BIOTECHNO 2021

Analysis of the tertiary protein structure for Protein Misfolding Diseases

Panagiotis Vlamos

vlamosp@gmail.com
Professor, Department of Head of the
Director of Bioinformatics and Human Electrophysiology Lab (BiHELab)
Chair of Hellenic Initiative Against Alzheimer's (HIAAD)

Panagiotis Vlamos

Panagiotis Vlamos is a Full **Professor,** Chairman of the University Research Center of the Ionian University and Director of the Laboratory of Bioinformatics and Human Electrophysiology **(BiHELab)** of the Ionian University.

He completed his undergraduate studies in Mathematics at the University of Athens and obtained his Doctorate in Applied Mathematics from the National Technical University of Athens. He has received many awards and has authored more than 250 papers in scientific journals and conference proceedings, as well as 16 educational books. His research goal is to help bridge the translation gap from data to models and from models to drug discovery and personalized therapy, developing novel approaches to biological and clinical problems, using high-performance computational methods.

Professor Vlamos is a pioneer in the field of Computational Biomarkers and he was the first to introduce the notion of Meta-biomarkers. In addition, Professor Vlamos is president of the Hellenic Initiative against Alzheimer's (HIAAD), a joint collaboration between researchers from Johns Hopkins University in the USA and scientists from BiHELab who combine their knowledge and expertise in order to tackle the "epidemic" of Alzheimer's Disease and related disorders in Greece. He is the founder and Head of the Program Committee of the biannual World Congress on "Genetics, Geriatrics and Research in Neurodegenerative Diseases" **(GeNeDis).**

Professor Vlamos is one of the three editors of the Handbook of Computational Neurodegeneration to be published by Springer International Publishing and the founder of the MSc Program “Bioinformatics & Neuroinformatics”.

BiHeLab’s Main Interests:

- **BiHELab’s** goal is to help bridge the translational gap from data to models and from models to drug discovery and personalized therapy by fostering collaborations and developing original quantitative approaches to biological and clinical problems.

- The ***Bioinformatics and Human Electrophysiology Lab (BiHELab)*** focuses on recent advances in geriatrics and neurodegeneration, ranging from basic science to clinical and pharmaceutical developments.

- The **main objective** of the laboratory is the creation of new and effective protocols for the diagnosis of various types of dementia and neurological disorders through identification, mapping, biological analysis as well as mathematical modelling and simulation of all factors associated with these diseases, in order to improve existing techniques and to design new targeted treatments.

Protein Misfolding Diseases

Human diseases caused by defects in protein folding, stability and aggregation

Disease	Protein affected	Description
Cystic fibrosis	Cystic fibrosis transmembrane conductance regulator (CFTR)	The ΔPhe508 mutant has wild-type activity, but impaired folding in the endoplasmic reticulum leads to degradation.
α1 Antitrypsin deficiency	α1 Antitrypsin (also known as SERPINA1)	80% of Glu342Lys mutants misfold and are degraded. Pathology is due to aggregation in patients with a reduced degradation rate.
SCAD deficiency	Short-chain acyl-CoA dehydrogenase (SCAD)	Impaired folding of Arg22Trp mutants leads to rapid degradation.
Alzheimer disease	Presenilin, γ-secretase	Mutations cause incorrect cleavage by the γ-secretase protease to produce the amyloid β-peptide; this aggregates into extracellular amyloid plaques.
Parkinson disease	α-Synuclein	Oxidative damage causes misfolding and aggregation. Hereditary forms are linked to deficiency in ubiquitin-mediated degradation.
Huntington disease	Huntingtin	CAG expansions in the Huntingtin gene lead to an abundance of polyglutamine fragments that aggregate and associate non-specifically with other cellular proteins.
Sickle cell anaemia	Haemoglobin	The Glu6Val mutation leads to aggregation in red blood cells.

