EMBnet Conference 2008: 20th Anniversary Celebration
September 18-20, 2008 - Martina Franca (Puglia), Italy

# IMGT®, an ontology and a system that bridge the gap between sequences and 3D structures

**Patrice Duroux**

PhD, Research Engineer in Computer Science
IMGT, IGH, CNRS UPR 1142

# Summary

1. Overview of IMGT®
2. Brief biological context
3. From sequences
4. From structures
5. The encounter
6. Perspectives

# IMGT a reference in immunoinformatics

➔ Provide data and tools for immunoglobulin (IG) and T cell receptor (TR) sequences and more recently structures

- data curation and management
- bioinformatics development

➔ Ensure quality and consistancy with respect to biological standardize background

- from IMGT Scientific Chart
- to Formal IMGT-ONTOLOGY

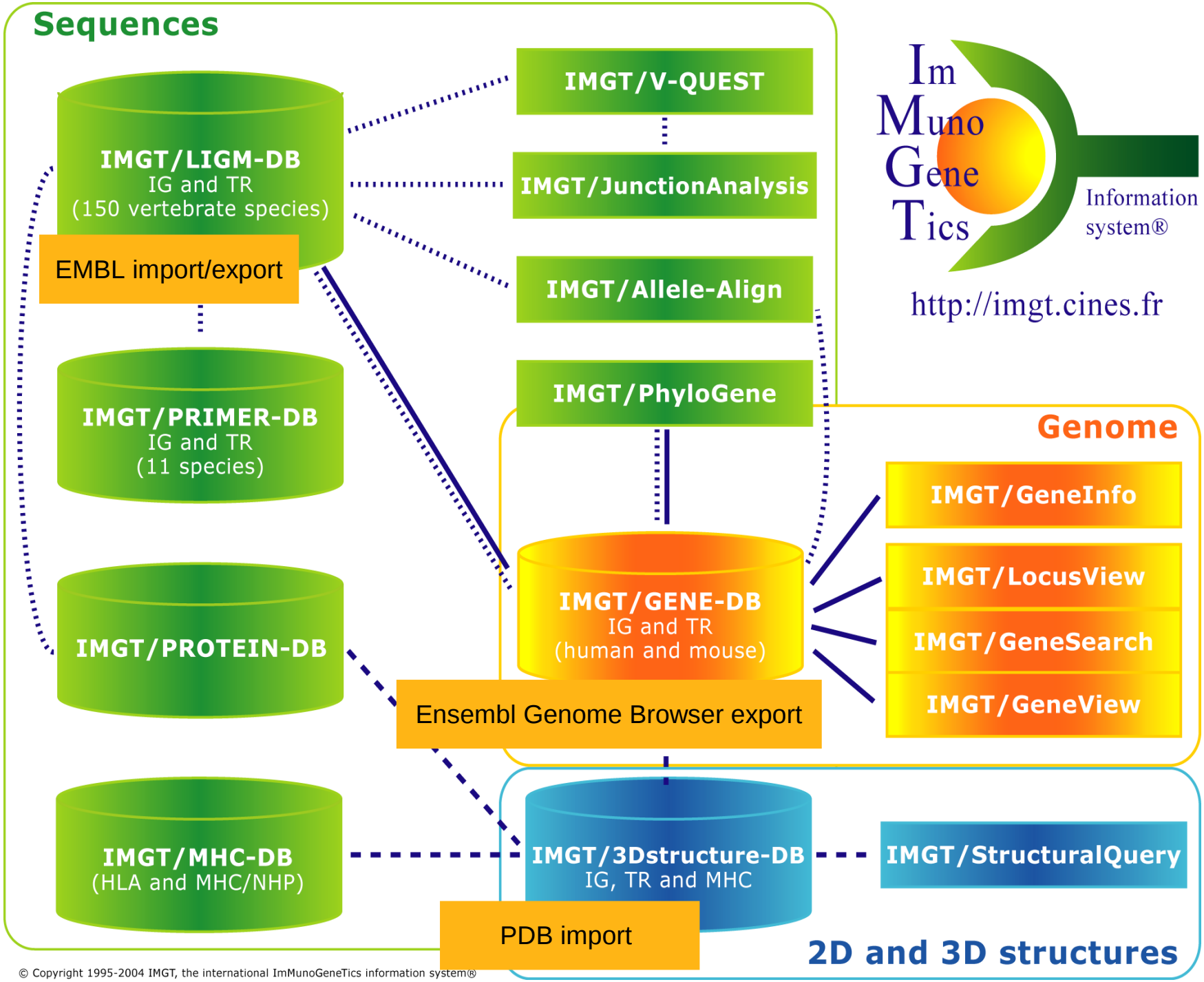
# IMGT a reference in immunoinformatics

➔ Reinforce relevance by integrating knowledge from different approaches

- genetic
- genomic
- more recently structure

➔ Provide a WWW access to all the resources for the largest community

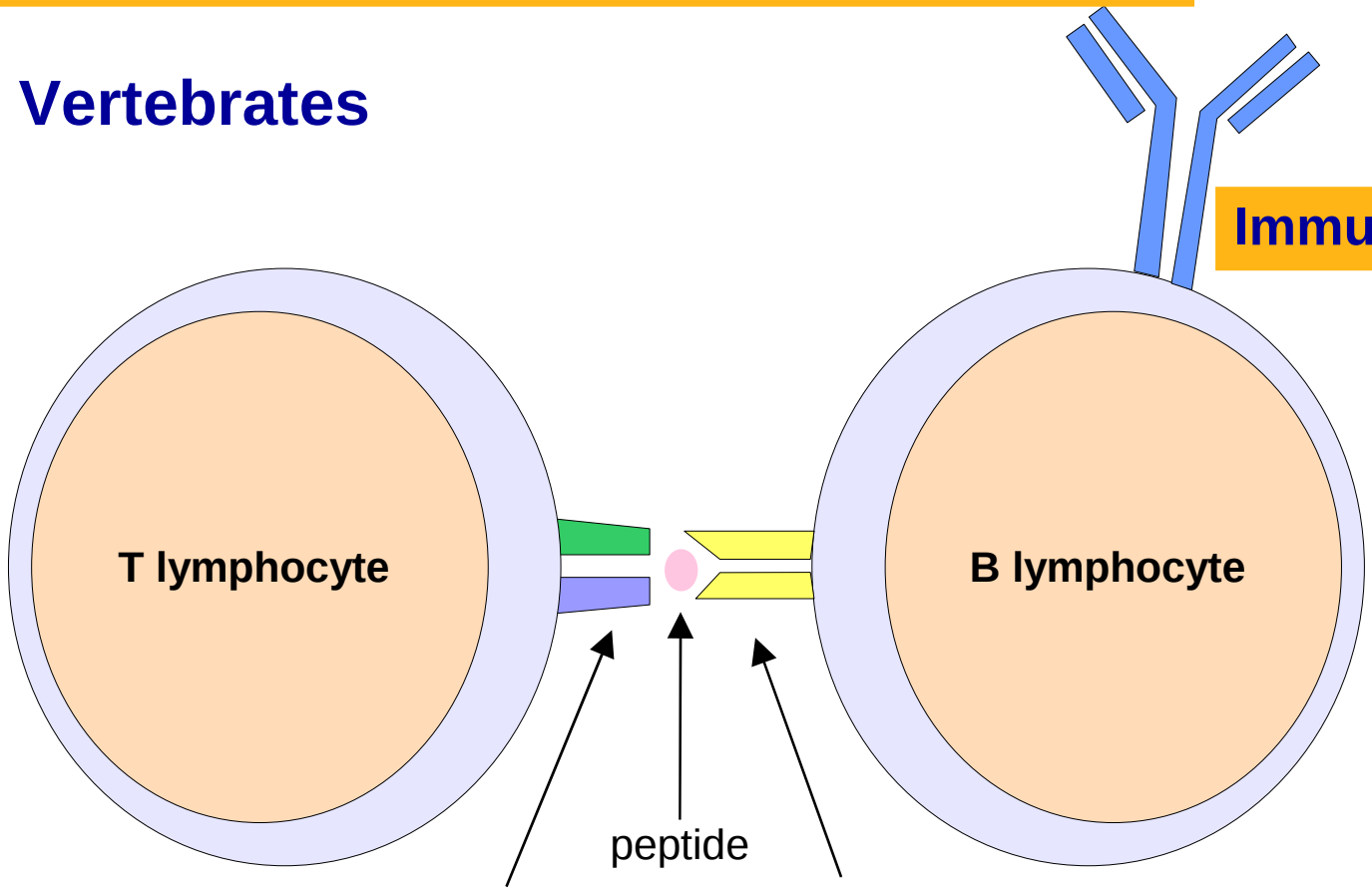
- indexes, Repertoire, tutorials, ... (HTML pages)
- data querying, specific tools (Perl CGI, Java Servlet)



# Adaptative immune response

Proteins which specifically recognize foreign antigens:
•in a native form for IG
•as processed as a peptide and presented by MHC for TR
Function of IG and TR is to bind specifically

## Vertebrates



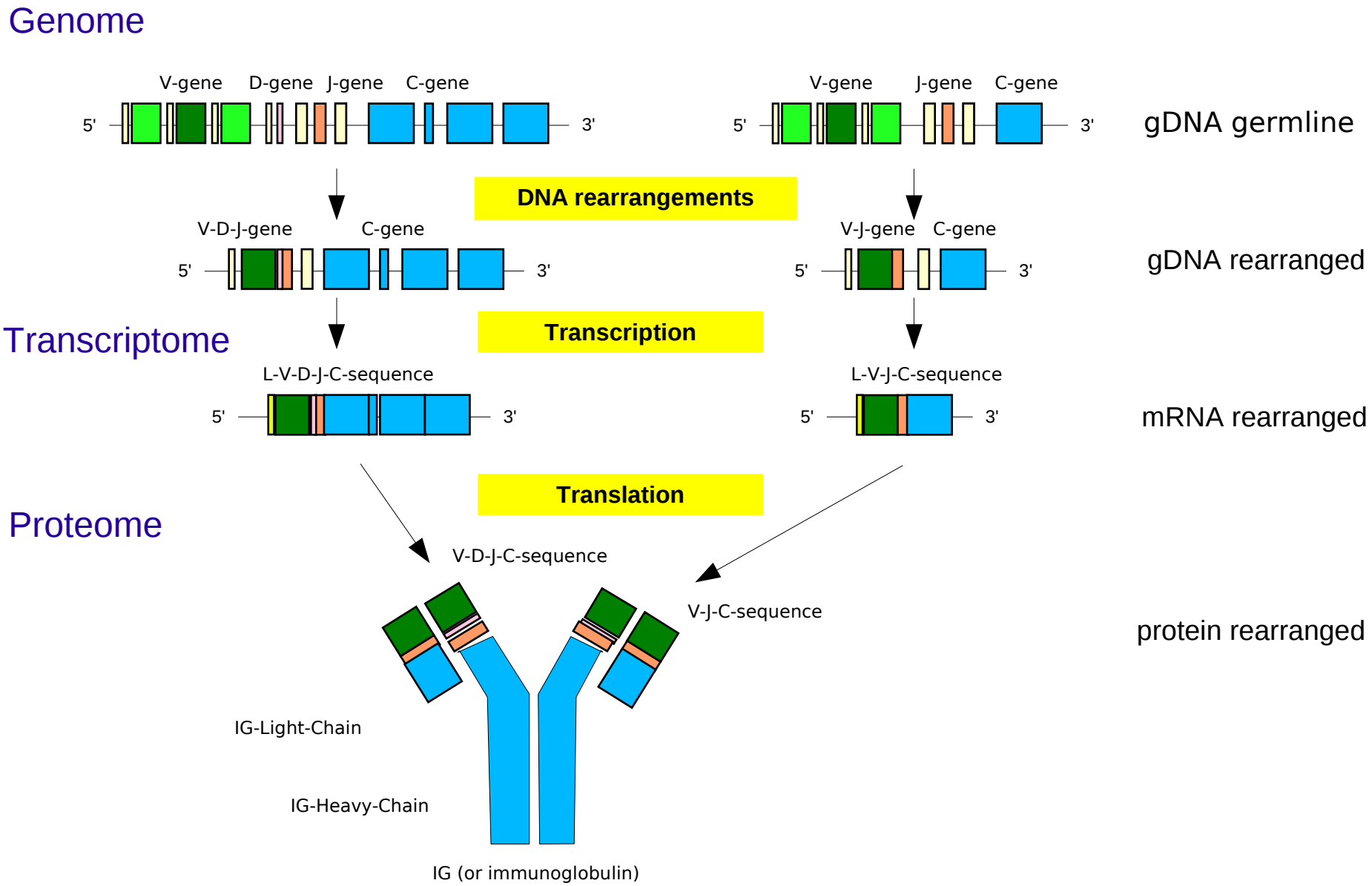

**Immunoglobulin (IG)**

**T cell receptor (TR)**

**Major histocompatibility complex (MHC)**

Trimolecular complex

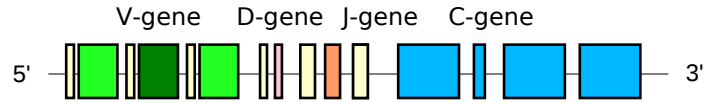
# IG: synthesis and complexity

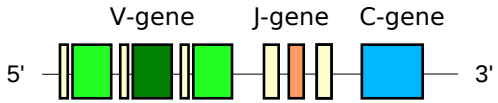


# IG: repertoire and diversity (Human)

Genome

150
**functional IG genes**



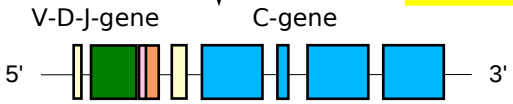

gDNA germline

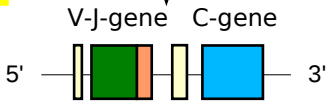
38-46 x 23 x 6

**DNA rearrangements**

30-35 x 5 (Kappa)
29-33 x 4-5 (Lambda)

6300



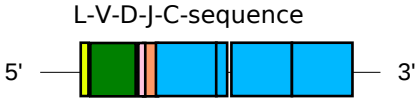



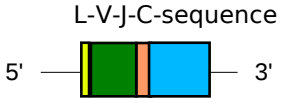

gDNA rearranged

Transcriptome

**Transcription**

185 (Kappa)
165 (Lambda)



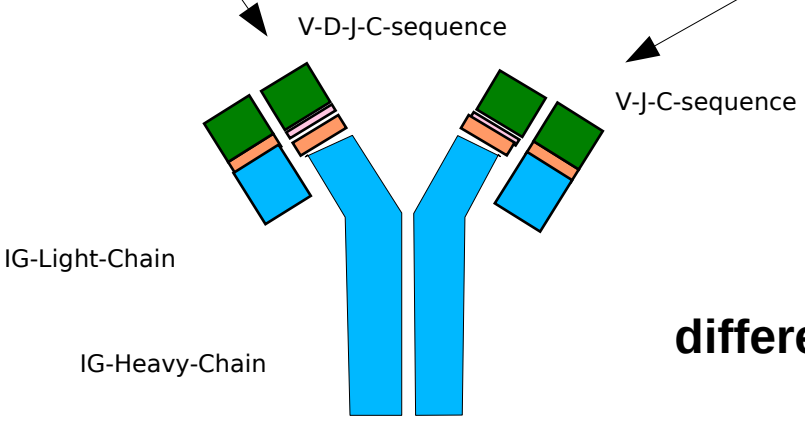

mRNA rearranged

Proteome

**Translation**

63 x $10^5$

3.5 x $10^5$




IG and TR not dealt in generalist databases owing to their particular structure

2 x $10^{12}$
**different antibodies**

# IMGT-ONTOLOGY

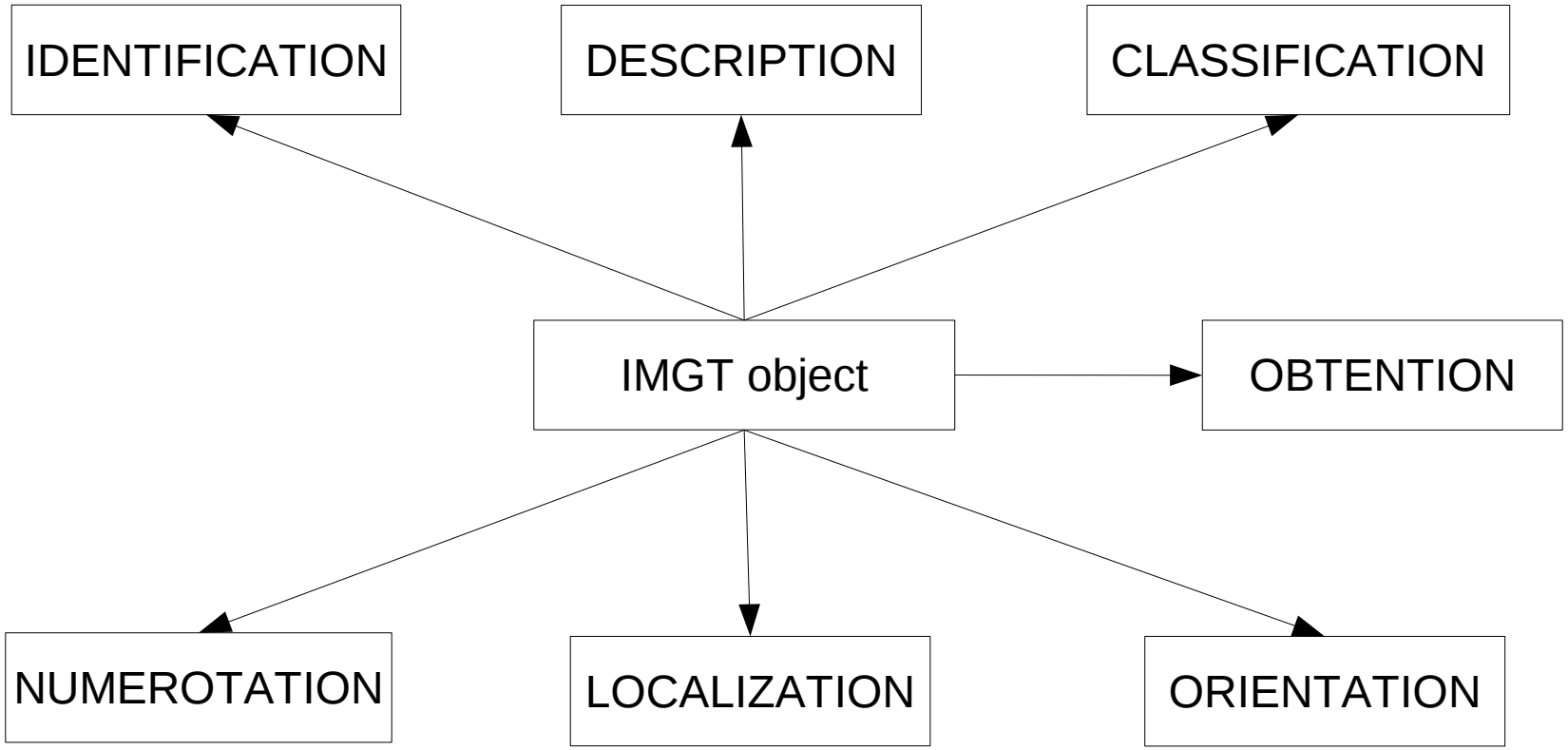
➔ Provides for main IMGT objects:

