

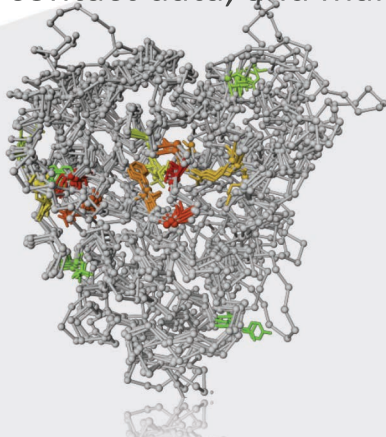
# 3DM: protein superfamily analysis system

## Intro to the Antibody Suite

- Structure visualization
- 3D alignments
- Literature analysis
- Mutation prediction
- Protein engineering
- Patent analysis
- Antibody analysis



# 3DM speeds up protein R&D

- **Make sense out of the flood of data available for proteins**
    - 3DM integrates data for complete protein superfamilies, such as mutation data from literature and patents, alignment data, correlation mutation data, ligand binding data and other intermolecular contact data, and many many more.
  - **3DM Tools give insight in your protein**
    - Use superfamily data to analyse your protein
  - **Solves the protein numbering problem**
    - Seamlessly transfer data between homologous proteins
  - **High-quality alignments**
    - ML optimized structure-based multiple sequence alignment
- 
- All data and all tools in 3DM are **connected** through a unifying 3D numbering scheme.
  - **3DM tools** analyse data for whole superfamilies, and help you gain insight in your protein.
  - Data from all proteins in the family can be **transferred** to your protein.