*Nature Reviews Genetics 2005;***6**:678-687

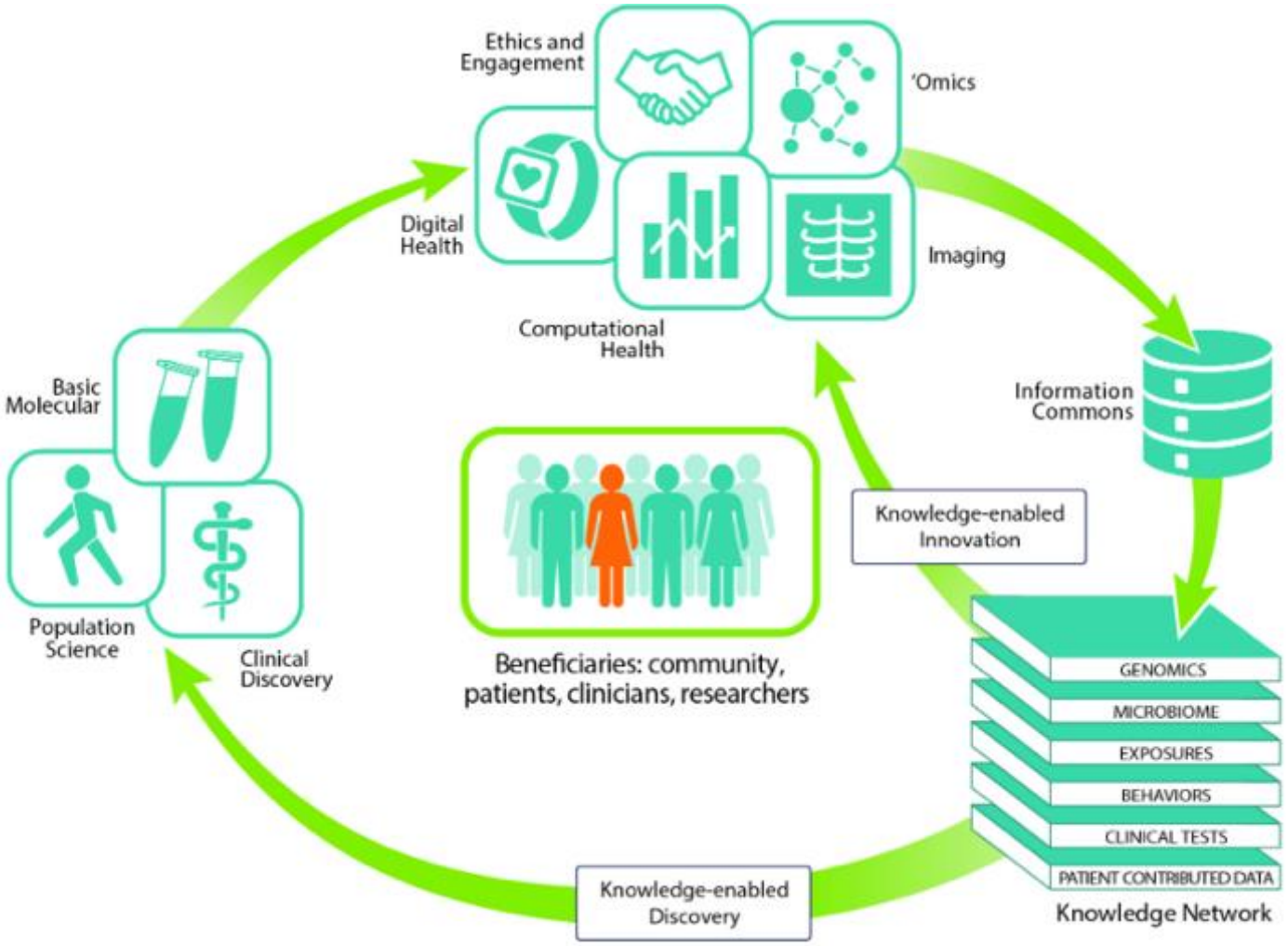
BiHE Lab

protein tangles that grow and connections between nerve cells shrivel

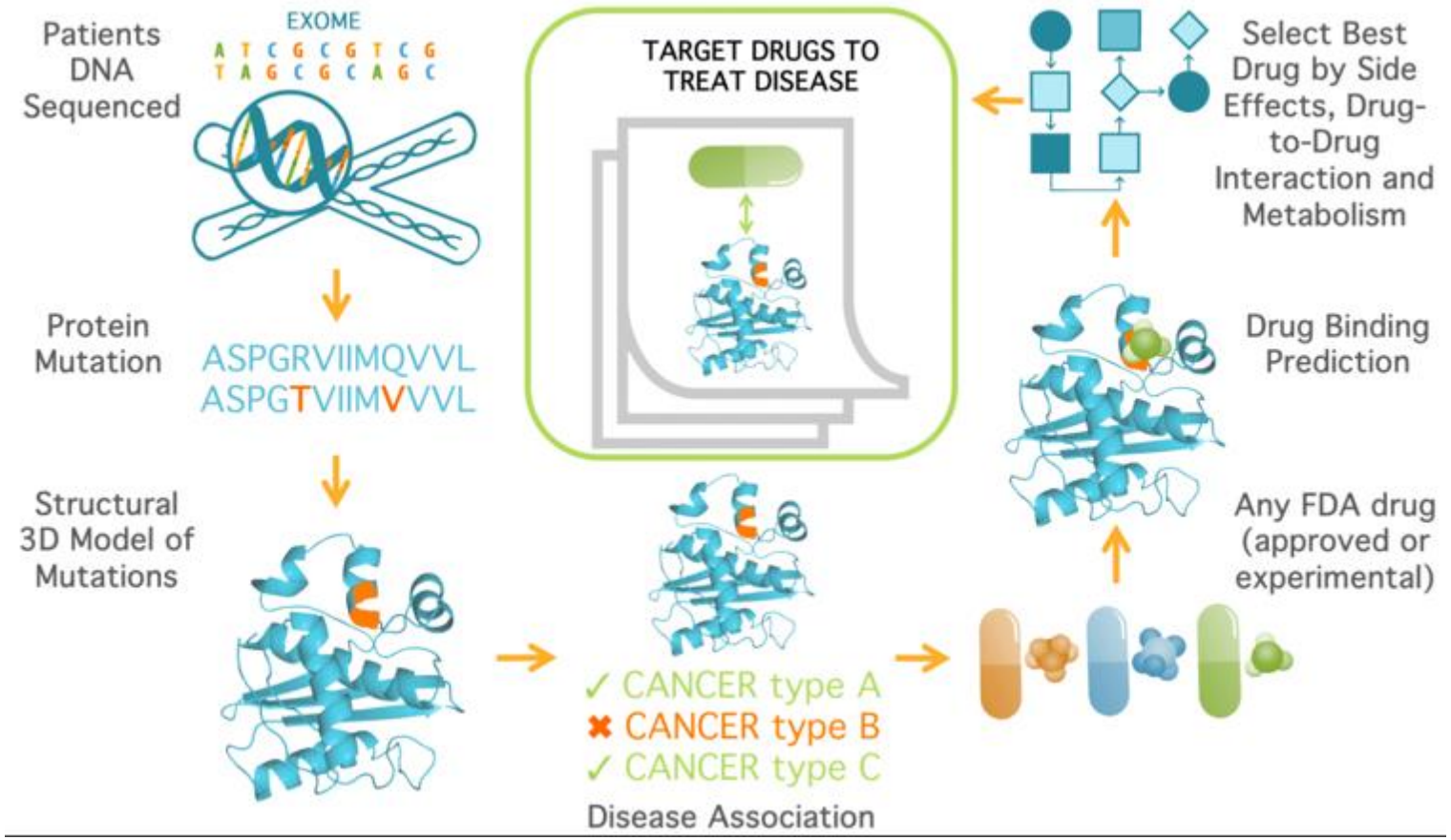
Precision medicine

- Precision medicine (PM) is a medical model that proposes the customization of healthcare, with medical decisions, treatments, practices, or products being tailored to the individual patient, instead of a one-drug-fits-all model.

