3DM: protein superfamily analysis system

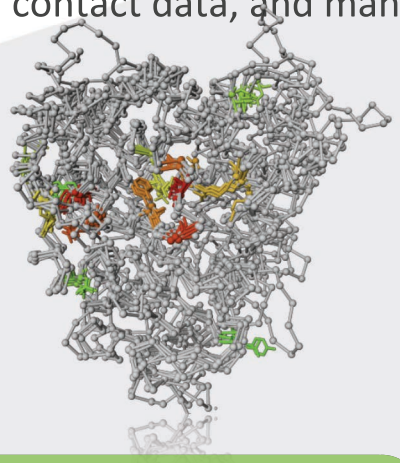
Intro to the Antibody Suite

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

bio-prodict
protein predictions

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

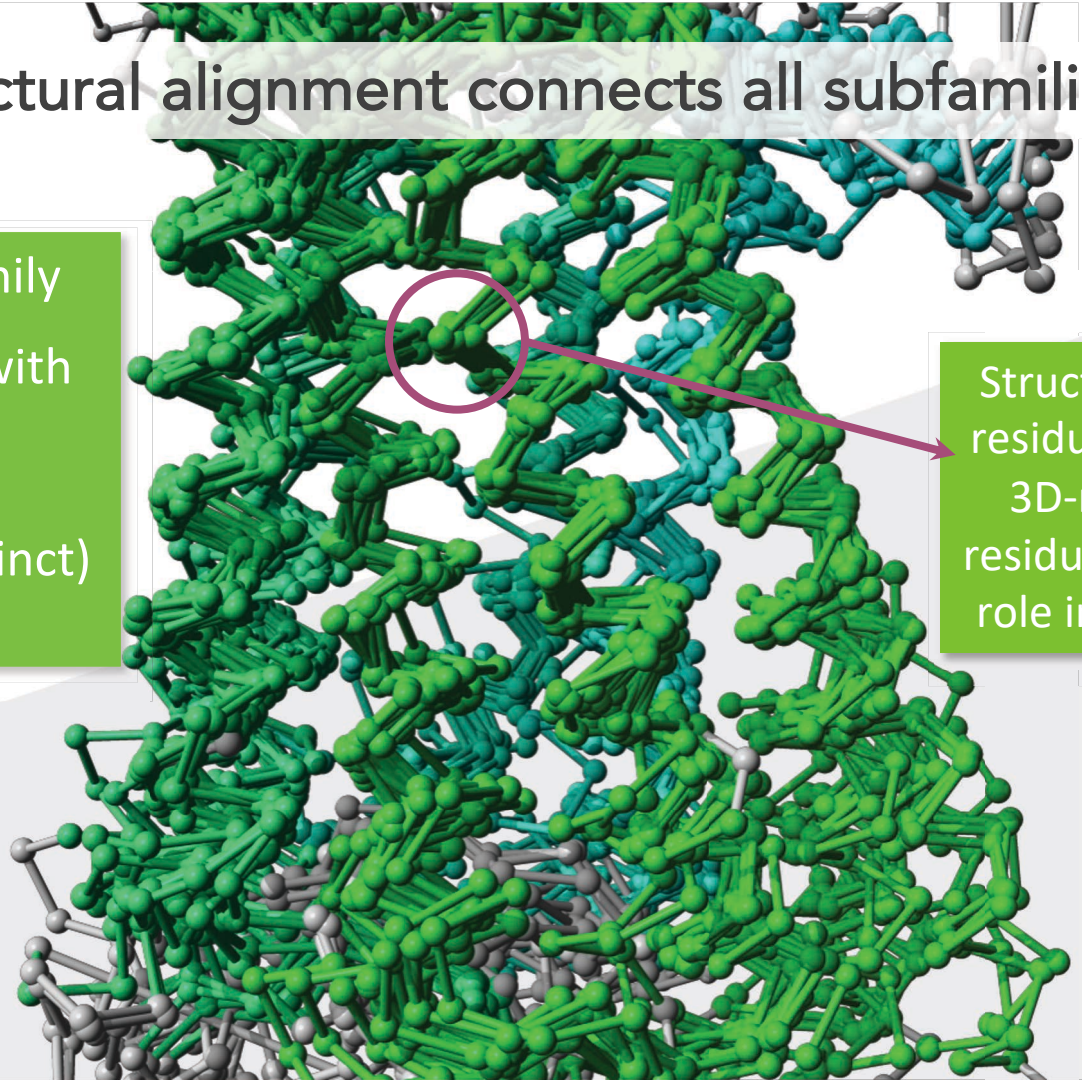


- All data and all tools in 3DM are **connected** through a unifying 3D numbering scheme.
- **3DM tools** analyse data for whole superfamilies, and help you gain insight in your protein.
- Data from all proteins in the family can be **transferred** to your protein.