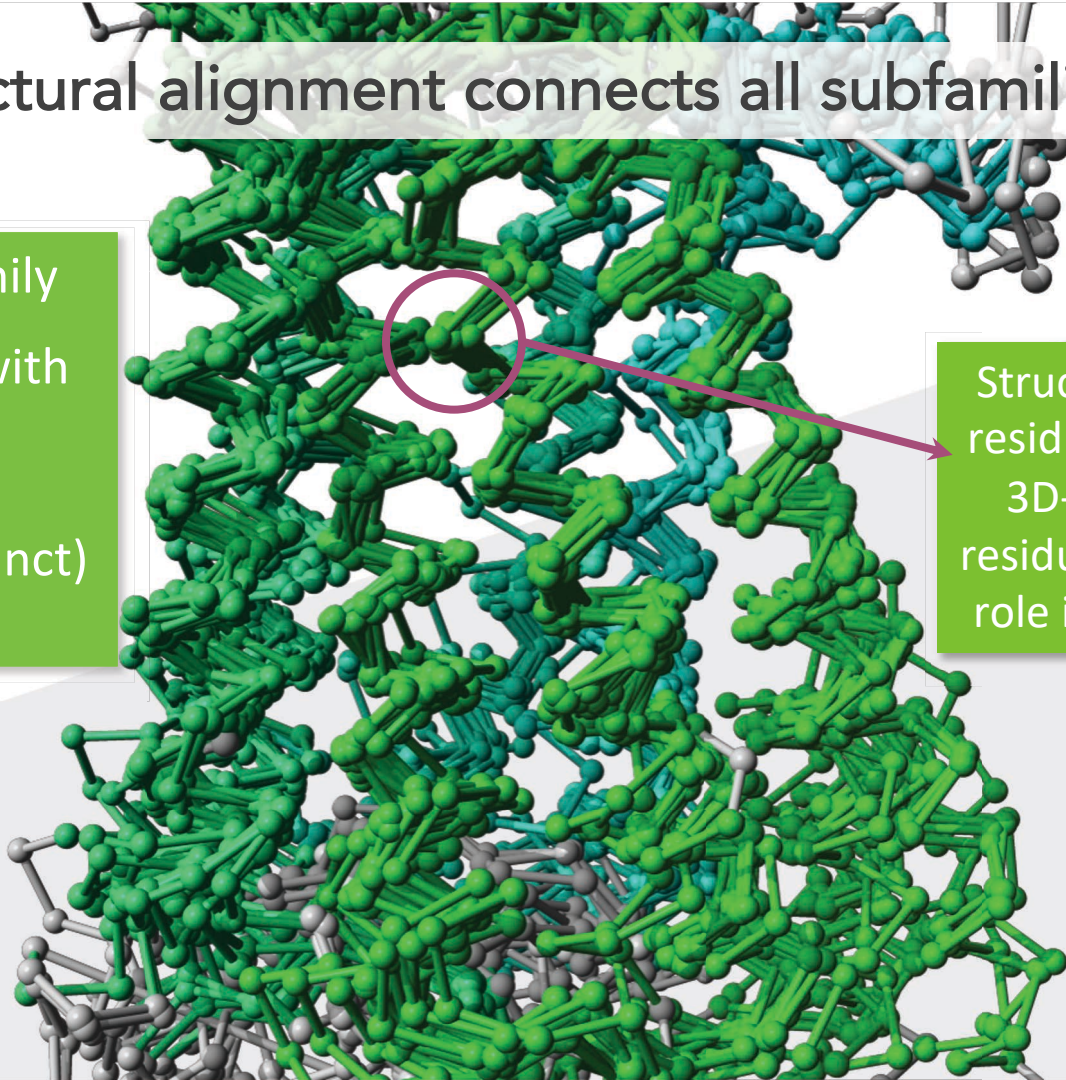


# A structural alignment connects all subfamilies

## GPCR protein family

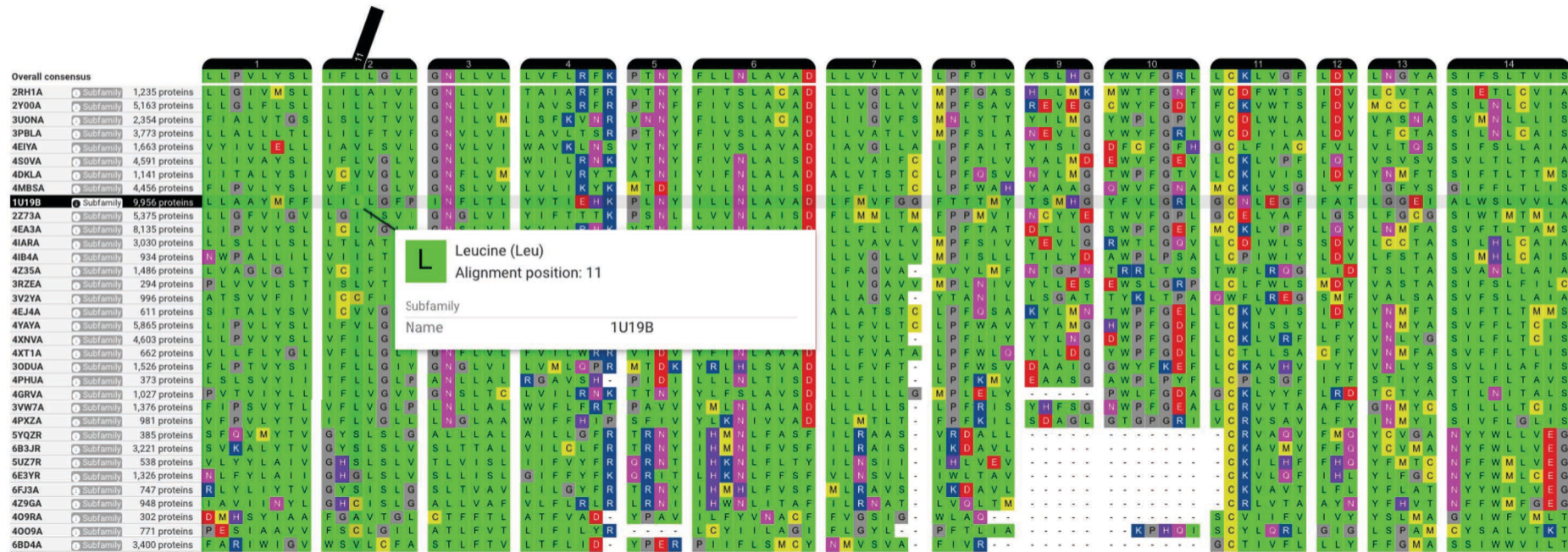
- 1104 structures with GPCR fold
- grouped into 43 (sequentially distinct) subfamilies

Structural equivalent residues get the same 3D-number. Such residues have a similar role in these proteins



# The structural alignment is combined with sequence alignments to cover the whole family

## Alignment



>85.000 WT sequences (from all available species) divided over 43 subfamilies and >29.000 patented sequences.



# Protein analysis

Visualize any type of data from the family on top of your target protein  
(here conservation data from the alignment and contacts between the different domains are shown)

## Sequence projection

MRGKKVWISLLFALALIFTMAFGSTSSAQAAGKSNGEKKYIVGFKQTMSTMSAAKKKDVISEKGGKVQKQFKYVDAASATLNEKAVKELKKDPSVAYVEE 100  
DHVAHAYAQSVPYGVSGIKAPALHSQGYTGSNVKKVAVIDSGIDSSHPDLKVVAGGASMVPSETNPFQDNNSHGTHVAGTVAAALNNSIGVLGVAPSASLYAV 200  
KVLGADGSGQYSWIINGIEWAIANNDVINNSLGSPSGSAALKAADVKAASGVVVVAAAAGNEGTSGSSSTVGYPGKYPSVIAVGAVDSSNQGRASFSSVG 300  
PELDVMAPGVS IQSTLPGNKYGAYNGTSMASPHVAGAAALILSKHPNWTNTQVRSSLENTTTKLGDSFYFGKGLINVQAAAQ



233

Leucine (Leu)

Alignment position: 116 (aligned in core region)

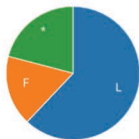
AMINO ACID DISTRIBUTION  
Protein/family data

PROTEIN MUTATIONS  
Protein data: 9

FAMILY CONTACTS  
Family data: 102

### Subfamily 1SUBA

Conservation of L  
62.04% (1,198 out of 1,931)



Based on subfamily

### Superfamily

Conservation of L  
47.31% (43,191 out of 91,299)