- identification (molecule type, configuration type, chain type)
- description (prototypes, labels)
- classification for gene and allele nomenclature (groups, subgroups, gene name, allele number, locus)
- numerotation (position in domains, IMGT Collier-de-Perles)

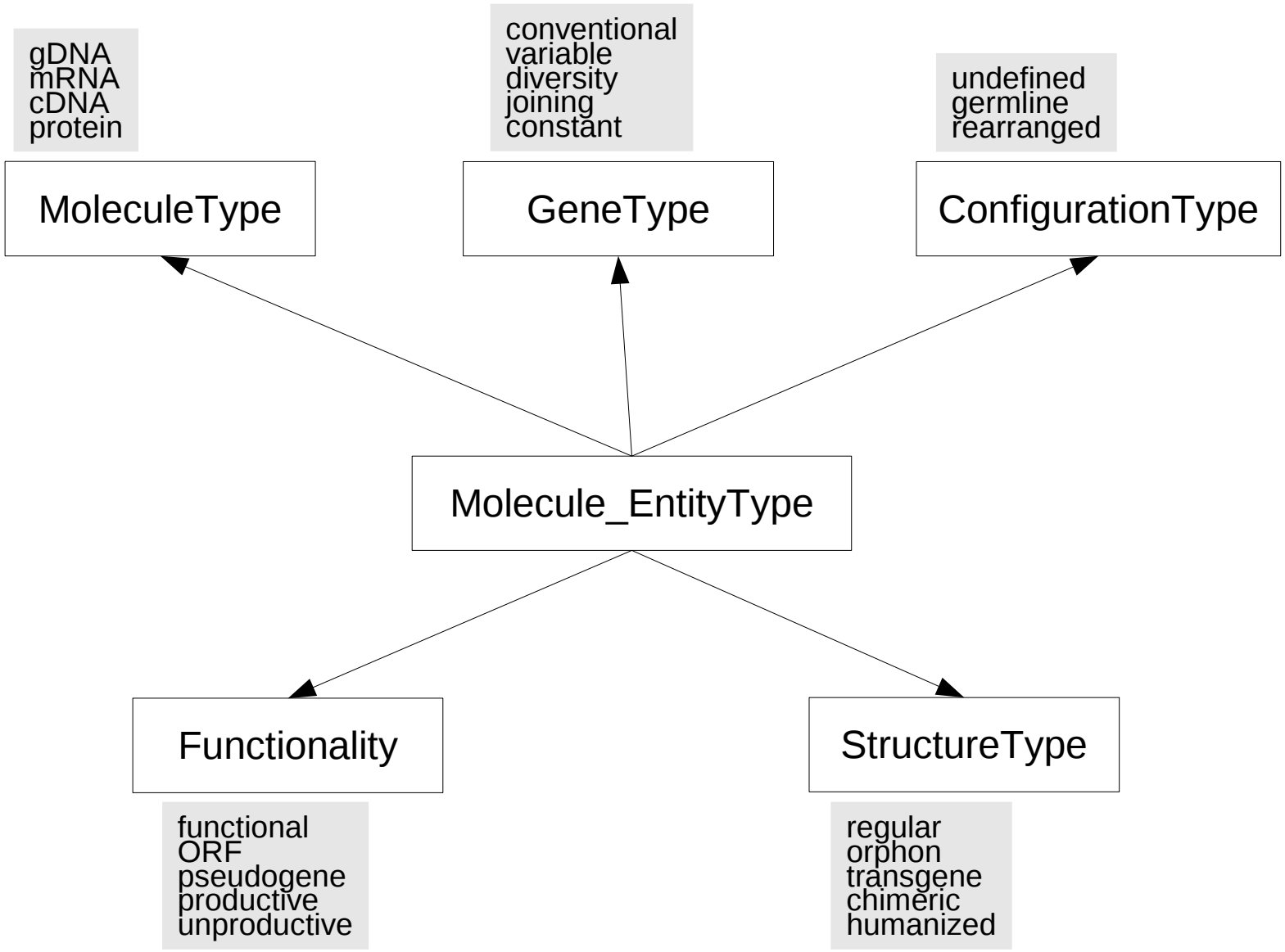
➔ Improve consistency and high quality in data and tools

➔ Develop better interoperability between IMGT® components for workflow perspective (IMGT-Choreography)

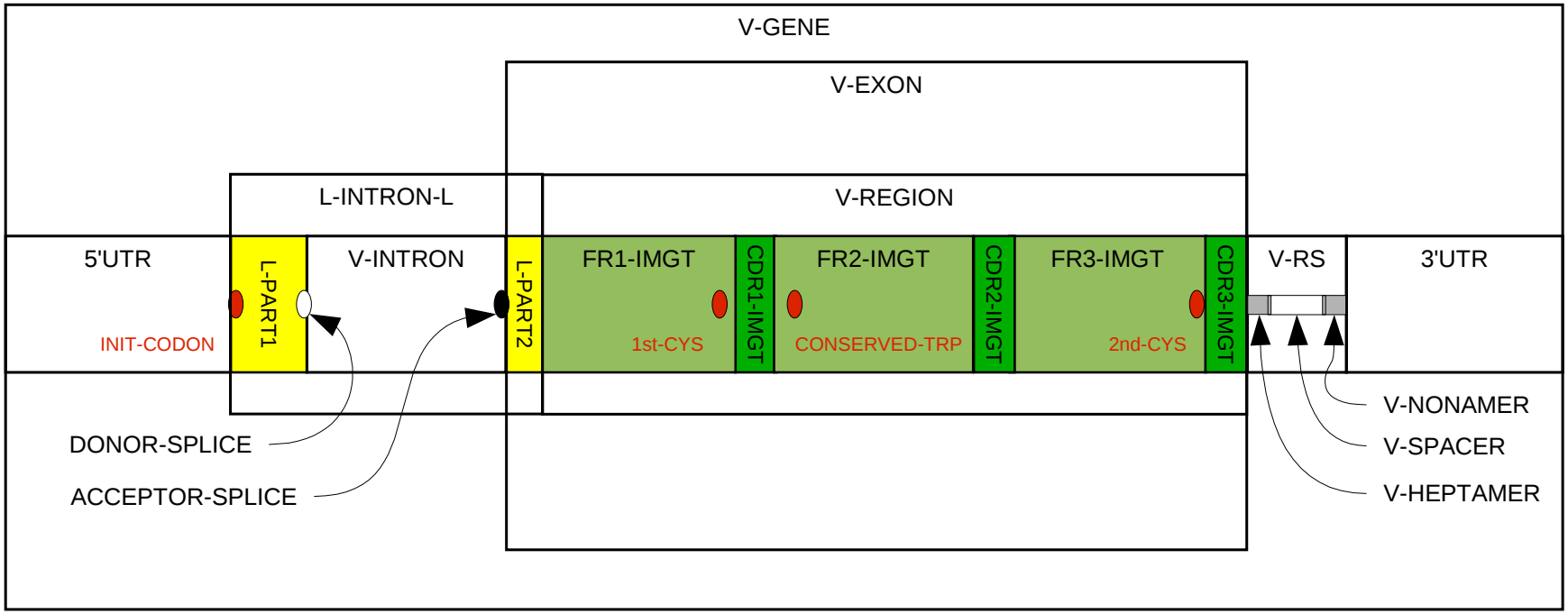
# Overview


IDENTIFICATION
DESCRIPTION
CLASSIFICATION
IMGT object
OBTENTION
NUMEROTATION
LOCALIZATION
ORIENTATION

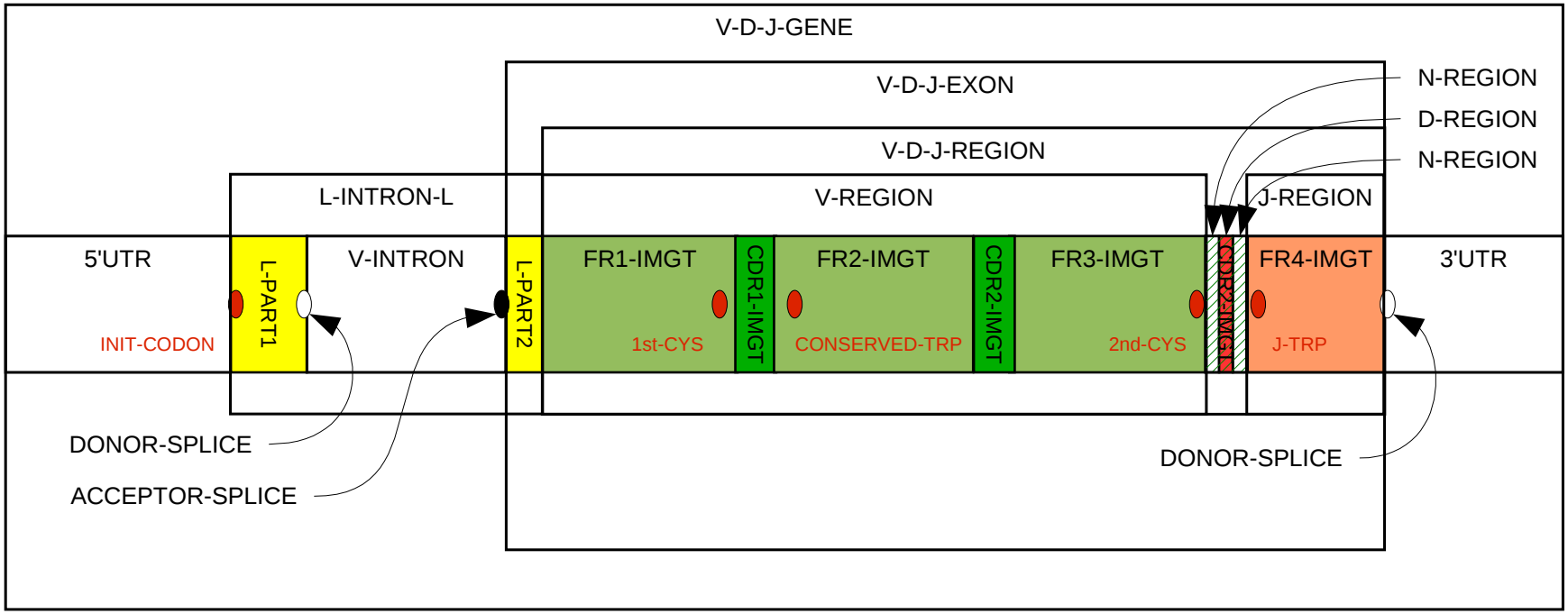
# IDENTIFICATION



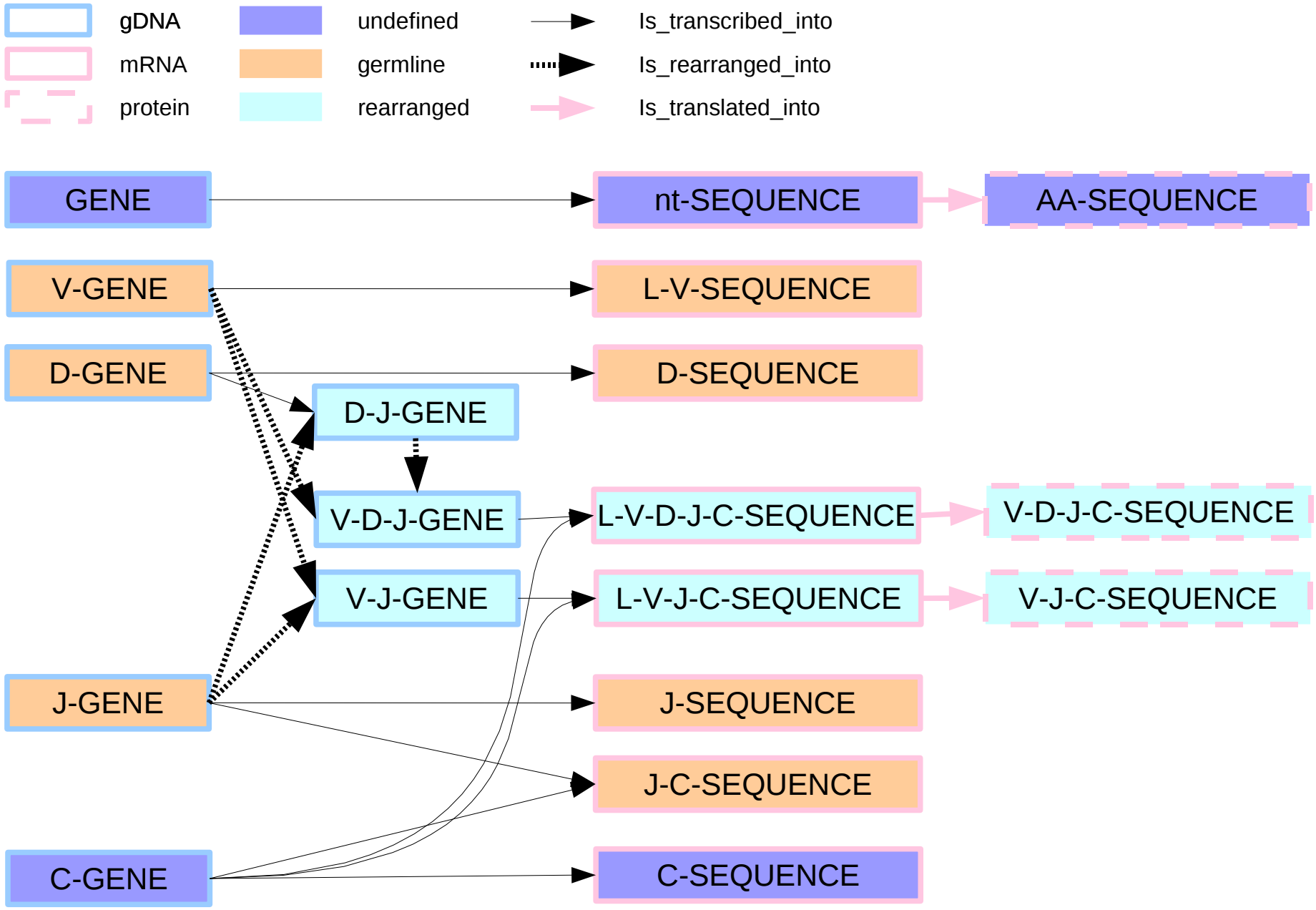
# DESCRIPTION: from V-GENE...



# DESCRIPTION: ....to V-D-J-GENE



# Some relations and dependencies



# About CLASSIFICATION

Name problem: genes and alleles naming

Taxonomy and phylogenetics concern

- Locus, group, subgroup, allele number

  IGHD1-7*01, TRAV24*02

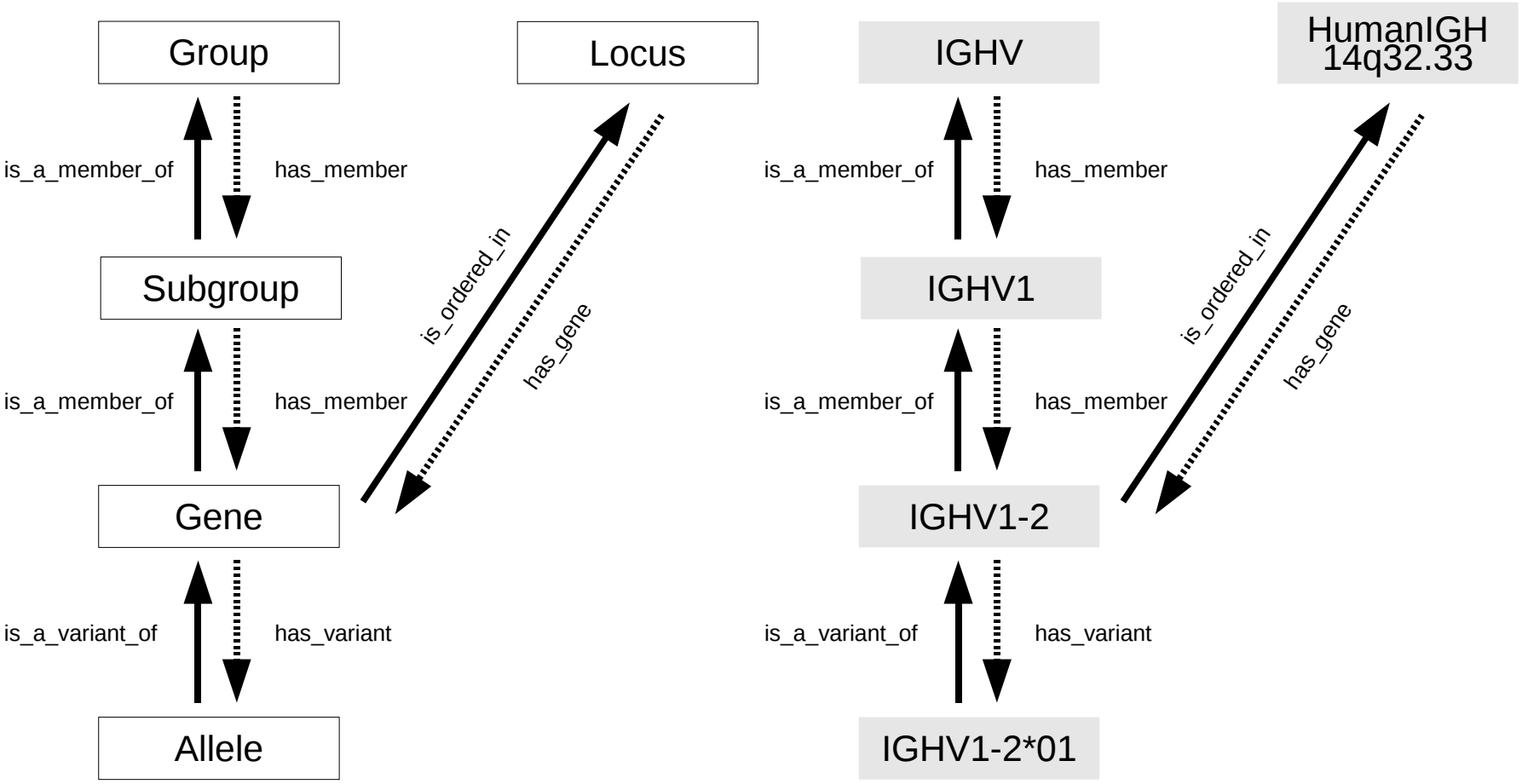
- Must be non-ambiguous

  like a (file)path? IG/H/V/1

- Have as less as possible redundancy

  canonical forms *versus* aliases: TRGV9 is in subgroup 2,
  IGHA1 is in group IGHC...