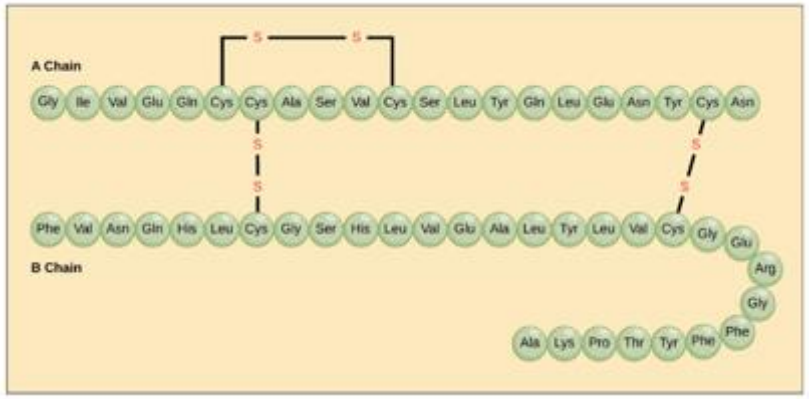
- Findings from basic, clinical, and social/behavioral research, data from digital health, 'omic technologies, imaging, and computational health sciences and ethical and legal guidelines are integrated into a knowledge network, which informs both science and care for individuals and populations.



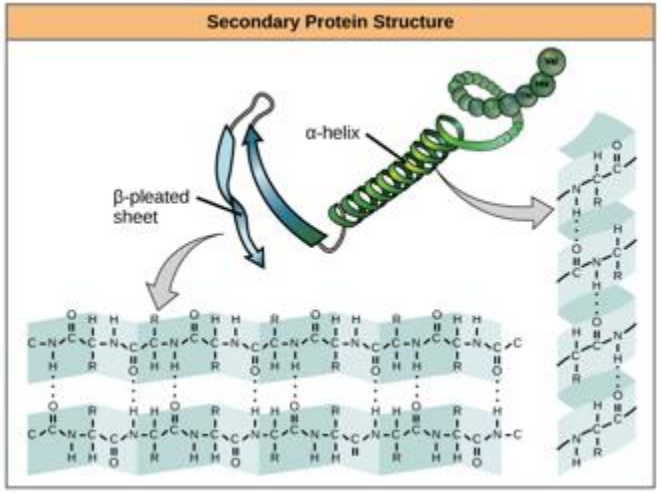
Proteins structures and precision medicine