Add



## Visualizations

Group on scope

Core regions



Domain annotations



### Family data

Conservation

Superfamily | Residue

Family contacts

Ligand

### Protein data

Protein mutations

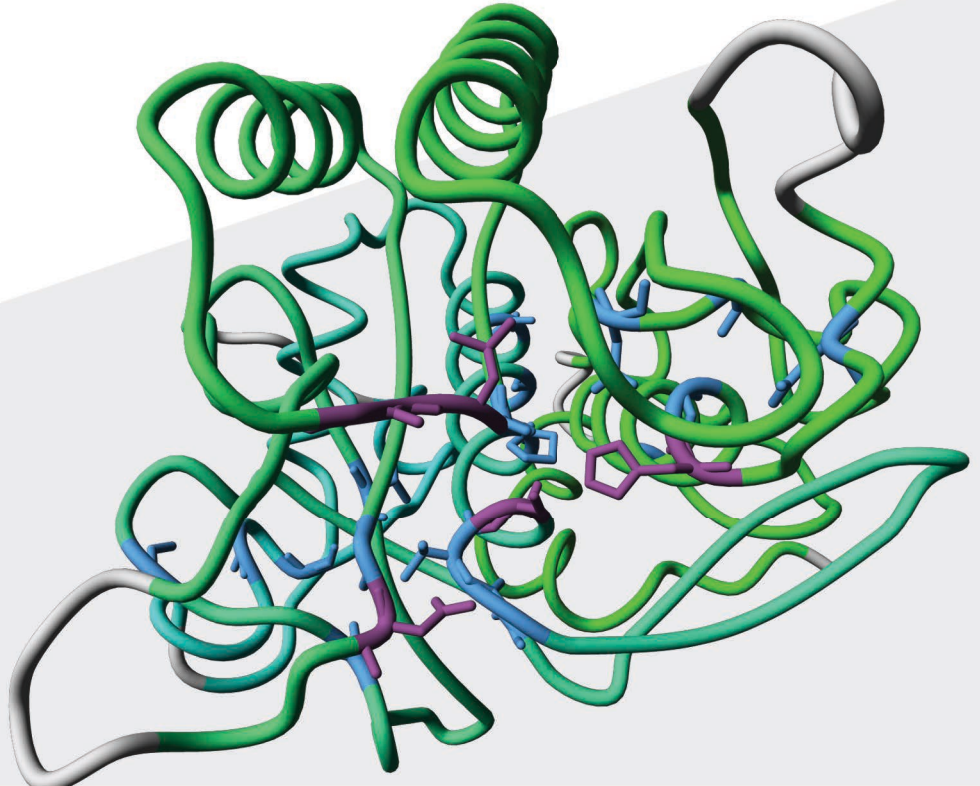


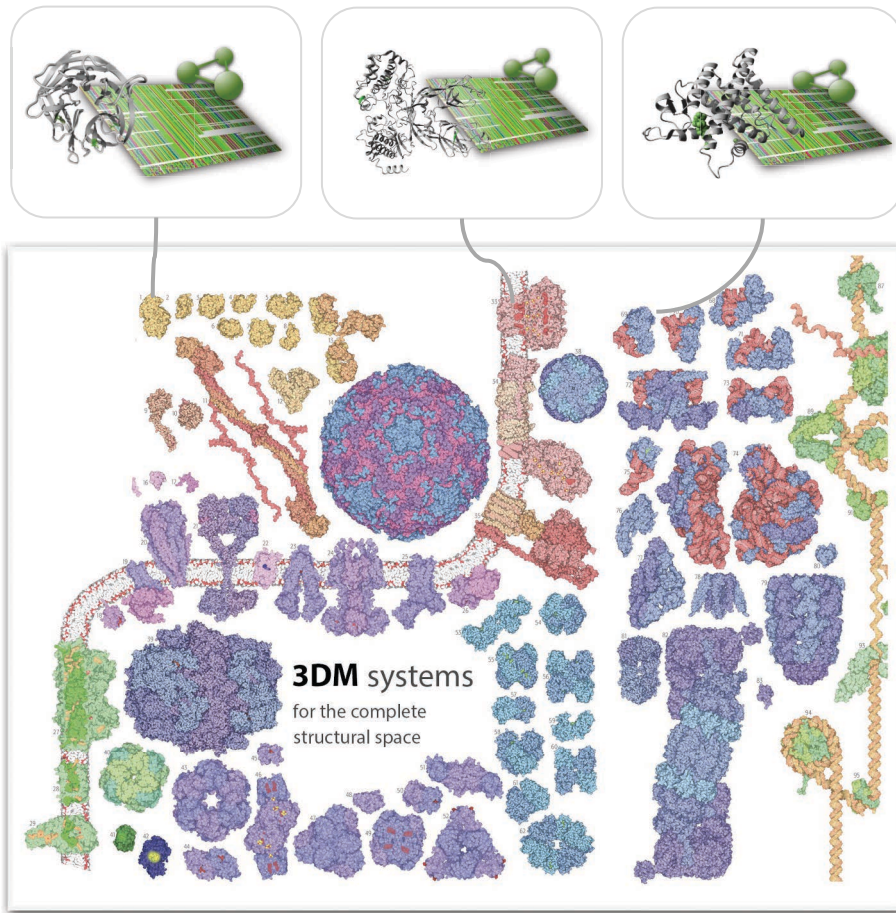
# Structure visualisations

Show any 3DM data directly in any protein structure

PyMol/Yasara integration:

- Generate scenes from 3DM
- Yasara/PyMol plugin
- Visualise data directly in any structure





# PDB-wide 3DM

## 3DM systems ready to go for any protein target

Total number of 3DM systems: 30,251

Protein (domains) aligned: over 130 million

Different alignment configurations focused on:

- Single domains

- Domain combinations

- Superfamilies (manually curated)

## Advantages

Use the 3DM web application and all tools to get instant insight into any protein (family)

- 3DM systems readily available for any protein target

- Ideal for a manual deep dive

- Ideal for antigen or target investigation

With over 30,000 families available, 3DM covers the complete structural space. providing access to the full power of 3DM instantly, including panel design, literature analysis, engineering hotspots, etc.