# CLASSIFICATION axiom



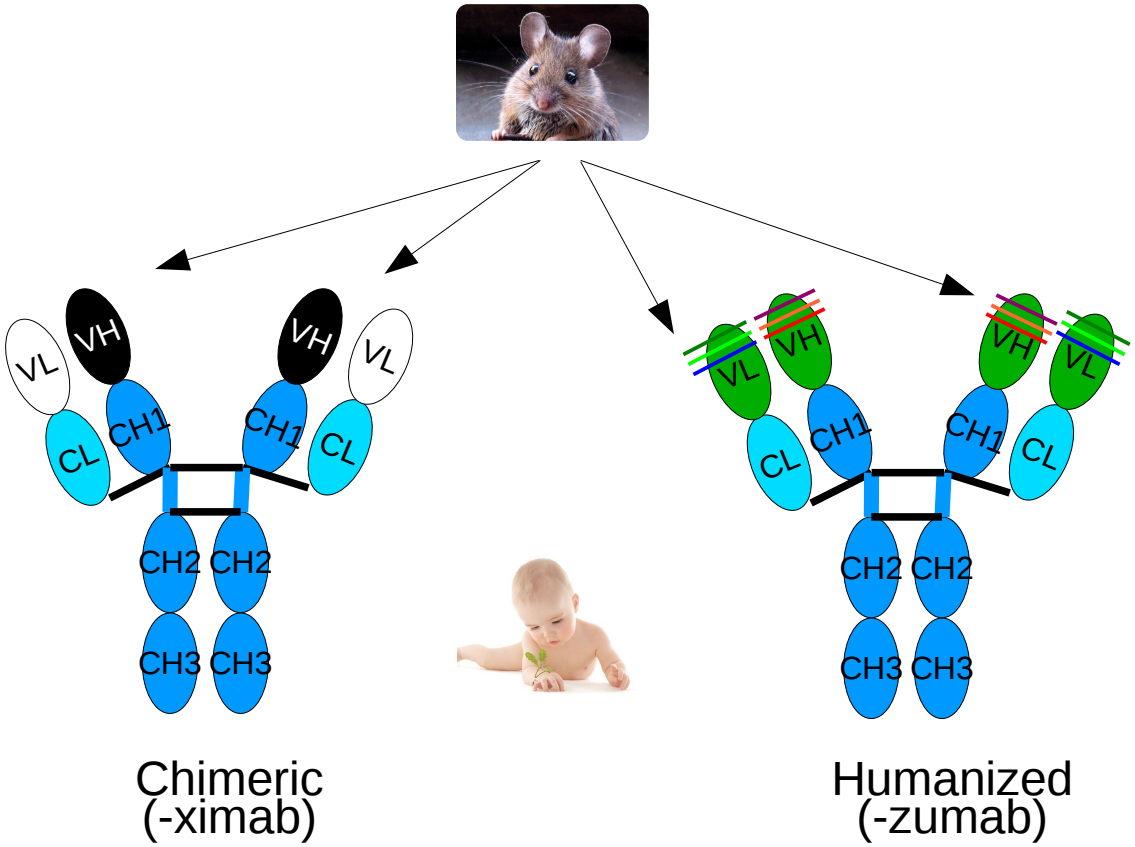
# Towards structural approach

Apply IMGT-ONTOLOGY to complex of proteins (from PDB)

Look at specificity and interaction

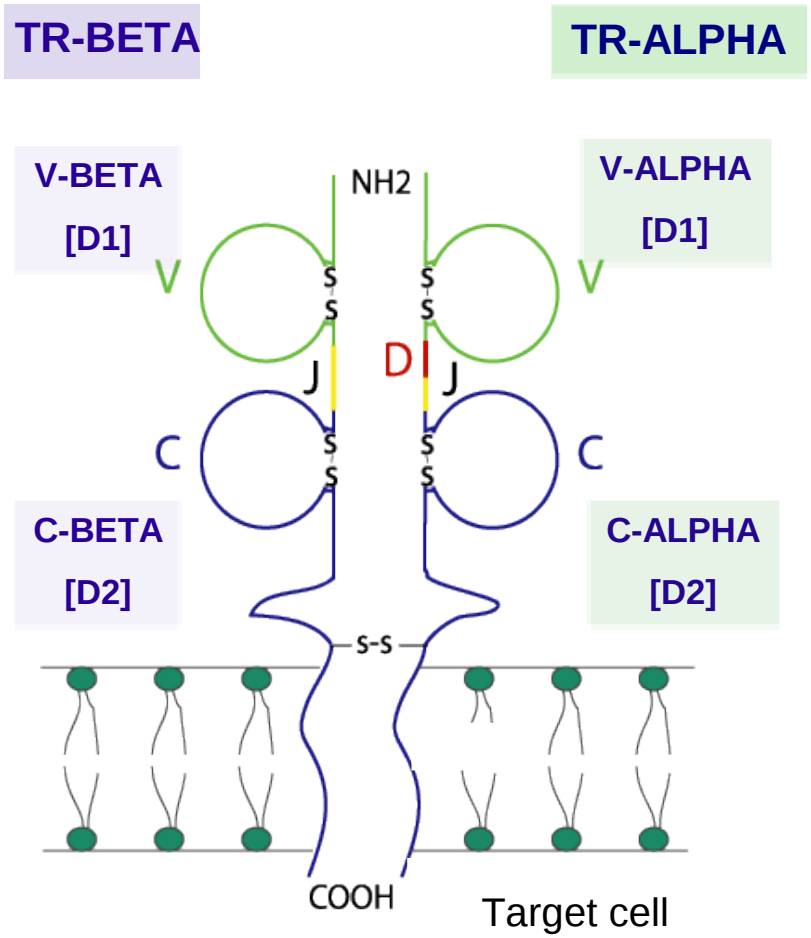
Help antibody engineering

Standardize.
A more clear view?



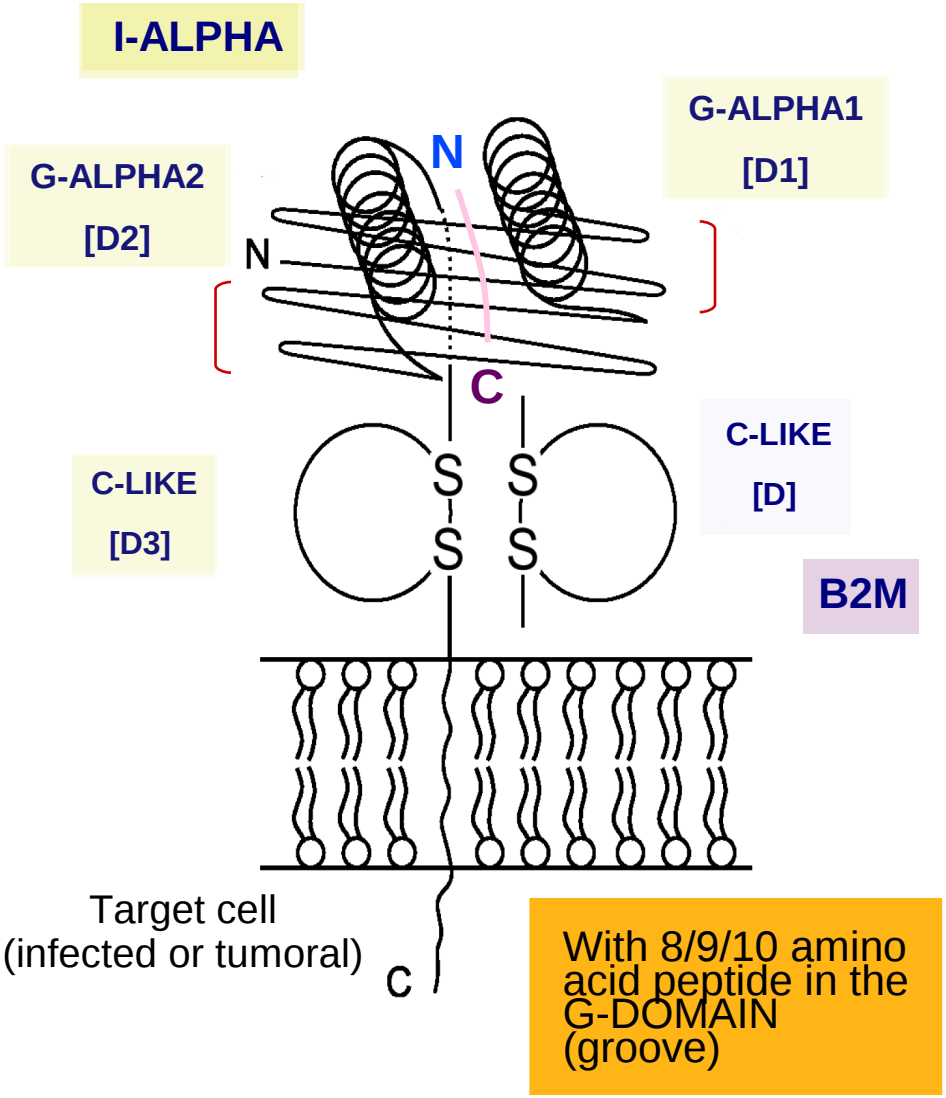
# TR and MHC-I chains and domains

TR-ALPHA_BETA



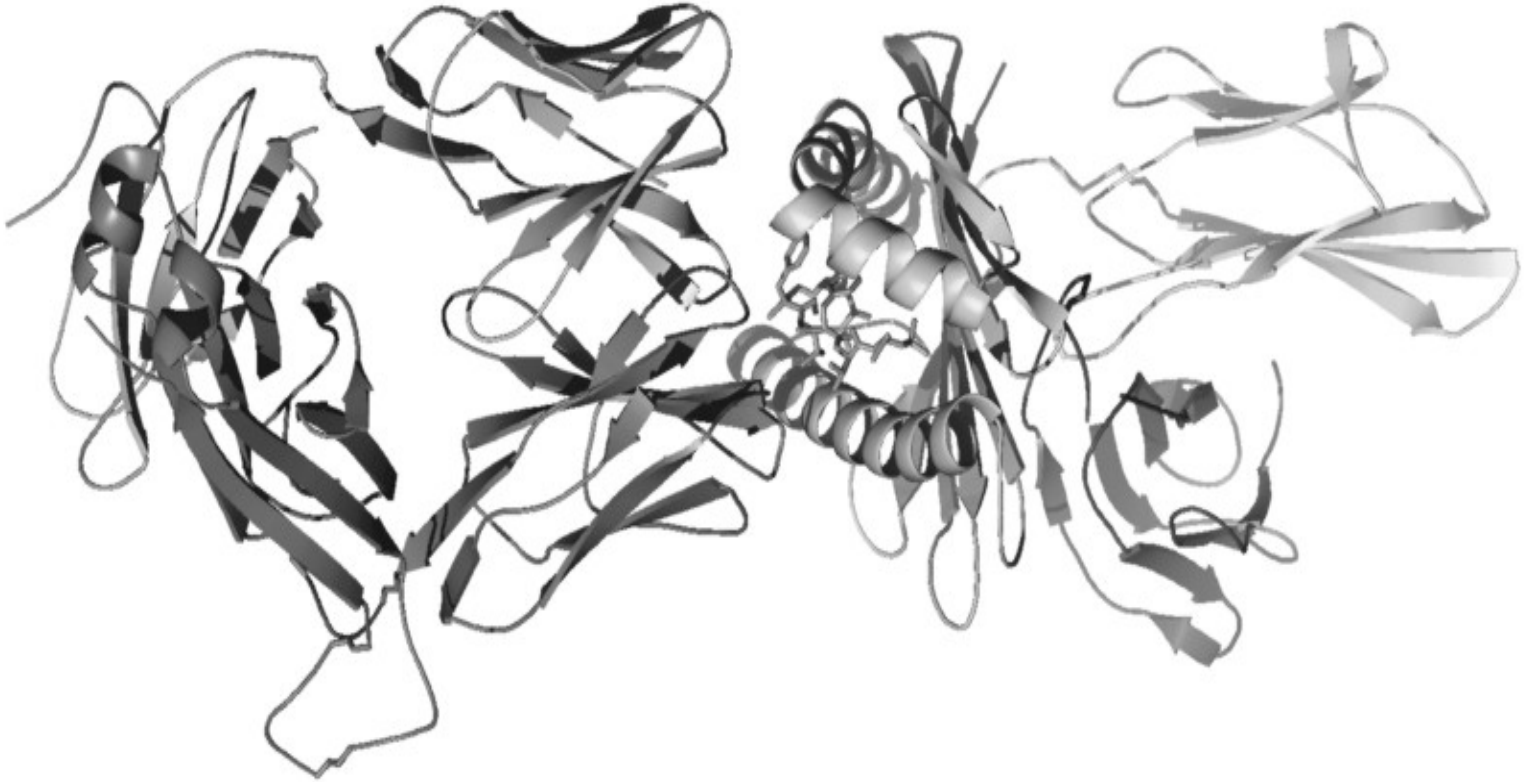

Contribution of the
2 V-DOMAINs
to the antigen binding site

MHC-I-ALPHA_B2M

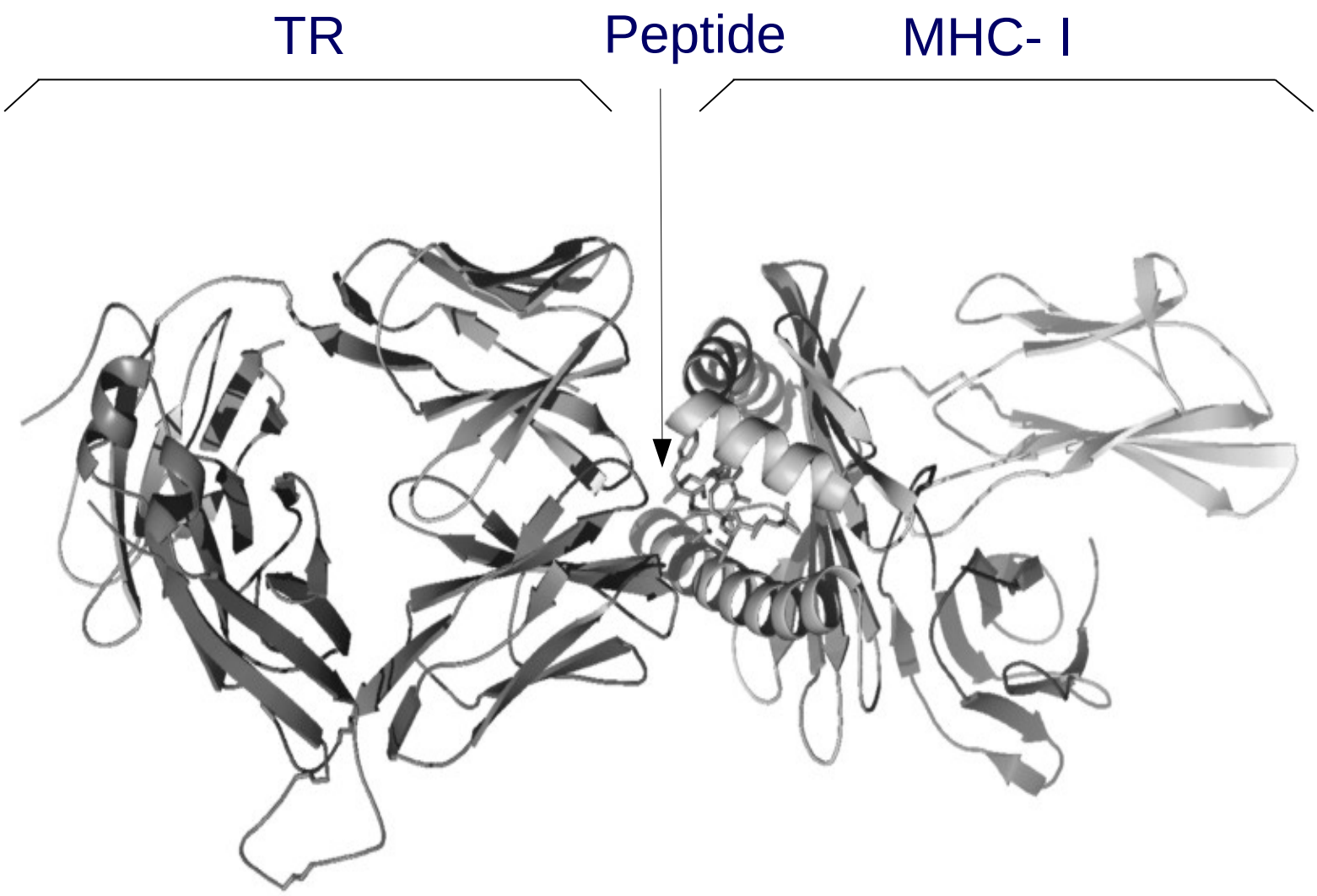



With 8/9/10 amino acid peptide in the G-DOMAIN (groove)

