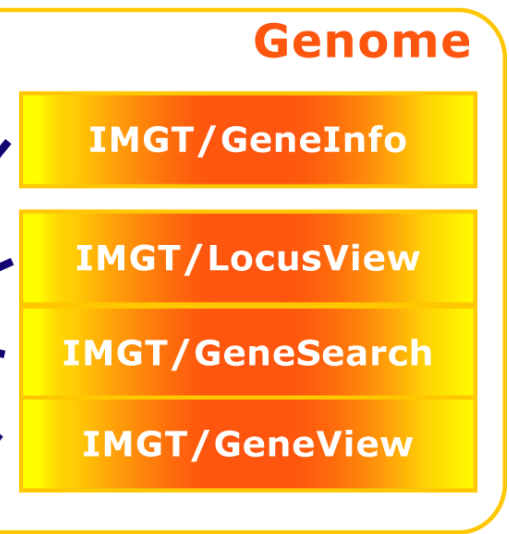
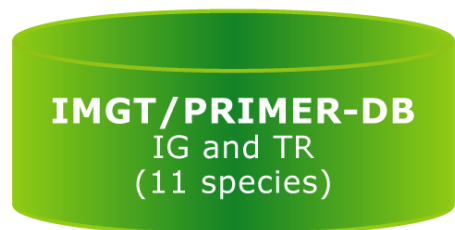
Laboratoire d'ImmunoGénétique Moléculaire
Université Montpellier, UPR CNRS 1142, IGH
Institut Universitaire de France

2006 Bioinformatics in Taiwan (BIT2006)
September 12-15, 2006

Sequences

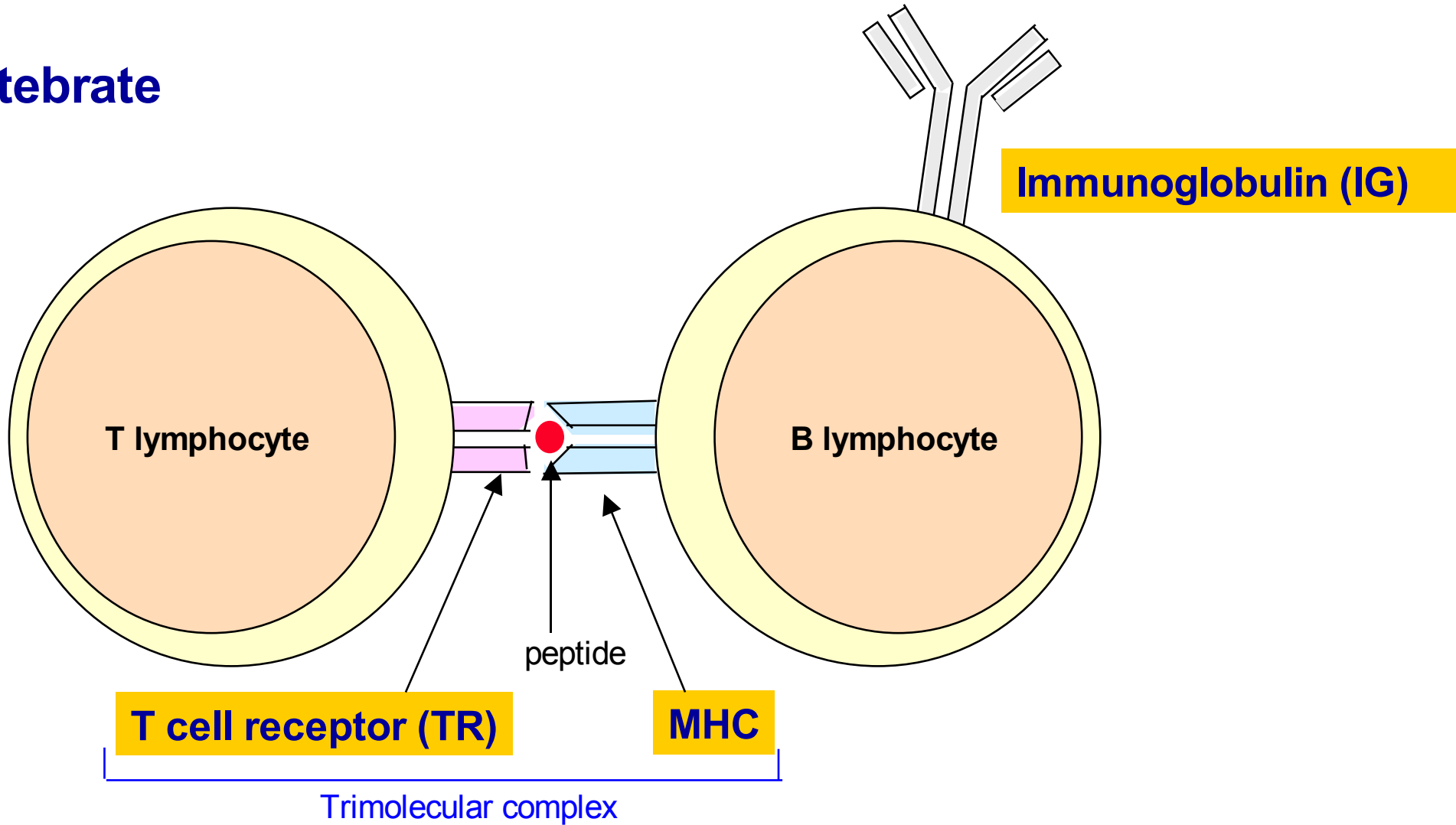


<http://imgt.cines.fr>



2D and 3D structures

Vertebrate S



When and why was IMGT created?

IMGT birth and objectives

IMGT was created in June 1989, following the Human Genome Meeting HGM, at New Haven, USA.

- * *Thomas Shows and Ken Kidd, meeting organizers*
- * *Lap Chee Tsui, chromosome 7 responsable*
- * *Claude Boucheix (CD nomenclature)*

1. At New Haven, entry of the 16 human T cell receptor gamma genes (TRG) in the genome database, Genome Databank (GDB).
2. IMGT was created to answer the need to manage the T cell receptor (TR) and immunoglobulin (IG) genes and to enter them in the generalist databases.

Why was IMGT necessary?

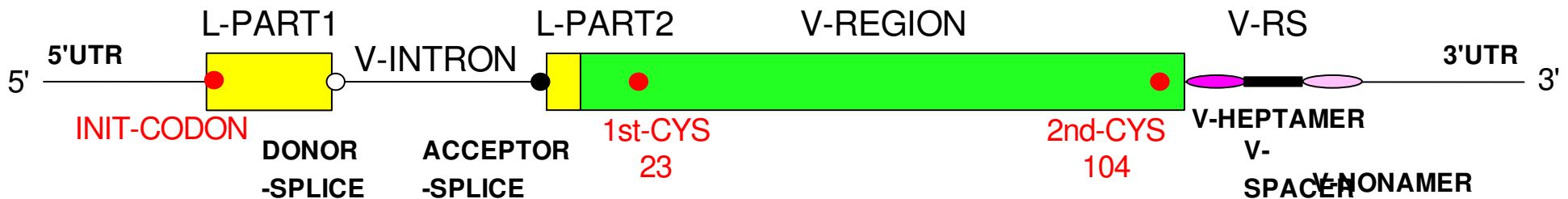
1. Unique structure of the immunoglobulin (IG) and T cell receptor (TR) genes.

Genomic DNA in germline configuration

V-GENE

>X62106.0|HSV12|*Homo sapiens* VI-2 gene for immunoglobulin heavy chain

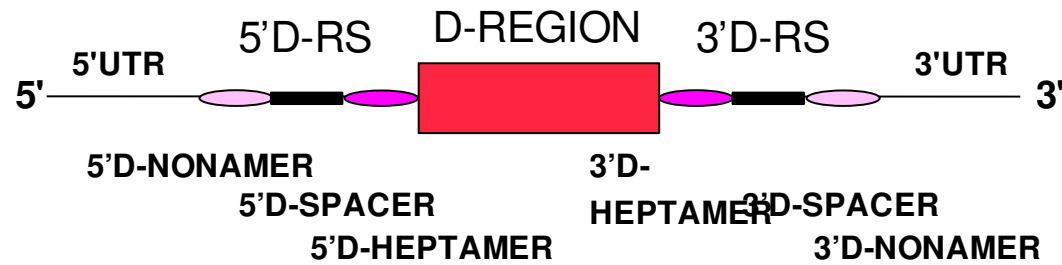
tgagagctcc	gttcctcacc	atggactgga	cctggaggat	cctcttcttg	gtggcagcag	60
ccacaaggtaa	gaggctccct	agtcccagtg	atgagaaaga	gattgagtc	agtccaggga	120
gatctcatcc	acttctgtgt	tctctcca	caaggccact	cccagggtgca	gctgggtgcag	180
tctggggctg	aggtgaagaa	gcctggggcc	tcagtgaagg	tctcctgcaa	ggcttctgga	240
tacaccttca	ccggctacta	tatgcactgg	gtgcgacagg	cccctggaca	agggcttgag	300
tggatgggat	ggatcaacc	taacagtgg	ggcacaact	atgcacagaa	gtttcagggc	360
agggtcacca	tgaccaggga	cacgtccatc	agcacagcct	acatgagct	gagagagctg	420
agatctgacg	acacggcctg	gtattactgt	gcgagagaca	cagtc	tgaaa	480
tgagggtg	tc agaaaccaa	gggaggaggc	ag		accacatcc	



Genomic DNA in germline configuration

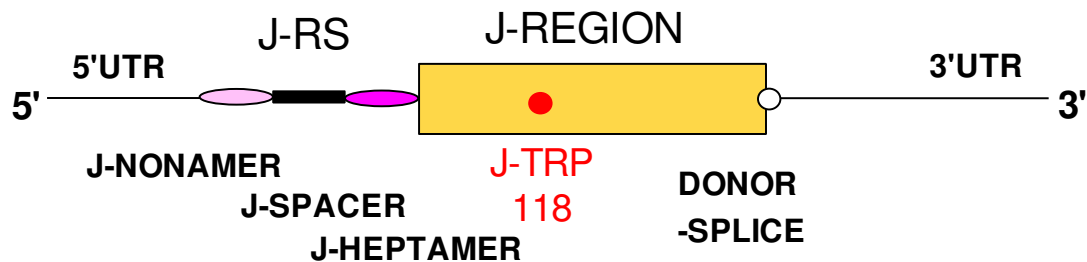
D-GENE

gggtttttgga c tgagctgaga ac cactgtgc taactgggga cacagtcatt
ggcagctct c caaaaacc



J-GENE

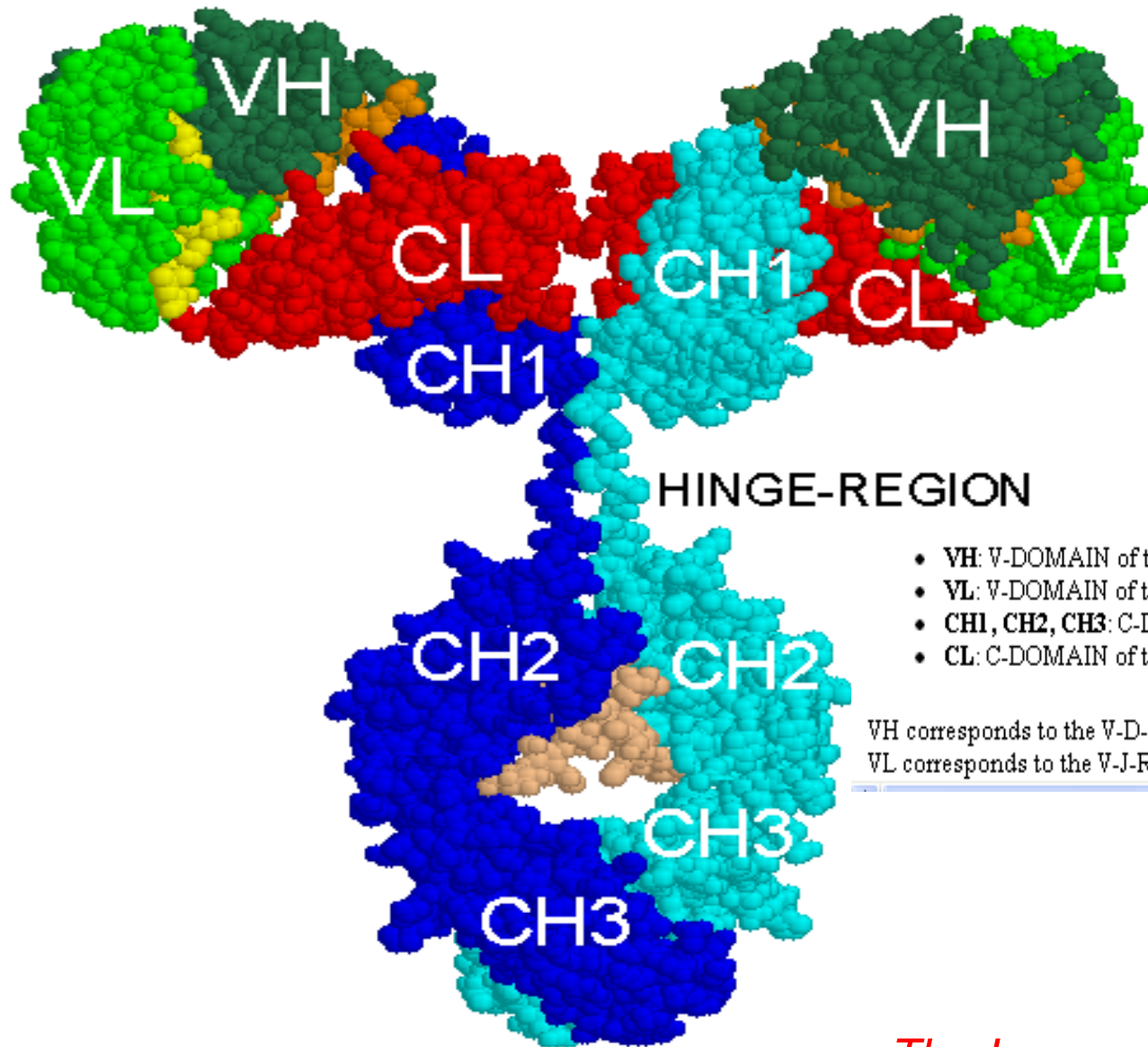
ggttttct gtgccctgg ctcagggtg actcaccgtg gctgaatact
ccagcactg ggccagggc acctgggtca cgtctcctc ag



Why was IMGT necessary?

2. Huge diversity of the immunoglobulin and T cell receptor repertoires.

Spacefill 3D representation of an IgG

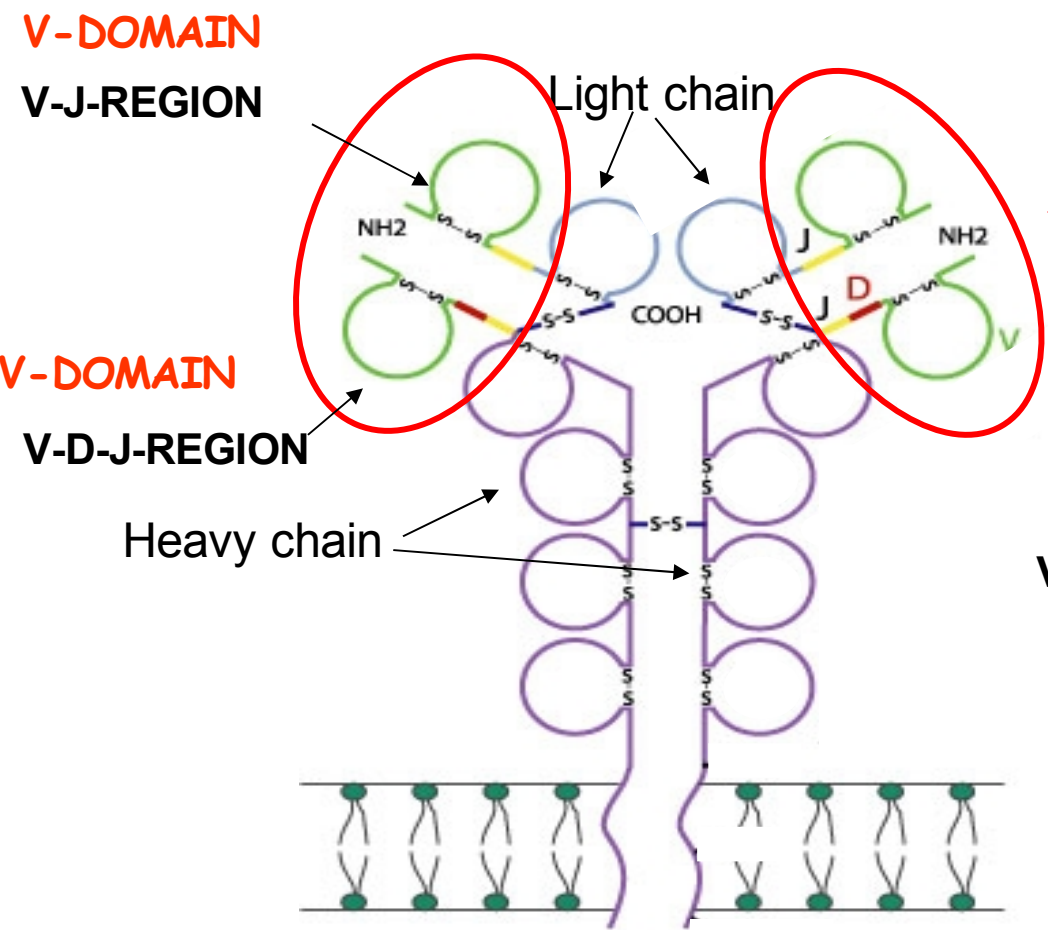


- **VH**: V-DOMAIN of the immunoglobulin heavy chain
- **VL**: V-DOMAIN of the immunoglobulin light chain
- **CH1, CH2, CH3**: C-DOMAIN of the immunoglobulin heavy chain
- **CL**: C-DOMAIN of the immunoglobulin light chain

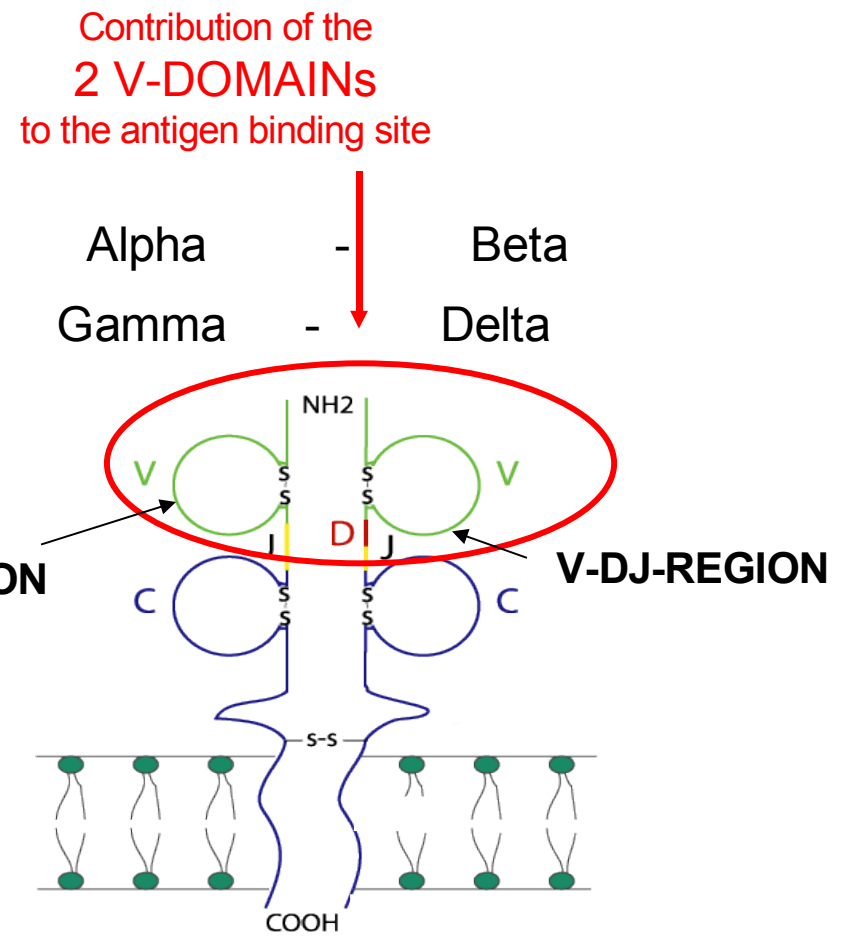
VH corresponds to the V-D-J-REGION (in green (V), orange (DJ)) of the heavy chain.
VL corresponds to the V-J-REGION (in green (V) and yellow (J)) of the light chain.