# 3DM capabilities

## Protein-focused

- protein R&D
- protein engineering
- protein selection panels
- patent analysis
- **antibody analysis suite**

## Compound-focused

- pocket analysis
- small molecule tractability R&D
- ligand fingerprinting

## Helix Predictions (AI-based products)

- Best in the world variant pathogenicity predictor for the human exome (<https://arxiv.org/abs/2104.01033>)
- Protein/antibody engineering
- Customized AI collaborative projects





# 3DM Antibody Suite



# Antibody 3DM systems

3DM systems for different Ab-regions to facilitate specific use cases

## Variable domain / Fab



• [Antibodies\\_heavy\\_chain Fab](#)

96k sequences, 18.9k mutations

• [Antibodies\\_light chain](#)

67k sequences, 6.5k mutations

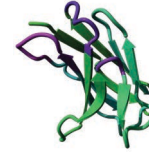
## Intact heavy chain / C-region



• [Antibodies\\_heavy\\_chain](#)

11.5k sequences, 6.9k mutations

## CDR-specific alignments



• [Antibodies heavy chain CDRs](#)

75.8k sequences across 33 alignments

• [Antibodies light chain CDRs](#)

45.1k sequences across 61 alignments

3DM systems

humanization

investigation of light chain -  
heavy chain interactions

stability engineering

analysis of interactions with  
the complement system

CDR classification

epitope binning

analysis of antibody-antigen  
interactions

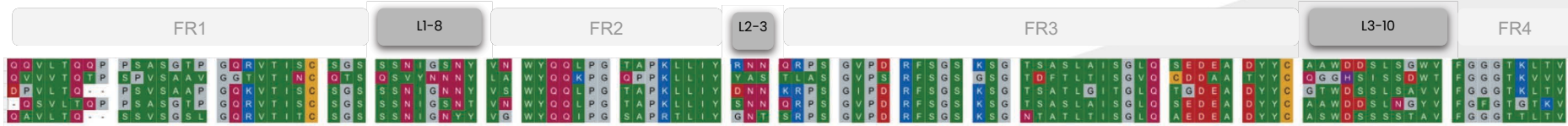
Use cases



# CDR System: CDR-specific alignments

Separate alignments for each CDR-length combination results in full CDR coverage – 94 alignments

CDR-L1 length 8, CDR-L2 length 3, CDR-L3 length 10



CDR-L1 length 6, CDR-L2 length 3, CDR-L3 length 9



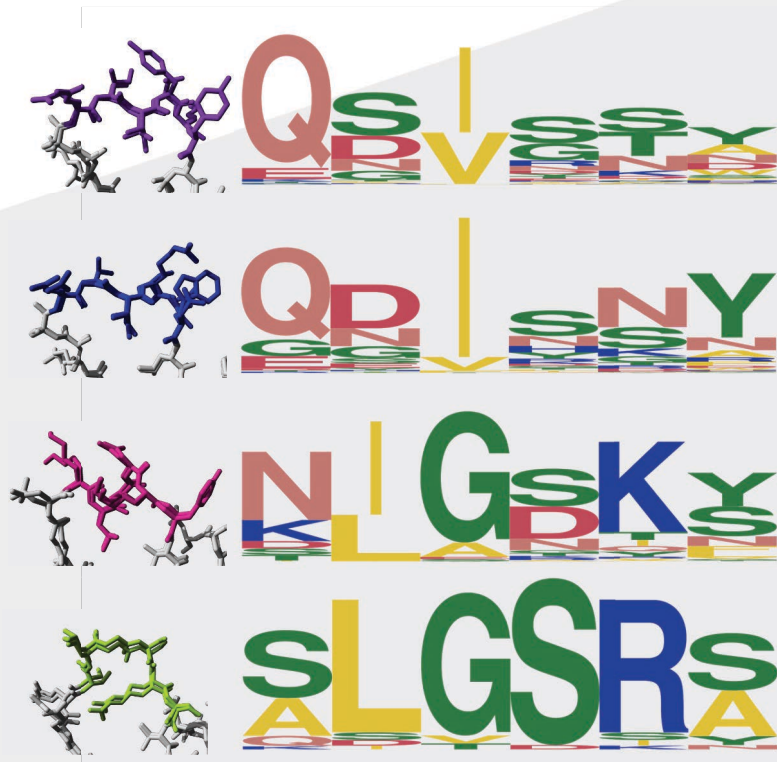
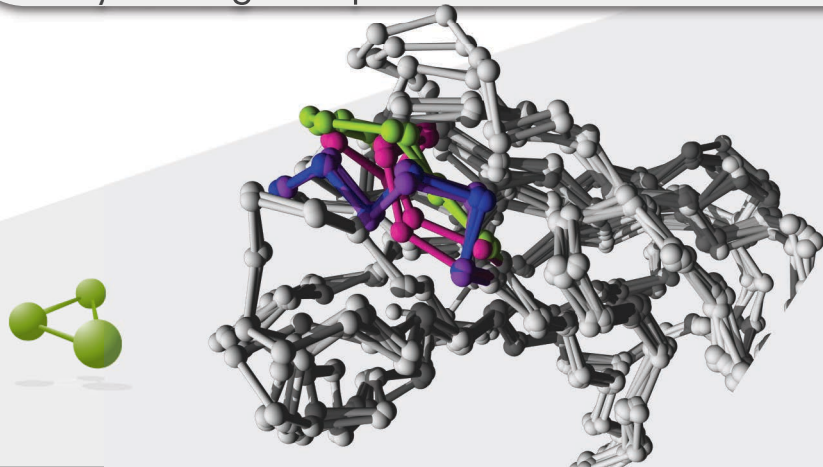
Enables analysis of residue statistics  
(e.g. conservation data) over **full CDR length**