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

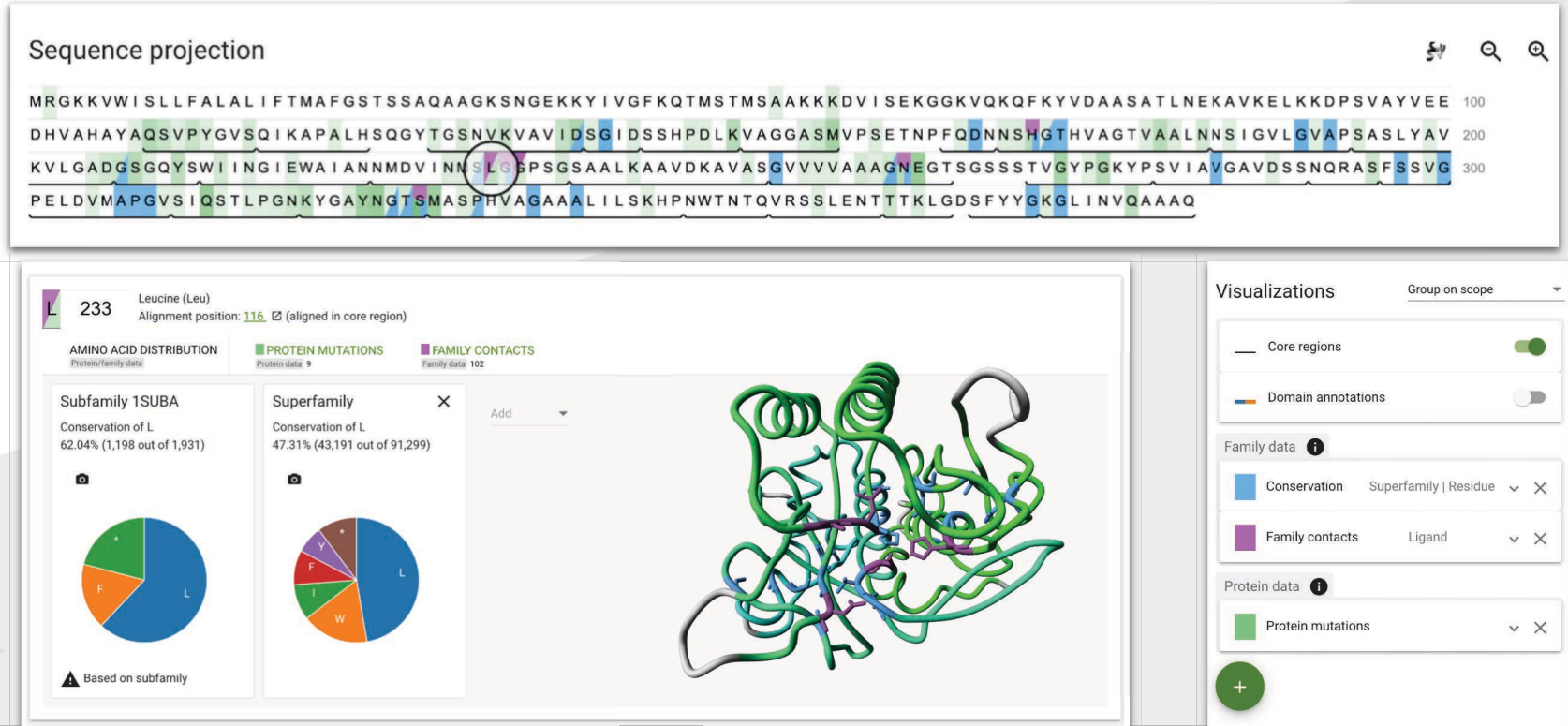
The structural alignment is combined with sequence alignments to cover the whole family

>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)