Immunoglobulin (IG)

T cell receptor (TR)



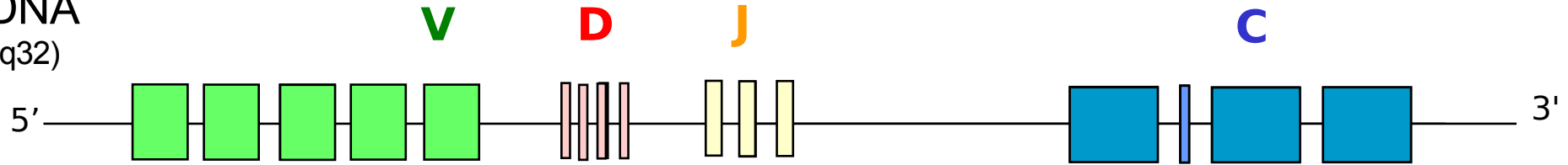
Membrane IgM



T cell receptor

Immunoglobulin (IG) synthesis

genomic DNA
(IGH Locus 14q32)

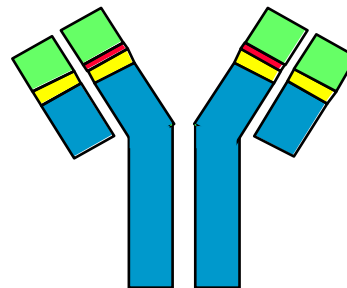


rearranged
DNA



mRNA

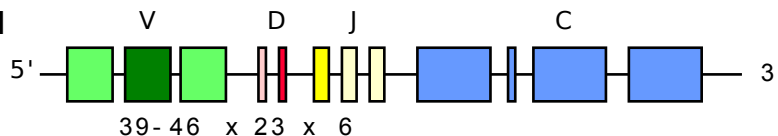
2×10^{12} different IG
per individual



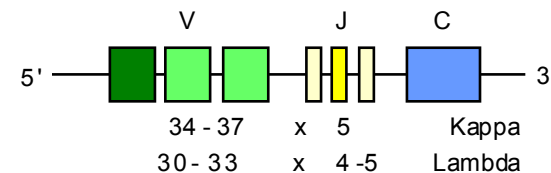
Immunoglobulin (IG) synthesis

150 FUNCTIONAL IG GENES

HEAVY CHAIN



LIGHT CHAIN



6300

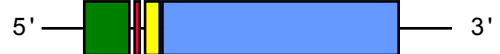
POTENTIAL RECOMBINATIONS

185 + 165

POTENTIAL RECOMBINATIONS

N-DIVERSITY
SOMATIC MUTATIONS

x 1000

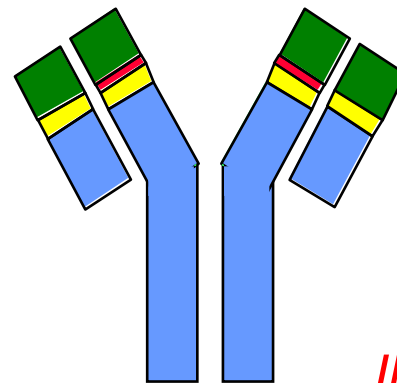


ABOUT 6.3×10^6 POSSIBILITIES

ABOUT 3.5×10^5 POSSIBILITIES

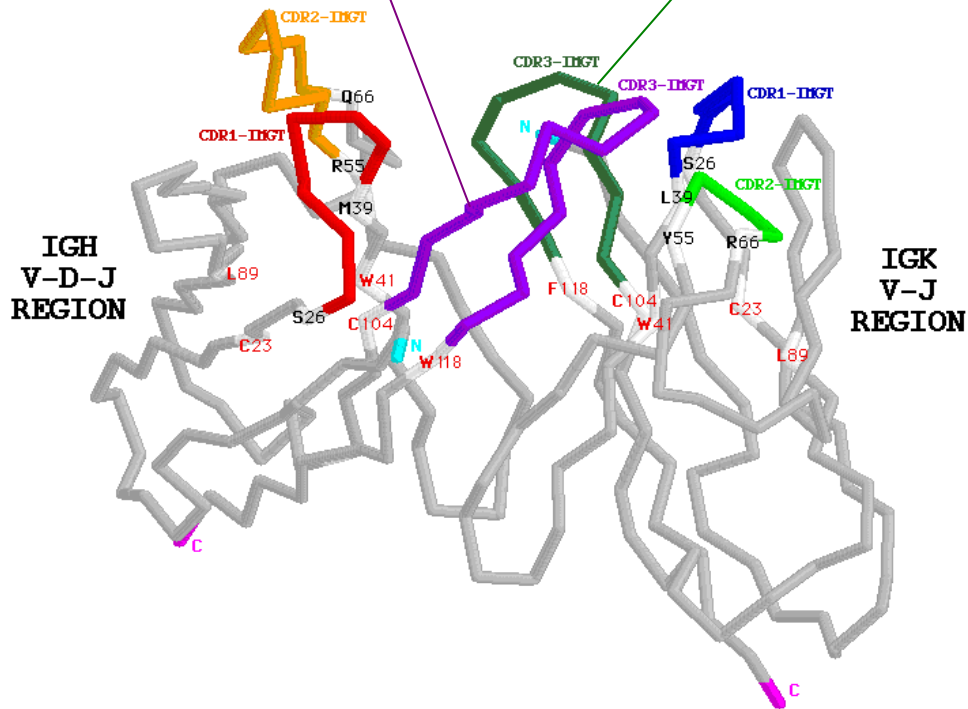
2×10^{12}

DIFFERENT ANTIBODIES



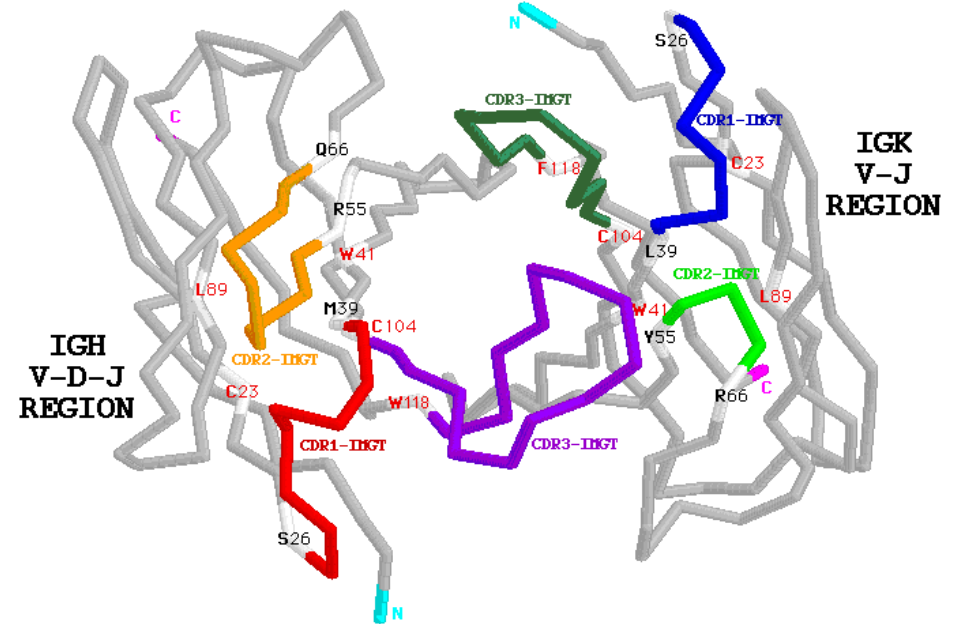
Junctions of the V-DOMAINS

VH V-D-J junction
V-KAPPA V-J junction



Side view

VH V-KAPPA



View from above

Mouse (*Mus musculus*) E5.2Fv

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

V-J junction (104-118)

V-D-J junction (104-118)