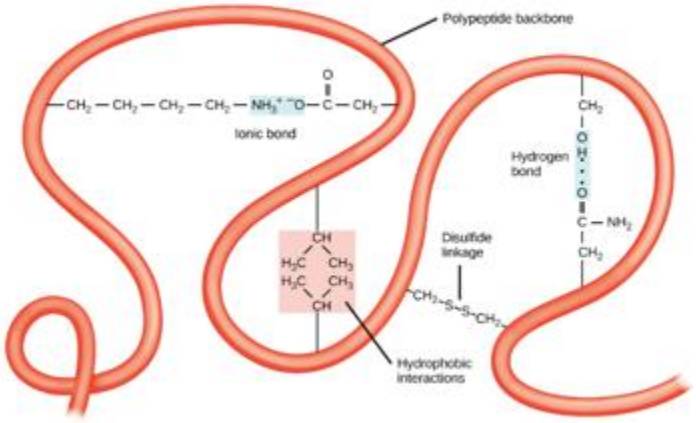
Protein Structure - Overview



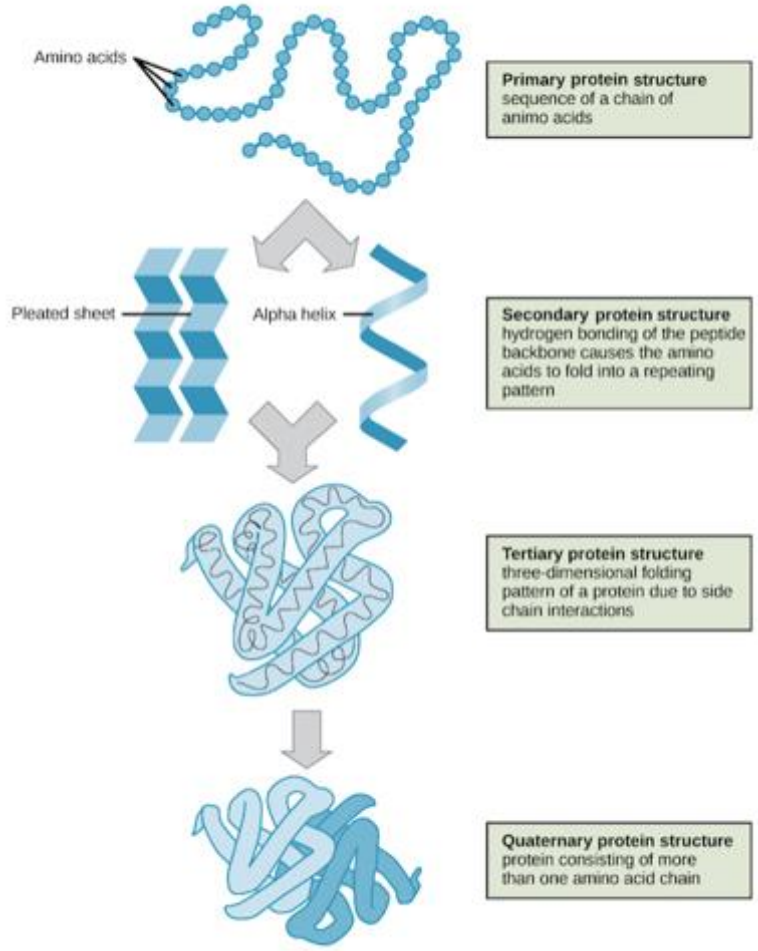
primary structure



Secondary structure



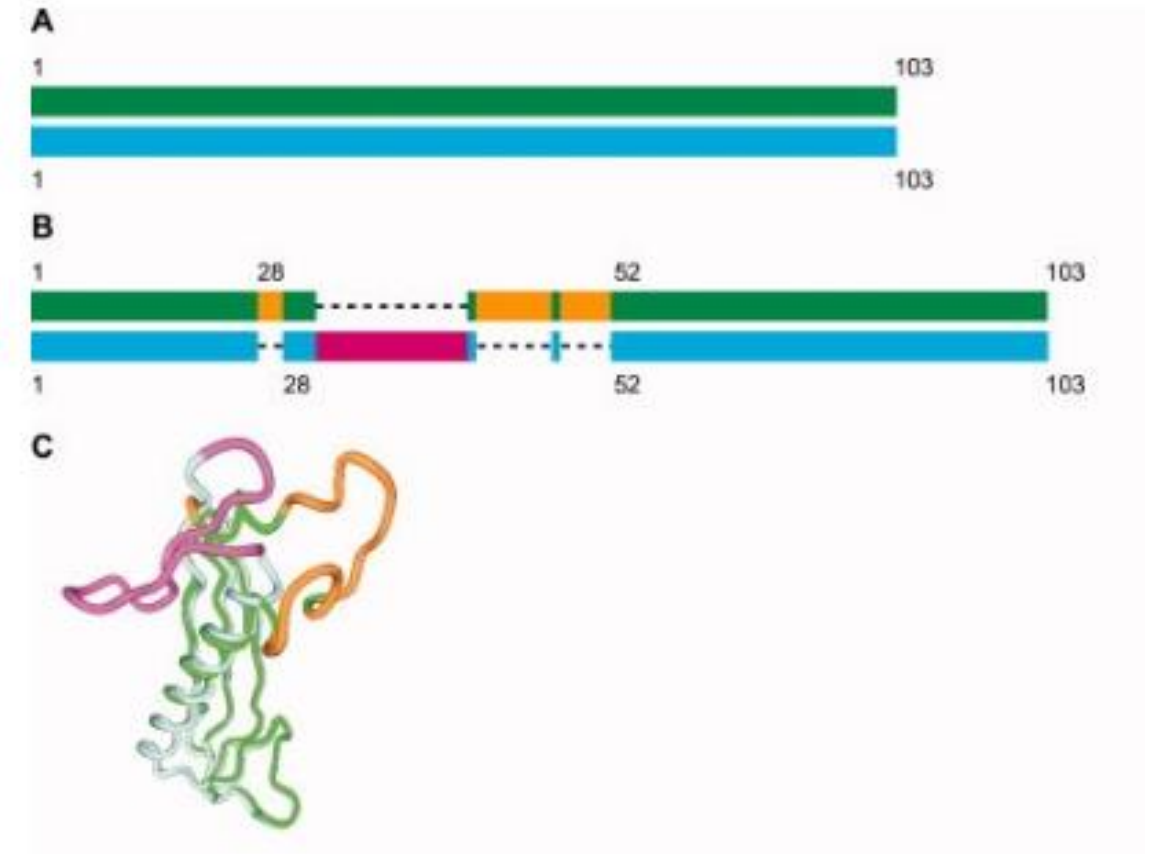
Tertiary structure



Quaternary structure

How primary and tertiary differs

- Schematic representation of the sequence alignment (**A**) versus structural alignment (**B**) of chain A versus chain D from PDB ID 1vr4. The two chains are 100% identical in sequence. The aligned parts are colored green (chain A) and cyan (chain D), while the unaligned parts are colored orange and magenta, respectively.



Kosloff, M., & Kolodny, R. (2008). Sequence-similar, structure-dissimilar protein pairs in the PDB. Proteins: Structure, Function, and Bioinformatics, 71(2), 891-902.