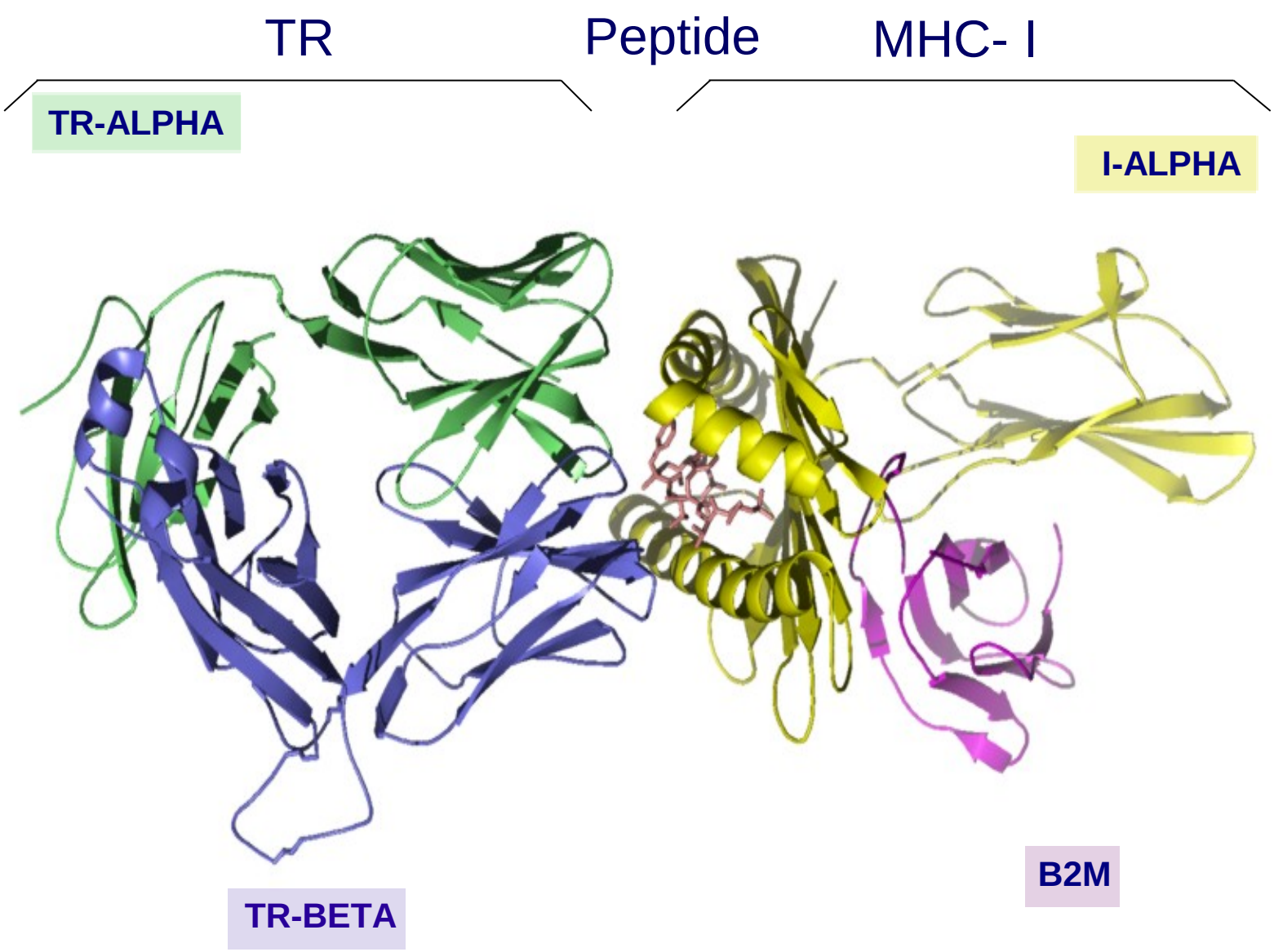
# TR/peptide/MHC-I complex (1ao7)



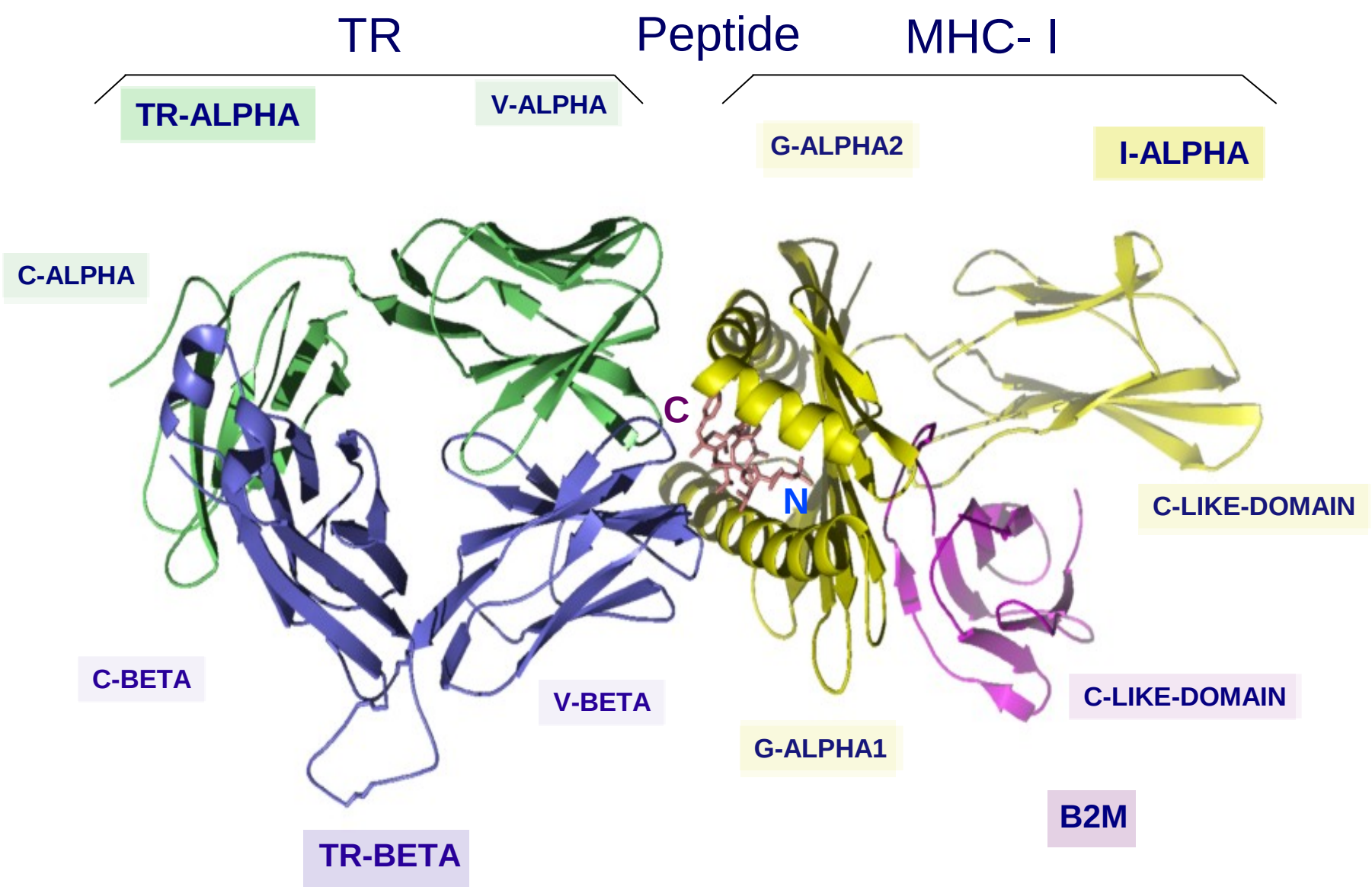
# TR/peptide/MHC-I complex (1ao7)



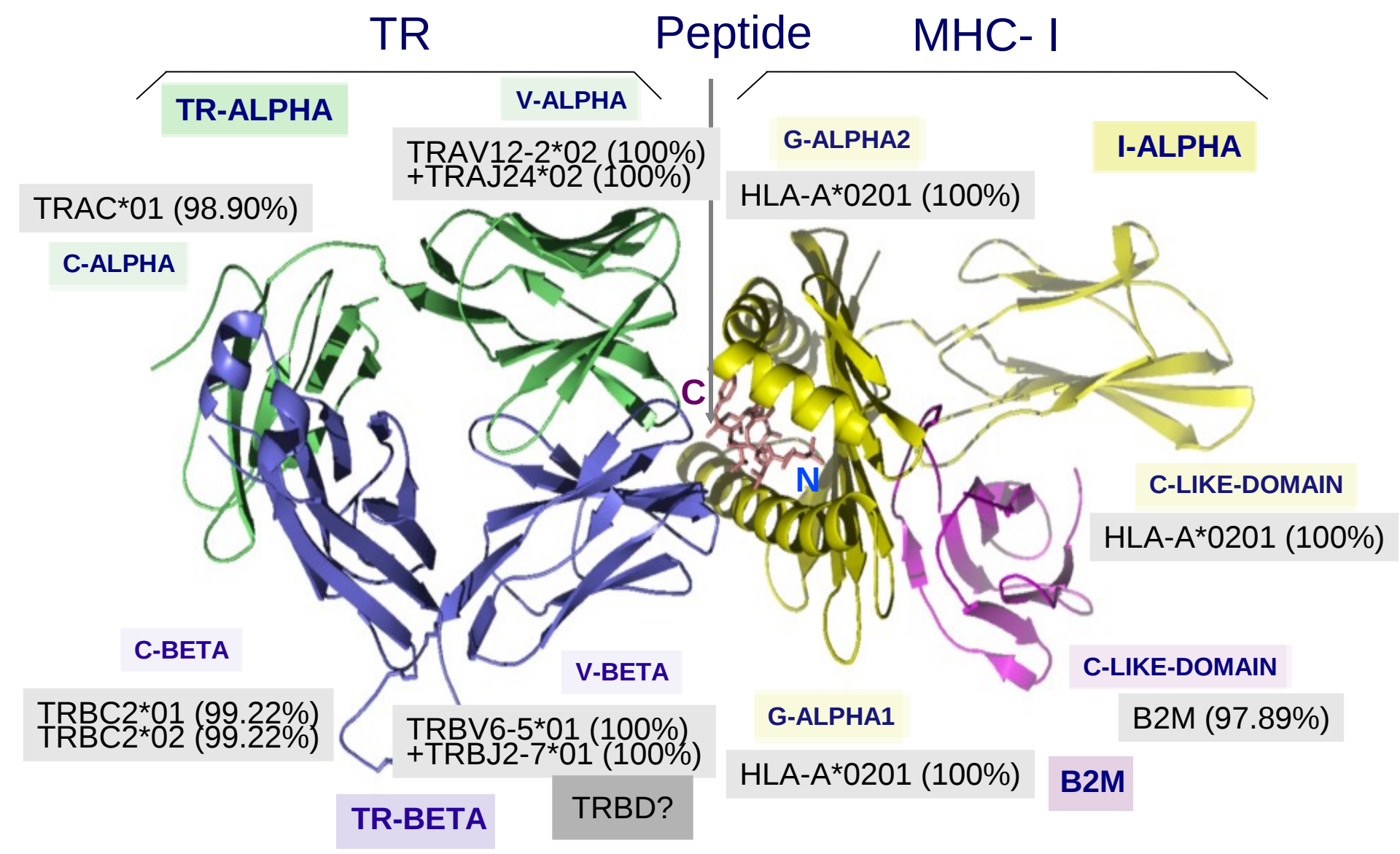
# TR/peptide/MHC-I complex (1ao7)



# TR/peptide/MHC-I complex (1ao7)

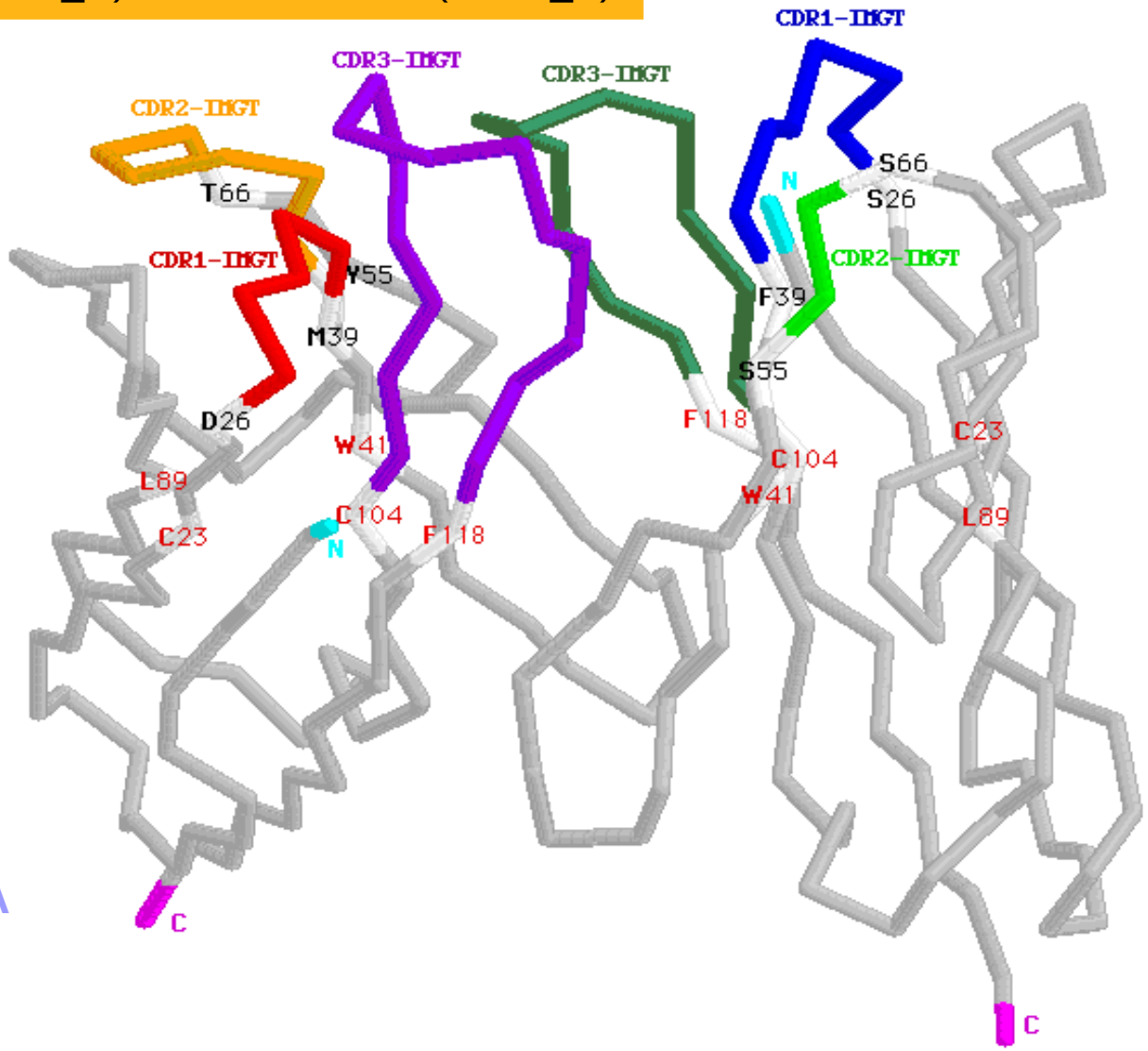


# TR/peptide/MHC-I complex (1ao7)



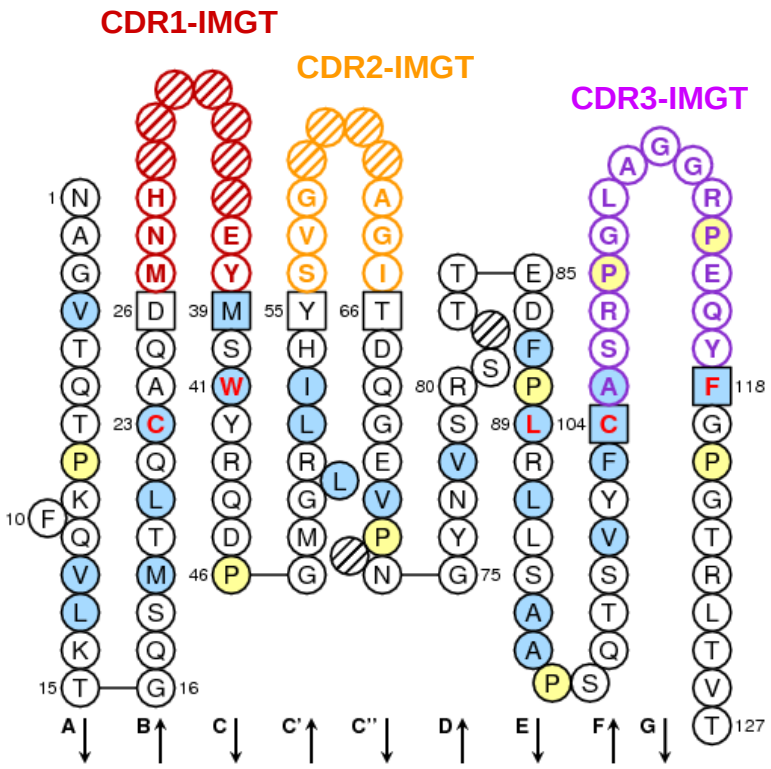
# V-DOMAINs in TR (Human A6)

Complementarity determining regions (CDR) in TR-ALPHA (1AO7_D) and TR-BETA-1 (1AO7_E)

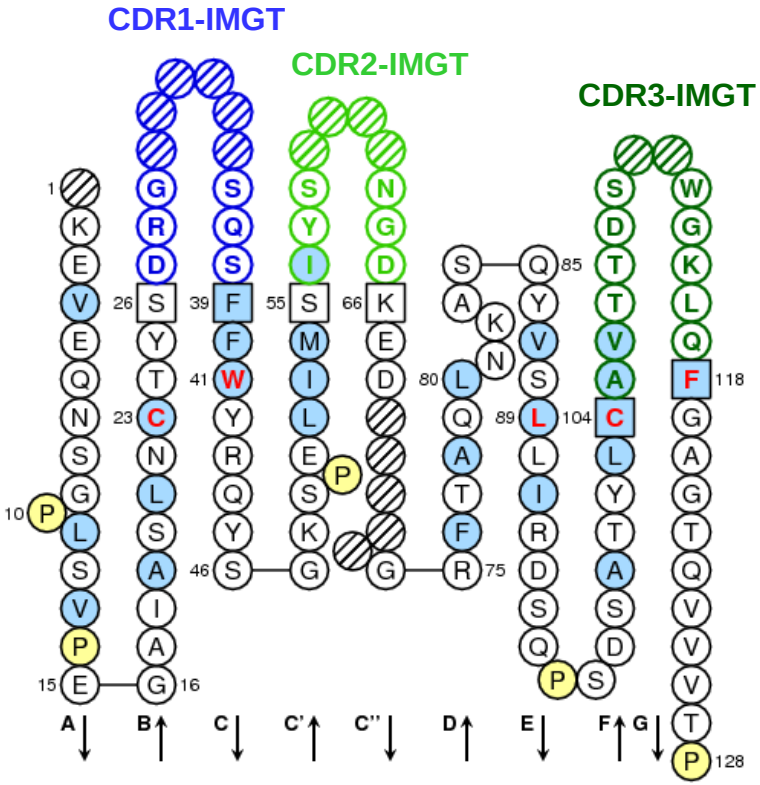


# NUMEROTATION: for V-DOMAIN

Based on IMGT unique numbering of V-DOMAIN




V-BETA
[5.6.14]