Generation of the JUNCTION diversity

3'V-REGION

N-REGION

D-REGION

N-REGION

5'J-REGION

tgtgcgaaaga



tacc



agcatattgtg

gtggtgactgctat

tcc



gat



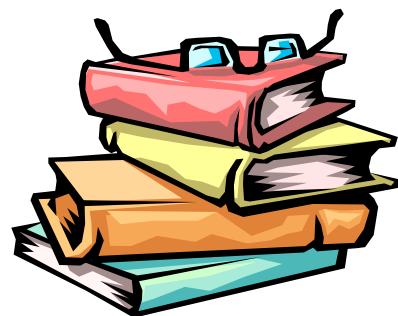
acaactgggtcg actcctgg

JUNCTION

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

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

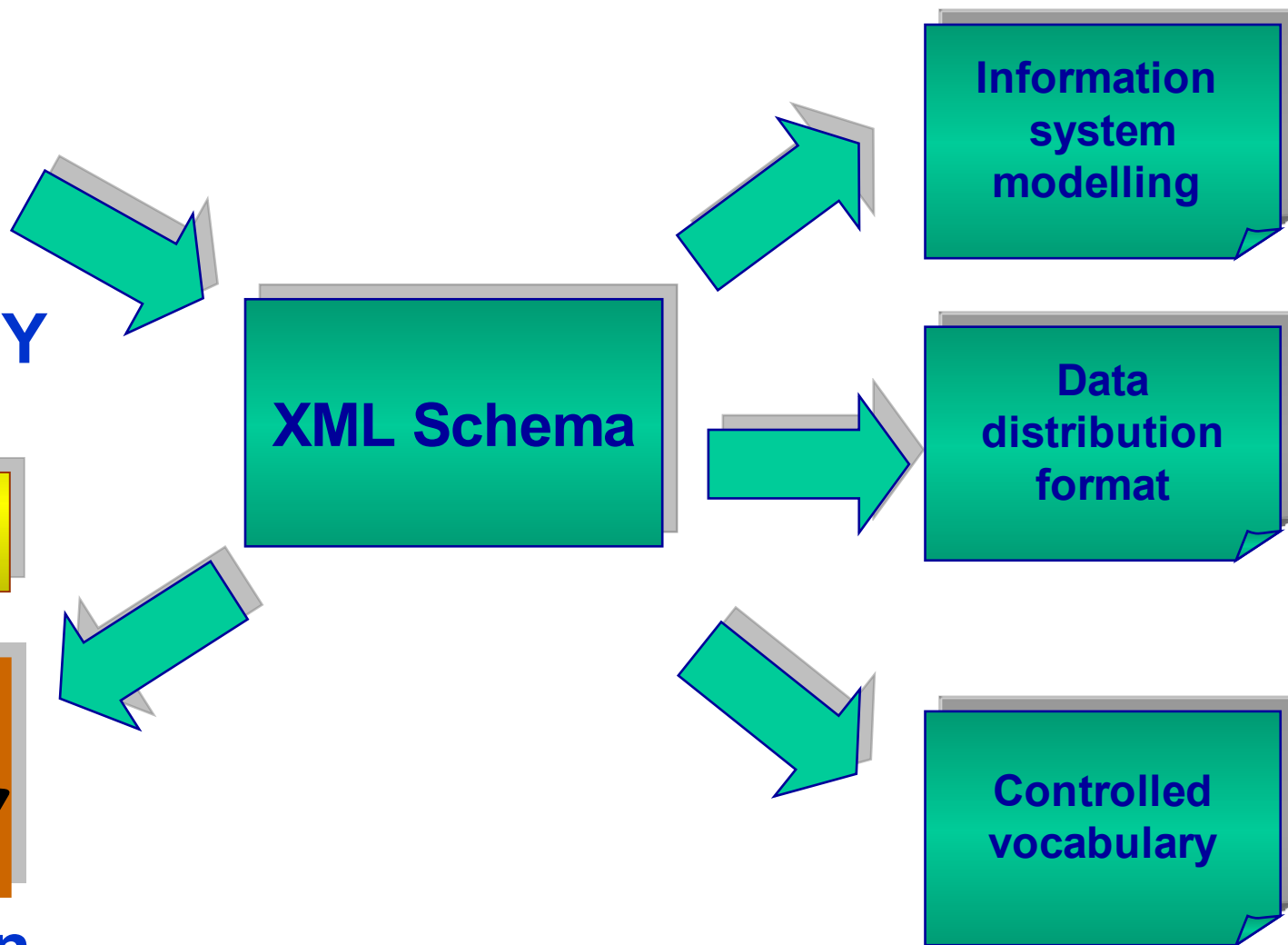
IMGT-ML architecture



IMGT-ONTOLOGY

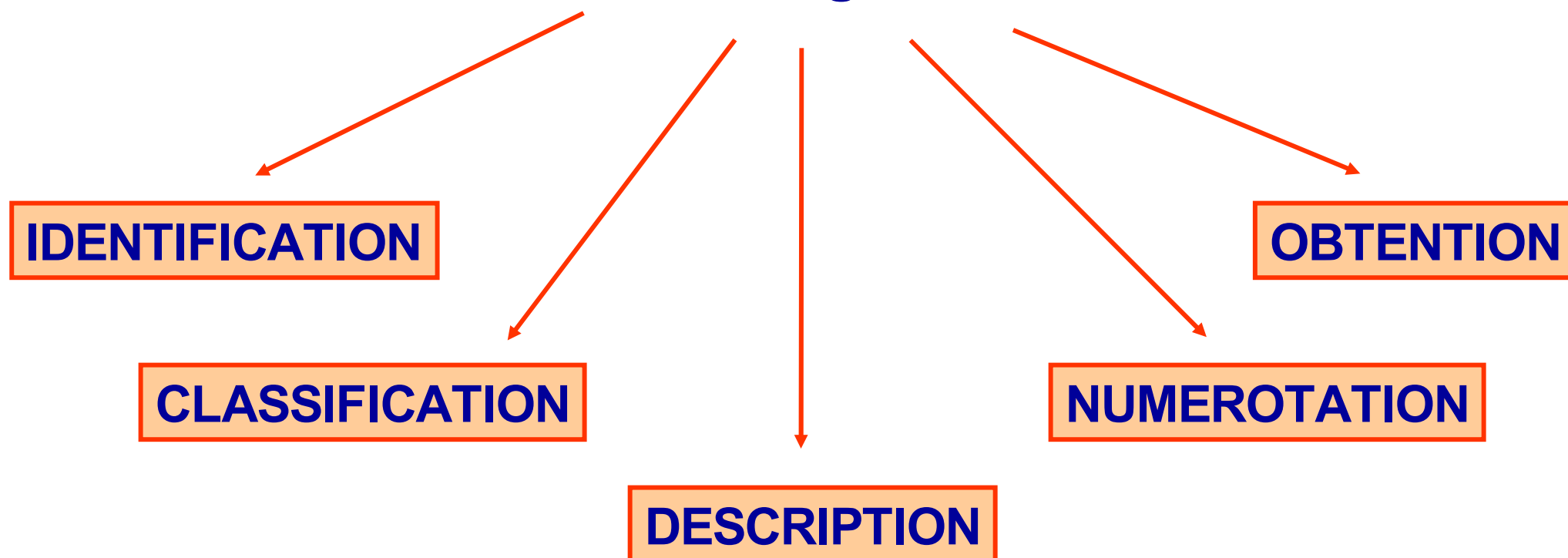


Documentation

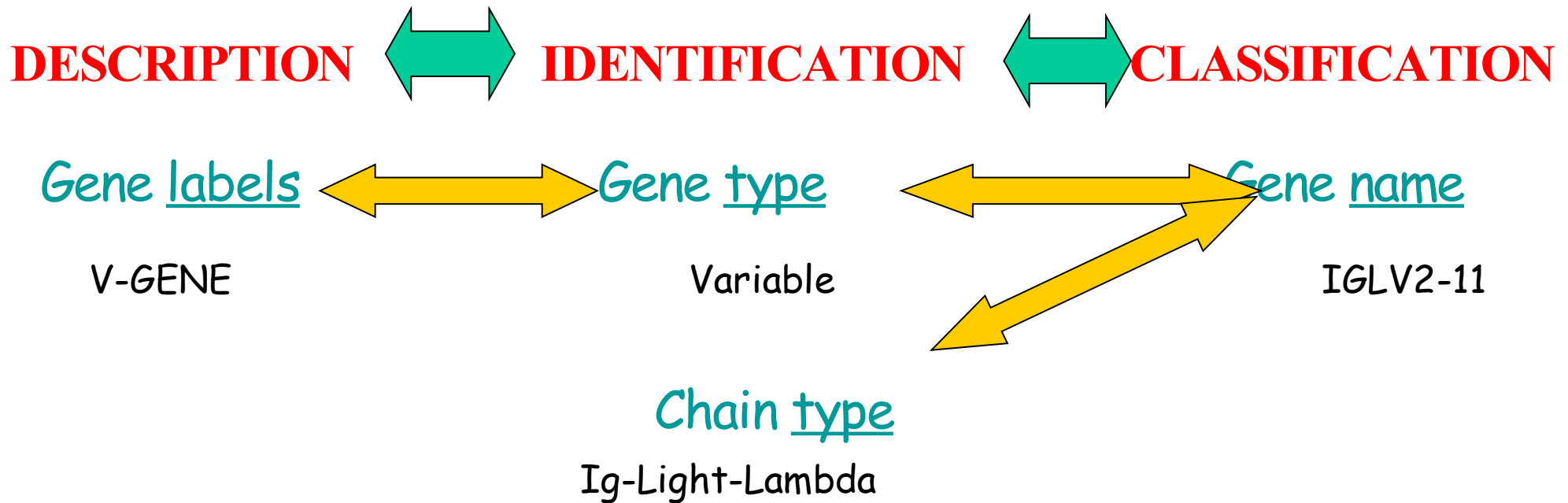


IMGT-ONTOLOGY five main concepts

to share, reuse and represent knowledge
in immunogenetics



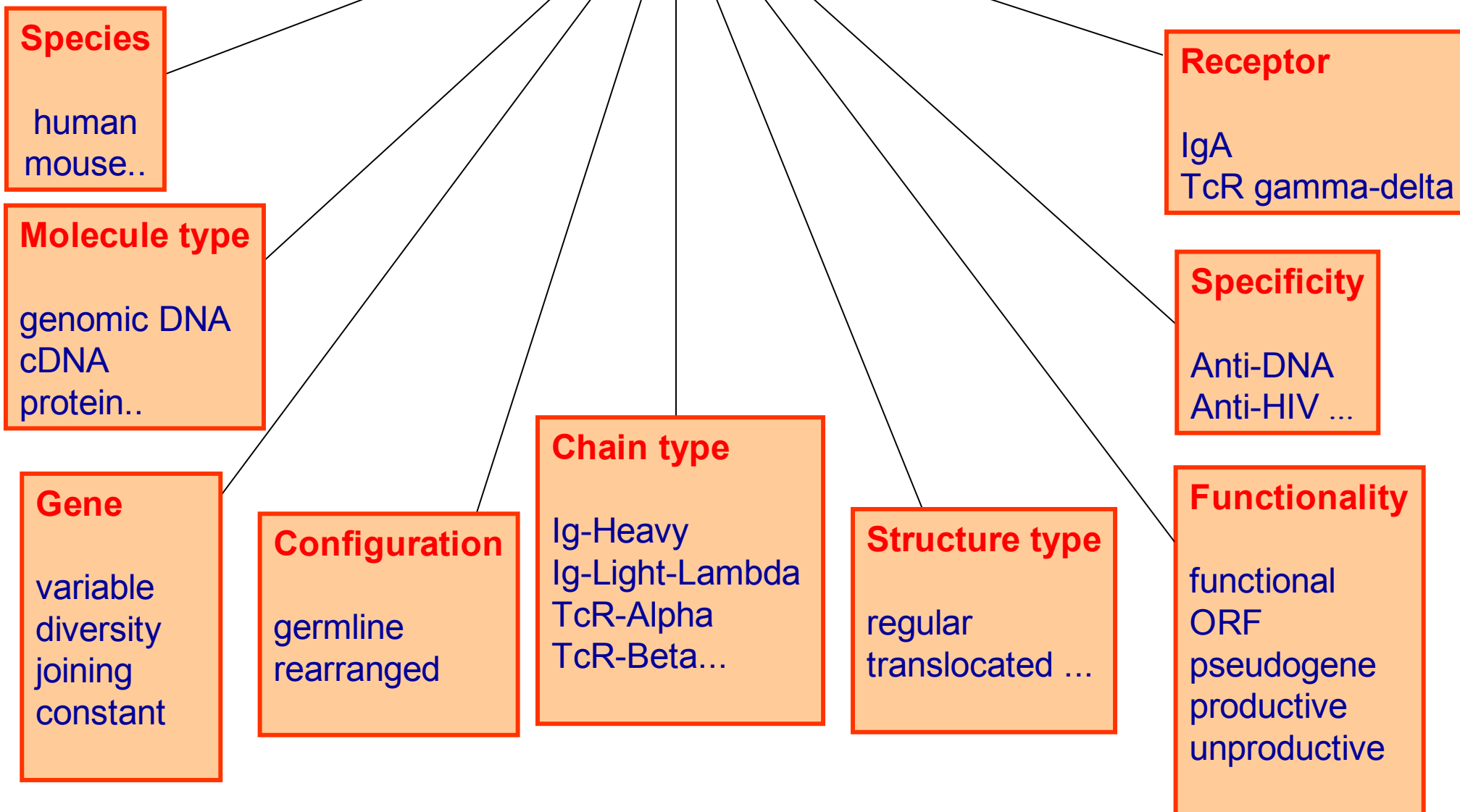
IMGT-ONTOLOGY concepts et relations



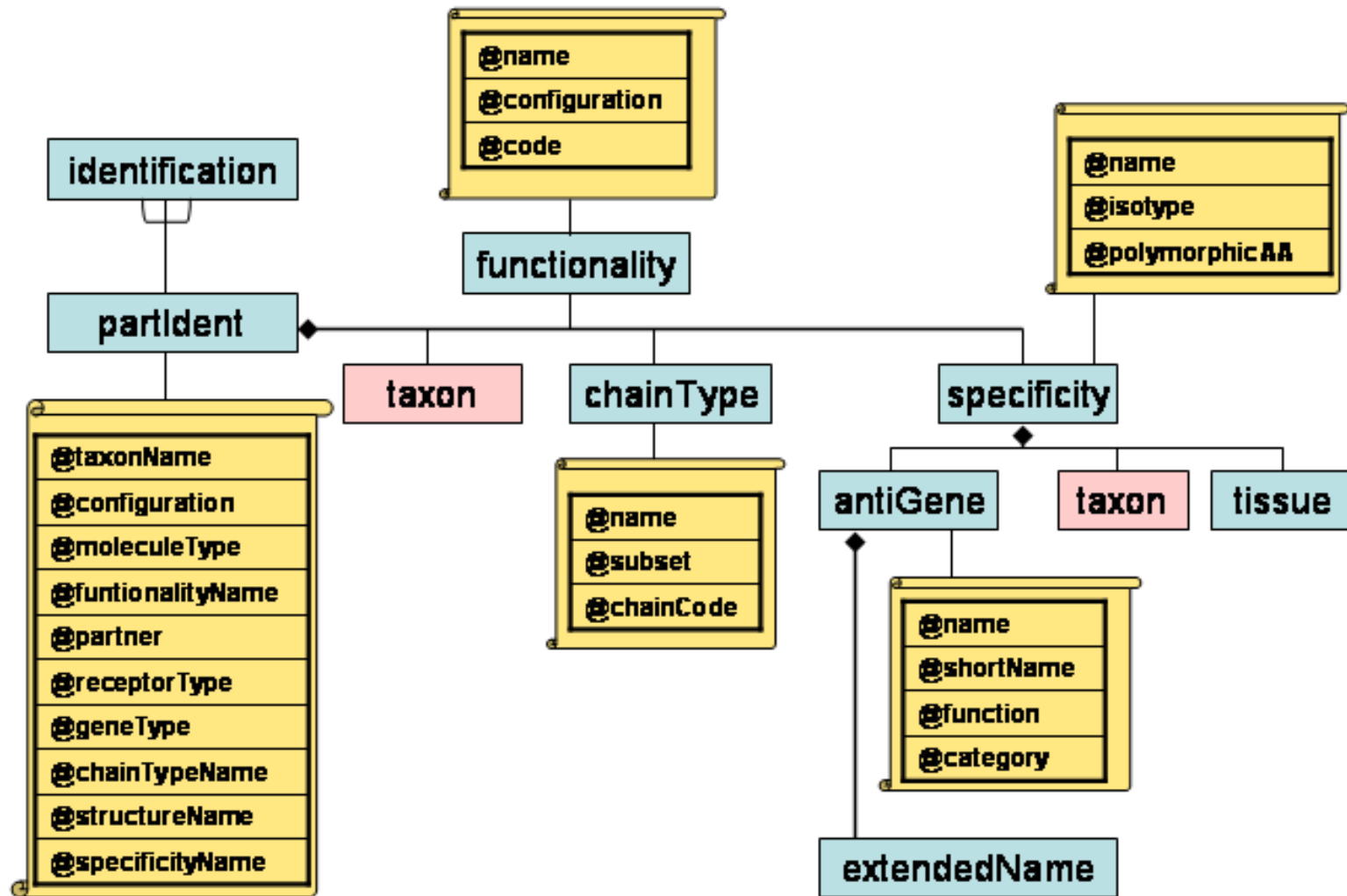
Problems we met:

- In many ontologies, no distinction between « concepts » and « instances »
- In GO (GeneOntology), only 2 types of relations: « is a », « is part of »

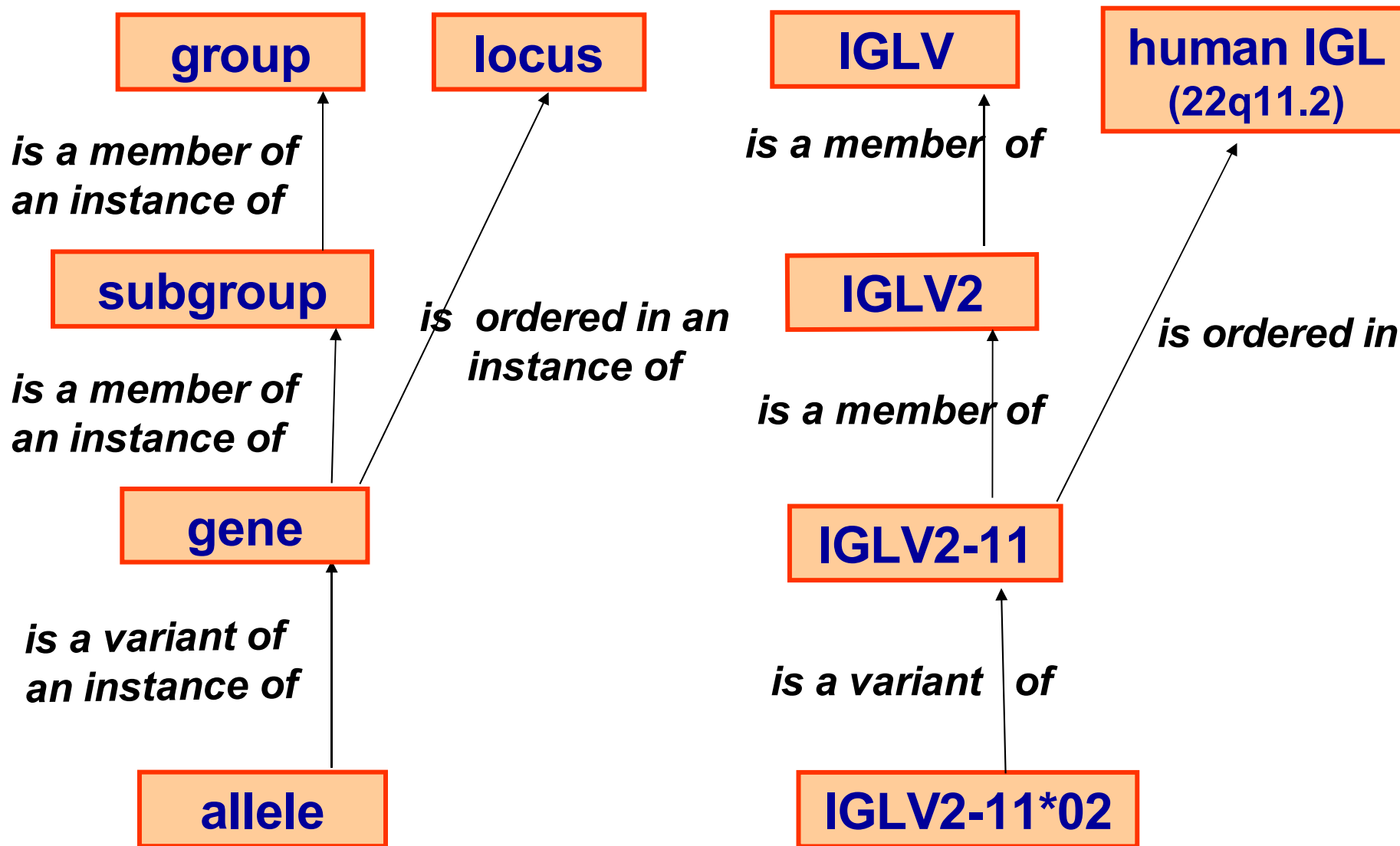
"IDENTIFICATION" concept



IMGT-IDENTIFICATION



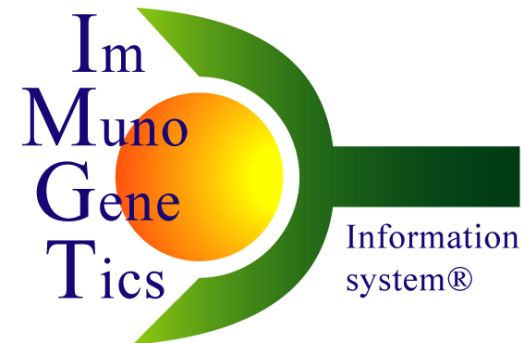
"CLASSIFICATION" concept



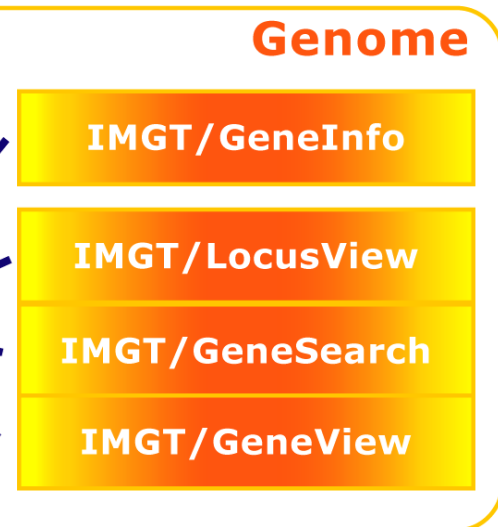
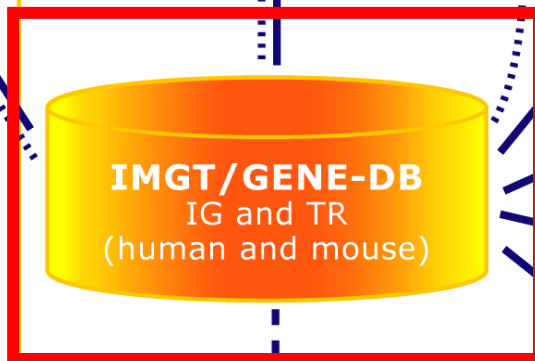
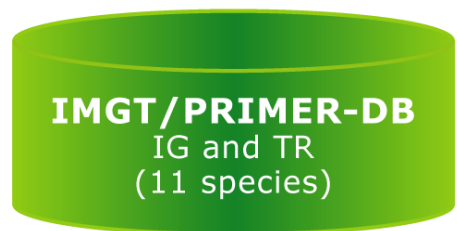
« Concepts »

« Instances »