How we use databases and online tools in building analytical pipelines

Nucleic Acids Research

Issues | Section browse | Advance articles | Submit | Purchase | About | All Nucleic Acids Re

Volume 49, Issue D1
8 January 2021

Article Contents

Comments (0)

Next >

The 2021 *Nucleic Acids Research* database issue and the online molecular biology database collection

Daniel J Rigden ✉, Xosé M Fernández

Nucleic Acids Research, Volume 49, Issue D1, 8 January 2021, Pages D1–D9,
https://doi.org/10.1093/nar/gkaa1216
Published: 23 December 2020

PDF | Split View | Cite | Permissions | Share

Abstract

The 2021 *Nucleic Acids Research* database Issue contains 189 papers spanning a wide range of biological fields and investigation. It includes 89 papers reporting on new databases and 90 covering recent changes to resources previously published in the Issue. A further ten are updates on databases most recently published elsewhere. Seven new databases focus on COVID-19 and SARS-CoV-2 and many others offer resources for studying the virus. Major returning nucleic acid databases include NONCODE, Rfam and RNAcentral. Protein family and domain databases include COG, Pfam, SMART and Panther. Protein structures are covered by RCSB PDB and dispersed proteins by PED and MobiDB. In metabolism and signalling, STRING, KEGG and WikiPathways are featured, along with returning KLIFS and new DKK and KinaseMD, all focused on kinases. IMG/M and IMG/VR update in the microbial and viral genome resources section, while human and model organism genomics resources include Flybase, Ensembl and UCSC Genome Browser. Cancer studies are covered by updates from canSAR and PINA, as well as newcomers CNCdatabase and Oncovar for cancer drivers. Plant comparative genomics is catered for by updates from Gramene and GreenPhylDB. The entire Database Issue is freely available online on the *Nucleic Acids Research* website (https://academic.oup.com/nar). The NAR

The NAR online Molecular Biology Database Collection has been substantially updated, revisiting nearly 1000 entries, adding 90 new resources and eliminating 86 obsolete databases, bringing the current total to 1641 databases. It is available at https://www.oxfordjournals.org/nar/database/c/.

OXFORD ACADEMIC | Journals

You are here: NAR Journal Home » Database Summary Paper Categories

NAR Database Summary Paper Category List

- Compilation Paper
- Category List
- Alphabetical List
- Category/Paper List
- Search Summary Papers

Nucleotide Sequence Databases
RNA sequence databases
Protein sequence databases
Structure Databases
Genomics Databases (non-vertebrate)
Metabolic and Signaling Pathways
Human and other Vertebrate Genomes
Human Genes and Diseases
Microarray Data and other Gene Expression Databases
Proteomics Resources
Other Molecular Biology Databases
Organelle databases
Plant databases
Immunological databases
Cell biology

- Compilation Paper
- Category List
- Alphabetical List
- Category/Paper List
- Search Summary Papers

Why to archive research data

Accessibility
- One-stop shop
- Uniform data representation/annotation

Persistence
- Typical websites have a half-life of 2 years
- Professional management

Context
- Comparisons against all other entries
- Validation
- Integration with other resources

Facilitate further analysis
- Database-wide studies and data-mining

The archive cost is ~1% of the cost of generating the data
Immeasurable scientific value