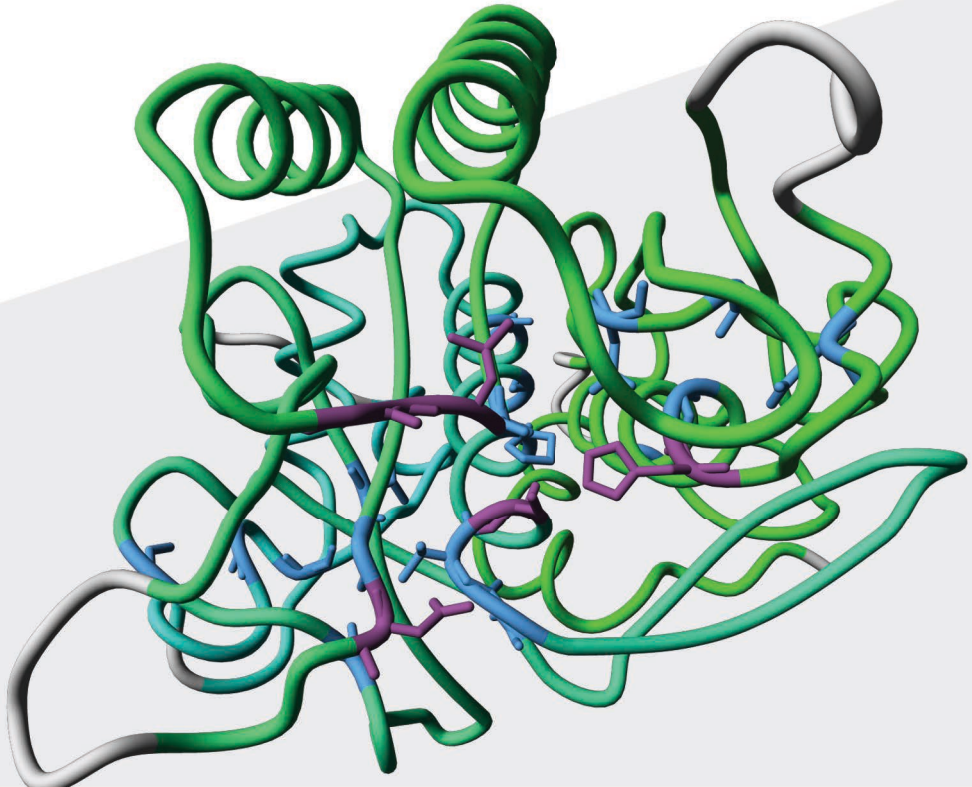


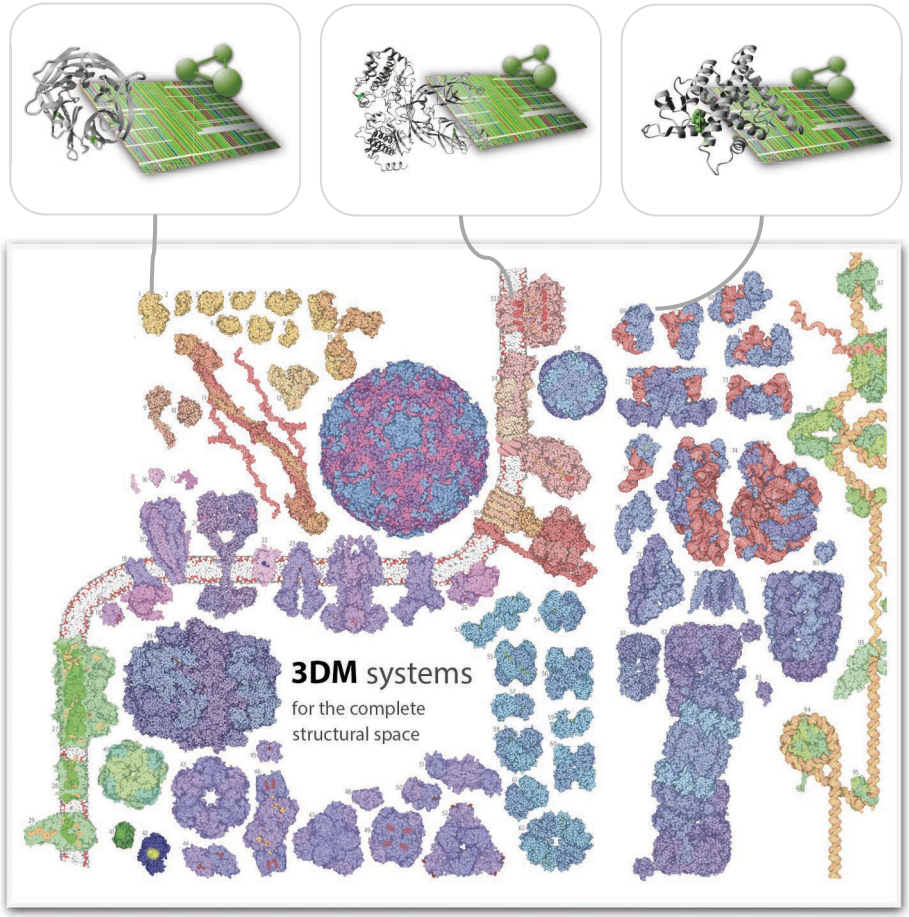
Structure visualisations

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





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.

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

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

bio-prodict
protein predictions

Antibody 3DM systems

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

3DM systems

Variable domain / Fab

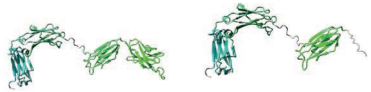
Antibodies, heavy chain Fab

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

Use cases

- humanization
- investigation of light chain - heavy chain interactions

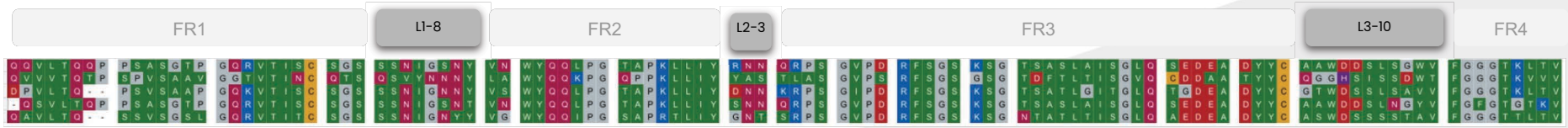
- stability engineering
- analysis of interactions with the complement system

- CDR classification
- epitope binning
- analysis of antibody-antigen interactions