Sequences



<http://imgt.cines.fr>



2D and 3D structures

Locus representation: Human IGL

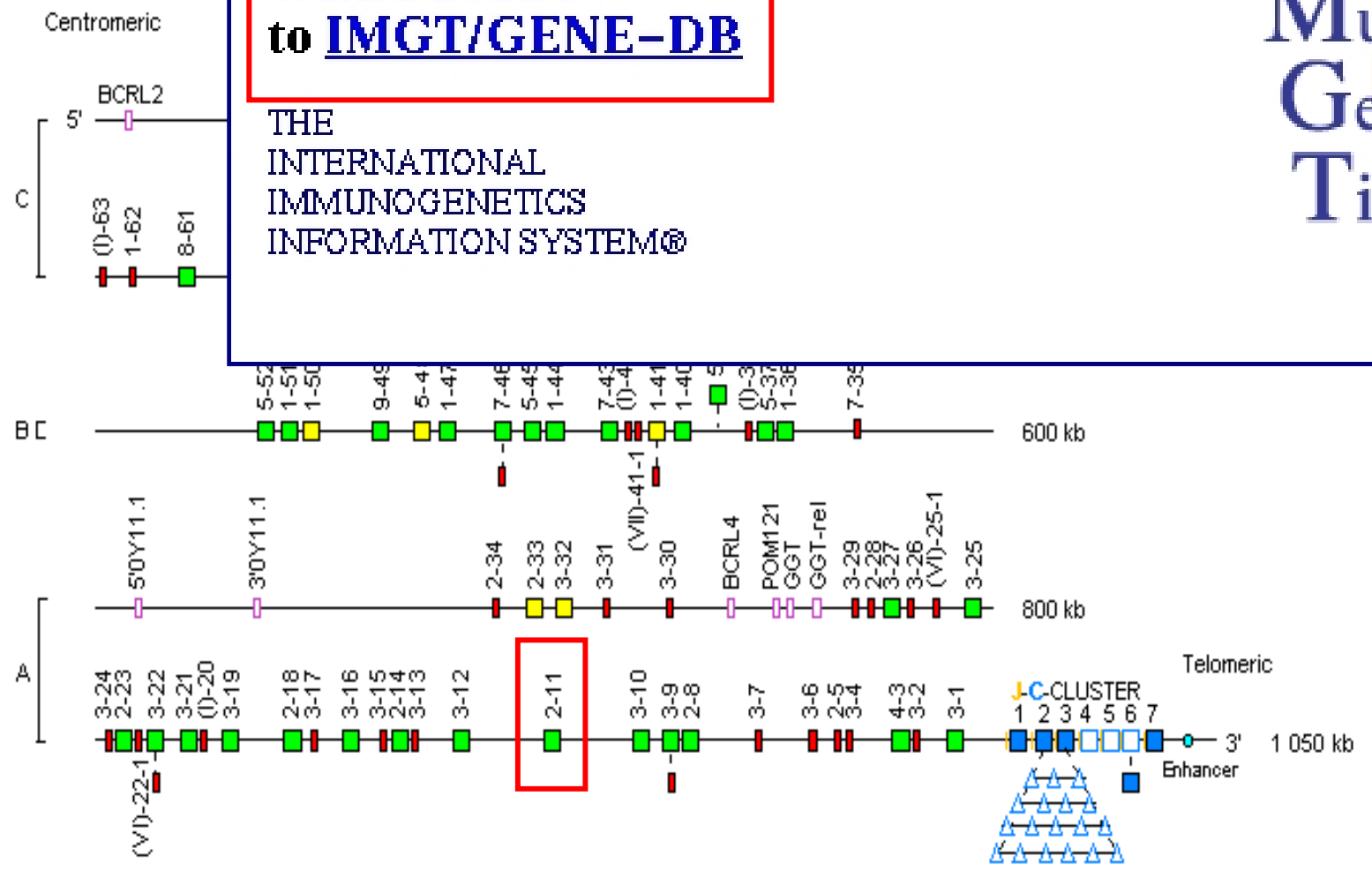
Human IGL 22q11.4

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

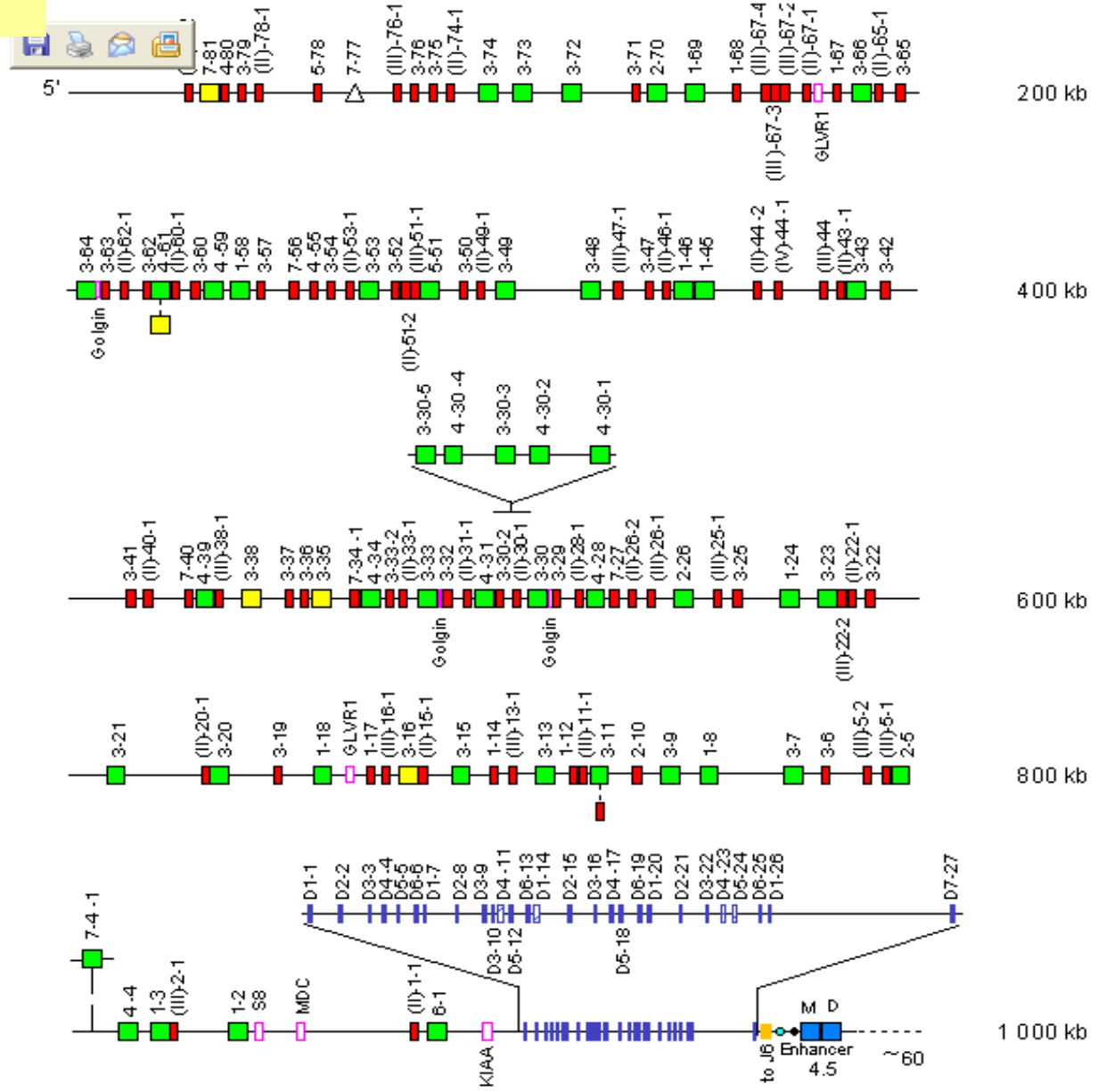
THE
INTERNATIONAL
IMMUNOGENETICS
INFORMATION SYSTEM®



<http://imgt.cines.fr>



Chromosome 14q32.33



1999 Entry of the 630 human IG and TR genes at NCBI

Cross-references between Entrez Gene and IMGT/GENE-DB

Gene - Mozilla Firefox

Entrez Gene

Search: Gene for [] Go Clear

Display: Full Report Show 5 Send to []

All: 1 Current Only: 1 Genes Genomes: 1 SNP GeneView: 1

1: TRDV1 T cell receptor delta variable 1 [*Homo sapiens*]
GeneID: 28518 Primary source: [HGNC:12262](#) updated 29-Aug-2006

Summary

Official Symbol: TRDV1 and Name: T cell receptor delta variable 1 provided by [HUGO Gene Nomenclature Committee](#)
See related: [IMGT/GENE-DB:TRDV1](#)

Gene name: TRDV1
Gene description: T cell receptor delta variable 1
RefSeq status: Reviewed
Organism: [Homo sapiens](#)
Lineage: *Eukaryota; Metazoa; Chordata; Cramiata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo*
Gene aliases: hDV101S1

Genomic regions, transcripts, and products

[RefSeq below](#)

NC_000014.7

Genomic context: chromosome: 14, Location: 14q11

Bibliography Gene References into Function (GeneRIF): [Submit](#)

[PubMed](#) links

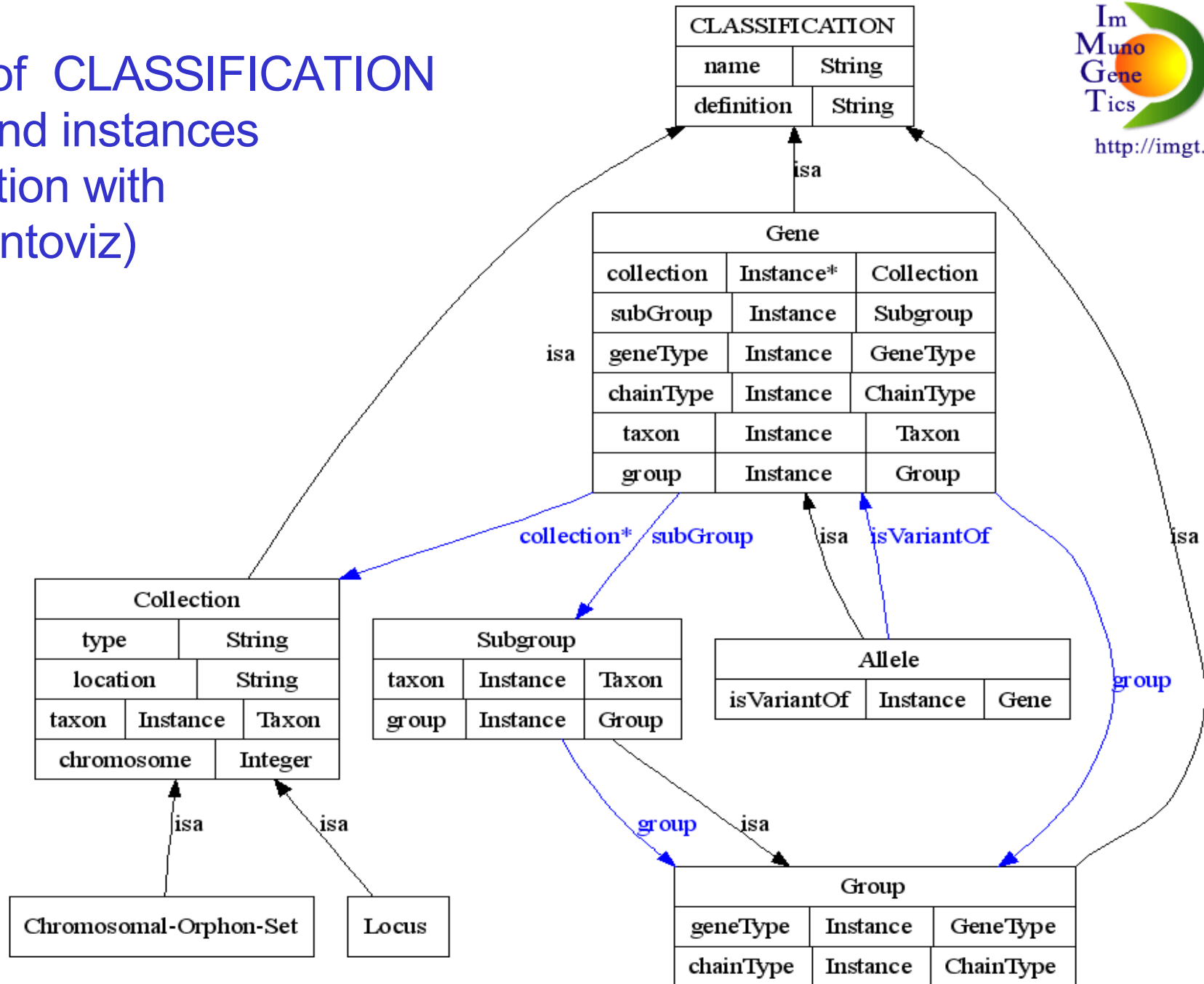
NCBI Reference Sequences (RefSeq)

Reference [NG_001332](#)

Related Sequences

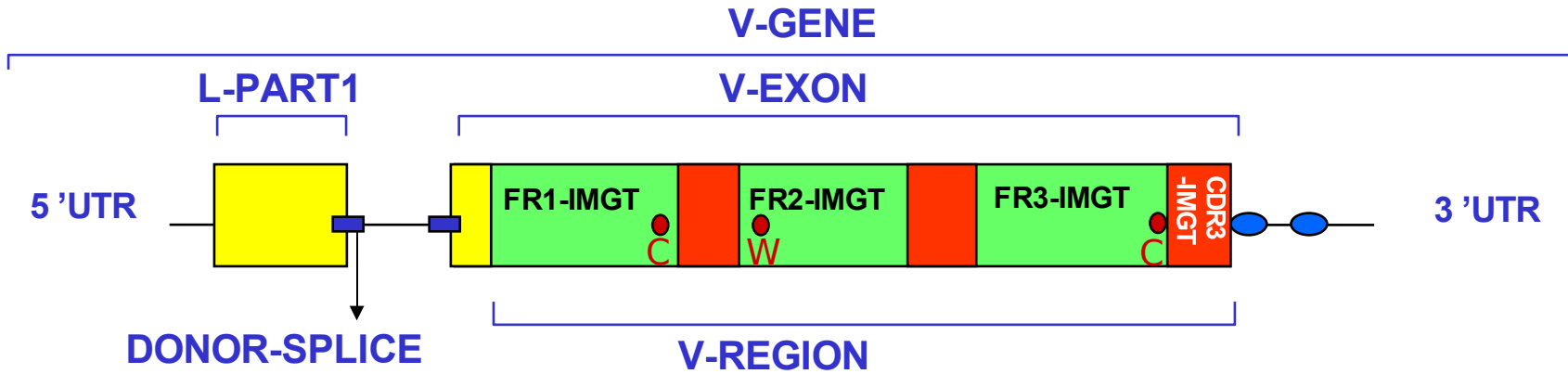
Nucleotide	Protein
Genomic AE000660 (100359..100928)	None
Genomic M22198	AAA61104
mRNA AF366516	AAK62930
mRNA U91153	AAB62513

Examples of CLASSIFICATION concepts and instances representation with Protégé (Ontoviz)



"DESCRIPTION" concept

PROTOTYPE for a V-GENE



Label 1

Label 2

Relations entre Labels

V-GENE

V-EXON



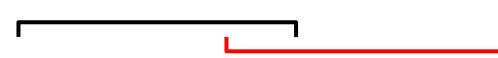
FR3-IMGT

CDR3-IMGT



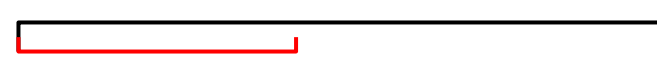
L-PART1

DONOR-SPLICE



V-REGION

FR1-IMGT

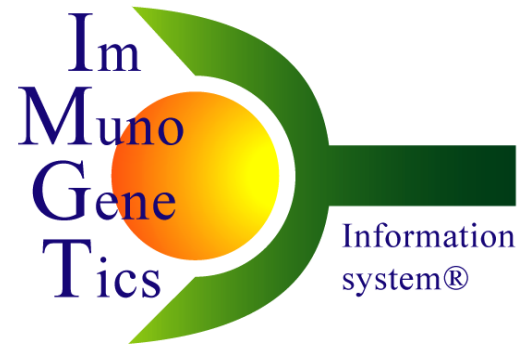
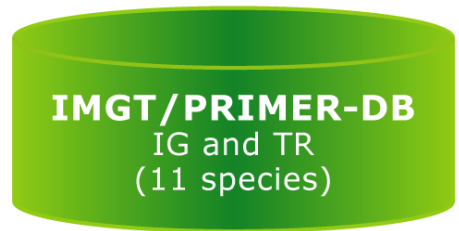


V-REGION

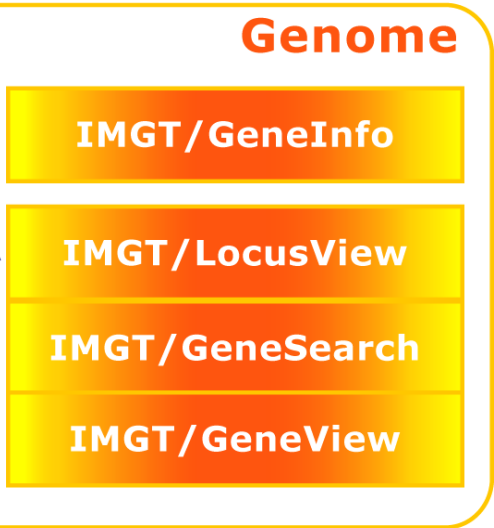
CDR3-IMGT



Sequences



<http://imgt.cines.fr>



2D and 3D structures

IMGT/LIGM-DB

DESCRIPTION

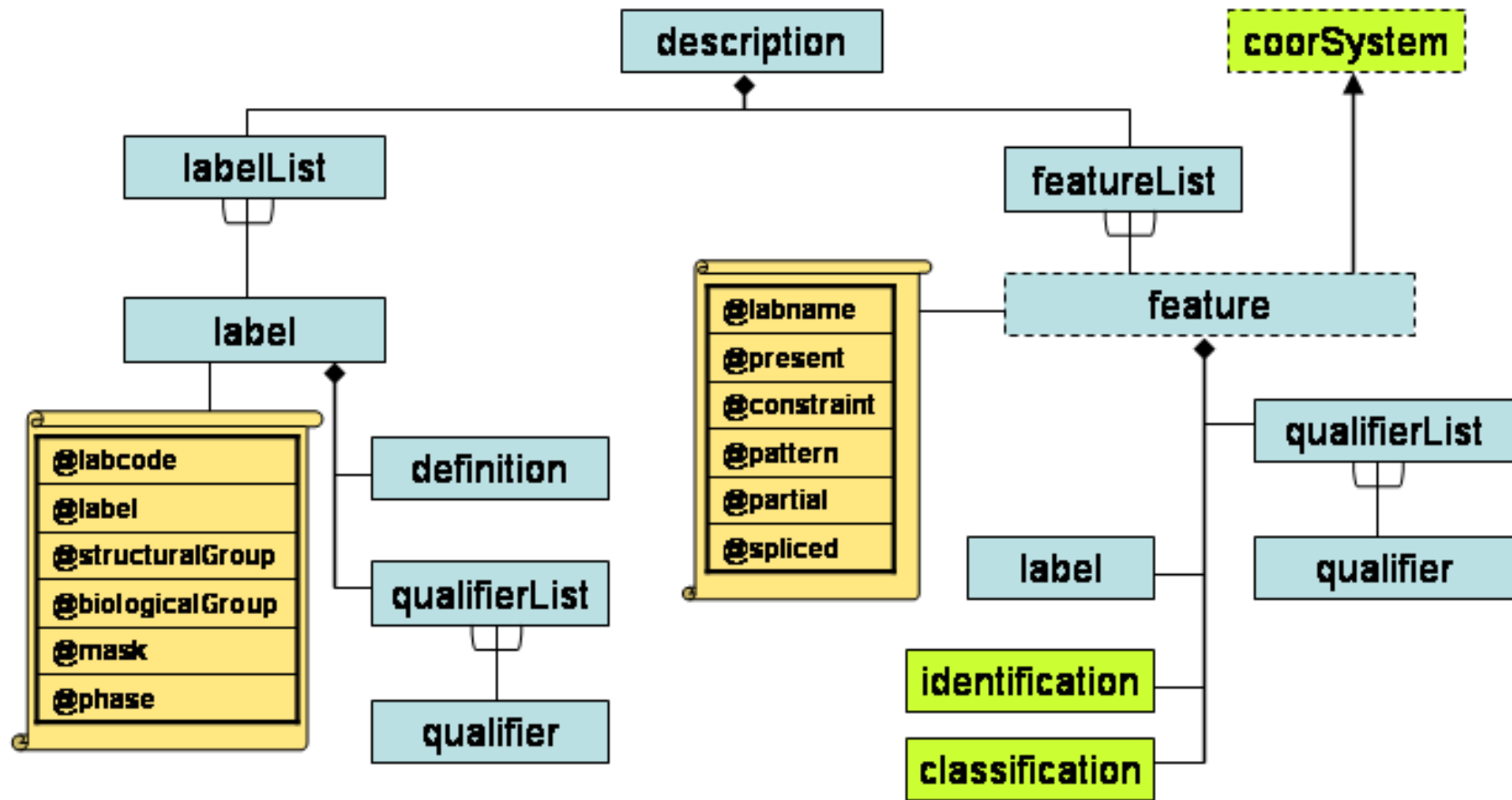
IMGT/LIGM-DB Consultation module v3 - Mozilla Firefox

```
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 /translation="QVHLVESGGAVFHPGRSLRLSRAASGFTFSSYGMHWVRQAP
FT AKGLEWVAVIWYDGSNKYYADSVKGRFTISRDN SKNTLYLQMN SLRAEDTAVYYC
FT AKHV TIAAAAGRRGAGMDVWGQGT TTVTVSS"
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 AKGLEWVAVIWYDGSNKYYADSVKGRFTISRDN SKNTLYLQMN SLRAEDTAVYYC
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="YYADSVKGRFTISRDN SKNTLYLQMN SLRAEDTAVYYC"
Done
```

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

SO (Sequence ontology):
67 IMGT labels

IMGT-DESCRIPTION



IMGT-ONTOLOGY in Protégé

Concepts of DESCRIPTION

CLASS BROWSER
For Project: ● IMGT-ONTOLOGY

Class Hierarchy

- DESCRIPTION
 - Entity
 - Cluster
 - Conserved_AA
 - Constant_exon
 - Core
 - FR_CDR
 - Intron_UTR
 - Junction
 - Peptide_signal
 - Splicing
 - Recombination-Signal
 - Nonamer
 - Spacer
 - Heptamer
 - Composed

CLASS EDITOR
For Class: ● DESCRIPTION (instance of IMGT-ONTOLOGY)

Name: DESCRIPTION

Documentation: The IMGT-ONTOLOGY 'DESCRIPTION' concept corresponds to the classification of terms and rules which are necessary to describe the organization and the components of the