# CDR Conformation clusters

Enables sequence-based comparisons within and between structural conformations

- **Structure-based clustering**  
based on torsion angles and coordinates, specifically fine-tuned for each CDR-length
- Sequence-based identification of structural CDR conformation class
- over 300 CDR class-specific structure-based subsets in 3DM with automatic classification of your target sequences.

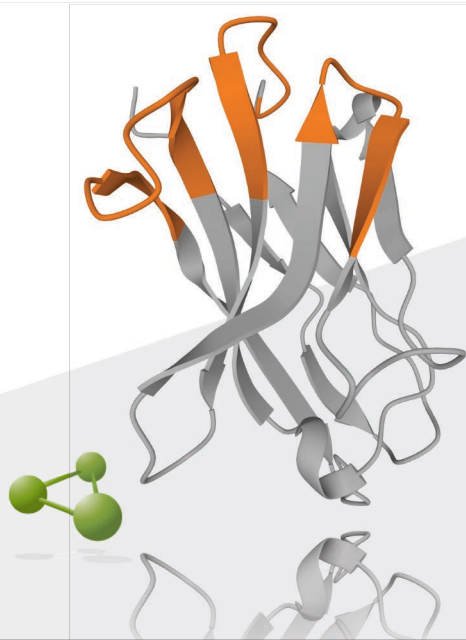




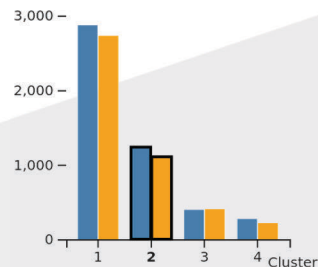
# CDR classification in protein analysis

annotate CDRs in your protein of interest

```
QVKLLLEQSGGGVVQPGRSLRLSCAASGIFTLTYG IHWVRQAPGKGLEWVA 50
GLWYGNTKNYAESVKGRFTISRDN SKDTLYLMNSLRADD TAVYYCARALQ 100
GLIHEMDDWGKGTA VTVSSASTK
```

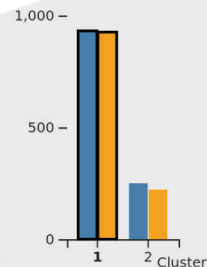


CDR 1 length 8



Assigned to conformation cluster 2  
(probability: 68%)

CDR 2 length 7



Assigned to conformation cluster 1  
(probability: 94%)

no. PDBs no. unique sequences Assigned cluster

Visualize CDR location in sequence and structure

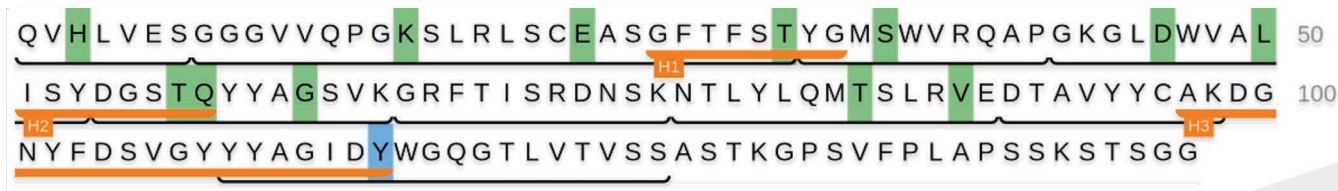
uses AlphaFold model when no PDB available

Assign CDRs to conformational cluster

Show distribution and statistics of the CDR clusters

with same CDR length as the assigned cluster

# Germline annotation in protein analysis



Identify the closest germline genes  
supports IMGT germlines for 16 species

Visualize differences with protein of  
interest and germline hit

To investigate affinity maturation

Humanisation: compare residues  
with human germline amino acid  
distribution



V gene:IGHV3-30\*18 (87.50%)



J gene:IGHJ5\*01 (92.31%)



QVQLVESGGGVVQPGKSLRLSCEASGFTFDSWGGQTLVTVSS  
SSYGMHWVRQAPGKGLEWVAVISYDGSN  
KYYADSVKGRFTISRDN SKNTLYLQMNSL  
RAEDTAVYYC

Gene segment  
Identity  
Similarity  
Hamming distance  
Length  
E-value  
Full gene segment  
sequence in IMGT