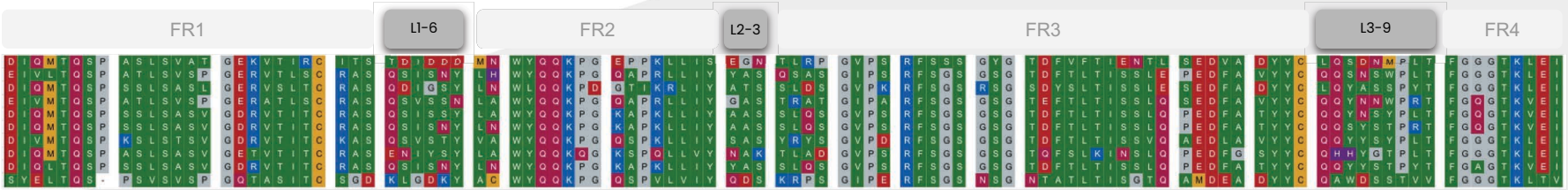
CDR System: CDR-specific alignments

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**

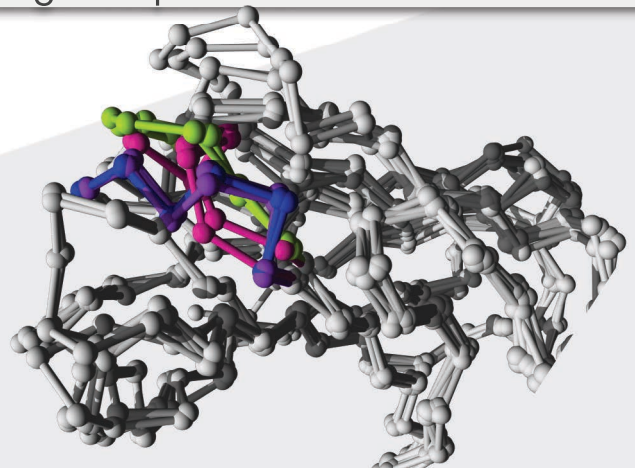


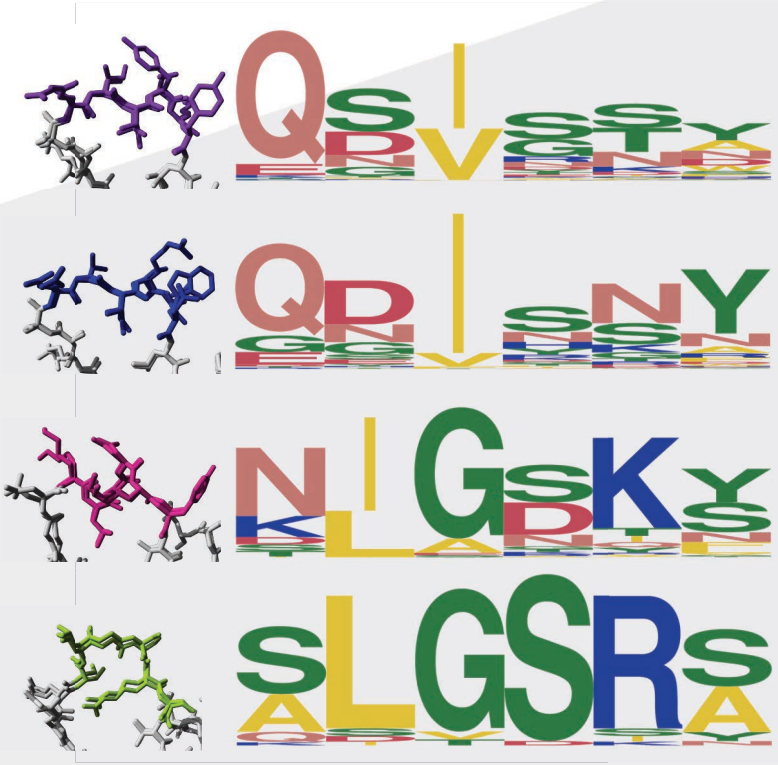
Enables analysis of residue statistics
(e.g. conservation data)over **full CDR length**

CDR Conformation clusters

Enables sequence-based comparisons within and between structural conformations

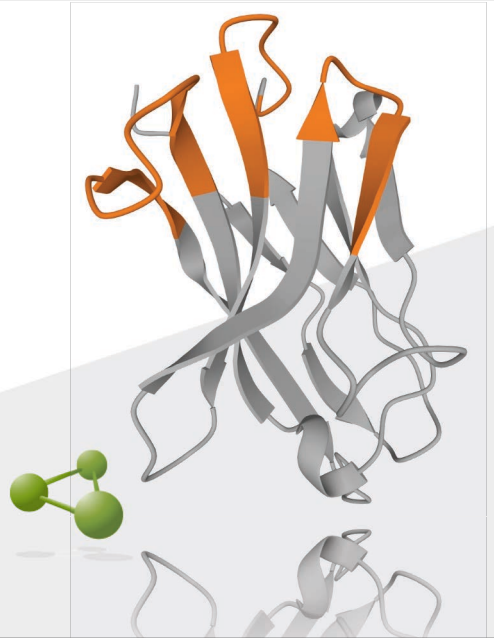
- Structure-based clustering
  based on torsion angles and coordinates, specifically fine-tuned for each CDR-length
- Sequence-based identification of structural CDR conformation class
- over 300 CDR class-specific structure-based subsets in 3DM with automatic classification of your target sequences.