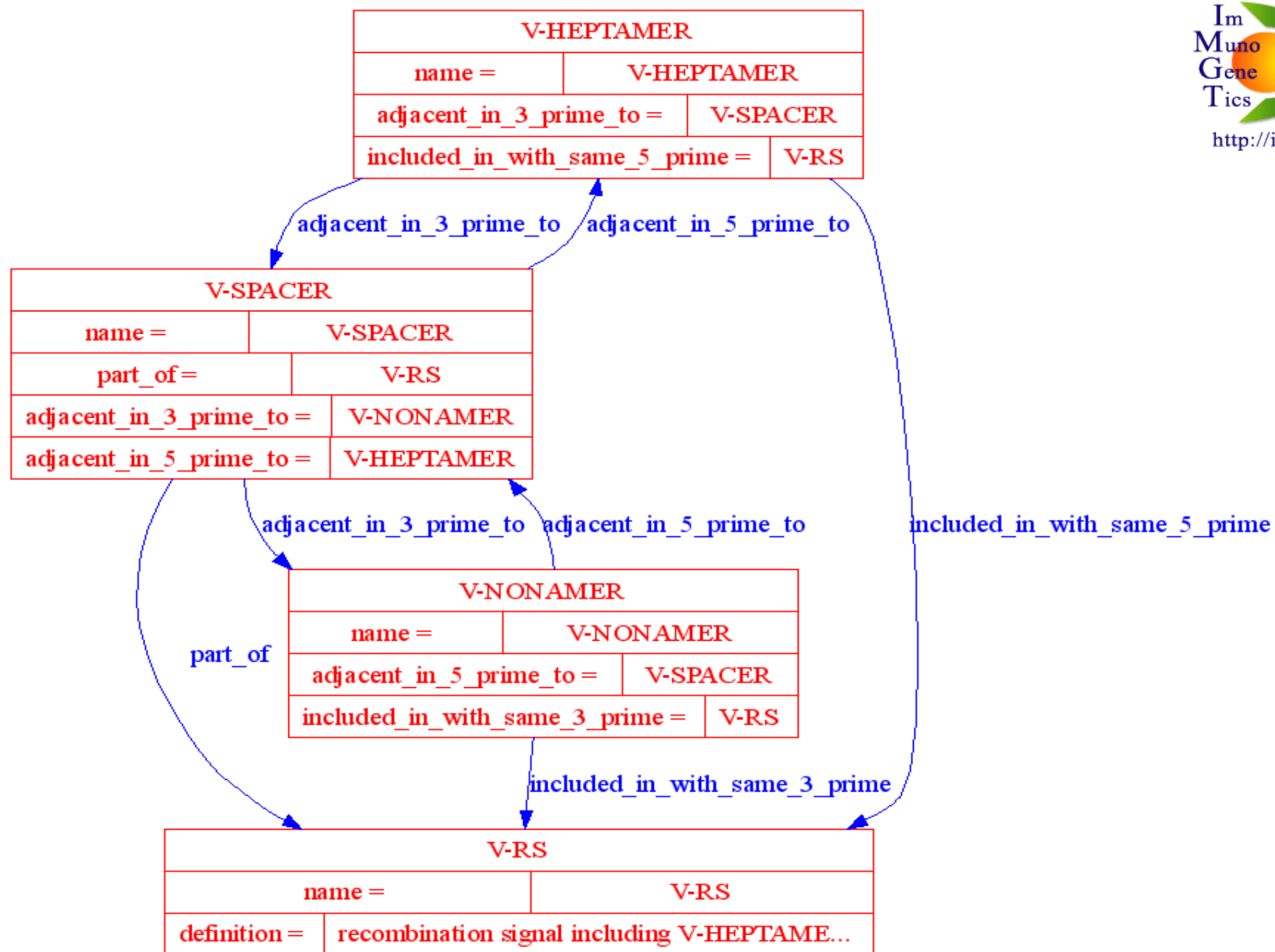
Role: Concrete ●

Constraints

Template Slots

Name	Cardinality	Type	Other Facets
adjacent_in_3_prime_to	multiple	Instance of DESCRIPTION	inverse-slot=adjacent_in_5_prime_to
adjacent_in_5_prime_to	multiple	Instance of DESCRIPTION	inverse-slot=adjacent_in_3_prime_to
adjacent_to	multiple	Instance of DESCRIPTION	
codingRegion	single	Boolean	default=false
definition	single	String	

Cui:
Broader:
Narrower:
Related Slot:
Synonym:
Semantic Type:

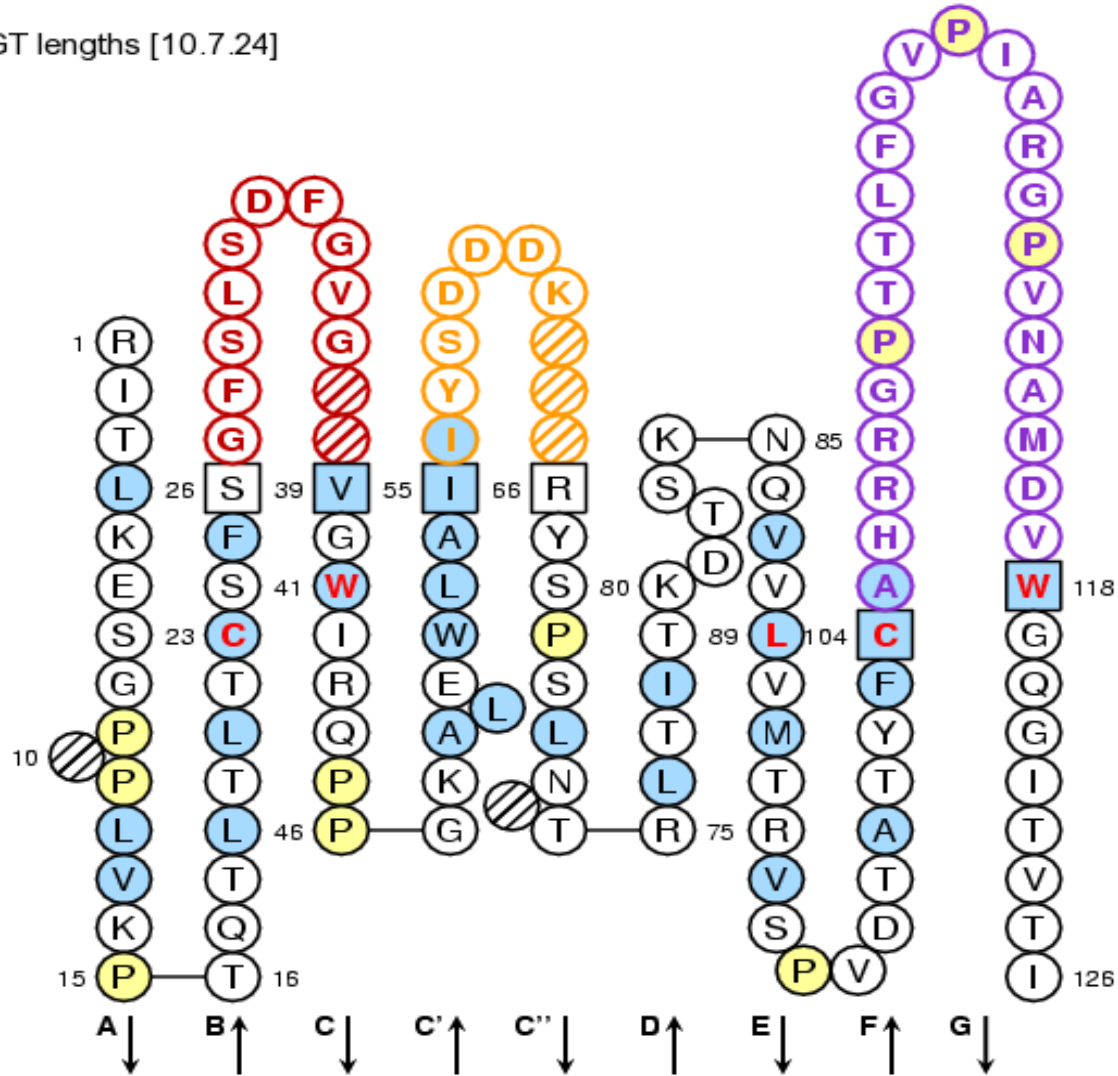


"NUMEROTATION" concept

IMGT Collier de Perles : *Homo sapiens* (Human) IGHV_1 V-DOMAIN from 2F5 (PDB: 1u8k_B)

CDR-IMGT lengths [10.7.24]

IMGT
Collier
de
Perles

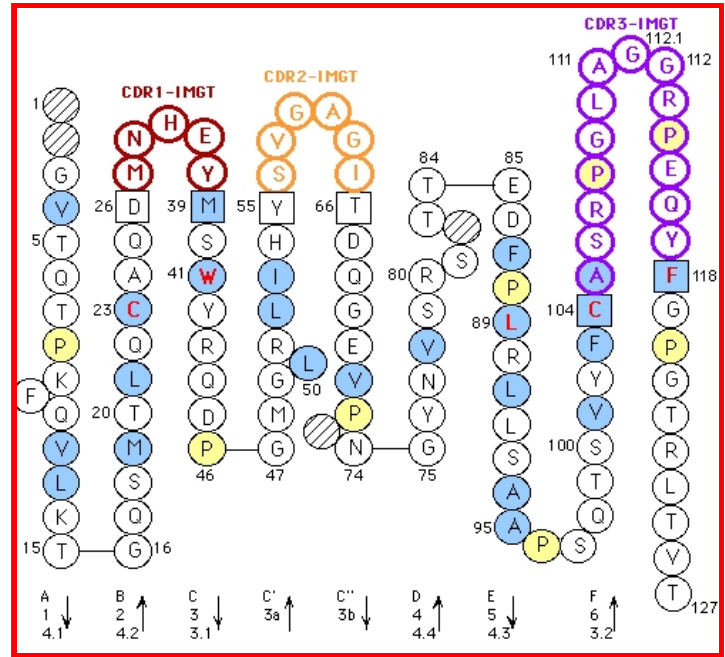


IMGT Web resources: 8 000 pages HTML



<http://imgt.cines.fr>

IMGT
Collier
de
Perles



Sequence alignment of alleles X02850, AE000659, and M86361 for TRAV8-6*01. The alignment shows the amino acid sequence and the corresponding DNA sequence for each region.

```

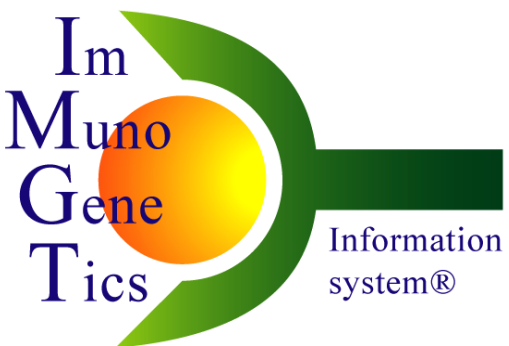
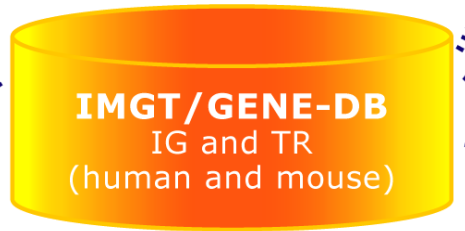
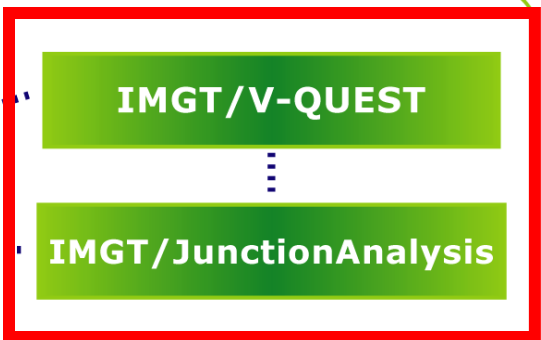
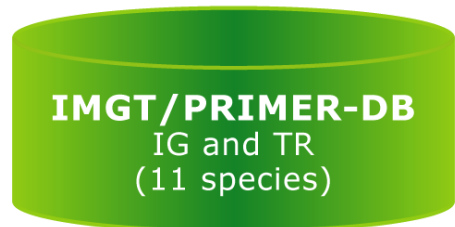
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
A 0 S V T 0 L D S 0 V P V F E E A P V E
GCC CAG TCT GTG ACC CAG CTT GAC AGC CAA GTC CCT TTT GAR GAR GGC CCT GTG GAG
-----
X02850 ,TRAV8-6*01
AE000659,TRAV8-6*02
M86361 ,TRAV8-6*02
-----
21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40
L R C N V S S V S V S V Y
CTG AGG TGC AAC TAC TCA TCG TCT GTT TCA GTG TAT ..... CTC TTC
-----
CDR1-IMGT
X02850 ,TRAV8-6*01
AE000659,TRAV8-6*02
M86361 ,TRAV8-6*02
-----
41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60
W Y V Q Y P N Q 6 L Q L L L K V L S G S
TGG TAT GTG CAA TAC CCC AAC CAA GGA CTC CAG CTT CTC CTG AAG TAT TTA TCA GGA TCC
-----
CDR2-
X02850 ,TRAV8-6*01
AE000659,TRAV8-6*02
M86361 ,TRAV8-6*02
-----
IMGT
61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80
T L L V E S I N
ACC CTG ..... GTT GAA AGC ATC AAC ..... GGT TTT GAG
-----
K G
R G
A G
-----
81 82 83 84 84A 84B 84c 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100
A E F H K S Q Q T S F H L R K P S V H I S D T A
GCT GAR TTT AAC AAG AGT CAA ACT TCC TTC CAC TTG AGG AAA CCC TCA CTC CAT ATA AGC GAC ACG GCT
-----
CDR3-IMGT
X02850 ,TRAV8-6*01
AE000659,TRAV8-6*02
M86361 ,TRAV8-6*02
-----
101 102 103 104 105 106 107 108
E Y F C A V S
GAG TAC TTC TGT GCT GTG AGT GA
-----
R
-----
#*
    
```

IMGT
Alignment
of alleles

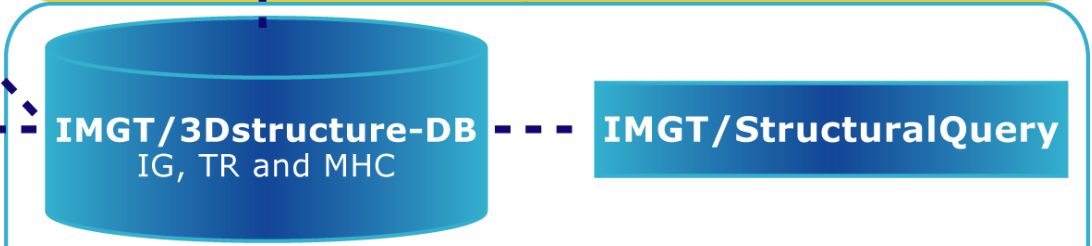
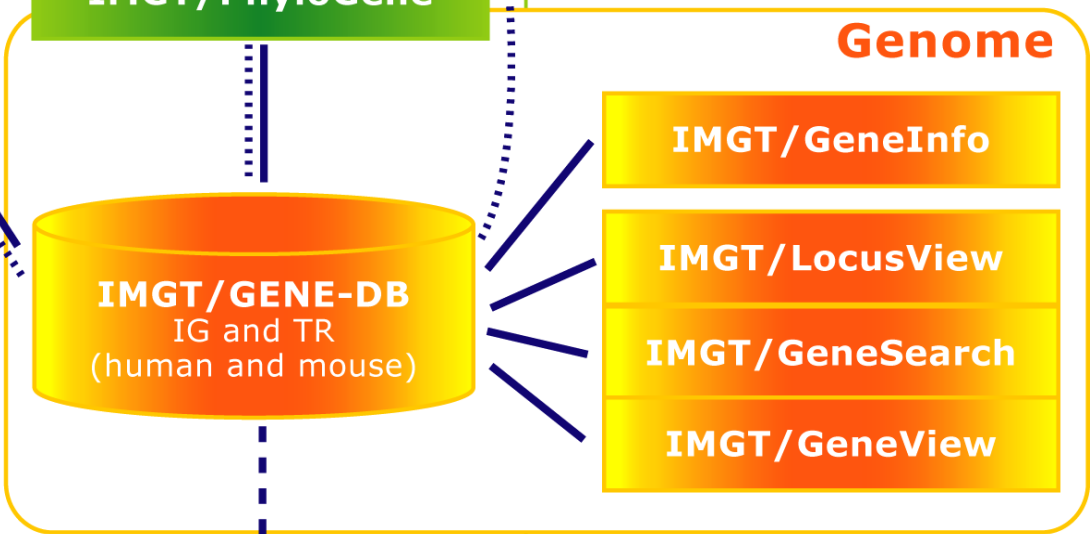
IMGT
Protein
Display

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.....	LSWYQQHGGAPITFLSY	NALDQ.....	LEETG.....	RFSSFLSRSDSYGYLLLQELQMKDSASYFC
AE000658, TRAV1-2	GQNIDQ.PIEMTATEGAIQINCTYQ	TSGFNG.....	LFWYQQHAGEAPITFLSY	NVLDD.....	LEEKQ.....	RFSSFLSRSGYSYLLLKELQMKDSASYLC
AE000658, TRAV2	KDQVFQ.PSTVASSEGAIVVEIFCNHS	VSNAYN.....	FFWYLFHPGCAPIRLLVK	GSK.....	PSQQG.....	RYNMTYER.FSSSLLILQVREDAAVYYC
AE000658, TRAV3	AQSVAQPEDQVNVVAEGNPLTVKCTYS	VSGNPY.....	LFWYVQYPNRLQLFLK	YITGDNL.....	VKGSY.....	GFEAEFNKSQTSFHLKPKSALVSDSALYFC
AE000658, TRAV4	LAKITQ.PISMDSYEGQEVNITCSHN	NIATNDY.....	ITWYQQFSPQGRFIIQ	GYKT.....	KVINE.....	VASLFIIPADRKSFTLSLPRVSLSDTAVYYC
AE000659, TRAV5	GEDVEQS.LFLSVREGDSSVINCTYT	DSSSTY.....	LYWYKQEPGAGLQLLTY	IFSNMD.....	MKQDQ.....	RLTIVLLNKKDKHLSLRIADTQTGDSAIYFC
AE000659, TRAV6	SQKIEQNSALNIQEGKATLTCNYT	NYSPAY.....	LQWYRQDPGRGPVFLLL	IRENEK.....	EKRKE.....	RLKVIPTDITLQKSLFHITASQPADSATYLC
AE000659, TRAV7	ENQVEHSPHFLGPQQGDVASMCTYS	VSRFNN.....	LQWYRQNTGMGPKHLLS	MYSAGY.....	EKQKG.....	RLNATLLK.NGSSLYITAVQPEDSATYFC
AE000659, TRAV8-1	AQSVSQHNHHVILSEASLELGCNYS	YGGTVM.....	LFWYVQYPGQHLQLLLK	YFSGDPL.....	VKGIK.....	GFEAEFIKSKFSPNLRKPSVQVSDTAEYFC
AE000659, TRAV8-2	AQSVTQLDSHVSVSEGTVPVLLRCNYS	SSYSPS.....	LFWYVQHPNKGQLQLLLK	YTSAAITL.....	VKGIN.....	GFEAEFKKSETSPHLTKPSAHMSDAAEYFC
AE000659, TRAV8-3	AQSVTPQPDIIHITVSEGALELRCNYS	YGATPY.....	LFWYVQSPGQGLQLLLK	YFSGDITL.....	VQGIK.....	GFEAEFKRSQTSFNLKPSVHWSDAAEYFC
AE000659, TRAV8-4	AQSVTQLGSHVSVSEGALVLLRCNYS	SSVPPY.....	LFWYVQYPNQLQLLLK	YTSAAITL.....	VKGIN.....	GFEAEFKKSETSPHLTKPSAHMSDAAEYFC
X02850, TRAV8-6	AQSVTQLDSQVPVFEAAPVELRCNYS	SSVSFY.....	LFWYVQYPNQLQLLLK	YLSGSL.....	VESIN.....	GFEAEFNKSQTSFHLKPSVHI SDTAEYFC
AE000660, TRAV8-7	TQSVTQLDGHITVSEEAPELRCNYS	YSGVPS.....	LFWYVQYSSQSLQLLLK	DLTEATQ.....	VKGRV.....	GFEAEFKKSETSPYLRKPSHVSDAEYFC
AE000659, TRAV9-1	GDSSVQTEGQVLPSEGDSLIVNCSYF	TTQVPS.....	LFWYVQVPPGEGPOLH.K	AMKAND.....	KGRNK.....	GFEAMVYRKFSTSPHLEKDSVQFSDSAVYFC

Sequences



<http://imgt.cines.fr>



2D and 3D structures

IMGT/V-QUEST

Alignment for V-GENE

Accession	Gene	Score	Sequence
AF402940			score GTGCAGCTGCTCGAGCAGTCTGGGGCT__GAGGTGAGCAAGCCTGGGGCCTCAGTAAAGGTTTCCTGCA
X62109	IGHV1-3*01	1146	CA.GTC.A...T.T.....AG.....G.....
X62107	IGHV1-3*02	1110	CA.GTT.A...G.T.....AG.....G.....
M99637	IGHV1-8*01	957	CA.GT..A...G.T.....AG.....G...C.....
L06612	IGHV1-46*03	948	CA.GT..A...G.T.....AG.....G.....
X92343	IGHV1-46*01	948	CA.GT..A...G.T.....AG.....G.....