X92214\_IGHV3-3018  
88%  
89%  
12  
96  
5e-62  
QVQLVESGGGVVQPG  
RSLRLSCEASGFTFSS  
YGMHWVRQAPGKGL  
EWVAVISYDGSNKYY  
ADSVKGRFTISRDN SK  
NTLYLQMNSLRAEDT  
AVYYC

Gene segment  
Identity  
Similarity  
Hamming distance  
Length  
E-value  
Full gene segment  
sequence in IMGT

J00256\_IGHJ501  
92%  
87%  
1  
13  
0.00000753  
NWFDSWGGQTLVTV  
SS

# Antibody sequence annotation

annotate your in-house sequences with 3DM in bulk

in-house Ab  
sequences (100k+)



## 3DM data



• [Antibodies, heavy chain](#)

96k sequences, 18.9k mutations

• [Antibodies, light chain](#)

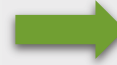
67k sequences, 6.5k mutations



CDR cluster assignment

closest germline annotation

covariance violations



Correlate with in-house features

Use 3DM Ab annotations to analyse your generated sequences  
and enhance candidate selection

Gain insight in how your data is distributed in respect to structural  
conformation and germline origin



# Antibody sequence annotation

future capabilities: additional modules to Ab annotation framework

in-house Ab  
sequences (100k+)



**3DM data**  
Fab targeted



• [Antibodies, heavy chain](#)

96k sequences, 18.9k mutations

• [Antibodies, light chain](#)

67k sequences, 6.5k mutations



CDR cluster assignment

closest germline annotation

covariance violations

**Alignment data**  
(e.g. conservation, PTMs)

**Patented sequences**

**Inter/intra domain  
contacts**

**Humanness,  
humanization target  
residues**

...



**Correlate with in-  
house features**





# Nanobody 3DM system

integrating VHH and nanobody data

## Variable domain / Fab



### [Antibodies, heavy chain](#)

96k sequences, 18.9k mutations

### [Antibodies, light chain](#)

67k sequences, 6.5k mutations

## Intact heavy chain / C-region



### [Antibodies, heavy chain](#)

11.5k sequences, 1.5k mutations

## CDR-specific alignments



### [Antibodies heavy chain CDRs](#)

75.8k sequences across 33 alignments

### [Antibodies light chain CDRs](#)

45.1k sequences across 61 alignments

## Nanobody (VHH)



### [Nanobodies \(2022\)](#)

1k+ structures (chains)

3DM systems

- VHH-targeted subsets (cameloid species, cameloid germlines)
- CDR conformational cluster assignment
- Germline annotation



# Custom development

3DM as a starting point



Finetune 3DM to your needs:

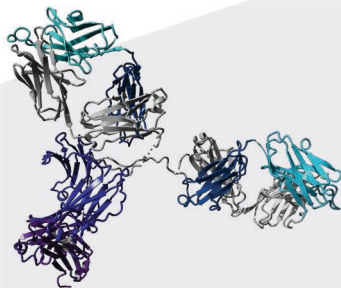
- Annotation modules
- TABS integration
- SabDab integration
- Molecular viewers (Schrodinger/ CCG)
- API calls, other features..



# Antibody complex analysis – in development

For analysis of domain type-specific contacts

In-depth investigation of antibody sequences  
in context of the whole antibody molecule



POC

Heavy chain 2VXVH - 3DM System Antibodies, heavy chain (2020) - Heavy chain C-region - Subset id

QVQLVQSGAEVKKPGASVKVSCQASGGRFSNFIHWVRQAPGQRFWMGWINPYNQNKEFSAKFDQRTF 70  
TADTSANTAYMELKSLRSADTAVYYCARVPSWQDSQQDNYMDVWGKQTTIVSSASTKGPSVFFLAP 140  
SSKSTSGGTAALGCLVKDYFPEPTVSWNSGALTSGVHTFPAVLQSSGLYSLSVVTVPSSSLGTQTYIC 210  
NVNHKPSNTKVDKKAEPKSCDKTHTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHED 280  
PEVKFNWYVDQVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTI SKAK 350  
GGPREFQVYITLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTFFVLDSDGSEFFLYSKL 420  
TYDKSRWQGGNVFSCSVMHEALHNHYTQKSLSLSPQK

Light chain 2VXVL - 3DM System Antibodies, light chain (2020) - Fab region - Subset id

EIVMTQSPATLSVSPGERATLSCRASEISSNLAWYQQKPGQAPRLFIYTASTRATDIPARFSGSGSGTE 70  
FTLTISLSQSEDAFVYYCQOYNWPSLTFGGGTRLEIKRTVAAPSVFIFPPSDEQLKSGTASVVCLLNNF 140  
YPREAKVQWKVDNALQSGNSQESVTEQDSKDSYLSSTLTLSKADYEKHKVYACEVTHOGLSPPVTKSF 210  
NRGE