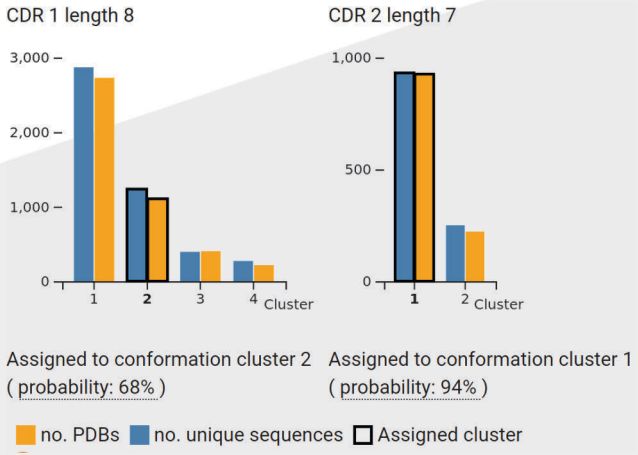




CDR classification in protein analysis

annotate CDRs in your protein of interest





Visualize CDR location in sequence and structure

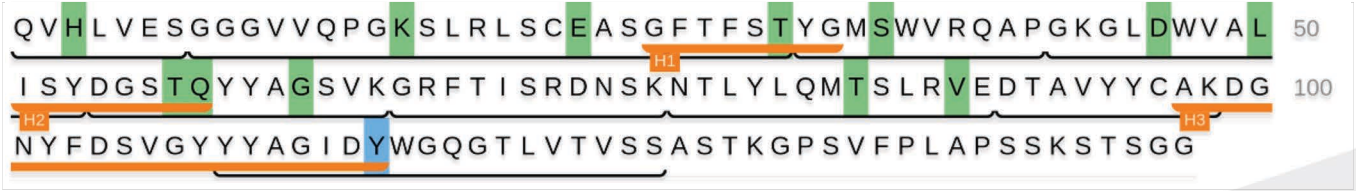
uses AlphaFold model when no PDB available

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

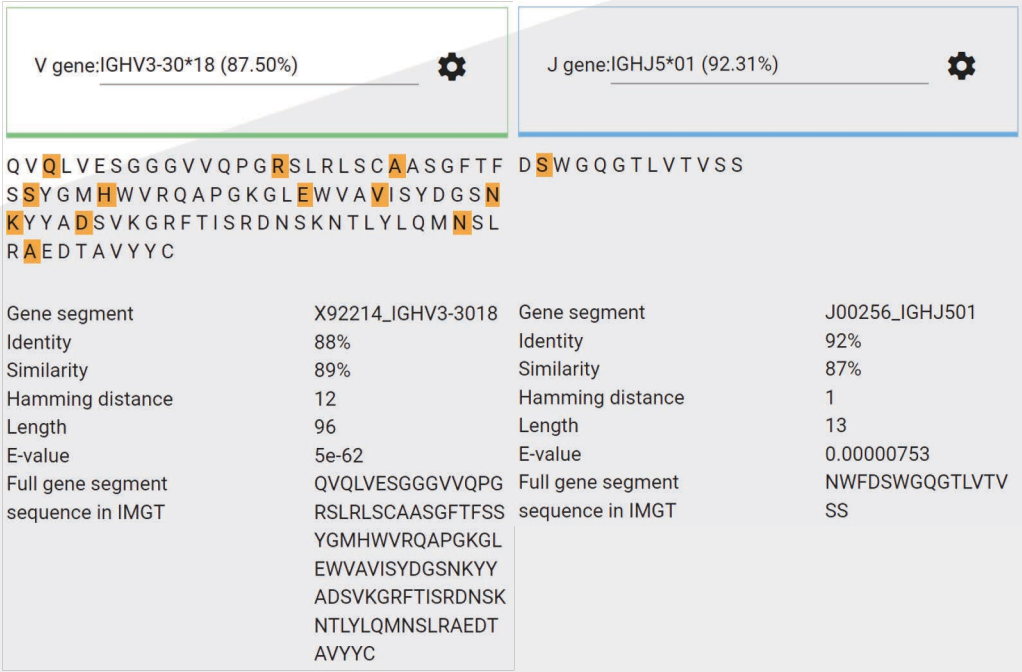


Identify the closest germline genes
supports IMGT germlines for 16 species

Visualize differences with protein of interest and germline hit

To investigate affinity maturation

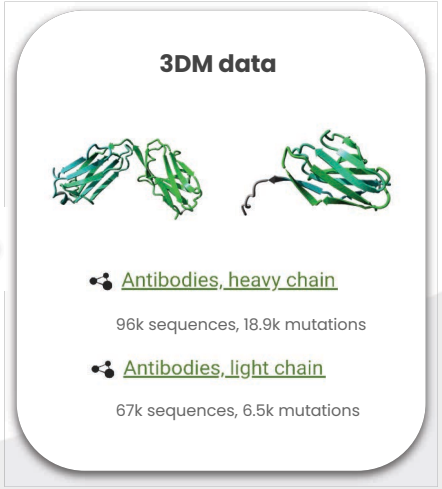
Humanisation: compare residues with human germline amino acid distribution



Antibody sequence annotation

annotate your in-house sequences with 3DM in bulk

in-house Ab sequences (100k+)