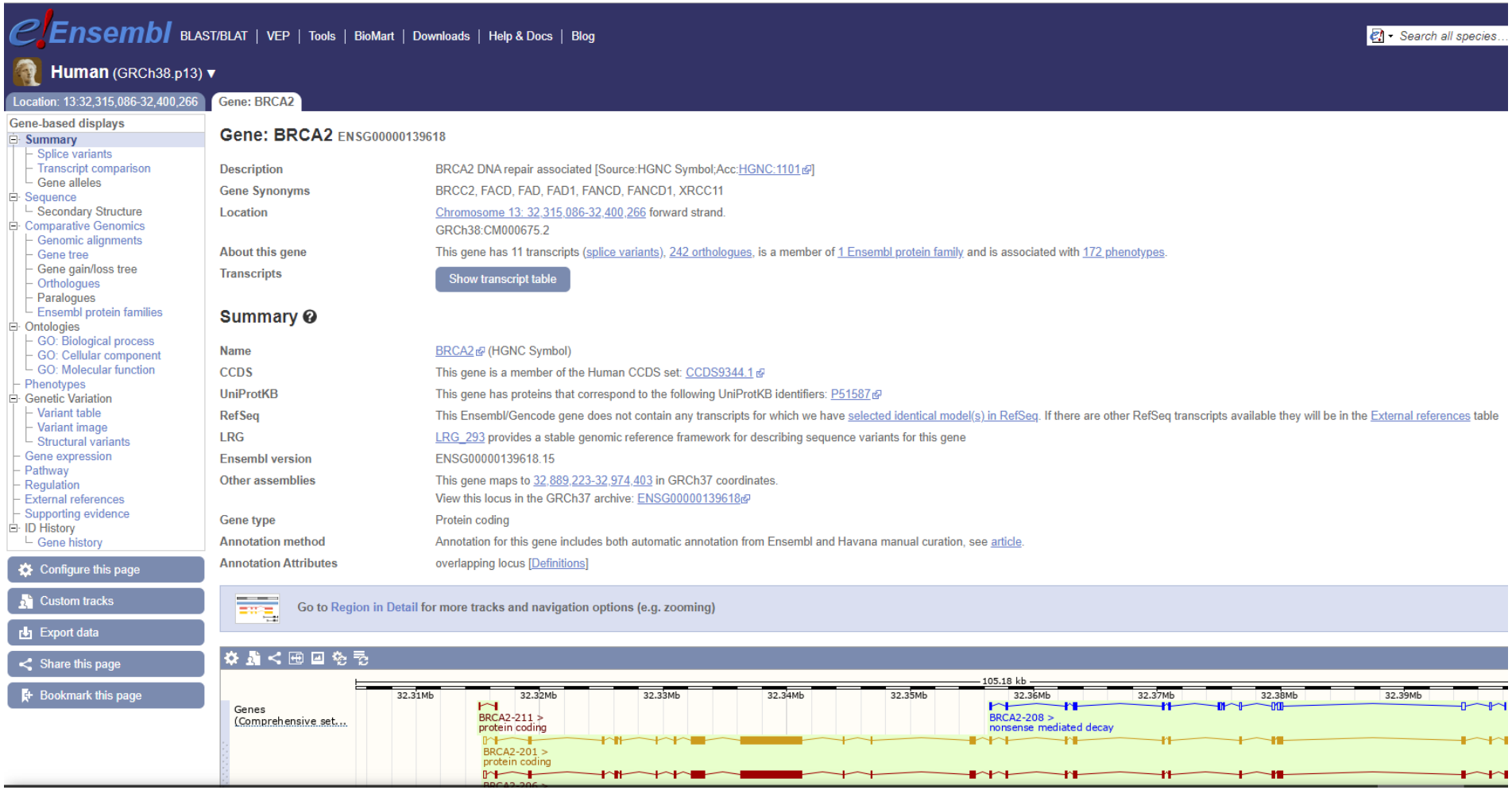
Ensembl

Ensembl is a genome browser for vertebrate genomes that supports research in comparative genomics, evolution, sequence variation and transcriptional regulation.

Components:

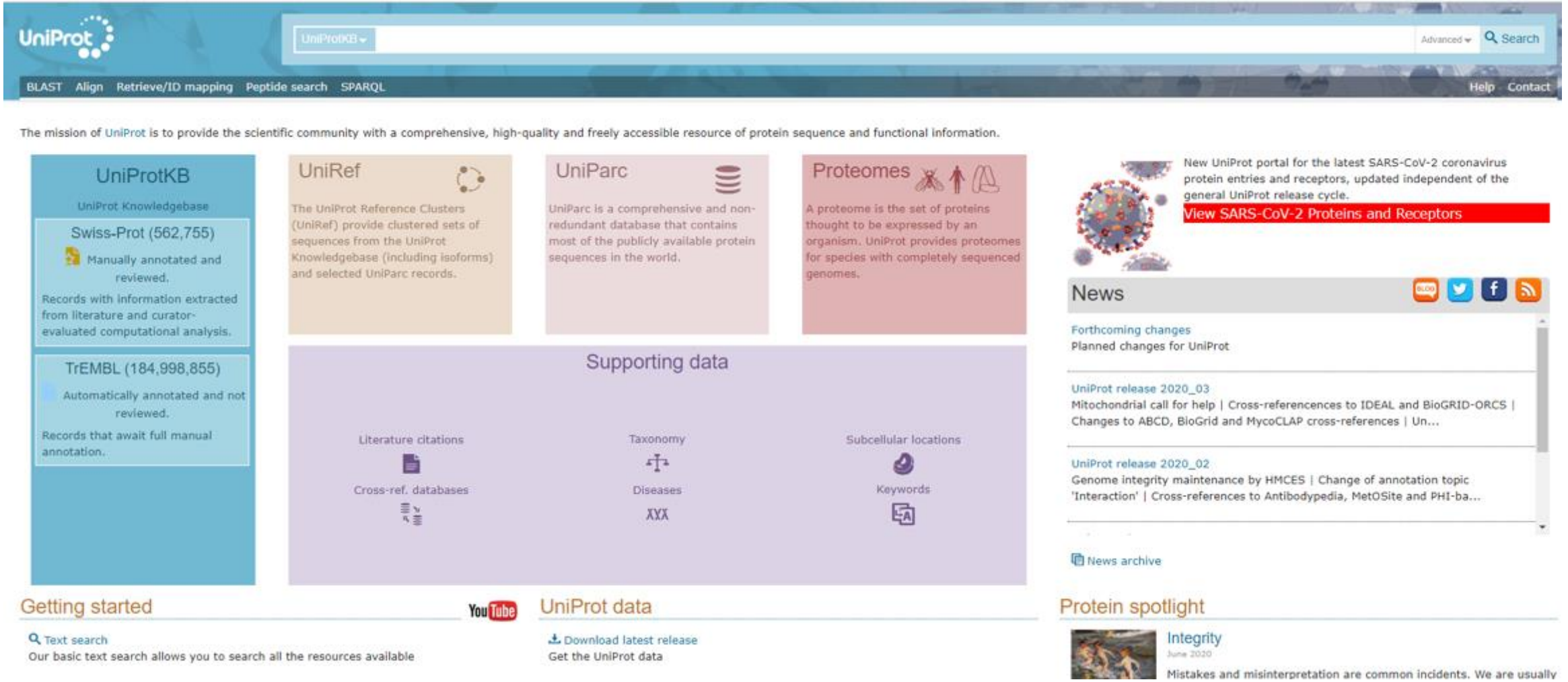
- genome browser
- tools
- databases
- APIs

Ensembl to provides a centralized resource for geneticists, molecular biologists and other researchers **studying the genomes**