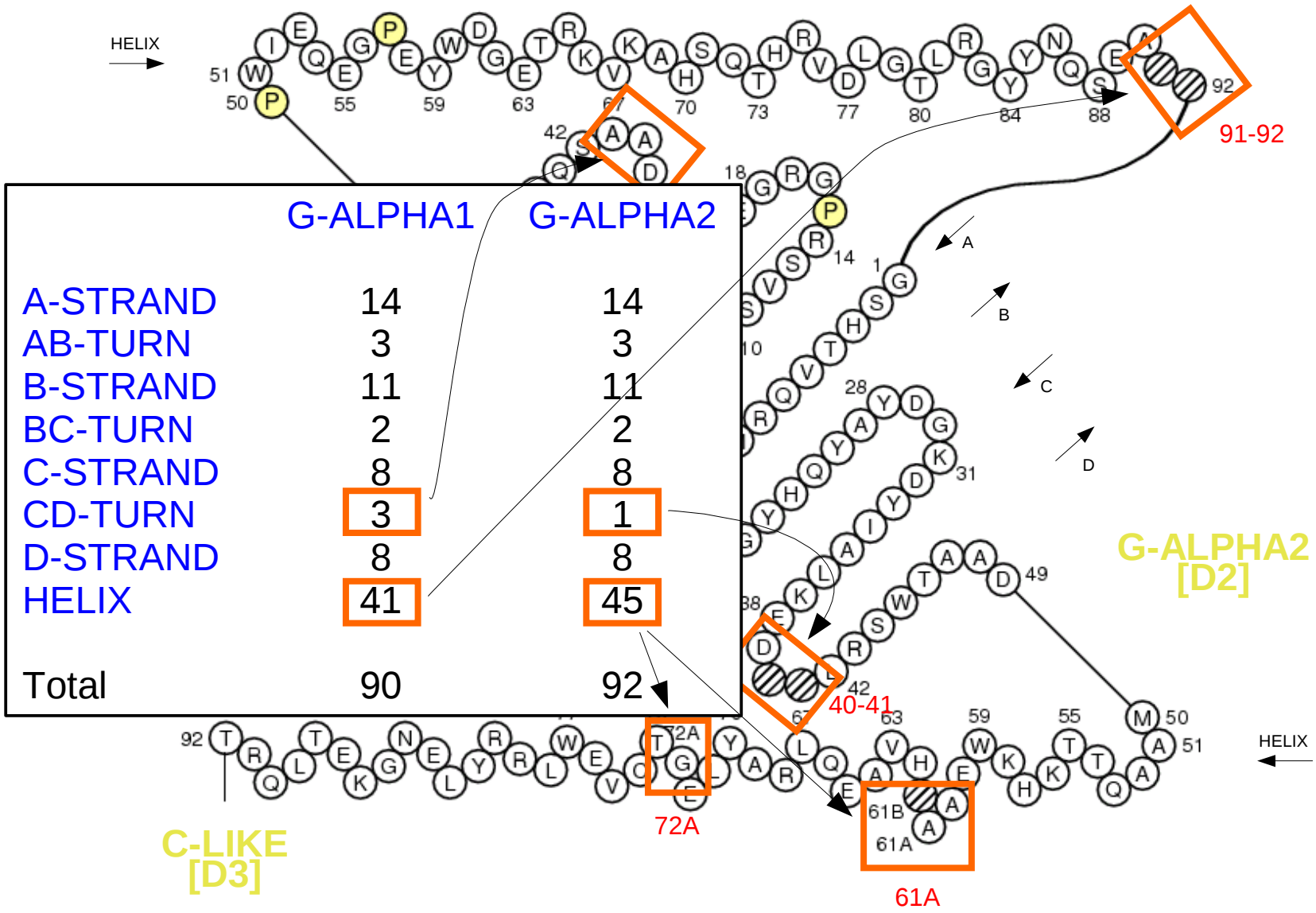



V-ALPHA
[6.6.11]

# V-DOMAIN: FR and CDR lengths

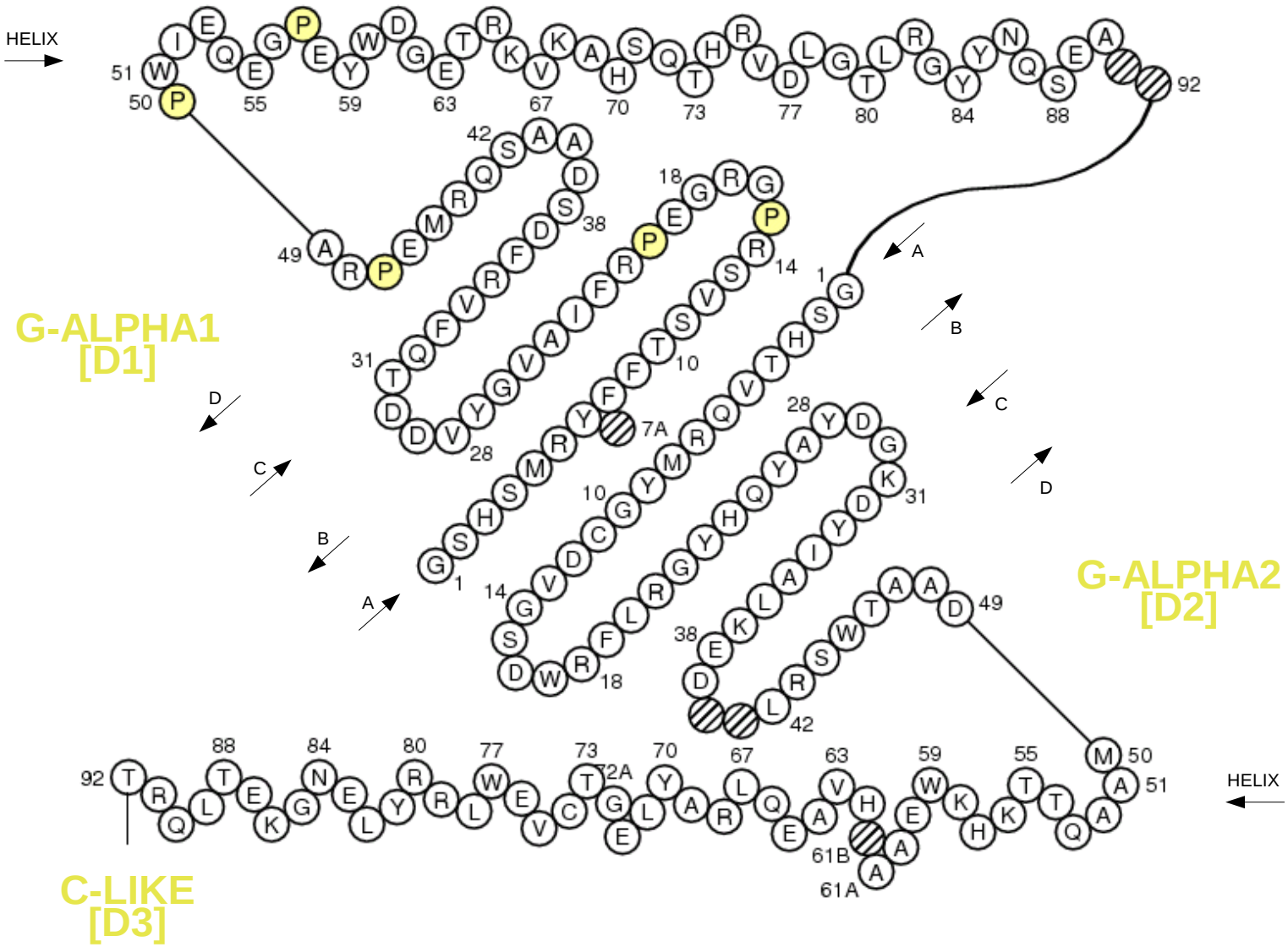
| | | V-BETA | V-ALPHA |
|---|---|---|---|
| A-STRAND | | 15 | 14 |
| AB-TURN | (FR1-IMGT) | 0 | 0 |
| B-STRAND | | 11 | 11 |
| BC-TURN | (CDR1-IMGT) | 5 | 6 |
| C-STRAND | | 8 | 8 |
| CC'-TURN | (FR2-IMGT) | 0 | 0 |
| C'-STRAND | | 9 | 9 |
| C'C''-TURN | (CDR2-IMGT) | 6 | 6 |
| C''-STRAND | | 8 | 4 |
| C''D-TURN | | 0 | 0 |
| D-STRAND | | 9 | 10 |
| DE-TURN | (FR3-IMGT) | 0 | 0 |
| E-STRAND | | 11 | 11 |
| EF-TURN | | 2 | 2 |
| F-STRAND | | 7 | 7 |
| FG-TURN | (CDR3-IMGT) | 14 | 11 |
| G-STRAND | | 10 | 11 |
| Total | | 115 | 110 |

# G-DOMAIN: strand, turn and helix lengths

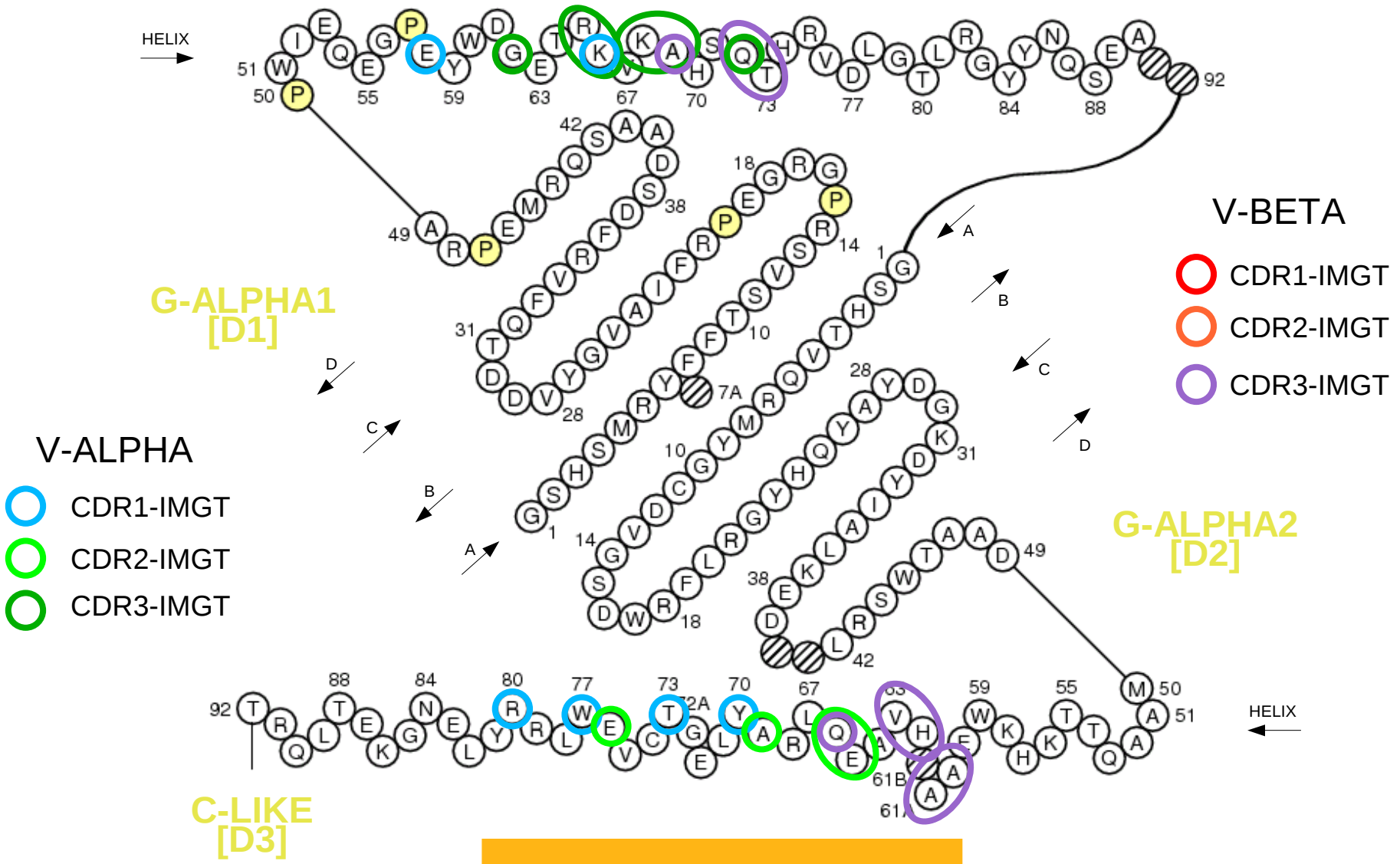



| | G-ALPHA1 | G-ALPHA2 |
|---|---|---|
| A-STRAND | 14 | 14 |
| AB-TURN | 3 | 3 |
| B-STRAND | 11 | 11 |
| BC-TURN | 2 | 2 |
| C-STRAND | 8 | 8 |
| CD-TURN | 3 | 1 |
| D-STRAND | 8 | 8 |
| HELIX | 41 | 45 |
| Total | 90 | 92 |