Numbering scheme  
Kabat

Core regions ☒  
Domain annotations ☒  
CDR IMGT ☒

SHOW IN STRUCTURE

+

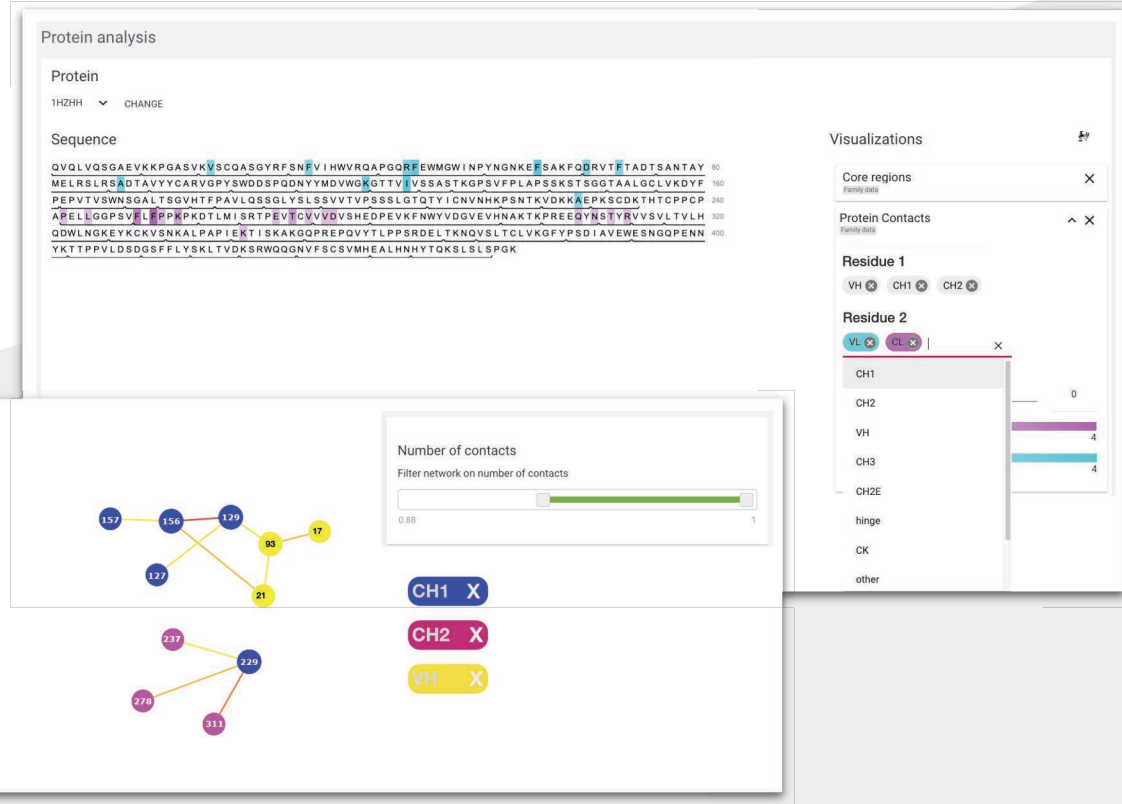
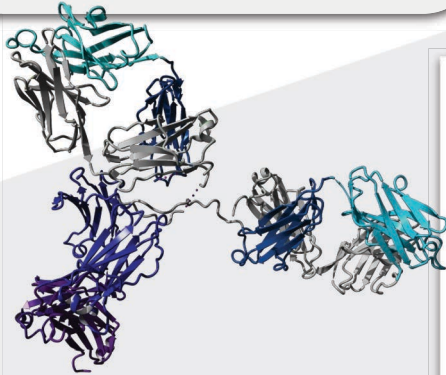


# Antibody domain contact statistics

For analysis of domain type-specific contacts

investigate interactions between specific domain combinations (e.g. VL with CH1 and VH)

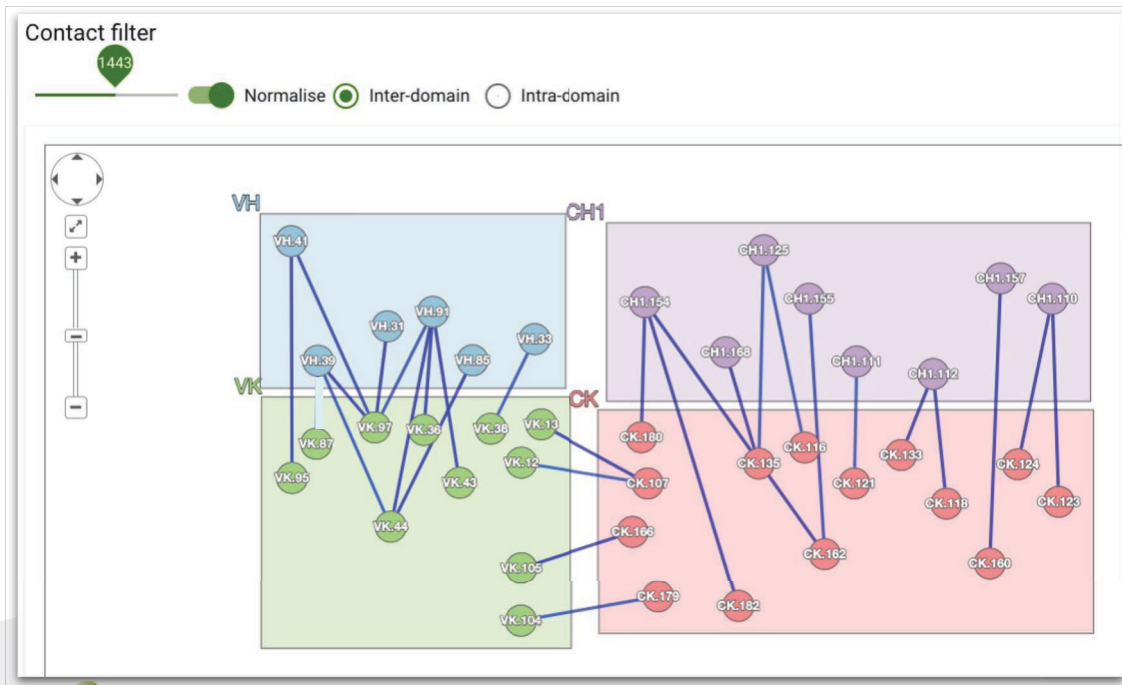
distinguish between antigen contacts vs light/heavy chain contacts