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

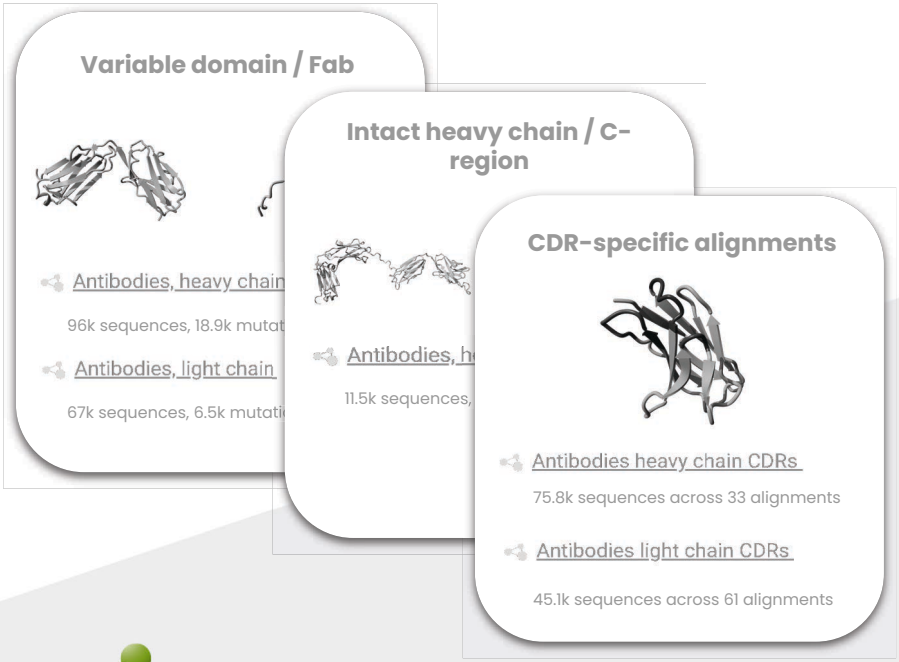
Humanness, humanization target residues

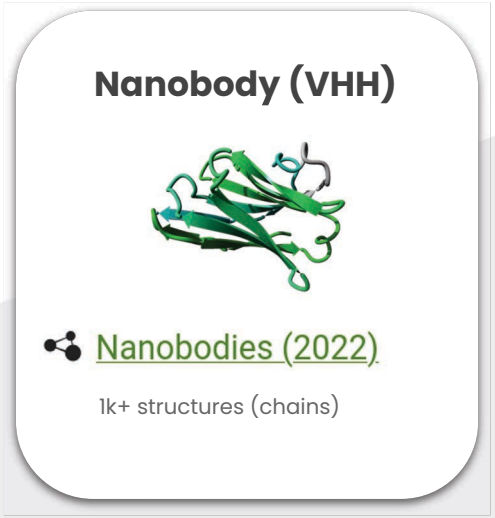
. . .

Correlate with in-house features

Nanobody 3DM system

integrating VHH and nanobody data





3DM systems

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

Custom development

3DM as a starting point

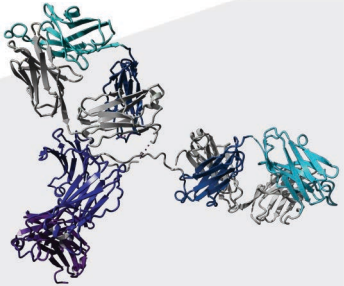
Finetune 3DM to your needs:

- 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

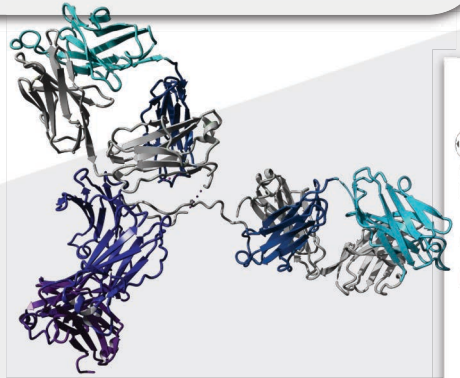


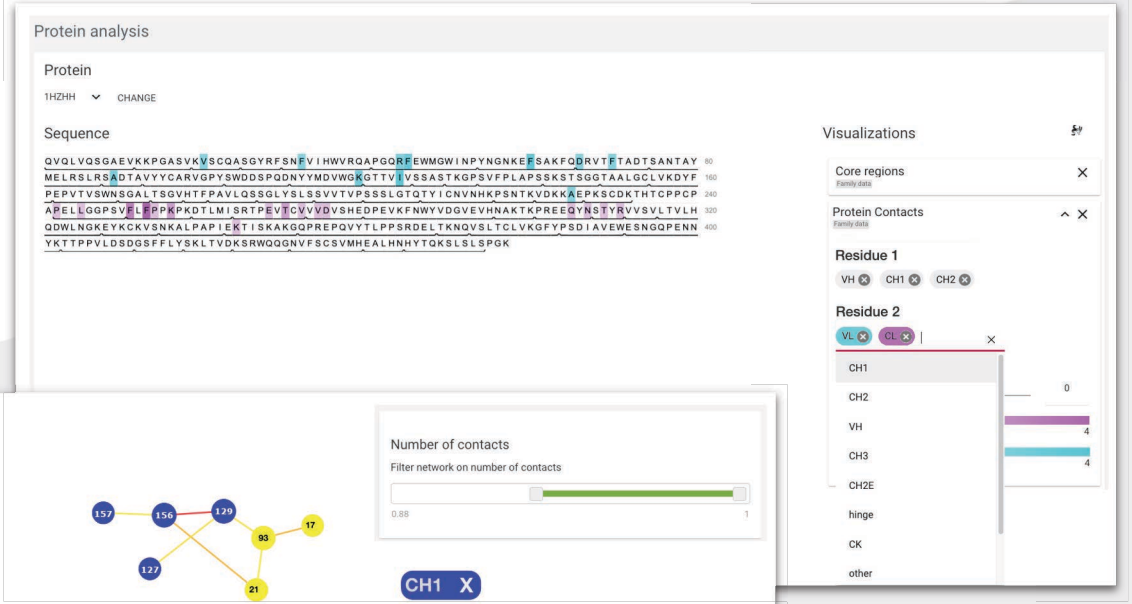
Antibody domain contact statistics

For analysis of domain type-specific contacts

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

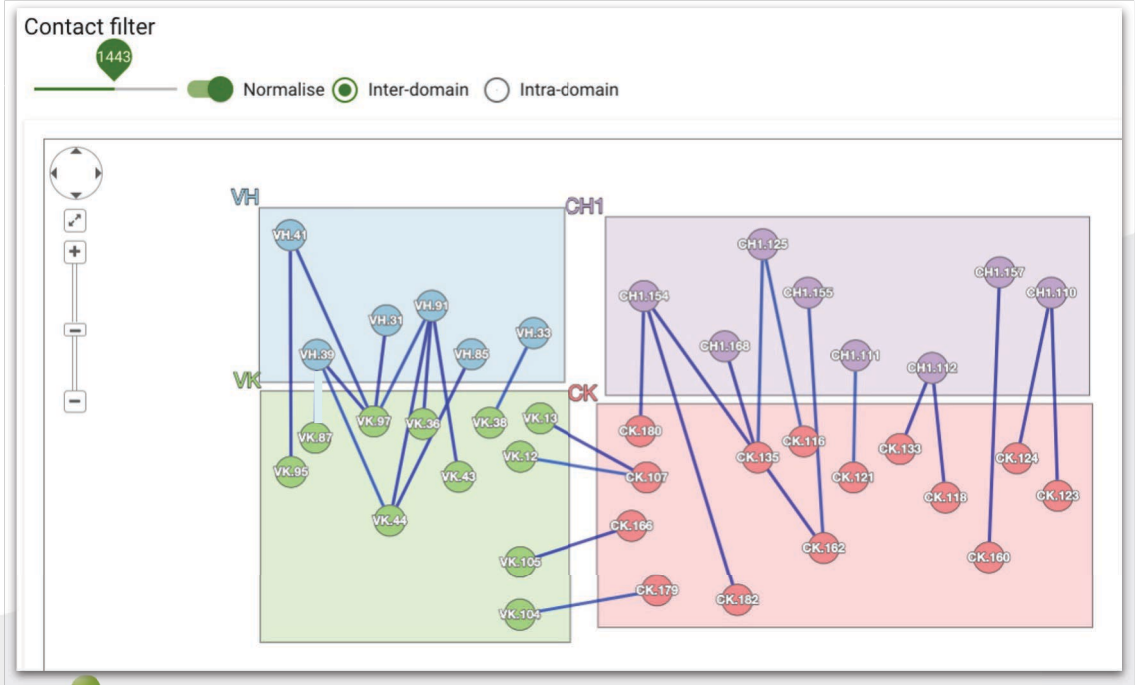
distinguish between antigen contacts vs light/heavy chain contacts