Alignment for J-GENE

Accession	Gene	Score	Sequence
AF402940			score CTTACGGGGCGGGACGCTTTGGACGTCTGGGGCCAAGGGACCACGGTCACCGTCTCCTCA
J00256	IGHJ3*01	181T.....T..T.....A.T.....T...G
X86355	IGHJ6*02	179	T.A.TACTACTACT...G.A.....
X86355	IGHJ3*02	172T.....T..TA.....A.T.....T...G



V-GENE

JUNCTION

J-GENE

Translation

```

<----- F R 1 - I M G T -----
1           5           10           15
input      E V Q L L E S G G G L V Q P G G S
M99660 IGHV3-23*01 GAG GTG CAG CTG TTG GAG TCG GGG GGA ... GGC TTG GTA CAG CCT GGG GGG TCC

----->
                CDR1 - IMGT
                20           25           30           35
input      L R L S C A A S G F T F S S F A
M99660 IGHV3-23*01 CTG AGA CTC TCC TGT GCA GCC TCT GGA TTC ACC TTT AGC AGC TTT GCC ... Y
                Y
                -A- ...

<----- F R 2 - I M G T -----
                40           45           50
input      M S W V R Q A P G K G L D W V S
M99660 IGHV3-23*01 ... ATG AGC TGG GTC CGC CAG GCT CCT GGG AAG GGG CTG GAC TGG GTC TCA
                E
                -G-

-->                CDR2 - IMGT
                55           60           65           70
input      E I S G S G G T T Y Y A G S V K
M99660 IGHV3-23*01 GAA ATT AGT GGT AGT GGC GGT ACC ACA ... TAT TAC GCA GGC TCC GTG AAG
                A S D
                -CT -T -G- ... -C -A-
    
```

IMGT/V-QUEST - Mozilla Firefox

Fichier Edition Affichage Aller à Marque-pages Outils ?

http://imgt.cines.fr/cgi-bin/IMGTdnap.jv

Collier de Perles

input

Legend:

A	B	C	C'	C''	D	E	F
1	2	3	3a	3b	4	5	6
4.1	4.2	3.1			4.4	4.3	3.2

Applet input [1]

IMGT/JunctionAnalysis

← → ↻ × 🏠 <http://imgt.cines.fr/cgi-bin/IMGTjcta.jv> OK

THANK YOU
for using **IMGT/JunctionAnalysis**

THE
INTERNATIONAL
IMMUNOGENETICS
INFORMATION SYSTEM®



<http://imgt.cines.fr>

Analysis of the JUNCTIONS

Input	V name	V-REGION	N1	D-REGION	N2	J-REGION	J name	
#1	M62724	IGHV7-4-1*02	tgtgcgagaga	aga	.tagcaatggctacaa....	aatatttgactactgg	IGHJ4*02
#2	Z47269	IGHV1-69*06	tgtgcgaga	gggggggctaaggtcgaatttttgagtggtt.....	tcatgggt	...actggttcgaccctgg	IGHJ5*02

Input	D name	Vmut	Dmut	Jmut	Ngc	
#1	M62724	IGHD5-24*01	0	2	0	1/7
#2	Z47269	IGHD3-3*02	0	2	0	14/21

Translation of the JUNCTIONS

	105	107	109	111	111.2	112.4	112.2	112	114	116	118	CDR3-IMGT													
	104	106	108	110	111.1	111.3	112.3	112.1	113	115	117	frame	length												
	C	A	R	E	D	S	N	G		Y	K	I	F	D	Y	W									
#1	M62724	tgt	gcg	aga	gaa	gat	agc	aat	ggc		tac	aaa	ata	ttt	gac	tac	tgg	+	13						
	C	A	R	G	G	A	K	V	E	F	L	E	W	F	H	G	Y	W	F	D	P	W			
#2	Z47269	tgt	gcg	aga	ggg	ggg	gct	aag	gtc	gaa	ttt	ttg	gag	tgg	ttt	cat	ggg	tac	tgg	ttc	gac	ccc	tgg	+	20

IMGT/JunctionAnalysis



THANK YOU
for using IMGT/JunctionAnalysis

THE
INTERNATIONAL
IMMUNOGENETICS
INFORMATION SYSTEM®

<http://imgt.cines.fr>

Analysis of the JUNCTIONS

Input	V name	V-REGION	N1	D-REGION	N2
#1	M62724	IGHV7-4-1*02	tgtgcgagaga	aga .taccaatgctacaa....	aata
#2	Z47269	IGHV1-69*06	tgtgcgagag.	ggggggctaaggtcgaatttttgagggtt.....	tcatgggt

Input	J-REGION	J name	D name	Vmut	Dmut	Jmut	Ngc	
#1	M62724tttgactactgg	IGHJ4*02	IGHD5-24*01	0	2	0	1/7
#2	Z47269	...actggttcgaccctgg	IGHJ5*02	IGHD3-3*02	0	2	0	13/20

Translation of the JUNCTIONS

	104	105	106	107	108	109	110	111	111.1	111.2	111.3	112.4	112.2	112.1	112	113	114	115	116	117	118	frame	CDR3-IMGT length		
#1	M62724	C	A	R	E	D	S	N	G						Y	K	I	F	D	Y	W	+	13		
		tgt	gcg	aga	gaa	gat	agc	aat	ggc						tac	aaa	ata	ttt	gac	tac	tgg				
#2	Z47269	C	A	R	G	G	A	K	V	E	F	L	E	W	F	H	G	Y	W	F	D	P	W	+	20
		tgt	gcg	aga	ggg	ggg	gct	aag	gtc	gaa	ttt	ttg	gag	tgg	ttt	cat	ggg	tac	tgg	ttc	gac	ccc	tgg		

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	tgtgt <u>acg</u>		tgttgtgcagc <u>gcct</u> gggtac	ccaaatatc		...actttgacc <u>act</u> gg	IGHJ4*02	IGHD6-13*01	1	2	1	5/15
#2	Z70257	IGHV3-7*02	tgtgcgag.		ggatggcagctc <u>ttat</u> gcc	cgccc		ctactggacttccgatctctgg	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.		cagagtaacgattttggagtggttatt.....	ccccggggga		..atgcttttgatctctgg	IGHJ3*02	IGHD3-3*01	0	0	0	12/17
#5	Z70610	IGHV4-34*09	tgtgcgagag.		tcgggagcgatttttgagtggttatt.....	cccga	ca	tgatgcttttgatctctgg	IGHJ3*02	IGHD3-3*01	0	0	0	9/12
#6	Z70611	IGHV4-59*01	tgtgcgaga..		catggt <u>aact</u> ataa.	tgccggcggtg		...actggttcgaccctgg	IGHJ5*02	IGHD3-9*01	0	2	0	9/13
#7	Z70613	IGHV4-59*01	tgtgcgagag.		cagcagctgggtac	ctccct		...ctttgactactgg	IGHJ4*02	IGHD6-13*01	0	0	0	4/6
#8	Z70614	IGHV4-59*01	tgtgcgaga..		cactataattcggggacttat.....	ccctc	gactactgg	IGHJ4*02	IGHD3-16*01	0	2	0	7/14
#9	Z70615	IGHV4-59*01	tgtgcgagag.		ggctg	gtaagaggg.....	tttcggaa		.tactggacttccgatctctgg	IGHJ2*01	IGHD5-24*01	0	2	0	7/13
#10	Z70616	IGHV4-34*01	tgtgcgagag.		cgggtttggg.....	ttccc		...actggttcgaccctgg	IGHJ5*02	IGHD3-16*01	0	0	0	6/8
#11	Z70620	IGHV4-30-4*01	tgtgcgagaga		ccggggcgggatggtt.....	cgg		.gatgcttttgatctctgg	IGHJ3*02	IGHD3-16*01	1	4	0	5/5
#12	Z70621	IGHV4-39*01	tgtgcgagaca		ccacgatttatggttcggggagtt.....	tgaccccc	ttgactactgg	IGHJ4*02	IGHD3-16*01	0	1	0	12/21
#13	Z70622	IGHV4-39*06	tgtgcgagaga	t	tgccccgctcctgccaaaat	gtattaactatggttcgggga.....	tatgtacg	tttactactgg	IGHJ4*03	IGHD3-10*01	0	0	0	15/28

IMGT/JunctionAnalysis

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

Addition

Addition

Mutation

Délétion

Délétion

Délétion

#	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.....		tgttgtgcagcgcctgttac	ccaaatac		...actttgaccactgg	IGHJ4*02	IGHD6-13*01	1	2	1	5/15
#2	Z70257	IGHV3-7*02	tgtgcgag.....		ggatggcagctccttatgcc	cgecc		ctactggtacttcgatctctgg	IGHJ2*01	IGHD2-2*01	0	2	0	9/11
#3	Z70606	IGHV4-31*03	<u>c</u>		c	.gactacg.....	cact		..atgcttttgatgtctgg	IGHJ3*01	IGHD4-17*01	0	0	0	3/5
#4	Z70608	IGHV4-39*05	cagagta		cagagtaacgatttttgagtggttatt.....	ccccggggga		..atgcttttgatatctgg	IGHJ3*02	IGHD3-3*01	0	0	0	12/17
#5	Z70610	IGHV4-34*09	tgtgcgagag.		tcgggagcgatttttgagtggttatt.....	cccgga	ca	tgatgcttttgatatctgg	IGHJ3*02	IGHD3-3*01	0	0	0	9/12
#6	Z70611	IGHV4-59*01	tgtgcgaga..		catggttaactataa.	tgccggcgcttg		..actggttcgaccctgg	IGHJ5*02	IGHD3-9*01	0	2	0	9/13
#7	Z70613	IGHV4-59*01	ctgttac		ctgttacgactactgg	ctc		gactactgg	IGHJ4*02	IGHD6-13*01	0	0	0	4/6
#8	Z70614	IGHV4-59*01	actataa		actataagactactgg	ccc		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.		cgggtttggg.....	ttccc		...actggttcgaccctgg	IGHJ5*02	IGHD3-16*01	0	0	0	6/8
#11	Z70620	IGHV4-30-4*01	tgtgcgagaga		ccggggcgggatggtt.....	cgg		.gatgctttgatctctgg	IGHJ3*02	IGHD3-16*01	1	4	0	5/5
#12	Z70621	IGHV4-39*01	tgtgcgagaca		ccacgattatggttcggggaggtt.....	tgaccccc	ttgactactgg	IGHJ4*02	IGHD3-16*01	0	1	0	12/21
#13	Z70622	IGHV4-39*06	tgtgcgagaga	t	tgccccgctcctgccaaat	gtattactatggttcgggga.....	tatgtacg	ttgactactgg	IGHJ4*03	IGHD3-10*01	0	0	0	15/28

The eleven IMG T amino acid classes according to the physico-chemical properties

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

JUNCTION alignments

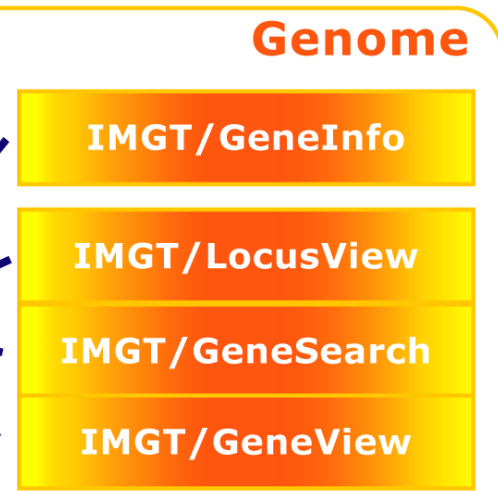
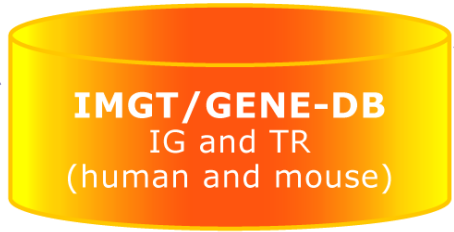
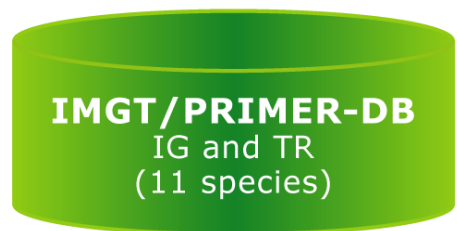
Click on mutated (underscored) amino acid to see the original one:

	104	105	106	107	108	109	110	111	111.1	111.2	111.3	111.4	112.5	112.4	112.3	112.2	112.1	112	113	114	115	116	117	118	Frame	CDR3-IMGJ length	Molecular mass			
#1 Z70256	C	<u>V</u>	R	V	V	Q	<u>R</u>	L	V									P	K	Y	H	F	D	<u>H</u>	W	+	15	2,438.76		
#2 Z70257	C	A	R	D	G	S	<u>S</u>	Y	A									R	P	Y	W	Y	F	D	L	W	+	16	2,256.49	
#3 Z70606	C	A	R	A	T	T	H												Y	A	F	D	V	W	+	11	1,604.77			
#4 Z70608	C	A	R	V	T	I	F	G	V	V								I	P	R	G	N	A	F	D	I	W	+	18	2,426.78
#5 Z70610	C	A	R	V	G	S	D	F	<u>W</u>	S	G							Y	S	R	H	D	A	F	D	I	W	+	19	2,539.73
#6 Z70611	C	A	R	H	G	<u>N</u>	Y	N	A									G	V	D	W	F	D	P	W	+	15	2,072.21		
#7 Z70613	C	A	R	A	A	A	G											T	S	L	F	D	Y	W	+	12	1,531.71			
#8 Z70614	C	A	R	H	Y	N	<u>S</u>	G										T	Y	P	L	D	Y	W	+	13	1,846.01			
#9 Z70615	C	A	R	G	L	V	<u>K</u>	<u>R</u>	V									S	E	Y	W	Y	F	D	L	W	+	16	2,419.75	
#10 Z70616	C	A	R	A	G	L	G											S	H	W	F	D	P	W	+	12	1,602.8			
#11 Z70620	C	A	R	D	R	G	<u>G</u>	H										<u>V</u>	R	D	A	F	D	I	W	+	14	1,932.17		
#12 Z70621	C	A	R	H	H	D	L	W	F									G	E	F	D	P	L	D	Y	W	+	16	2,307.53	
	C	A	R	D	C	P	A	P	A	K	M	Y	Y	Y	G	S	G	I	C	T	F	D	Y	W	+					

Sequences



<http://imgt.cines.fr>

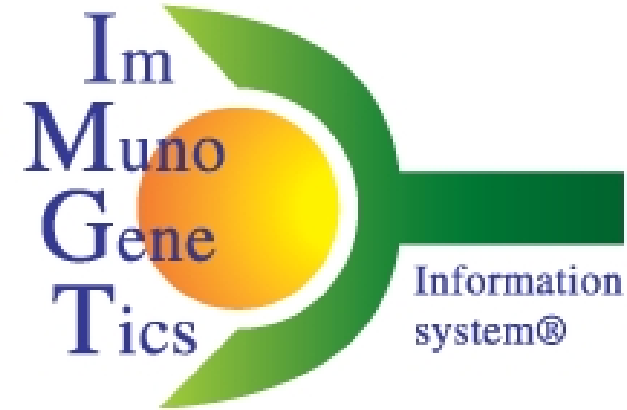


2D and 3D structures

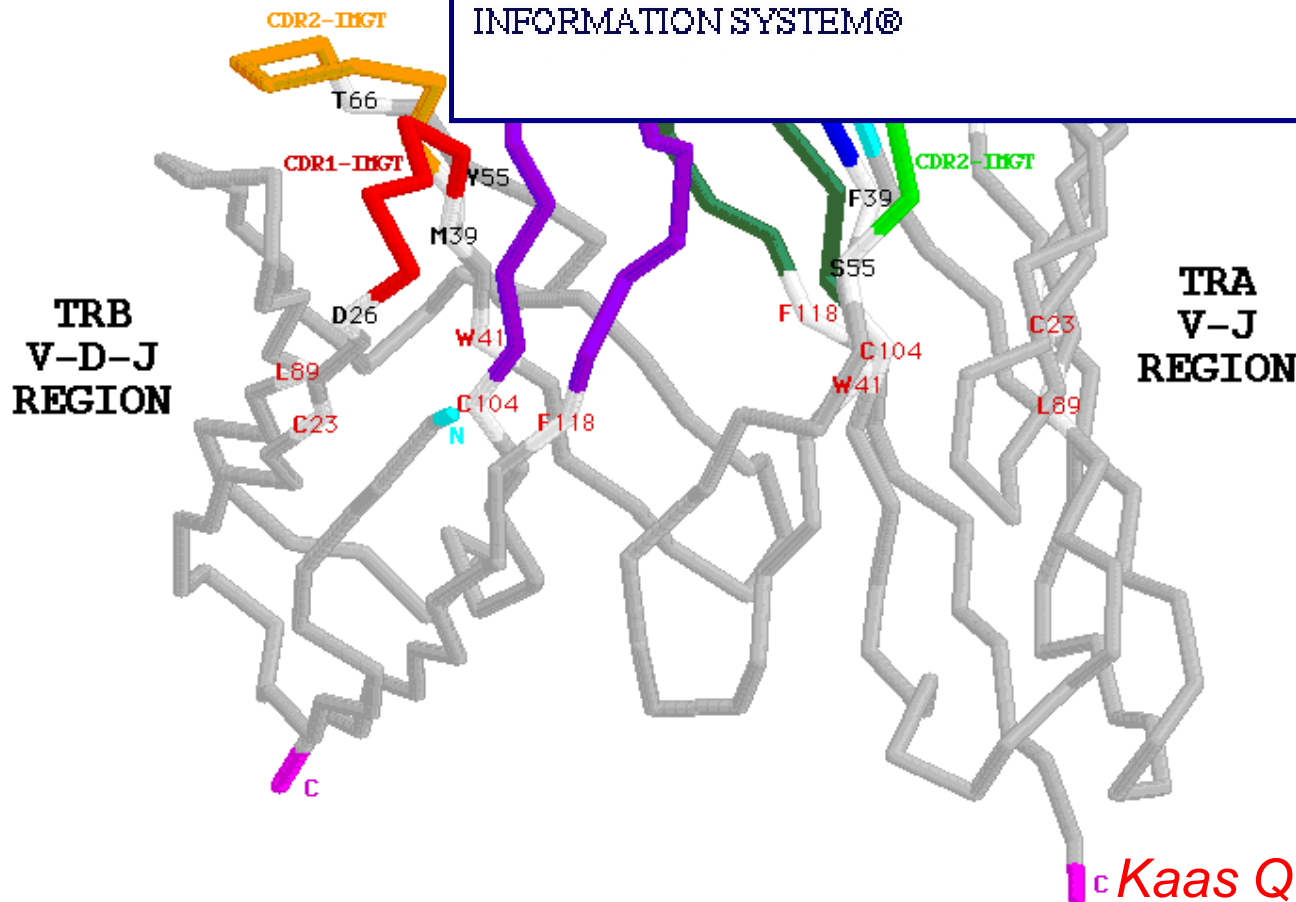
WELCOME !

