

3DM: protein superfamily analysis system

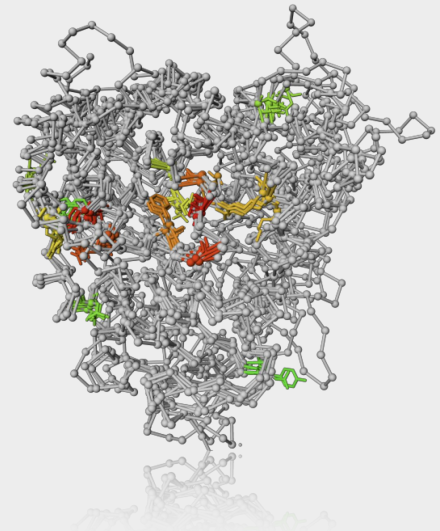
Catalyzing protein R&D

- Protein discovery
- Protein engineering
- Patent 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
- **3DM Tools give insight in your protein**
 - Use superfamily data to analyse your protein
- **Solves the numbering problem**
 - Seamlessly transfer data between homologous proteins
- **High-quality alignments**
 - Structure-based multiple sequence alignment



Example: Subtilisin

Highlight positions described in literature to affect specificity

Sequence projection

MMRKSSFVLGMLTALMLVFTMAFSDSASAAQPAKNVEKDYIVGFKSGVKTASVKKDI IKESGGKVDKQFR 70
IINAAKAKLDKEALKEVKNDPDVAYVEEDHVAHALAQTPYGIPLIKADKVQAQGFKGANVKVAVLDTGI 140
QASHPDLNVVGAS FVAGEAYNTDGNHGHGTHVAGTVAALDNTTGVVLGVAPSVSLYAVKVLNSSGSGSYSG 210
IVSGIEWATTNGMDVINMSLGGASG STAMKQAVDNAYARGVVVVAAAGNSGSSGNTNTIGYPAKYDSVIA 280
VGAVDNSNRASFSSVGALEVMAPGAGVYSTYPTNTYATLNGTSMASPHVAGAAALILSKHPNLSASQV 350
RNRLSSTATYLGSSFFYYGKGLINVEAAQ

T 237 Threonine (Thr)
Alignment position: [123](#) (aligned in core region)

AMINO ACID DISTRIBUTION
Protein/family data

PROTEIN MUTATIONS
Literature data: 1

KEYWORD MUTATIONS
Literature data

Subfamily 1AF4A

Conservation of T
13.09% (300 out of 2,292)



▲ Based on subfamily

Superfamily

Conservation of T
6.22% (4,194 out of 67,463)



Add



Visualizations

Group on data type

Core regions



Domain annotations



Literature data

Protein mutations



Keyword mutations

specificity



Keyword
specificity

Top

All



Search scope

☒ Superfamily

☐ This protein

0 1

13

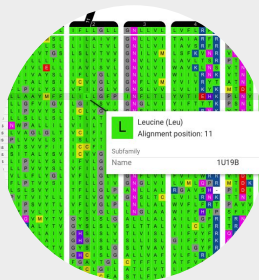
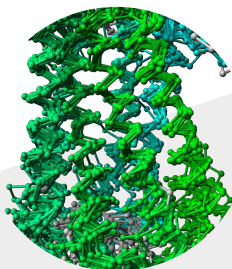
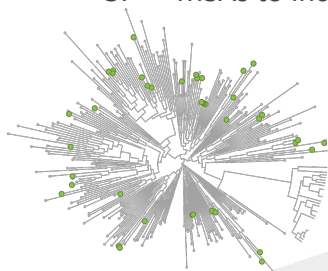
Mutations