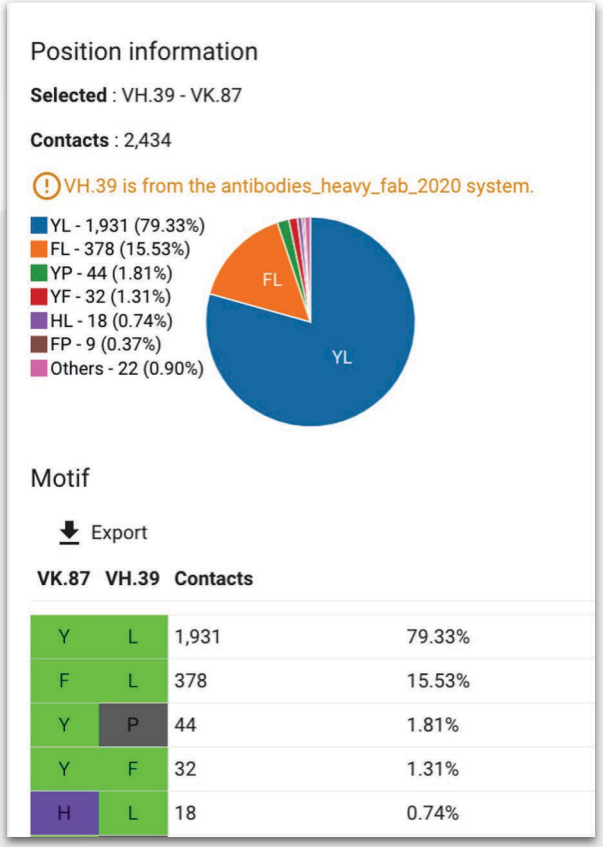




Antibody domain contacts

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





HARMONY: Antibody codon optimizer

Sequence optimization on complex constructs

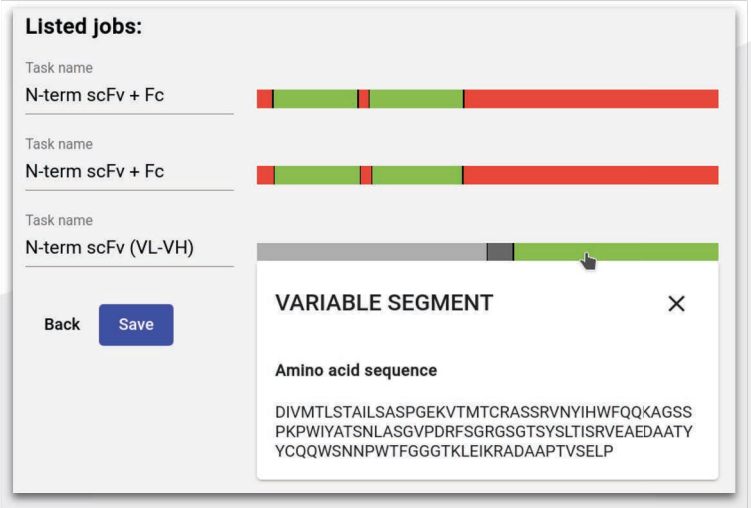
Bio-Prodict

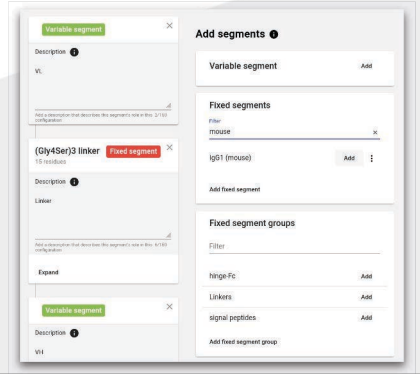
bio-prodict
protein predictions

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





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

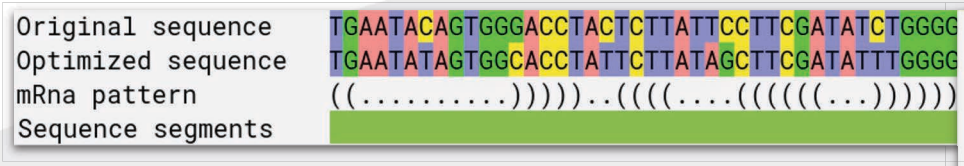
Batch optimize sequences for 100's of constructs

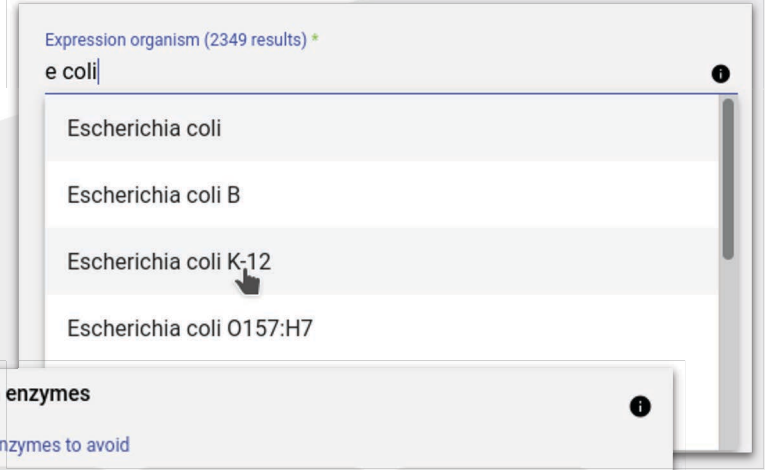
Data import/export and **API support**

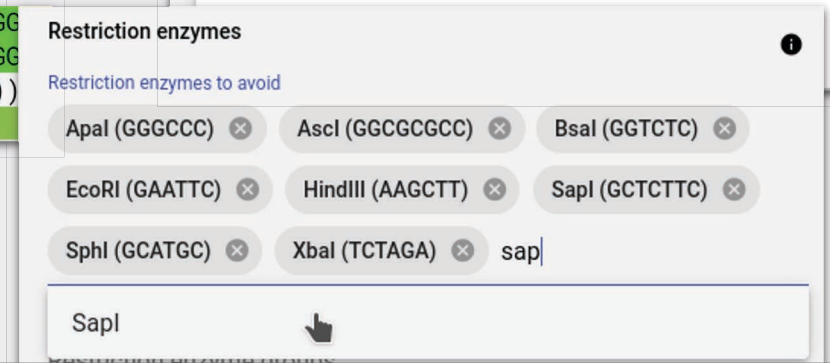
HARMONY: Antibody codon optimizer

Sequence optimization on complex constructs

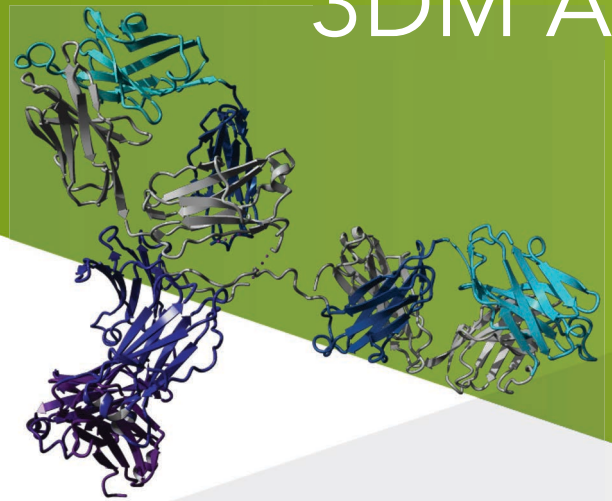
- Harmonize codon usage to match frequency of expression host (~60k organisms supported)
- Optimize GC content & avoid repeats
- Identification and removal of restriction sites
- Removal of mRNA hairpins (using MFE)
- Simultaneously optimized in a single run







3DM Antibody Suite



bio-prodict
protein predictions