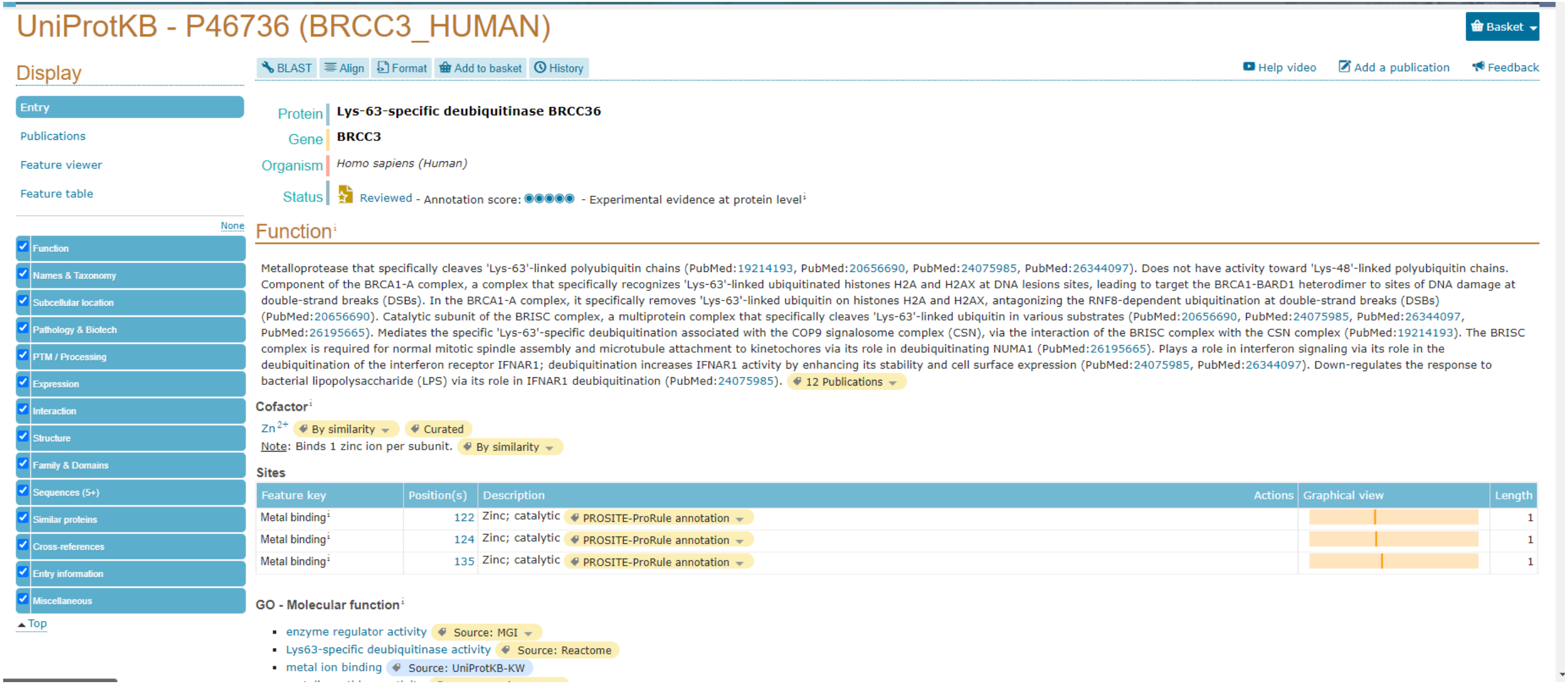


UniProt

The Universal Protein Resource (UniProt) is a comprehensive resource for protein sequence and annotation data. The UniProt databases are the UniProt Knowledgebase (UniProtKB), the UniProt Reference Clusters (UniRef), and the UniProt Archive (UniParc).



UniProt


UniProtKB - P46736 (BRCC3_HUMAN)
Basket
Display
BLAST
Align
Format
Add to basket
History
Help video
Add a publication
Feedback
Entry
Publications
Feature viewer
Feature table
None
Function
Names & Taxonomy
Subcellular location
Pathology & Biotech
PTM / Processing
Expression
Interaction
Structure
Family & Domains
Sequences (5+)
Similar proteins
Cross-references
Entry information
Miscellaneous
Top
Protein
Lys-63-specific deubiquitinase BRCC36
Gene
BRCC3
Organism
Homo sapiens (Human)
Status
Reviewed - Annotation score: - Experimental evidence at protein level
Function
Metalloprotease that specifically cleaves 'Lys-63'-linked polyubiquitin chains (PubMed:19214193, PubMed:20656690, PubMed:24075985, PubMed:26344097). Does not have activity toward 'Lys-48'-linked polyubiquitin chains. Component of the BRCA1-A complex, a complex that specifically recognizes 'Lys-63'-linked ubiquitinated histones H2A and H2AX at DNA lesions sites, leading to target the BRCA1-BARD1 heterodimer to sites of DNA damage at double-strand breaks (DSBs). In the BRCA1-A complex, it specifically removes 'Lys-63'-linked ubiquitin on histones H2A and H2AX, antagonizing the RNF8-dependent ubiquitination at double-strand breaks (DSBs) (PubMed:20656690). Catalytic subunit of the BRISC complex, a multiprotein complex that specifically cleaves 'Lys-63'-linked ubiquitin in various substrates (PubMed:20656690, PubMed:24075985, PubMed:26344097, PubMed:26195665). Mediates the specific 'Lys-63'-specific deubiquitination associated with the COP9 signalosome complex (CSN), via the interaction of the BRISC complex with the CSN complex (PubMed:19214193). The BRISC complex is required for normal mitotic spindle assembly and microtubule attachment to kinetochores via its role in deubiquitinating NUMA1 (PubMed:26195665). Plays a role in interferon signaling via its role in the deubiquitination of the interferon receptor IFNAR1; deubiquitination increases IFNAR1 activity by enhancing its stability and cell surface expression (PubMed:24075985, PubMed:26344097). Down-regulates the response to bacterial lipopolysaccharide (LPS) via its role in IFNAR1 deubiquitination (PubMed:24075985).
12 Publications
Cofactor
Zn2+
By similarity
Curated
Note: Binds 1 zinc ion per subunit.
By similarity
Sites
Feature key
Position(s)
Description
Actions
Graphical view
Length
Metal binding 122 Zinc; catalytic PROSITE-ProRule annotation 1
Metal binding 124 Zinc; catalytic PROSITE-ProRule annotation 1
Metal binding 135 Zinc; catalytic PROSITE-ProRule annotation 1
GO - Molecular function
enzyme regulator activity
Source: MGI
Lys63-specific deubiquitinase activity
Source: Reactome
metal ion binding
Source: UniProtKB-KW