The power of 3DM

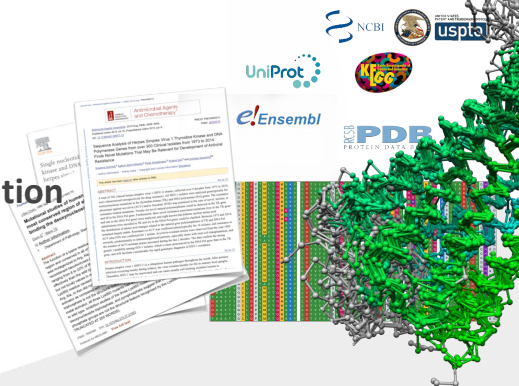
Structure-based MSA

1. Select structural templates to represent subfamilies
2. Structural alignment to connect sequentially diverse subfamilies
3. MSAs to include all proteins



Protein data integration

- Literature
- Patents
- Structure data
- Alignment data



Example: GPCR 3DM System

43 subfamilies connected:

- >900 integrated protein structures
- >85,000 WT proteins aligned
- >20,000 patented sequences
- >50,000 mutations described in the literature
 - searchable for their effects!

This data can be visualized in any protein or structure!

Powerful information system that **enhances your protein R&D**

Enhance your protein R&D with 3DM

Applications and Capabilities

- **Protein Insights**

- Conservation, contacts and literature

- **Protein discovery**

- Find new enzymes

- **Protein engineering**

- Smart mutation libraries

- **Patent analysis**

- Determine IP & FTO

- **Strain engineering**

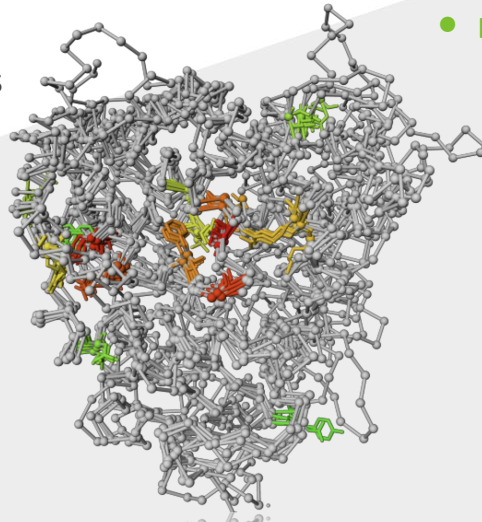
- Investigate top variants

- **Small molecule tractability**

- Compare (human) pockets in the family

- **mAb therapeutics**

- antibody sequence analysis



Protein analysis

Visualise data from the family on top of any protein

Sequence projection

MRGKKVWISLLFALALIFTMAFGSTSSAQAAGKSNGEKKYIVGFKQTMSTMSAAKKDVI SEKGGKVQKQFKYVDAASATLNEKAVKELKKDPSVAYVEE 100
DHVAHAYAQSVPYGVSIKAPALHSQGYTGSNVKVAVIDSGIDSSHPDLKVAGGASMVPSETNPFQDNNSHGTHVAGTVAALNNSIGVLGVAPSASLYAV 200
KVLGADGSGQYSWINGIEWAIANNMDVINNSLGSPSGSAALKAAVDKAVASGVVVVAAAGNEGTSGSSSTVGYPGKYPSVIAVGAVDSSNQRAFSSVG 300
PELDVMAPGVSIQSTLPGNKYGAYNGTSMASPHVAGAAALILSKHPNWTNTQVRSSLENTTTTKLGDSFYFGKGLINVQAAQAQ

L 233 Leucine (Leu)
Alignment position: 116 (aligned in core region)

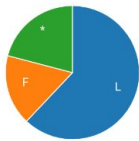
AMINO ACID DISTRIBUTION
Protein/family data

PROTEIN MUTATIONS
Protein data: 9

FAMILY CONTACTS
Family data: 102

Subfamily 1SUBA

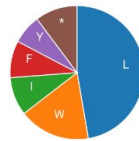
Conservation of L
62.04% (1,198 out of 1,931)



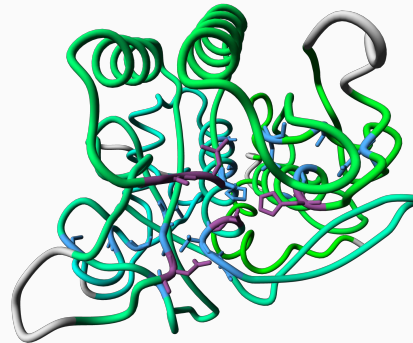
Based on subfamily

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



Patent analysis

Determine FTO

Quickly assess IP, including:

- Patented mutations
- Patent sequence variations

```
MRGKKVWISLLFALALIFTMAFGSTSSAQAAGKSNGEKKYIVGFKQTMSTMSAAKKKDV I 60
SEKGGKVQKQFKYVDAASATLNEKAVKELKKDPSVAYVEEDHVAHAYAQSVPYGVSIKA 120
PALHSQGYTGSNVKVAVIDSGIDSSHPDLKVAGGASMVPSETNPFQDNNSHGTHVAGTVA 180
ALNNSIGVLGVAPSASLYAVKVLGADGSGQYSWIINGIEWAIIANNMDV NMS GGPSSGA 240
ALKAAVDKAVASGVVVVAAAGNEGTS GSSSTVGYPGKYPSVIAVGAVDSSNQRAFSSVG 300
PELDVMAPGVS IQSTLPGNKYGAYNGTSMASPHVAGAAALILSKHPNWTNTQVRSSLENT 360
TTKLGDSFYYGKGLINVQAAAQ
```

Methods for cleaning using a variant protease derived from subtilisin

The present invention provides protease variants, compositions comprising protease variants, and methods of using such protease variants and compositions.

Variants

SEQ ID NO: 14	Identity: 97%	A I A N N M D V I N M S L G G P S G S A A M231I
		A I A N N M D V I N I S L G A P S G S A A



SEQ ID NO: 14	Identity: 97%	A I A N N M D V I N M S L G G P S G S A A M231I
		A I A N N M D V I N I S L G A P S G S A A



Simple family

4 Patents

[US9157052 B2](#)

Publication: 2015-10-13

Applicant: DANISCO US INC

✓ Granted

[US8728790 B2](#)

Publication: 2014-05-20

Applicant: DANISCO US INC,

✓ Granted



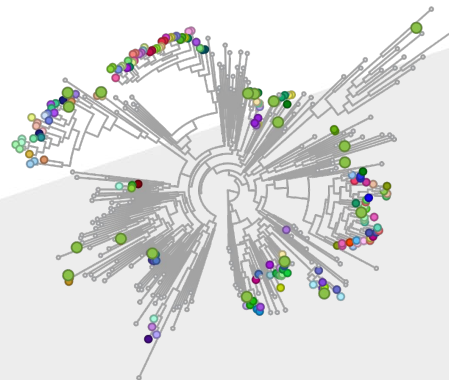
Intelligent in-silico panel design with 3DM

Use the alignment to determine diversity

A single active-site residue can change specificity

- Smart diversity: focus on diversity at key positions
- Select likely to express proteins
- Evaluate FTO