# G-DOMAIN: IMGT Collier de Perles

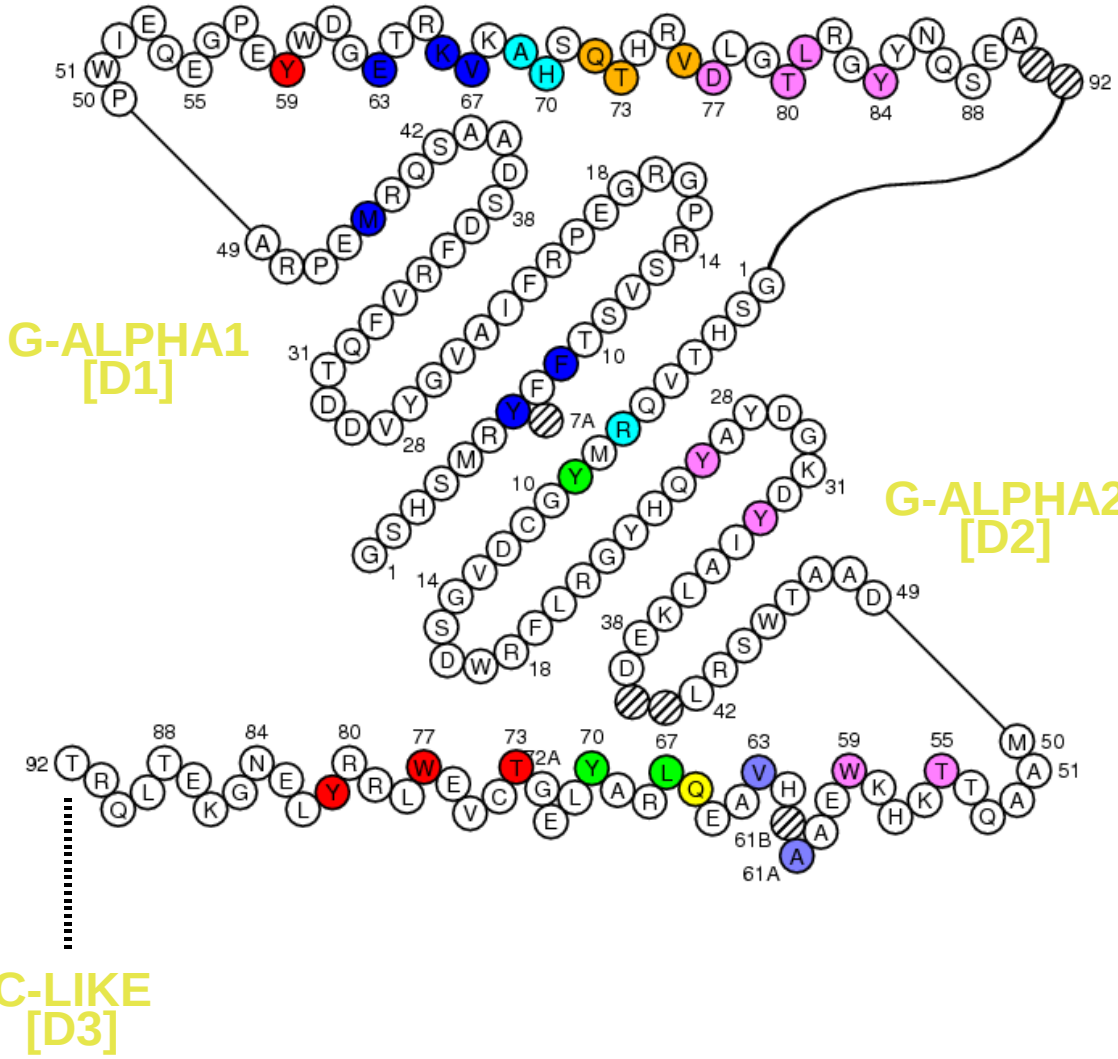


# G-DOMAIN: contacts with TR (A6)



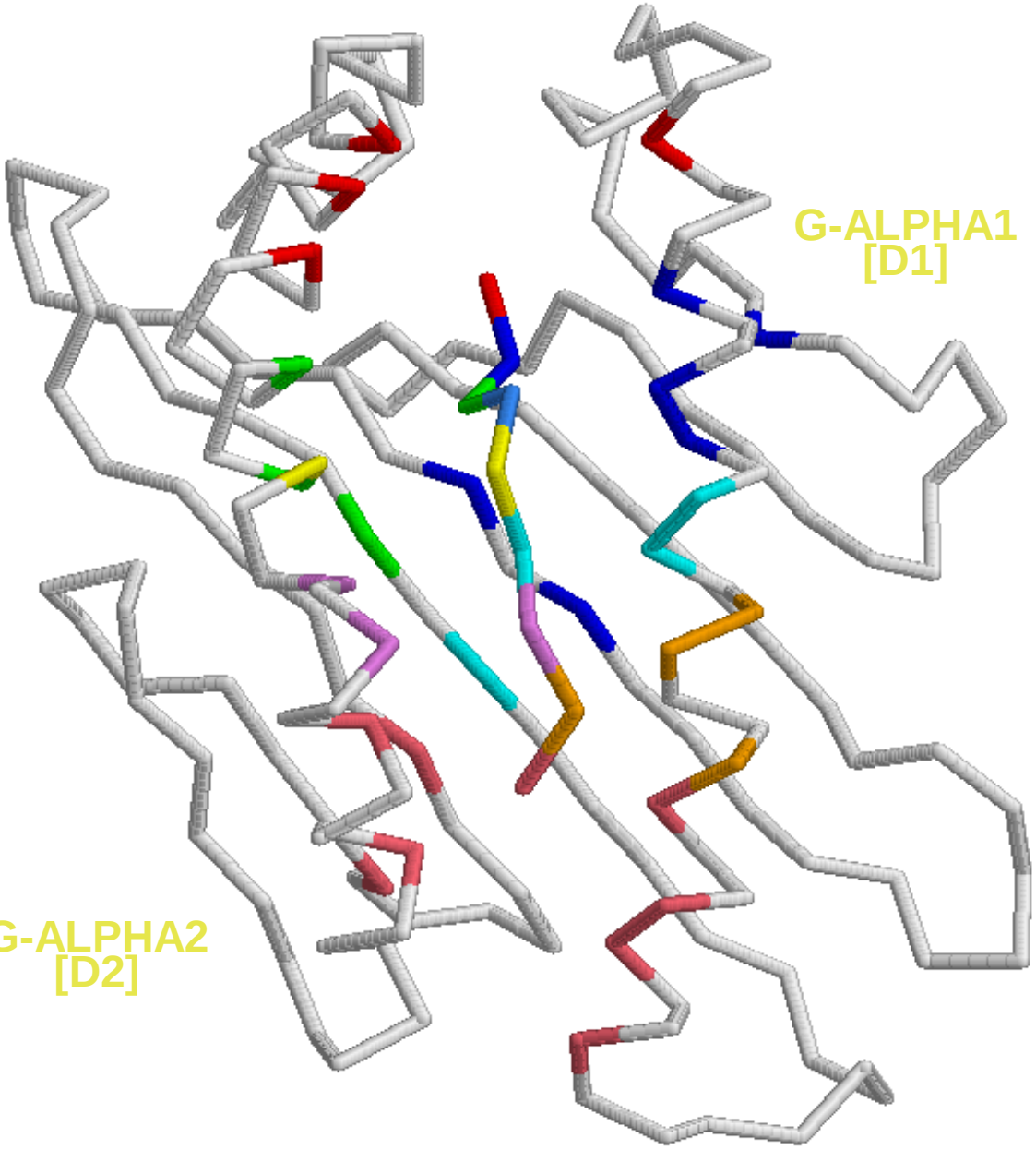

No contact with CDR1-IMGT and CDR2-IMGT of V-BETA

# G-DOMAIN: contacts with (9 AA) peptide




| | 9-amino acid peptides | Peptide |
|---|---|---|
| C1 | 1 | 1 L |
| C2 | - | -- |
| C3 | 2 | 2 L |
| C4 | 3 | 3 F |
| C5 | 4 | (4 G) |
| C6 | 5 | 5 Y |
| C7 | - | -- |
| C8 | 6 | 6 P |
| C9 | 7 | 7 V |
| C10 | 8 | 8 Y |
| C11 | 9 | 9 V |

# HLA-A*0201 contacts with 9 AA peptide




| | 9-amino acid peptides | Peptide |
|---|---|---|
| C1 | 1 | 1 L |
| C2 | - | -- |
| C3 | 2 | 2 L |
| C4 | 3 | 3 F |
| C5 | 4 | (4 G) |
| C6 | 5 | 5 Y |
| C7 | - | -- |
| C8 | 6 | 6 P |
| C9 | 7 | 7 V |
| C10 | 8 | 8 Y |
| C11 | 9 | 9 V |

# TRAV12-2*02 allele in sequence (M81774)

IMGT/LIGM-DB (forthcoming)

| | | |
|---|---|---|
| V-REGION | | [67..345] |
| | **/translation** | QKEVEQNSGPLSVPEGAIASLNCTYSDRGSQSFFWYRQYSGKSPELIMSIYSNGDKEDGRFTAQLNKASQYVSLLIRDSQPSDSATYLCAVYH |
| | | **cagaaggaggtggagcagaattctggacccctcagtgttccagagggagccattgcctctctcaactgcacttacagtgaccgaggttcccagtccttcttctggta**<br>Q K E V E Q N S G P L S V P E G A I A S L N C T Y S D R G S Q S F F W Y<br>R R R W S R I L D P S V F Q R E P L P L S T A L T V T E V P S P S S G<br>E G G G A E F W T P Q C S R G S H C L S Q L H L Q * P R F P V L L L V<br>**Caution:** translation of partial subregions can be erroneous. |
| | **/gene** | TRAV12-2 |
| | **/allele** | TRAV12-2*02 |

## Sequence

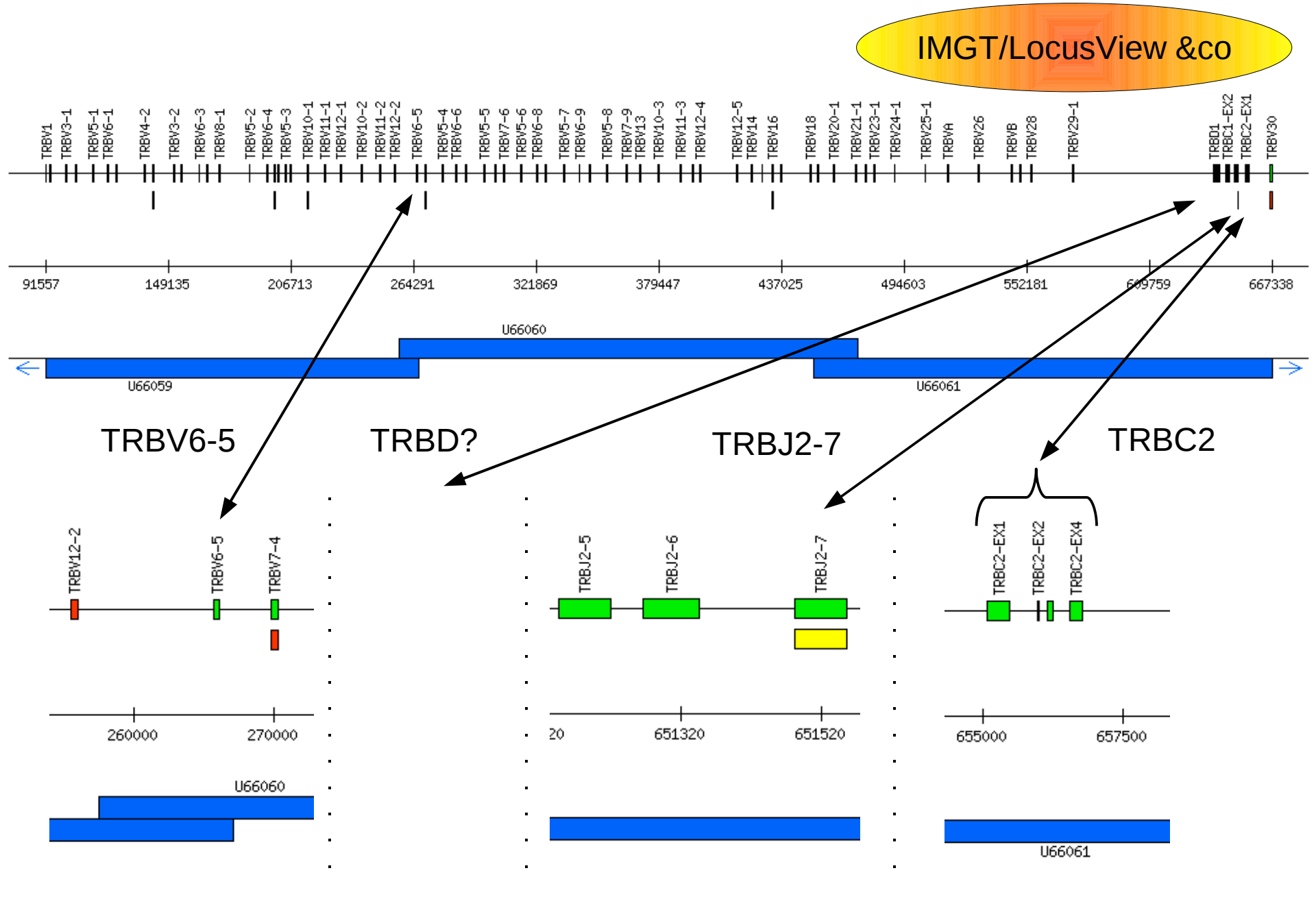
```
1       atgatgaaat ccttgagagt tttactagtg atcctgtggc ttcagttgag ctgggtttgg
61      agccaacaga aggaggtgga gcagaattct ggacccctca gtgttccaga gggagccatt
121     gcctctctca actgcactta cagtgaccga ggttcccagt ccttcttctg gtacagacaa
181     tattctggga aaagccctga gttgataatg tccatatact ccaatggtga caaagaagat
241     ggaaggttta cagcacagct caataaagcc agccagtatg tttctctgct catcagagac
301     tcccagccca gtgattcagc cacctacctc tgtgccgtgt accactctgg ttctgcaaag
361     caactgacct ttggatctgg gacacaattg actgttttac ctgatatcca gaaccctgac
421     cctgccgtgt ac
```

L-V-J-C-REGION [1..432>
L-V-J-REGION [1..402]
L-V-REGION [1..345]
V-J-C-REGION [67..432>
V-J-REGION [67..402]
V-REGION [67..345]
FR2-IMGT [163..213]
CONSERVED-TRP [169..171]

## Literature References

[1] -1..-1 MEDLINE: 86253078
Yoshikai Y., Kimura N., Toyonaga B., Mak T.W.,
*"Sequences and repertoire of human T cell receptor alpha chain variable region genes in mature T lymphocytes"*;
Journal: J. Exp. Med. 164(1) [1986]

# Human TRB locus at 7q34



# TR/peptide/MHC-I available complexes

IMGT/3Dstructure-DB

| | IMGT entry ID | IMGT protein name | IMGT receptor description | Species | Ligand(s) | Experimental technique | Resolution | PDB release date |
|---|---|---|---|---|---|---|---|---|
| 1 | 1ao7 | A6<br>HLA-A*0201 | TR-ALPHA_BETA-1<br>MHC-I-ALPHA_B2M | *Homo sapiens* | Tax peptide 11-19 (Q82235) [Human T lymphotropic virus type 1] | X-ray diffraction | 2.6 | 17-SEP-97 |
| 2 | 1bd2 | HLA-A*0201<br>B7 | MHC-I-ALPHA_B2M<br>TR-ALPHA_BETA-1 | *Homo sapiens* | Tax peptide 11-19 (Q82235) [HTLV-1] | X-ray diffraction | 2.5 | 19-AUG-98 |
| 3 | 1fo0 | BM3.3<br>H2-K1b | FV-ALPHA_BETA<br>MHC-I-ALPHA_B2M | *Mus musculus* | pBM1 peptide [Mouse] | X-ray diffraction | 2.50 | 02-OCT-00 |
| 4 | 1g6r | 2C<br>H2-K1b | TR-ALPHA_BETA-1<br>MHC-I-ALPHA_B2M | *Mus musculus* | Superantagonist peptide SIYR [Chimeric] | X-ray diffraction | 2.80 | 15-NOV-00 |
| 5 | 1jtr | H2-K1b<br>2C | MHC-I-ALPHA_B2M<br>TR-ALPHA_BETA-1 | *Mus musculus* | NADH-ubiquinone oxidoreductase MLRQ subunit peptide 61-68 (Q62425) [Mouse] | X-ray diffraction | 2.40 | 15-MAY-02 |
| 6 | 1kj2 | H2-K1b<br>KB5-C20 | MHC-I-ALPHA_B2M<br>FV-ALPHA_BETA | *Mus musculus* | GTP-binding protein 1 peptide 161-168 pKB1 (O08582) [Mouse] | X-ray diffraction | 2.71 | 27-MAR-02 |
| 7 | 1lp9 | HLA-A*0201<br>12.2 | MHC-I-ALPHA_B2M<br>TR-ALPHA_BETA-1 | *Homo sapiens*<br>*Mus musculus* | Self peptide P1049 [Human] | X-ray diffraction | 2.00 | 11-NOV-03 |
| 8 | 1mi5 | LC13<br>HLA-B*0801 | TR-ALPHA_BETA-1<br>MHC-I-ALPHA_B2M | *Homo sapiens* | EBNA-3A peptide 193-201 (P12977), I9>L [EBV] | X-ray diffraction | 2.50 | 04-FEB-03 |
| 9 | 1mwa | H2-K1b<br>2C | MHC-I-ALPHA_B2M<br>TR-ALPHA_BETA-1 | *Mus musculus* | NADH-ubiquinone oxidoreductase MLRQ subunit peptide 61-68 (Q62425) [Mouse] | X-ray diffraction | 2.40 | 27-NOV-02 |
| 10 | 1nam | BM3.3<br>H2-K1b | FV-ALPHA_BETA<br>MHC-I-ALPHA_B2M | *Mus musculus* | Nucleocapsid protein VSV8 peptide 52-59 (P11212) [Stomatitis] | X-ray diffraction | 2.70 | 11-MAR-03 |
| 11 | 1oga | JM22<br>HLA-A*0201 | TR-ALPHA_BETA-1<br>MHC-I-ALPHA_B2M | *Homo sapiens* | Matrix protein M1 peptide 58-66 (Q66PA1) [Influenza A virus] | X-ray diffraction | 1.4 | 11-JUL-03 |
| 12 | 1qrn | A6<br>HLA-A*0201 | TR-ALPHA_BETA-1<br>MHC-I-ALPHA_B2M | *Homo sapiens* | Tax peptide 11-19 (Q82235), P6>A [HTLV-1] | X-ray diffraction | 2.80 | 08-JUN-01 |
| 13 | 1qse | HLA-A*0201<br>A6 | MHC-I-ALPHA_B2M<br>TR-ALPHA_BETA-1 | *Homo sapiens* | Tax peptide 11-19 (Q82235), V7>R [HTLV-1] | X-ray diffraction | 2.80 | 21-DEC-99 |
| 14 | 1qsf | A6<br>HLA-A*0201 | TR-ALPHA_BETA-1<br>MHC-I-ALPHA_B2M | *Homo sapiens* | Tax peptide 11-19 (Q82235), Y8>A [HTLV-1] | X-ray diffraction | 2.80 | 21-DEC-99 |
| 15 | 1ypz | G8<br>H2-T22 | TR-GAMMA-1_DELTA<br>MHC-I-ALPHA_B2M | *Mus musculus*<br>*Homo sapiens* | | X-ray diffraction | 3.40 | 12-APR-05 |
| 16 | 2ak4 | SB27<br>HLA-B*3508 | TR-ALPHA_BETA-1<br>MHC-I-ALPHA_B2M | *Homo sapiens* | BZLF1 trans-activator protein peptide 52-64 (P03206) [EBV] | X-ray diffraction | 2.50 | 11-OCT-05 |
| 17 | 2ckb | H2-K1b<br>2C | MHC-I-ALPHA_B2M<br>TR-ALPHA_BETA-1 | *Mus musculus* | NADH-ubiquinone oxidoreductase MLRQ subunit peptide 61-68 (Q62425) [Mouse] | X-ray diffraction | 3.2 | 09-SEP-98 |
| 18 | | KK50.4 | TR-ALPHA_BETA-1 | *Homo sapiens* | PEPTIDE FROM CMV GPUL40 [Cytomegalovirus] | X-ray diffraction | 2.60 | 21-MAR-06 |
| 19 | | | | *Homo sapiens* | NY-ESO-1 tumor-associated antigen | X-ray diffraction | 2.10 | 25-APR-06 |
| 20 | | | | *Homo sapiens* | NY-ESO-1 tumor-associated antigen | X-ray diffraction | 2.70 | 25-APR-06 |
| 21 | | | | *Homo sapiens* | EBV PEPTIDE | X-ray diffraction | 2.70 | 27-FEB-07 |

| | |
|---|---|
| Human | 12 |
| Mouse | 7 |
| Human/Mouse | 2 |
| Total | 21 |

# One residue contact analysis

IMGT/3Dstructure-DB

## IMGT Residue@Position card

Residue@Position: 113 - ARG (R) - V-BETA - 1ao7_E

CDR3-IMGT

**General information:**

| | |
|---|---|
| PDB file numbering | 102 |
| IMGT file numbering | 113 |
| Residue full name | Arginine |
| Formula | C6 H15 N4 O2 1+ |

**IMGT LocalStructure@Position**

| | |
|---|---|
| Secondary structure | Coil |
| Phi (in degrees) | -89.71 |
| Psi (in degrees) | 111.56 |
| ASA (in square angstrom) | 73.2 |

| IMGT Num | Residue | | Domain | Chain | Atom contacts | Polar | Hydrogen Bond | Non Polar |
|---|---|---|---|---|---|---|---|---|
| 61 | ALA | A | G-ALPHA2 | 1ao7_A | 5 | 2 | 1 | 3 |
| 61A | ALA | A | G-ALPHA2 | 1ao7_A | 24 | 6 | 0 | 18 |
| 62 | HIS | H | G-ALPHA2 | 1ao7_A | 12 | 2 | 0 | 10 |
| 66 | GLN | Q | G-ALPHA2 | 1ao7_A | 2 | 1 | 0 | 1 |
| 5 | TYR | Y | | 1ao7_C | 1 | 0 | 0 | 1 |
| 108 | PRO | P | V-BETA | 1ao7_E | 15 | 1 | 0 | 14 |
| 111 | ALA | A | V-BETA | 1ao7_E | 6 | 2 | 0 | 4 |
| 112.1 | GLY | G | V-BETA | 1ao7_E | 24 | 5 | 0 | 19 |
| 115 | GLU | E | V-BETA | 1ao7_E | 17 | 3 | 0 | 14 |