to **IMGT/3Dstructure-DB**

THE
INTERNATIONAL
IMMUNOGENETICS
INFORMATION SYSTEM®

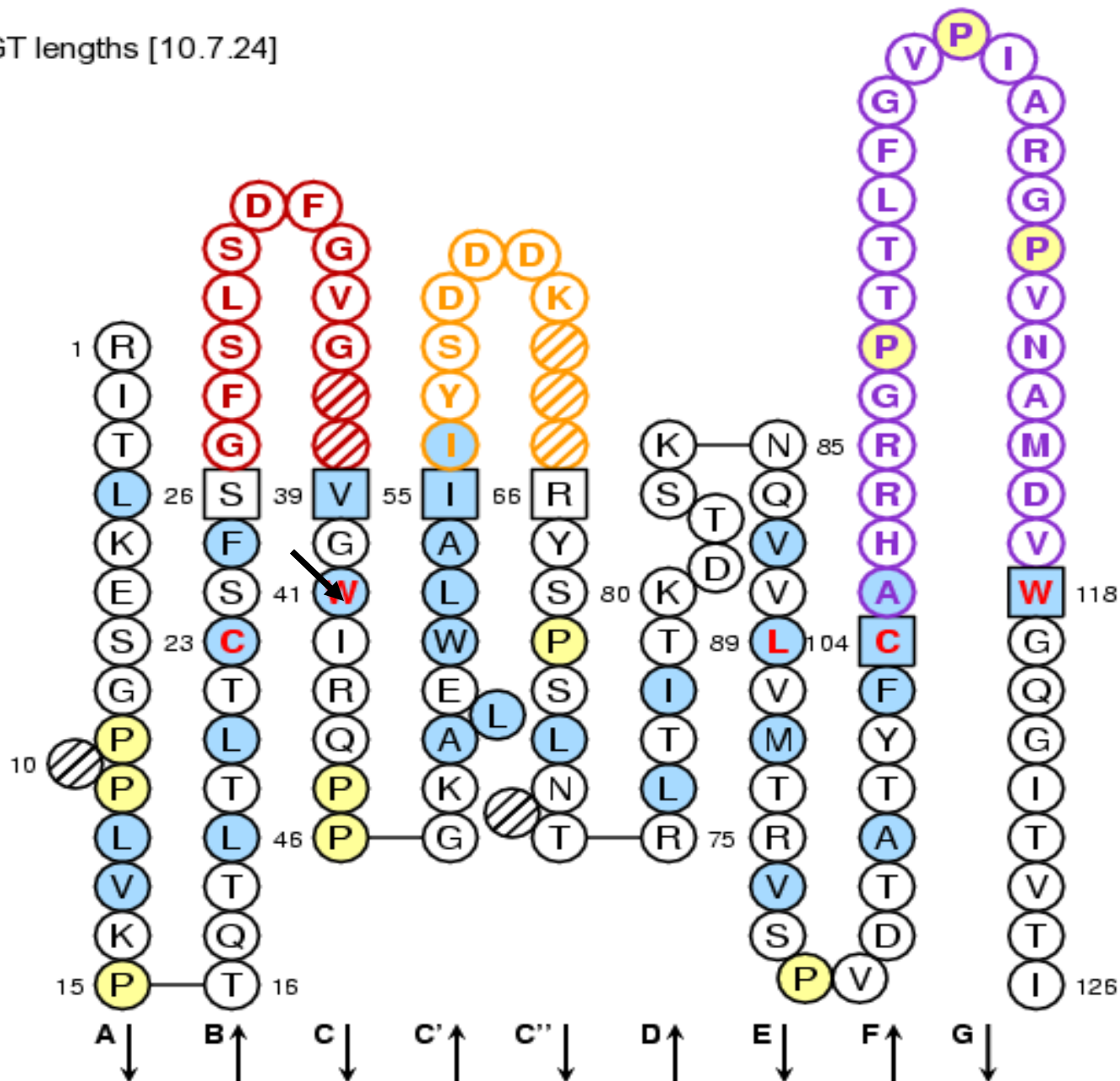


<http://imgt.cines.fr>



Kaas Q. et al. NAR 32, D208-D210 (2004)

CDR-IMGT lengths [10.7.24]



IMGT Residue@Position contacts

41V - TRP (W)

chain : 1u8k_B

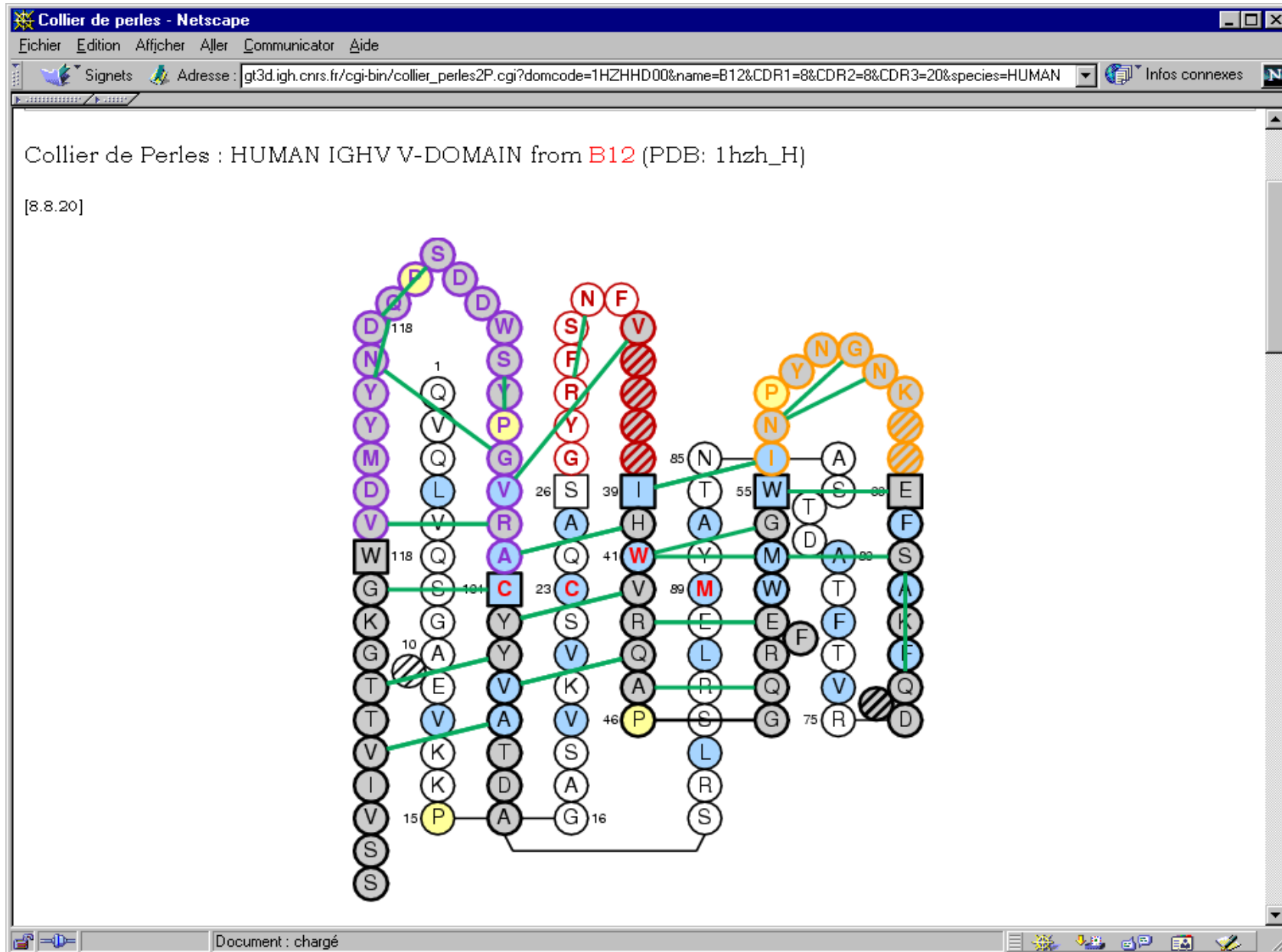
Secondary structure	Extended conformation
Phi	-122.64
Psi	137.12
ASA	0.0

Residue local structure

	IMGT Num	Code 1C	Code 3C	Chain	Domain	Tot	NCov	Pol	HB	NPol	Cov	SS
Info	6V1	E	GLU	1u8k_B	VH	6	6	0	0	6	0	0
Info	21V1	L	LEU	1u8k_B	VH	17	17	0	0	17	0	0
Info	22V1	T	THR	1u8k_B	VH	8	8	0	0	8	0	0
Info	23V1	C	CYS	1u8k_B	VH	10	10	0	0	10	0	0
Info	39V1	V	VAL	1u8k_B	VH	2	2	1	0	1	0	0
Info	43V1	R	ARG	1u8k_B	VH	2	2	1	0	1	0	0
Info	51V1	E	GLU	1u8k_B	VH	2	2	1	0	1	0	0
Info	52V1	W	TRP	1u8k_B	VH	11	11	3	0	8	0	0
Info	53V1	L	LEU	1u8k_B	VH	15	15	1	1	14	0	0
Info	54V1	A	ALA	1u8k_B	VH	18	18	4	2	14	0	0
Info	55V1	I	ILE	1u8k_B	VH	2	2	1	0	1	0	0
Info	78V1	I	ILE	1u8k_B	VH	5	5	0	0	5	0	0
Info	87V1	V	VAL	1u8k_B	VH	11	11	1	0	10	0	0
Info	88V1	V	VAL	1u8k_B	VH	6	6	2	0	4	0	0
Info	89V1	L	LEU	1u8k_B	VH	33	33	1	0	32	0	0
Info	102V1	Y	TYR	1u8k_B	VH	5	5	0	0	5	0	0
Info	103V1	F	PHE	1u8k_B	VH	18	18	2	0	16	0	0
Info	104V1	C	CYS	1u8k_B	VH	26	26	0	0	26	0	0
Info	105V1	A	ALA	1u8k_B	VH	1	1	1	0	0	0	0

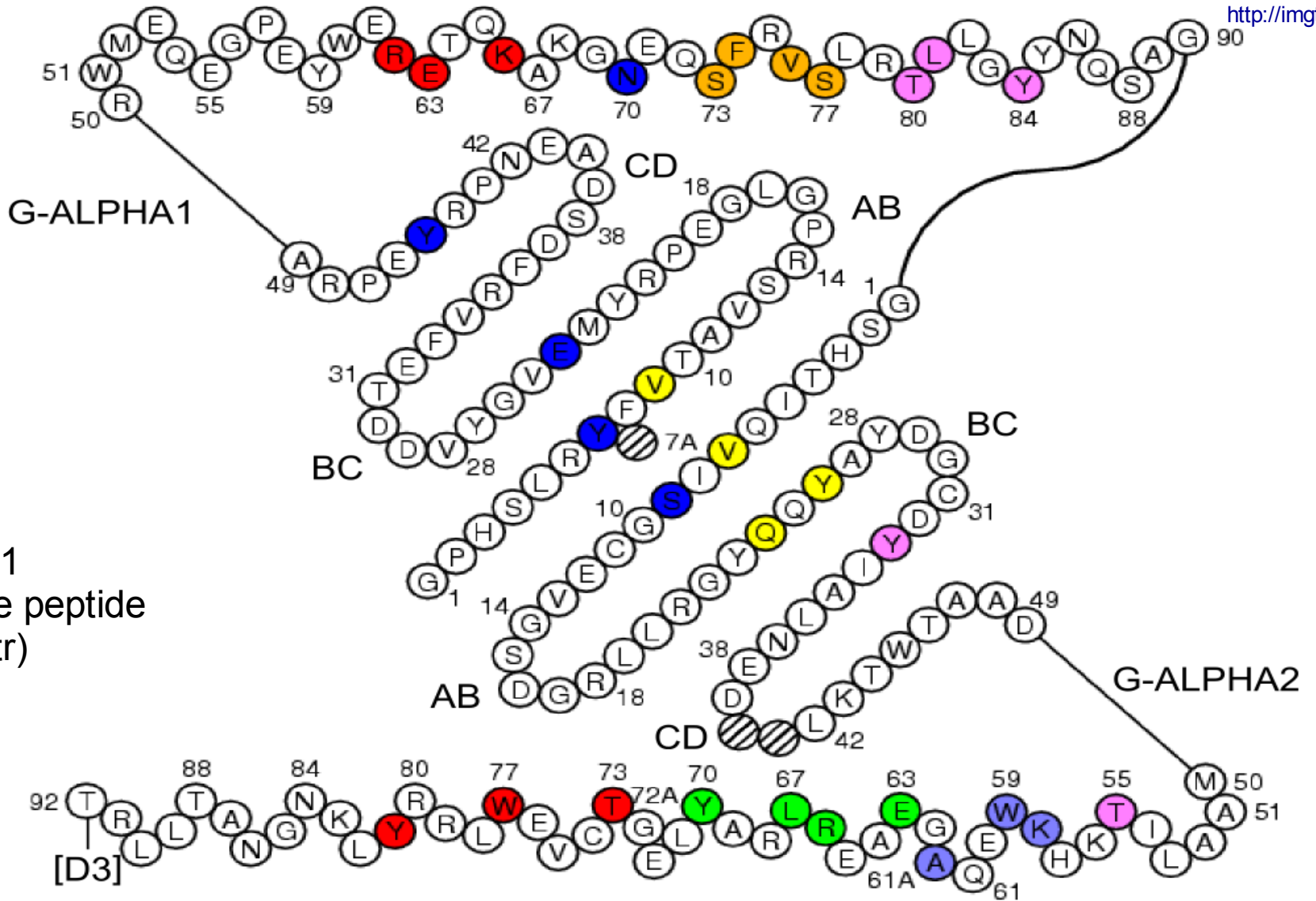
Tot	Total number of atomic pair contacts
NCov	Number of non covalent atomic
Pol	Number of polar atomic pair contacts
HB	Number of hydrogen bonds
NPol	Number of non polar atomic pair contacts
Cov	Number of covalent links (other than chain covalent links)
SS	Number of disulfide bridges

IMGT Collier de Perles on two layers



IMGT pMHC contact sites

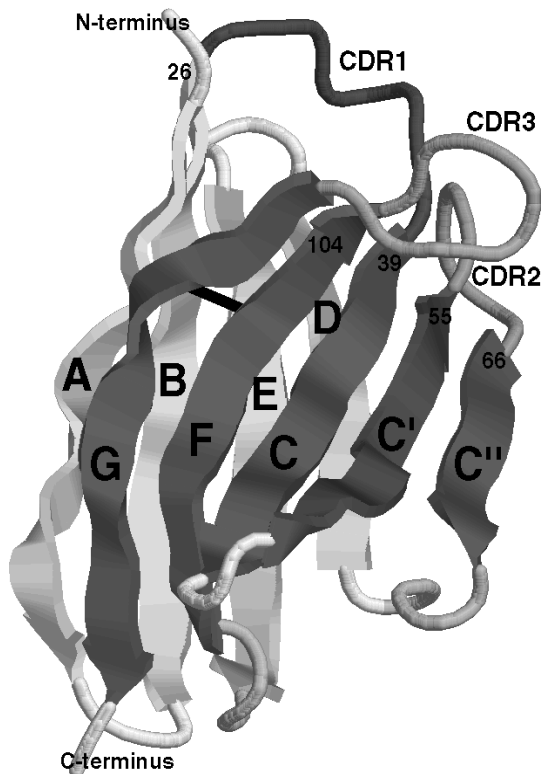
C1	1 E
C2	--
C3	2 Q
C4	3 Y
C5	(4 K)
C6	5 F
C7	--
C8	--
C9	6 Y
C10	7 S
C11	8 V



H2-K1*01
8 residue peptide
(code 1jtr)