>90% expression rates achieved by our partners



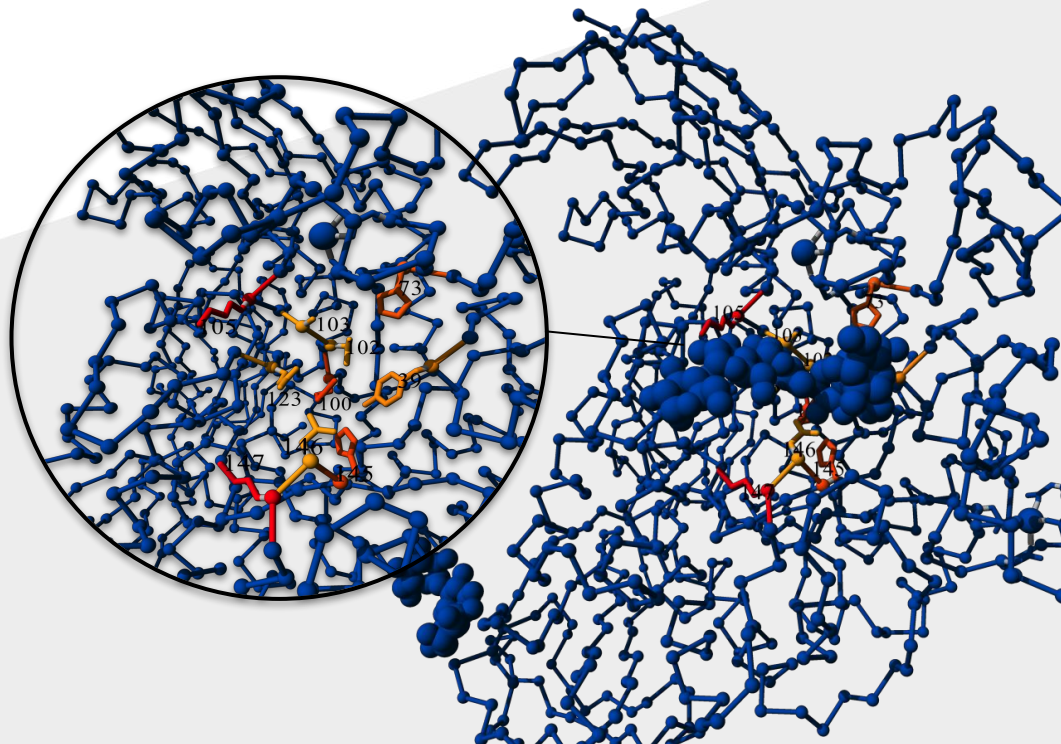
Number	50	54	58	113	207	Proteins
1.	D	H	V	N	S	28,639
2.	D	H	T	S	S	7,999
3.	P	H	C	S	S	4,471
4.	D	H	V	S	S	4,173
5.	D	H	C	S	S	3,492
6.	D	H	T	N	S	2,843



Pocket analysis

- Compare pocket residues between human proteins of the same fold
- Compare pockets with model organisms
- Analyse compounds present in pockets

	39	40	73	100	102	103	105	123	145	146
P19961	Y	Q	H	R	D	A	K	E	H	D
P04746	Y	Q	H	R	D	A	K	E	H	D
P04745	Y	Q	H	R	D	A	K	E	H	D
Q04446	Y	Q	H	R	D	G	T	E	H	D
Q07837	Y	G	H	S	D	A	K	-	P	D
P17050	-	-	N	K	D	G	F	S	I	G
P06280	D	S	N	K	D	G	Y	E	I	G
P08195	V	A	Y	Q	R	D	E	G	-	-