G-ALPHA2

Peptide

V-BETA

# Exploring known Human cDNA TRAV12-2

IMGT/V-QUEST

Search for some specificities with known data

(no D)

| Sequence | V-GENE and allele | Functionality | V Score | V Identity | J-GENE and allele | CDR-IMGT lengths | AA JUNCTION | JUNCTION frame |
|---|---|---|---|---|---|---|---|---|
| M27369\|M27369\|Homo sapiens (cl | TRAV12-2*01 | Unproductive (stop codons) | 791 | 80,99% (196/242 nt) | TRAJ52*01 | [12,10,14] | CAVKPAGGTSYGKLTF | in frame |
| S60781\|S60781\|Homo sapiens T-c | TRAV12-2*01, or TRAV12-2*02 | No rearrangement found | 636 | 99,22% (128/129 nt) | - | [12,X,X] | - | - |
| S82064\|S82064\|V alpha 2.1-J al | TRAV12-2*01 | Productive | 1075 | 100,00% (216/216 nt) | TRAJ39*01 | [12,10,12] | CAVNAGNAGNMLTF | in frame |
| S82066\|S82066\|V alpha 2.1-J al | TRAV12-2*01 | Productive | 955 | 100,00% (192/192 nt) | TRAJ45*01 | [12,10,10] | CAVNEGADGLTF | in frame |
| X92783\|HSXPMS2A\|H.sapiens mRNA | TRAV12-2*01 | Productive | 1142 | 99,14% (231/233 nt) | TRAJ57*01 | [12,10,15] | CAVNIVGTQGGSEKLVF | in frame |
| AF020651\|AF020651\|Homo sapiens | TRAV12-2*01 | Productive | 1325 | 99,63% (266/267 nt) | TRAJ29*01 | [12,10,12] | CSVMSNSGNTPLVF | in frame |
| AF327017\|AF327017\|Homo sapiens | TRAV12-2*01 | Productive | 1000 | 100,00% (201/201 nt) | TRAJ43*01 | [12,10,11] | CAVDAADNNDMRF | in frame |
| M17652\|HSTCAYN\|Human T-cell re | TRAV12-2*02 | Productive | 1245 | 100,00% (250/250 nt) | TRAJ15*01 | [12,10,13] | CAVNIPNQAGTALIF | in frame |
| M17653\|HSTCAYO\|Human T-cell re | TRAV12-2*02 | Productive | 1245 | 100,00% (250/250 nt) | TRAJ15*01 | [12,10,10] | CAPKPGGTALIF | in frame |
| M81774\|HSIGTCACA\|Homo sapiens | TRAV12-2*02 | Productive | 1330 | 100,00% (267/267 nt) | TRAJ22*01 | [12,10,12] | CAVYHSGSARQLTF | in frame |
| U40464\|HS404641\|Human T cell r | TRAV12-2*02, or TRAV12-2*03 | Productive | 875 | 100,00% (176/176 nt) | TRAJ41*01 | [12,10,13] | CALKGRSNSGYALNF | in frame |
| X58746\|X58746\|Human mRNA for T | TRAV12-2*02 | No rearrangement found | 1330 | 100,00% (267/267 nt) | - | [12,10,X] | - | - |
| X92883\|HSPHC46A1\|H.sapiens mRN | TRAV12-2*02 | Productive | 1142 | 99,14% (231/233 nt) | TRAJ34*01 | [12,10,11] | CAVPFYNTDKLIF | in frame |
| AF532854\|AF532854\|Homo sapiens | TRAV12-2*02 | Productive | 1161 | 99,57% (234/235 nt) | TRAJ32*02 | [12,10,10] | CADGGATNKLIF | in frame |
| M13724\|HSTCAXH\|Human T-cell re | TRAV12-2*03 | Rearranged sequence (but no junction found) | 1210 | 100,00% (243/243 nt) | TRAJ16*01 | [12,10,X] | - | - |
| M13725\|HSTCAXI\|Human T-cell re | TRAV12-2*03 | Productive | 1210 | 100,00% (243/243 nt) | TRAJ18*01 | [12,10,14] | CAVNYPRGTTLGRLYF | in frame |
| X04946\|HSTCRA12\|Human mRNA for | TRAV12-2*03 | Productive | 1210 | 100,00% (243/243 nt) | TRAJ18*01 | [12,10,14] | CAVNYPRGTTLGRLYF | in frame |

# Corresponding available alignment

IMGT/V-QUEST

**6. V-REGION protein display (mutations displayed)**

```
                                  FR1-IMGT          CDR1-IMGT       FR2-IMGT      CDR2-IMGT                FR3-IMGT
                                  (1-26)            (27-38)         (39-55)       (56-65)                  (66-104)
                                  ____________________ ___________ _____________ _________ ________________________________________
                                  1       10        20   30        40        50    60        70        80  ABC   90        100
                                  .........|.........|..... ...|..... .|.........|.... ....|..... ....|.........|.....|.....|....
AE000659 TRAV12-2*01              QKEVEQNSGPLSVPEGAIASLNCTYS DRGSQS..... FFWYRQYSGKSPELIMF IYSNG..... DKEDG.......RFTAQLNKASQYVSLLIRDSQPSDSATYLC AVN
M13724|HSTCAXH|Human T-cell r ........------------------- --V---..... ----------------S ----..... -----.......------------------------------ -LRDGQKLLFARGTMLKVDL
M13725|HSTCAXI|Human T-cell r ........------------------- --V---..... ----------------S ----..... -----.......------------------------------ ---YPRGTTLGRLYFGRGTQLTVWP
M17652|HSTCAYN|Human T-cell r ......--------------------- ------..... ----------------S ----..... -----.......------------------------------ ---IPNQAGTALIFGKGTTLSVRS
M17653|HSTCAYO|Human T-cell r ......--------------------- ------..... ----------------S ----..... -----.......------------------------------ -PKPGGTALIFGKGTTLSVSS
M27369|M27369|Homo sapiens (c .........TPQCSR-SHCLSQLHLQ *PS---..... ---------------- ----..... -----.......------------------------------ --KPAGGTSYGKLTFGQGTILTVHP
M81774|HSIGTCACA|Homo sapiens ------------------------- ------..... ----------------S ----..... -----.......------------------------------ --YHSGSARQLTFGSGTQLTVLP
S60781|S60781|Homo sapiens T- ------------------------- --V---..... ----------
S82064|S82064|V alpha 2.1-J a ................---------- ------..... ---------------- ----..... -----.......------------------------------ ---AGNAGNMLTFGGGTRLMVKP
S82066|S82066|V alpha 2.1-J a ........................- ------..... ---------------- ----..... -----.......------------------------------ ---EGADGLTFGKGTHLIIQP
U40464|HS404641|Human T cell  ......................... .....-..... ----------------S ----..... -----.......------------------------------ -LKGRSNSGYALNFGKGTSLLVTP
X04946|HSTCRA12|Human mRNA fo ........------------------- --V---..... ----------------S ----..... -----.......------------------------------ ---YPRGTTLGRLYFGRGTQLTVWP
X58746|X58746|Human mRNA for  ------------------------- ------..... ----------------S ----..... -----.......------------------------------ -
X92783|HSXPMS2A|H.sapiens mRN ............----------S-- ------..... ---------------- ----..... -----.......----------M------------- ---IVGTQGGSEKLVFGKGTKLTVNP
X92883|HSPHC46A1|H.sapiens mR ............------------- ------..... --------R-------T ----..... -----.......------------------------------ --PFYNTDKLIFGTGTRLQVFP
AF020651|AF020651|Homo sapien ------------------------- ------..... ---------------- ----..... -X---.......------------------------------ S-MSNSGNTPLVFGKGTRLSVIX
AF327017|AF327017|Homo sapien ......................---- ------..... ---------------- ----..... -----.......------------------------------ --DAADNNDMRFGAGTRLTVKP
AF532854|AF532854|Homo sapien ..........-------V------- ------..... ----------------S ----..... -----.......------------------------------ -DGGATNKLIFGTGTLLAVQP
```

```
1ao7_D|TR-ALPHA .KEVEQNSGPLSVPEGAIASLNCTYS DRGSQS..... FFWYRQYSGKSPELIMS IYSNG..... DKEDG.......RFTAQLNKASQYVSLLIRDSQPSDSATYLC AVTTDSWGKLQFGAGTQVVVTP
```

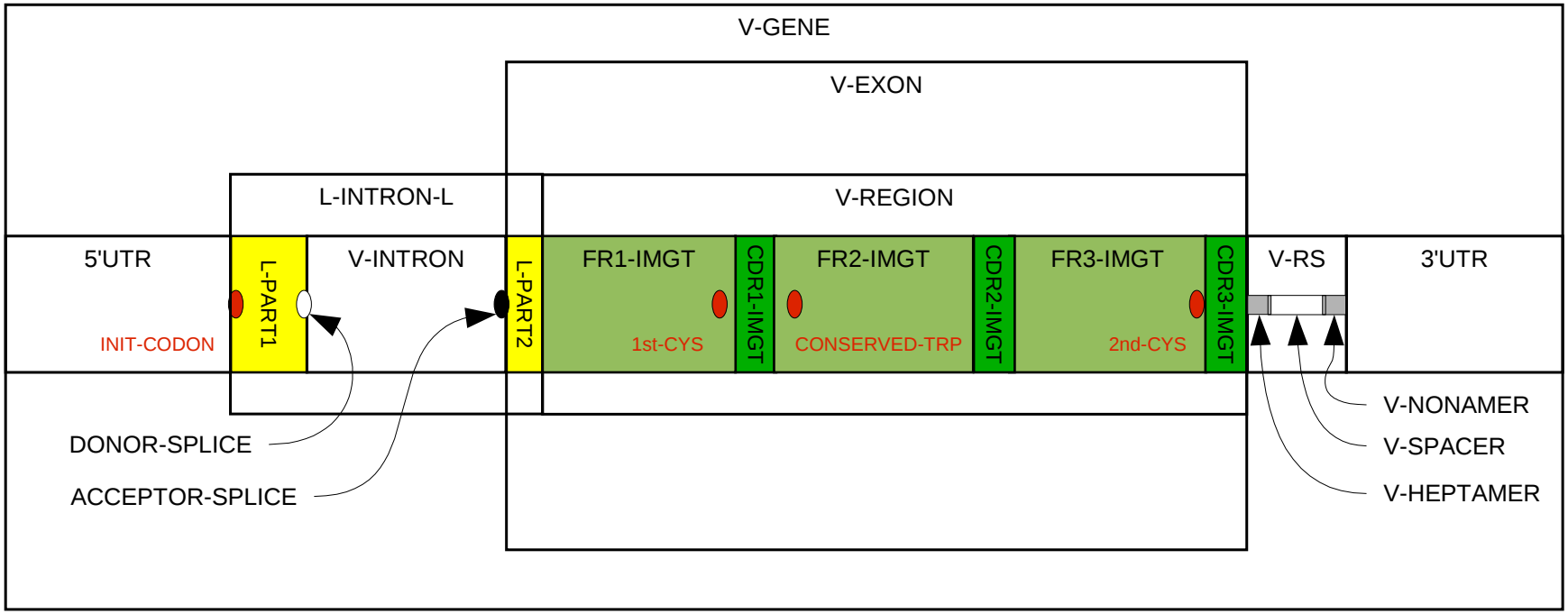
Remarks:
1. nucleotide sequence input only
2. no match with the other JUNCTIONs

# Some perspectives: facing evolution

- Improve consistency and high quality regarding data
  ...as well as tools!
  - size of data (sequence length, more complex structure)
  - number of data (more and more species, even individuals)
  - flow of data (new sequencing technologies)
- Extend the IMGT-ONTOLOGY by considering:
  - most of the IMGT Scientific Chart
  - other approaches (structural, clinical?, cellular?...)
  - other sorts of data (MHC and more generaly RPI)
- Provide a better interoperability between components inside IMGT® with IMGT-Choreography

**IMGT® team thanks you all!**

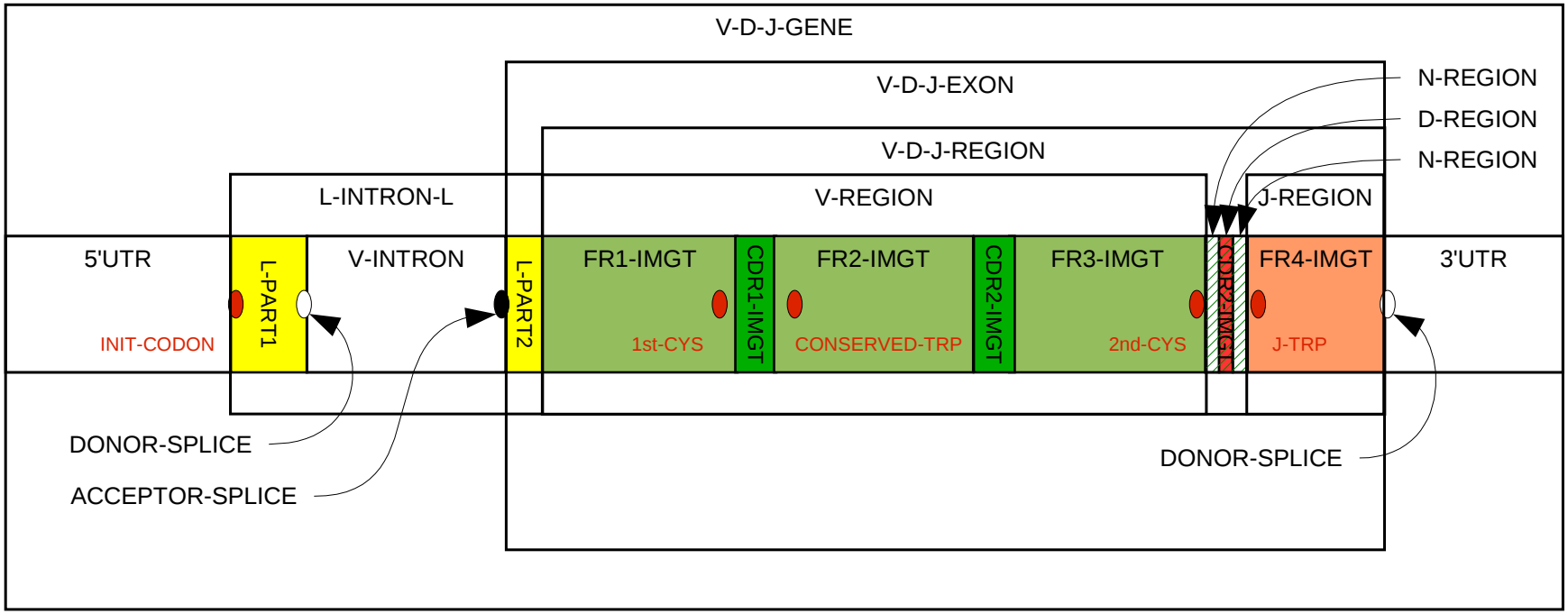
# DESCRIPTION: from V-GENE...




V-GENE(x) ↔ 5'UTR($x_1$) & L-INTRON-L($x_2$) & V-EXON($x_3$) & V-RS($x_4$) & 3'UTR($x_5$) & $x=x_1.x_2.x_3.x_4.x_5$

L-INTRON-L(x) ↔ L-PART1($x_1$) & V-INTRON($x_2$) & L-PART2($x_3$) & $x=x_1.x_2.x_3$

# DESCRIPTION: ....to V-D-J-GENE




CDR3-IMGT(x) ↔ N-REGION($x_1$) & D-REGION($x_2$) & N-REGION($x_3$) & $x=x_1.x_2.x_3$

# A step to specify IMGT tools

Input: *x*

Question: did exist $x_1$, $x_2$, ..., $x_n$ to solve a corresponding equation?

Algorithms:

- IMGT/JunctionAnalysis: V-J-GENE or V-D-J-GENE
- IMGT/V-QUEST: V-J-SEQUENCE or V-D-J-SEQUENCE
- IMGT/Automat: V-J-C-SEQUENCE or V-D-J-C-SEQUENCE
- IMGT/LIGMotif: other sequences (gDNA)
- IMGT/3Dstructure-DB pipeline: structures

➔ Provide revelant information to help the curators to check and decide.

# Why are IMGT Collier de Perles so useful?

- bridge the gaps between sequences and structures,
- are used whatever the MHC and whatever the species,

| | MHC-Ia | MHC-Ib | MHC-IIa | MHC-IIb |
|---|---|---|---|---|
|  | HLA-A,-B,-C | HLA-E,-F,-G | HLA-DPA,-DQA,- DRA<br>HLA-DPB,-DQB,-DRB | HLA-DMA,-DOA<br>HLA-DMB,-DOB |
|  | H2-D,-K,-L | H2-M,-Q,-T | H2-AA,-EA<br>H2-AB,-EB | H2-DMA,-DOA<br>H2-DMB,-DOB |

- have been extended to the MHC-I-like proteins (CD1, FCRN, RAET, HFE, MICA, AZGP1,…)

Interestingly, only one additional position **54A** in **G-ALPHA1-LIKE** was needed to extend the IMGT unique numbering for G-DOMAIN to the G-LIKE-DOMAIN

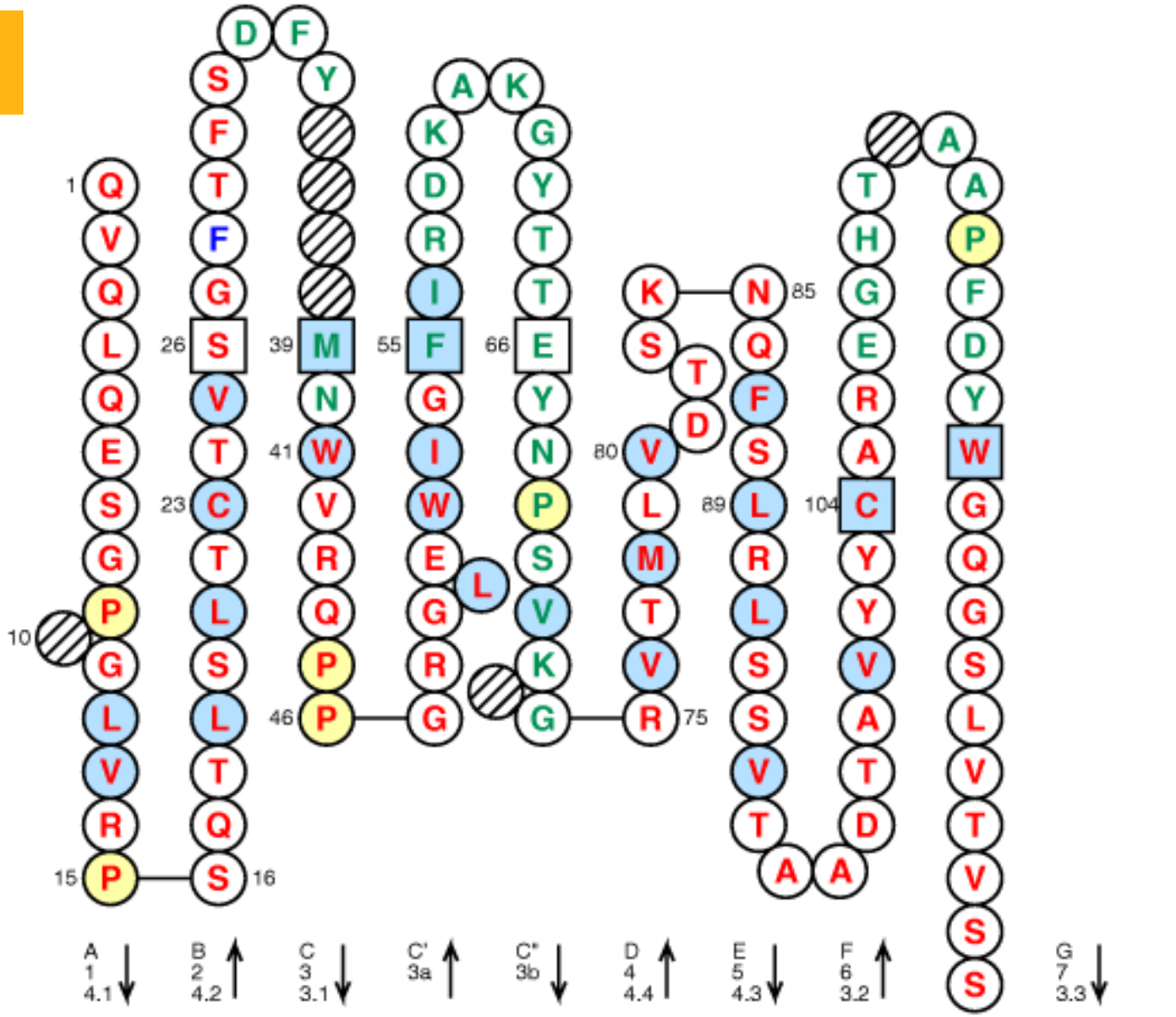
# Humanized CAMPATH-1H mutant 1

Mutant 1: S28>F
Mutant 2: S31>T
(alemtuzumab)