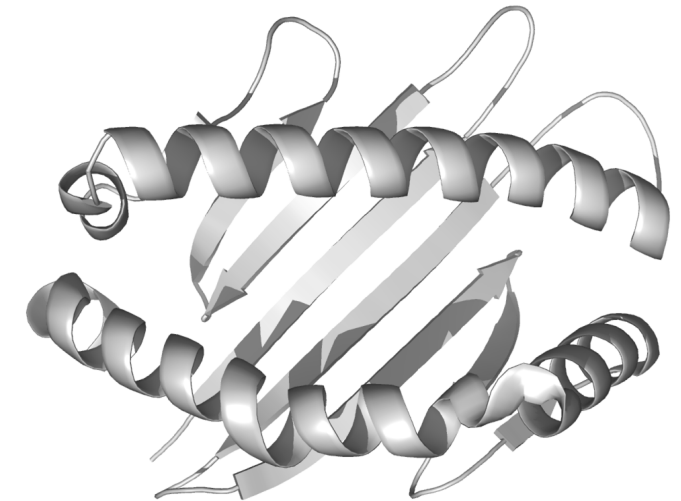
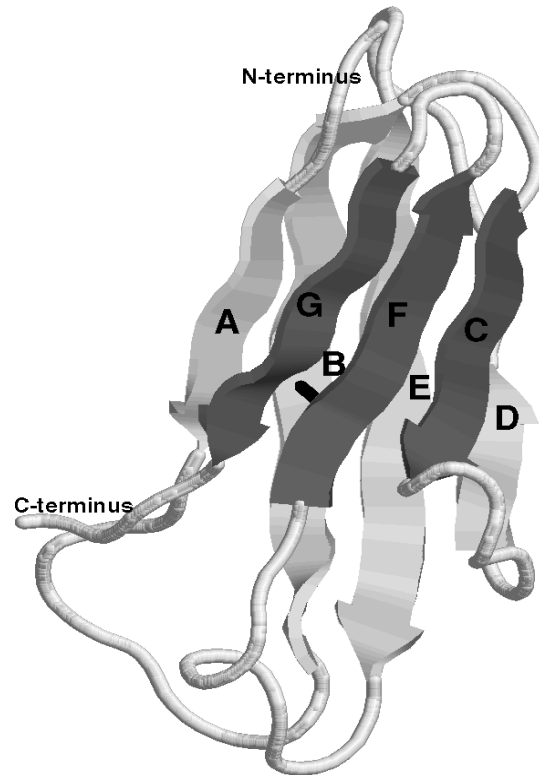
IMGT unique numbering

V-DOMAIN (IG,TR)
AND
V-LIKE-DOMAIN
(other than IG,TR)



Immunoglobulin superfamily (IgSF)

C-DOMAIN (IG,TR)
AND
C-LIKE-DOMAIN
(other than IG,TR)

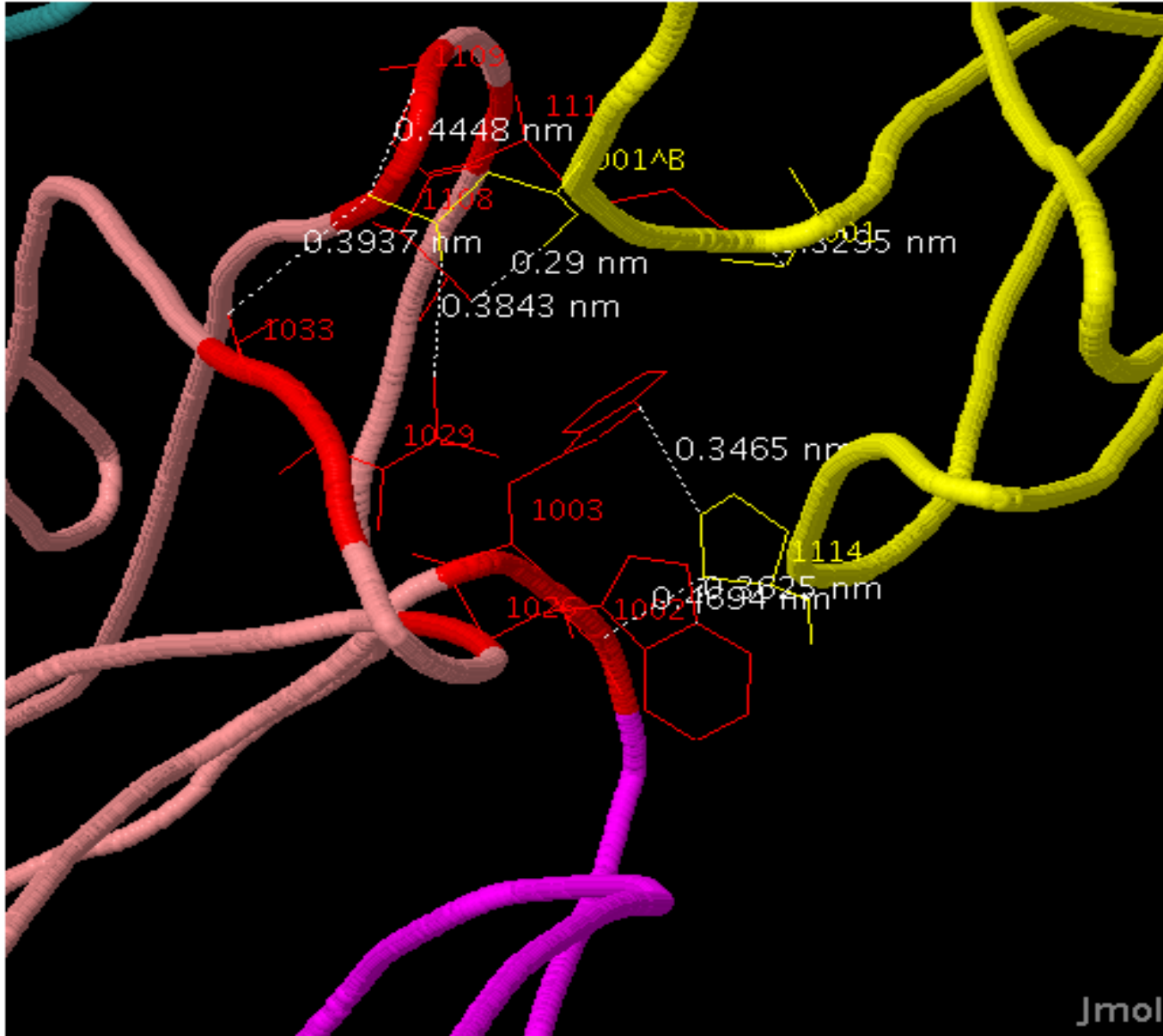


MHC superfamily (MhcSF)

Interactions between domains

FCGR3
B

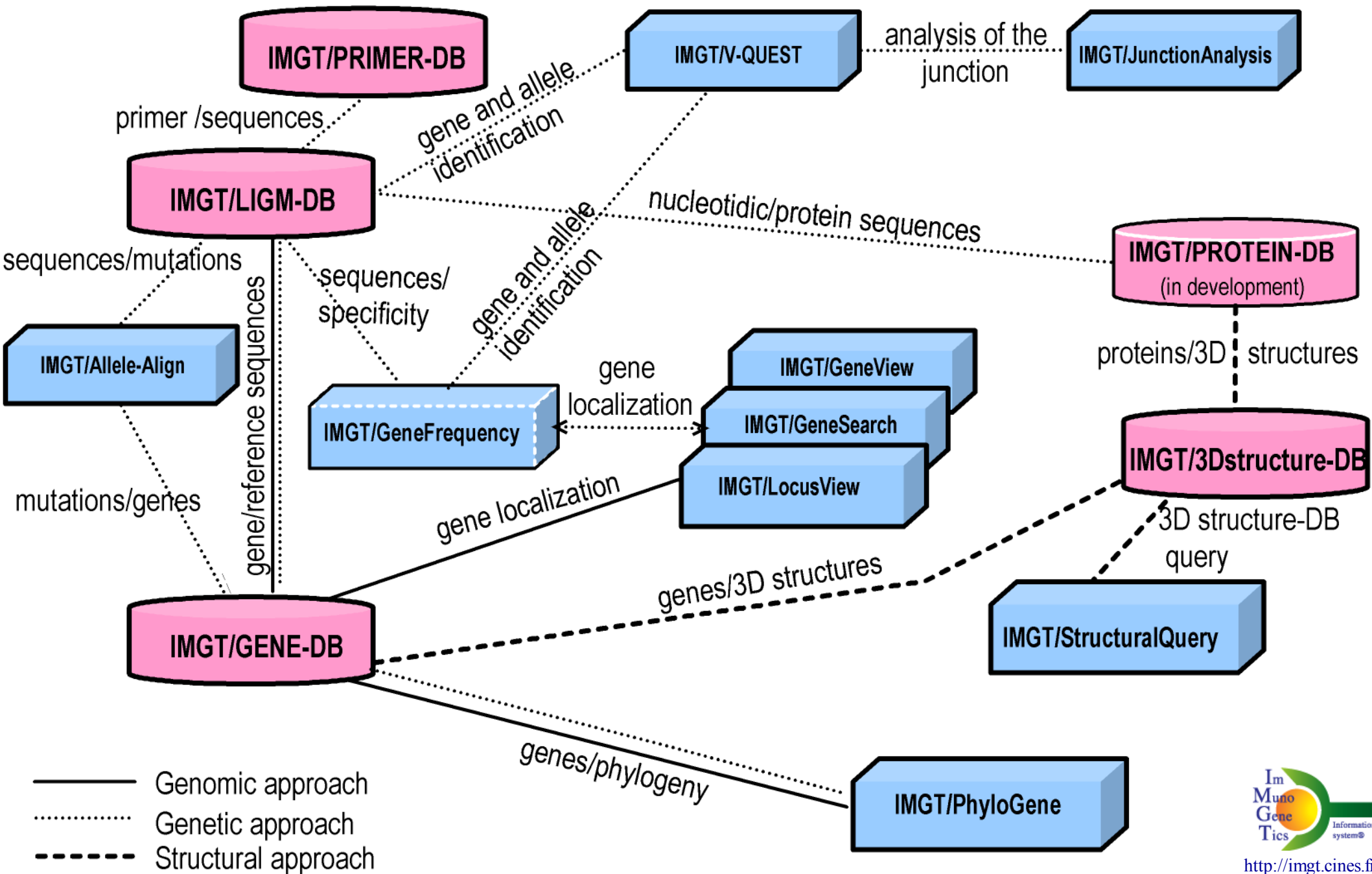
[D2]
D-LIKE-
DOMAIN



[D1] C-LIKE-DOMAIN

IGHG1
(FC-GAMMA1)

CH2
C-DOMAIN





IMGT-ONTOLOGY concepts: *IMGT Scientific chart rules*****
ImmunoGrid modelling of the IS at the molecular level.

- **IDENTIFICATION**: ***Keywords***
IG, TR, MHC nucleotide and amino acid description, 3D structure
- **DESCRIPTION**: ***Labels***
Nucleotide and amino acid sequences and 3D structures
- **CLASSIFICATION**: ***Nomenclature***
Gene and protein names
- **NUMEROTATION**: ***Numbering***
Amino acid positions in sequences, IMGT Colliers de Perles, 3D structures
- **ORIENTATION**: ***Genomic orientation***
Sequence orientation on chromosome, locus, clone, contig

Who is using IMGT?

Medical research:

allergies
autoimmune diseases
AIDS
leukemias
lymphomas
myelomas
translocations
detection of residual diseases

Biotechnology related to antibody engineering:

chimeric
humanized
human antibodies
scFv
combinatorial libraries
intrabodies

Veterinary research:

IG and TR repertoire
domestic and farm species
wild species

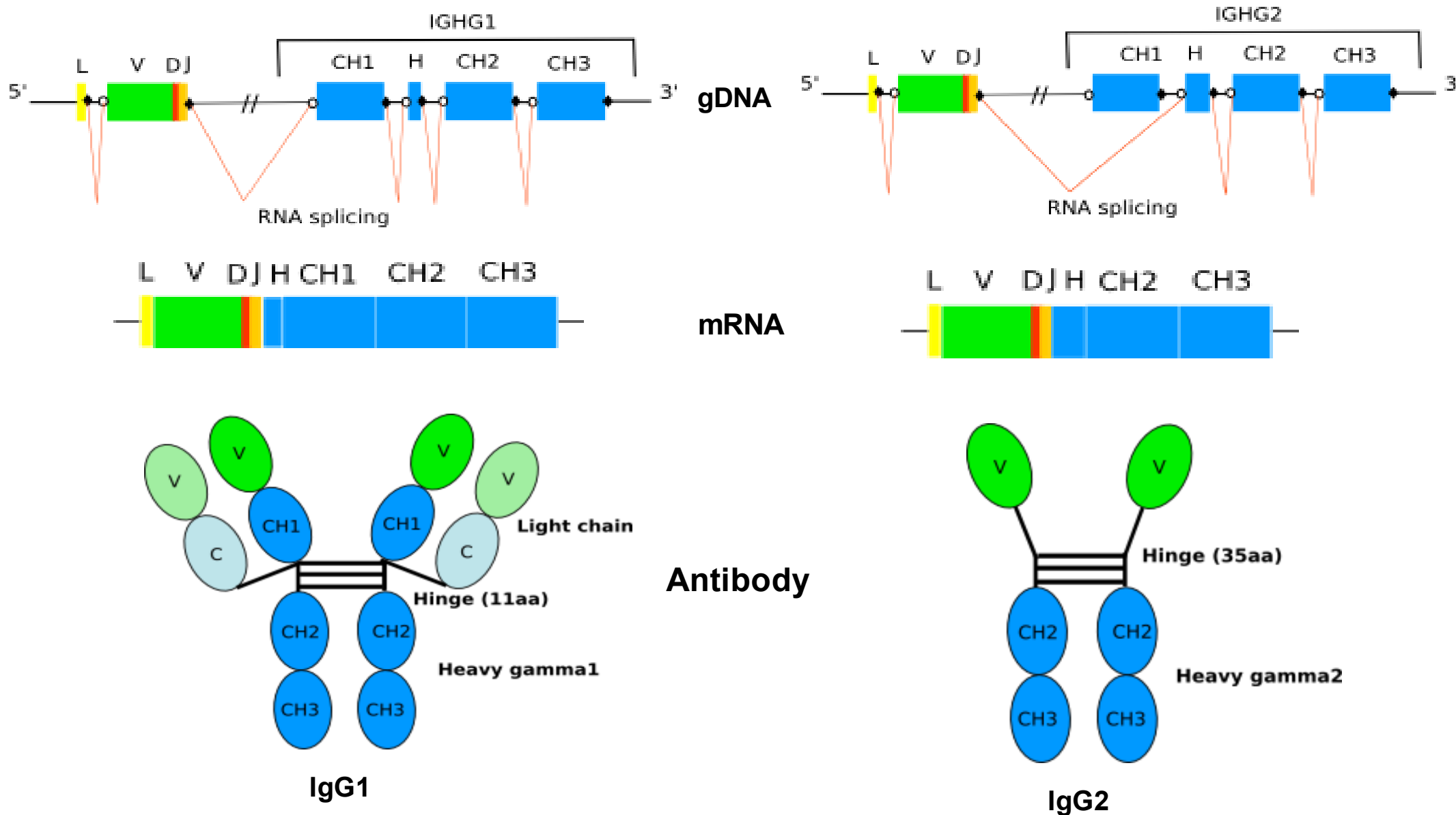
Comparative and developmental immunology:

evolution of the adaptive immune system



The IMGT team at Montpellier

Camelidae (camel, llama) antibody synthesis

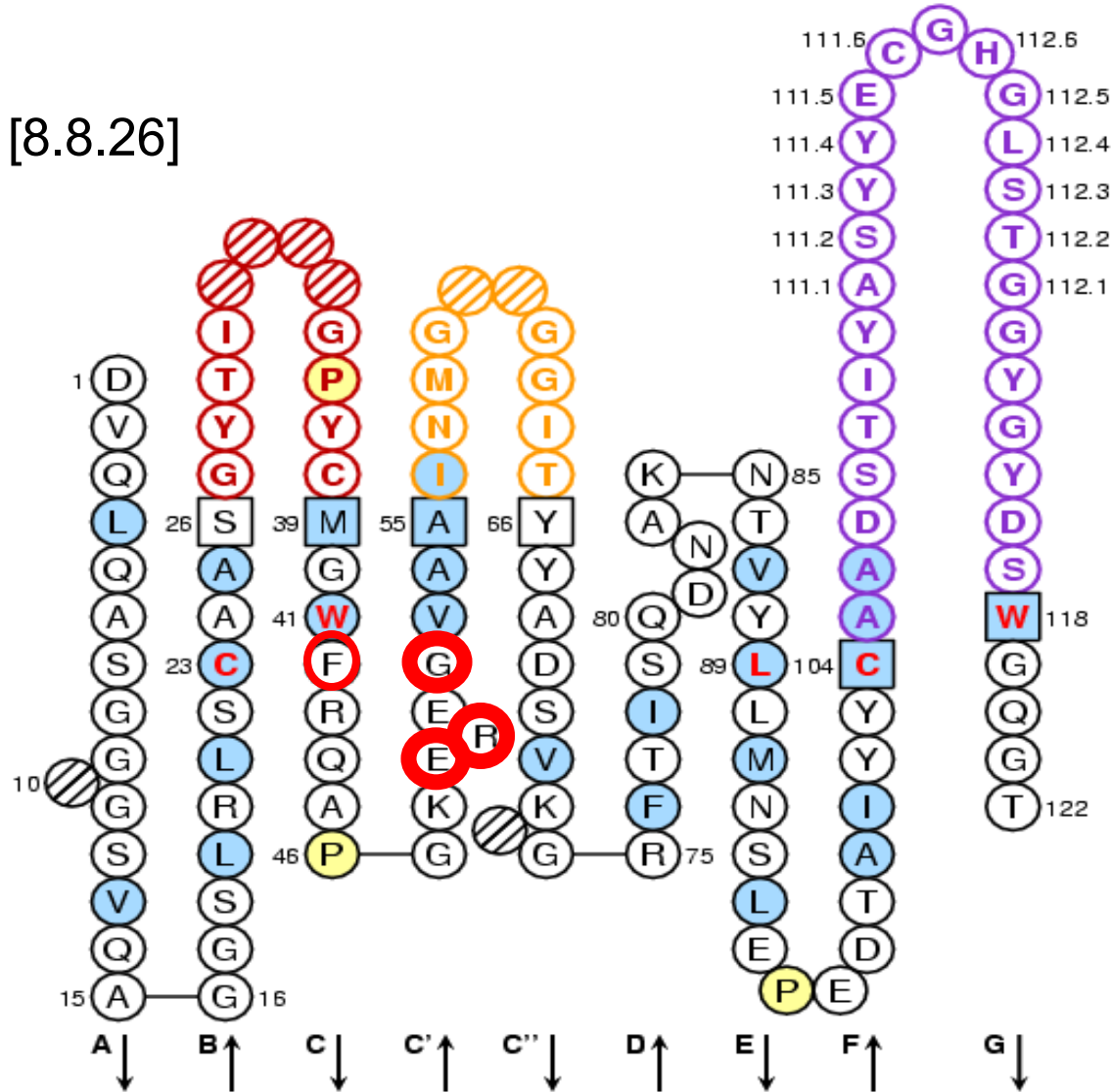


IMGT Collier de Perles

Camelus dromedarius (Camel) VH Single-Domain Antibody
(1jto_A)

CDR-IMGT lengths [8.8.26]

42: F, Y
49: E, Q
50: R, C
52: G, F, L, W



ImmunoGrid

The European Virtual Human Immune System Project



Immune system standardized concepts



ImmunoGrid is a computer model of the **Human Immune System** implemented with **Grid technologies**.

- It integrates processes at molecular, cellular and organ levels.
- ImmunoGrid is a project funded by the European Commission under the contract FP6-2004-IST-4, N°028069.

ImmunoGrid:

- simulates immune processes
- complements experimental work
- combines experimental and computational studies to advance our knowledge of immunology
- supports clinical applications