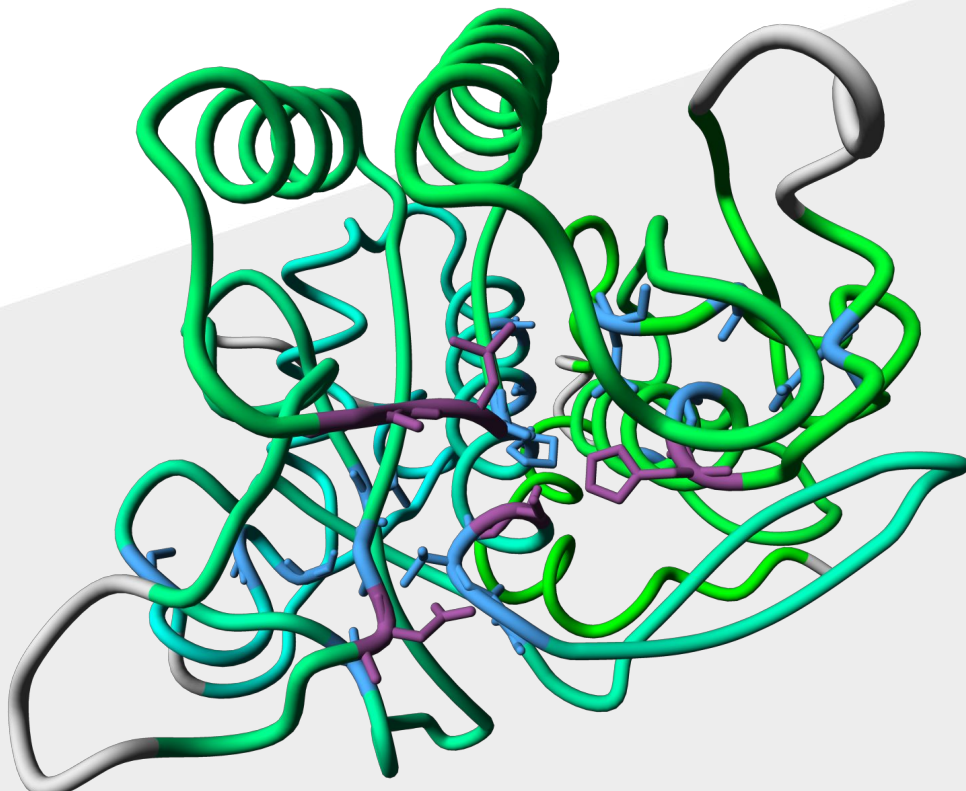


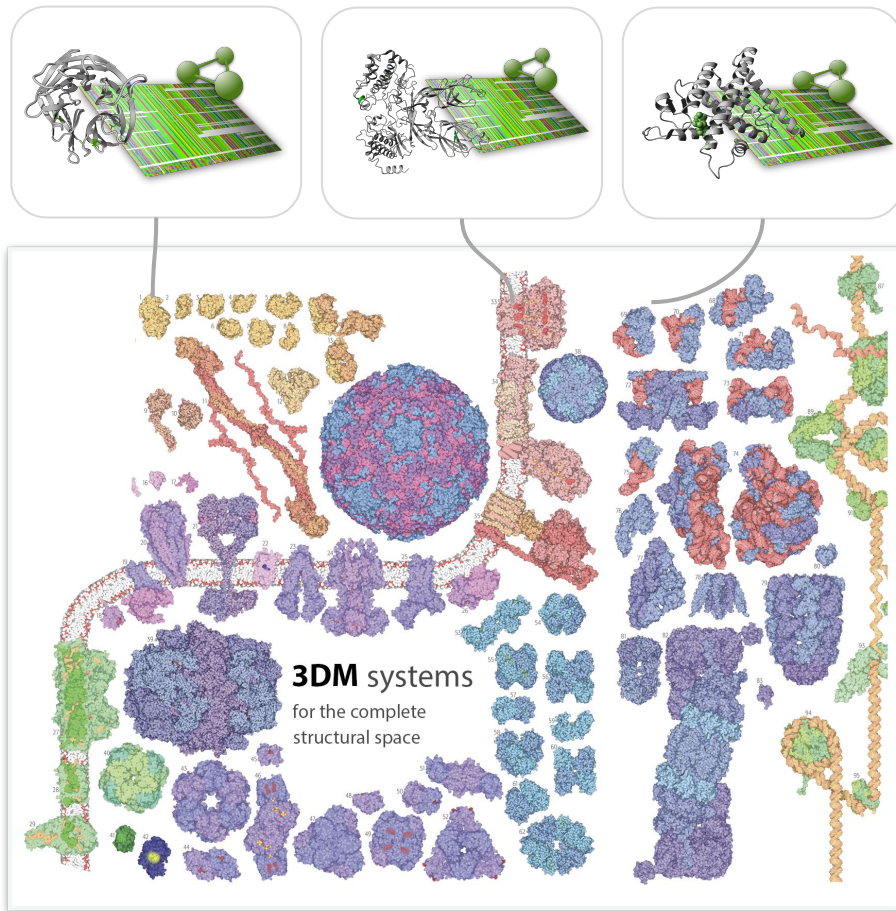
Structure visualisations

Show 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 50,000 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.

How we work

Custom 3DM systems (license)

- Custom build 3DM system
- We can include your own structures/sequences
- Alignments fine tuned for your use case

PDB-Wide system access (license)

- 3DM systems ready to go for any protein target
- Includes tickets for custom 3DM systems or custom development

Custom development (tickets or project based)

- Order bespoke features to get the most out of 3DM
- API access: Integrate 3DM data into inhouse computational biology pipelines
- Whole exome 3DM systems for your production strain

Q&A

More information:
www.bio-product.com