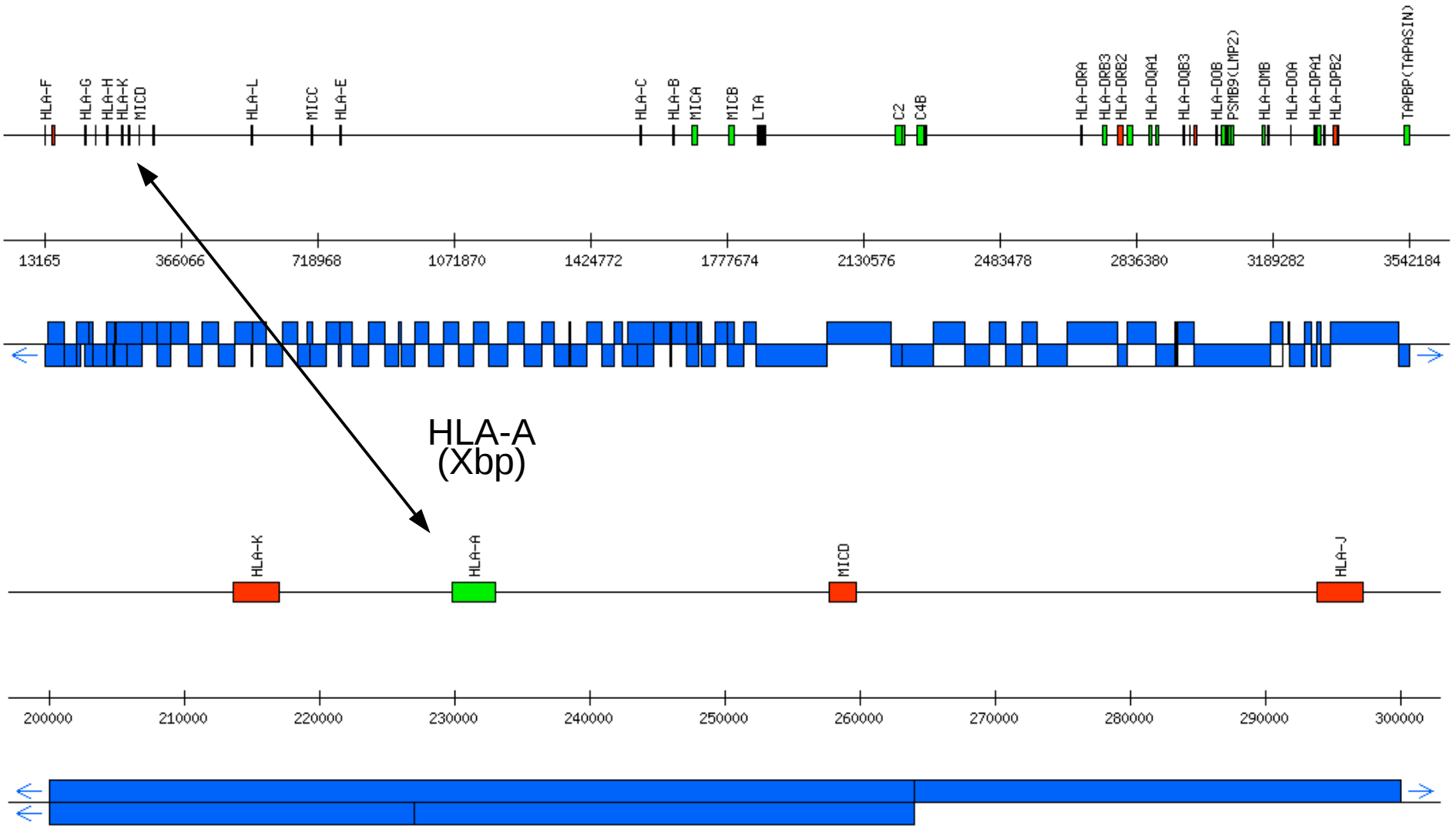
VH domain
(V-D-J-REGION)

[8.10.12]

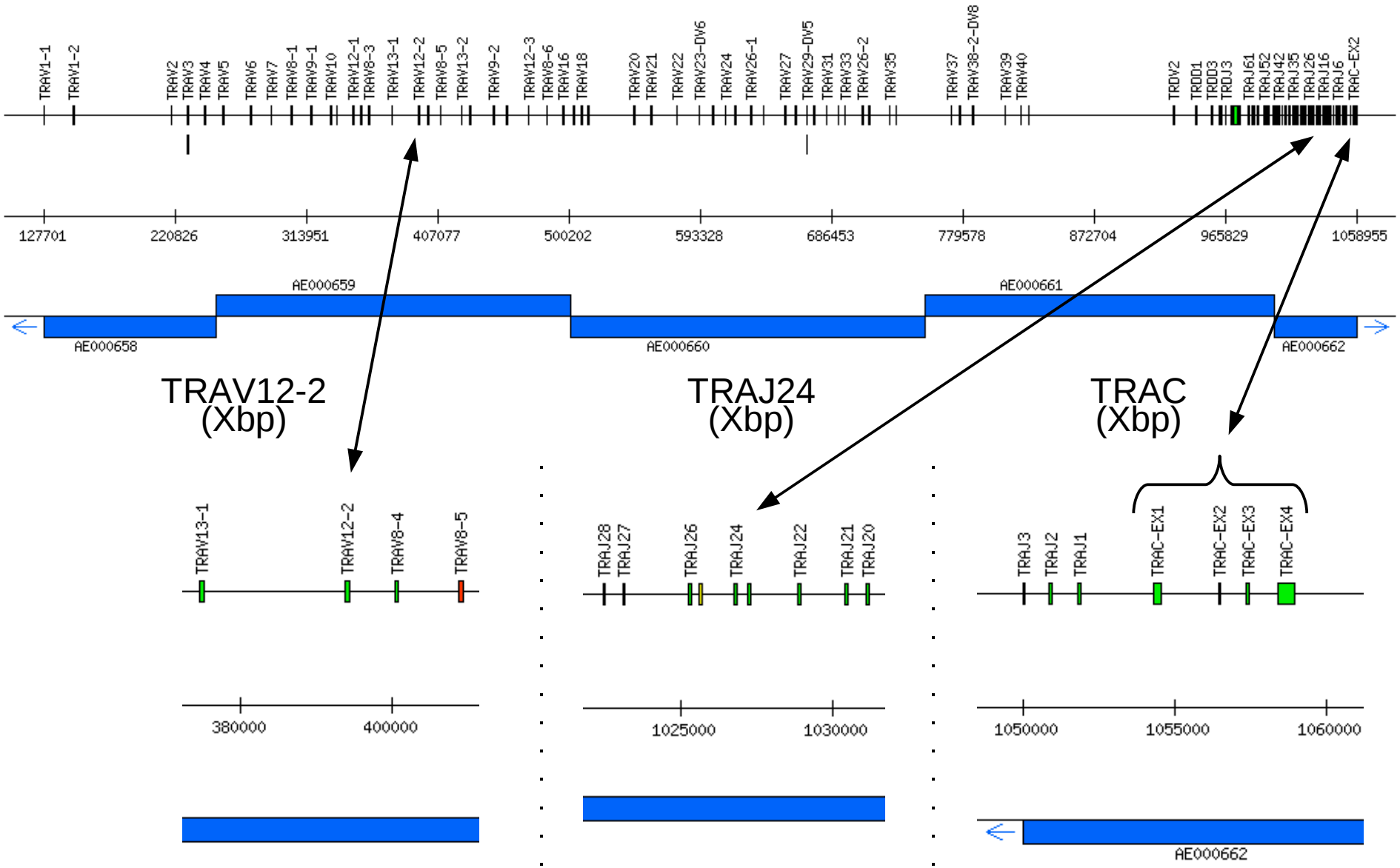
X human
X rat



# Human MHC locus at 6p21.3

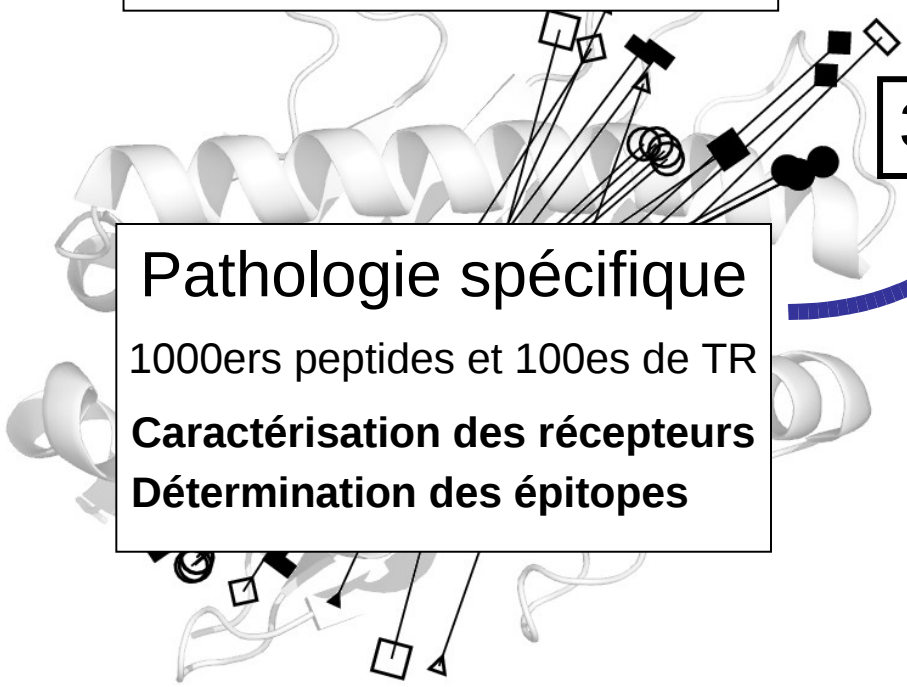


# Human TRA-TRD locus at 14q11.2



# Structural variability analysis

TR/peptide/MHC
21 3D-structures



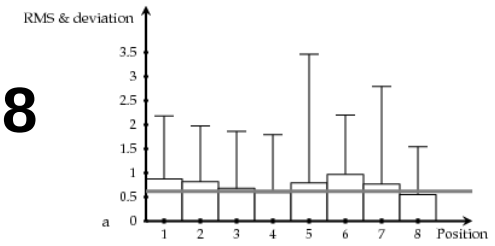
Pathologie spécifique

1000ers peptides et 100es de TR

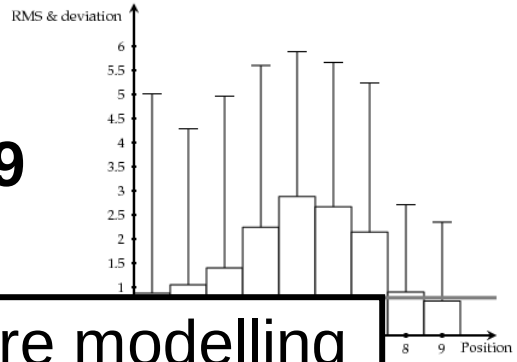
**Caractérisation des récepteurs**
**Détermination des épitopes**

3D-structure modelling

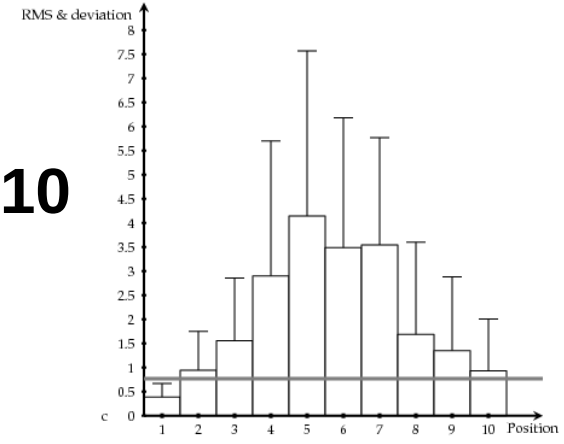
**MHC-I 8**



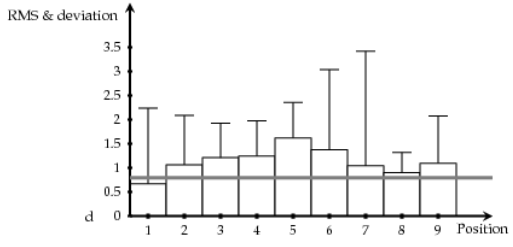

**MHC-I 9**




**MHC-I 10**

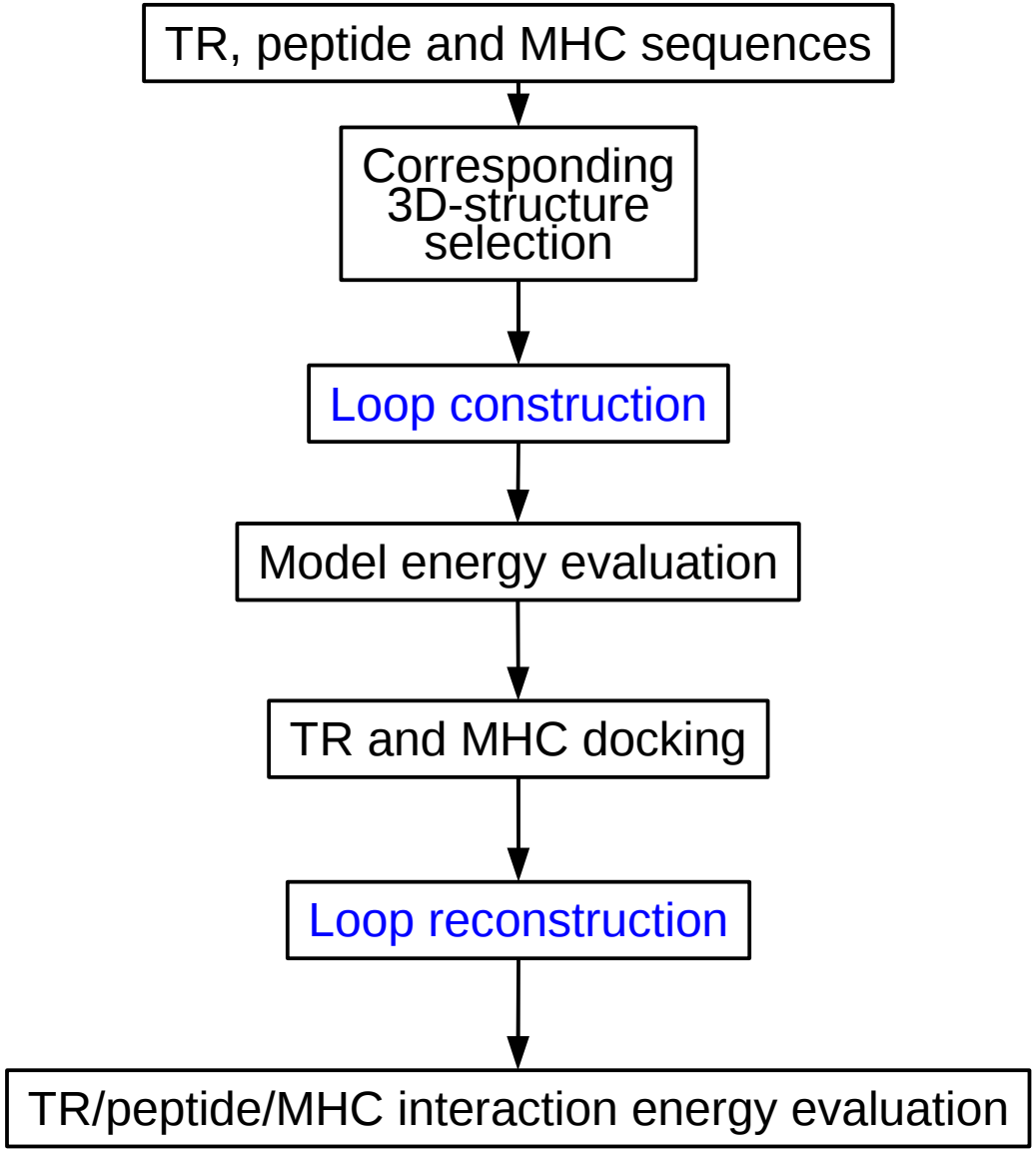



**MHC-II**




Im Muno Gene Tics Information system®

# TR/peptide/MHC interaction model protocol



Different ways to consider:

- synthesis, we know the allele components and try to imagine the possible resulting structure(s)

- analytic, we try to solve the protein history as we know the structure (also in terms of AA sequence) and try to see what may be the possible origin alleles.

But we have to consider model for both to frame the search. IMGT-ONTOLOGY is a way to have a formal construction (in the sens of formal language) "describing" such we can talk about sequence in different approaches () in term of combination of components to see how it fit.

What can be wrong with a structure complex?
May be due to 'bad' alleles or synthesis?