# Antibody domain contacts

investigate interactions between specific domains (e.g. VK with CK and VH)



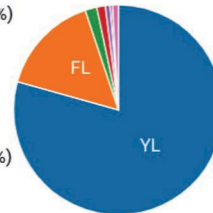
## Position information

**Selected :** VH.39 - VK.87

**Contacts :** 2,434

! VH.39 is from the antibodies\_heavy\_fab\_2020 system.

YL - 1,931 (79.33%)  
FL - 378 (15.53%)  
YP - 44 (1.81%)  
YF - 32 (1.31%)  
HL - 18 (0.74%)  
FP - 9 (0.37%)  
Others - 22 (0.90%)



## Motif

Export

**VK.87 VH.39 Contacts**

Y	L	1,931	79.33%
F	L	378	15.53%
Y	P	44	1.81%
Y	F	32	1.31%
H	L	18	0.74%

# HARMONY: Antibody codon optimizer

Sequence optimization on complex constructs

Bio-Product



# HARMONY: Antibody codon optimizer

Sequence optimization on complex constructs

- Combine **fixed** and **variable** segments
- Optimization:
  - Evaluates whole sequence
  - Optimizes variable segments
  - Checks segment boundaries
- Single optimization run for whole construct

## Listed jobs:

Task name

N-term scFv + Fc



Task name

N-term scFv + Fc



Task name

N-term scFv (VL-VH)



Back

Save

## VARIABLE SEGMENT



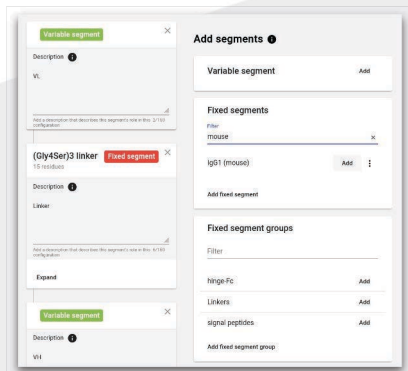
### Amino acid sequence

DIVMTLSTAILSASPGKVTMTCRASSRVNYIHWFQQKAGSS  
PKPWYATSNLASGVPDRFSGRSGTSYSLTISRVEADAATY  
YCQWSSNNPWTFGGGKLEIKRADAAPTVELP

Create **custom optimization templates** by combining fixed and variable building blocks

**Batch optimize** sequences for 100's of constructs

Data import/export and **API support**



# HARMONY: Antibody codon optimizer

Sequence optimization on complex constructs

- Harmonize codon usage to match frequency of expression host (~60k organisms supported)
- Optimize GC content & avoid repeats
- Identification and removal of restriction sites
- Removal of mRNA hairpins (using MFE)
- Simultaneously optimized in a single run

Original sequence	TGAATACAGTGGGACCTACTCTTATTCCCTTCGATATCTGGGG
Optimized sequence	TGAATATAGTGGCACCTATTCTTATAGCTTCGATATTTGGGG
mRNA pattern	((.....)))))..((((.....((((((.....))))))
Sequence segments	



Expression organism (2349 results) \*

e coli

Escherichia coli

Escherichia coli B

Escherichia coli K-12

Escherichia coli O157:H7

Restriction enzymes

Restriction enzymes to avoid

Apal (GGGCCC) x

AscI (GGCGCGCC) x

BsaI (GGTCTC) x

EcoRI (GAATTC) x

HindIII (AAGCTT) x

SapI (GCTCTTC) x

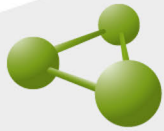
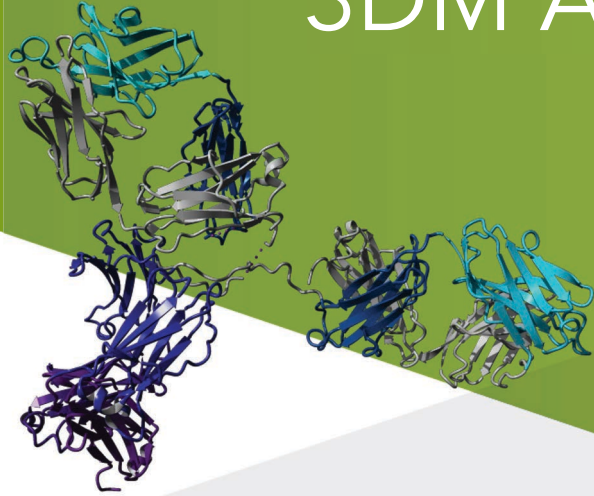
SphI (GCATGC) x

XbaI (TCTAGA) x

sap

SapI

# 3DM Antibody Suite



**bio-product**  
protein predictions