UniProt

Display

Entry

Publications

Feature viewer

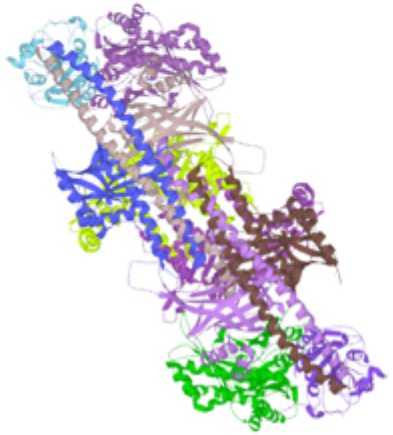
Feature table

None

- Function
- Names & Taxonomy
- Subcellular location
- Pathology & Biotech
- PTM / Processing
- Expression
- Interaction
- Structure
- Family & Domains
- Sequences (5+)
- Similar proteins
- Cross-references
- Entry information
- Miscellaneous

Top

Structure



PDB Entry	Method	Resolution	Chain	Positions	Links
6H3C	Electron mi...	3.90 Å	B/G	1-316	PDBe RCSB PDB PDBj PDBsum
6R8F	Electron mi...	3.80 Å	A/C	1-316	PDBe RCSB PDB PDBj PDBsum

3D structure databases

SMR	P46736
ModBase	Search...
PDBe-KB	Search...

Family & Domains

Domains and Repeats

Feature key	Position(s)	Description	Actions	Graphical view	Length
Domain	12 – 179	MPN PROSITE-ProRule annotation	Add BLAST		168

UniProt

Display

Entry

Publications

Feature viewer

Feature table

None

- Function
- Names & Taxonomy
- Subcellular location
- Pathology & Biotech
- PTM / Processing
- Expression
- Interaction
- Structure
- Family & Domains
- Sequences (5+)
- Similar proteins
- Cross-references
- Entry information
- Miscellaneous

Top

PROSITE | View protein in PROSITE
PS50249 MPN, 1 hit

Sequences (5+)

Sequence status: Complete.
Sequence processing: The displayed sequence is further processed into a mature form.
This entry describes **5** isoforms produced by **alternative splicing**. Align | Add to basket
This entry has 5 described isoforms and 6 potential isoforms that are computationally mapped. Show all | Align All

Isoform 2 (identifier: **P46736-1**) [UniParc] FASTA | Add to basket

This isoform has been chosen as the canonical *sequence. All positional information in this entry refers to it. This is also the sequence that appears in the downloadable versions of the entry.*

« Hide

Length: 316
Mass (Da): 36,072
Last modified: May 10, 2002 - v2
Checksum: 5720358C1A2F7421

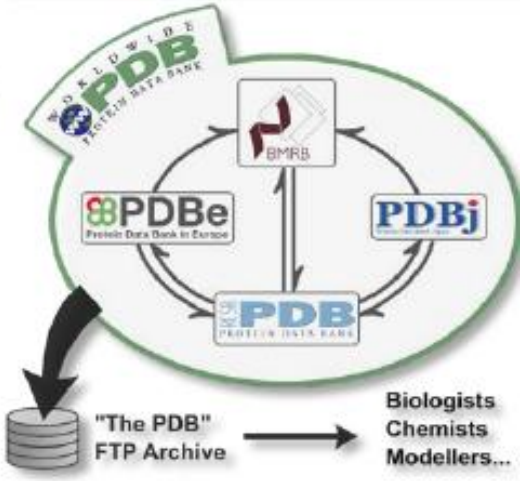
BLAST GO

```
        10         20         30         40         50
MAVQVVQAVQ AVHLESDAFL VCLNHALSTE KEEVMGLCIG ELNDDTRSDS
        60         70         80         90        100
KFAYTGTEMR TVAEKVDAVR IVHIHSVIIL RRSDKRKDRV EISPEQLSAA
       110        120        130        140        150
STEAERLAEL TGRPMRVVGW YHSHPHITVW PSHVDVRTQA MYQMMDQGFV
       160        170        180        190        200
GLIFSCFIED KNTKTGRVLY TCFQSIQAQK SSESLHGPRD FWSSSQHISI
       210        220        230        240        250
EGQKEEERYE RIEIPIHIVP HVTIGKVCLE SAVELPKILC QEEQDAYRRI
       260        270        280        290        300
HSLTHLDSVT KIHNGSVFTK NLCSQMSAVS GPLLQWLEDR LEQNQQHLQE
       310
LQQEKEELMQ ELSSLE
```

PDB – PROTEIN DATA BANK

PDB – Protein Data Bank

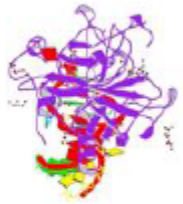
- Archives atomistic experimental models (X-ray, NMR, EM) plus supporting experimental data since 1971
- > 134,000 entries (Oct-2017)
- EMBL-EBI involved since 1999
- Managed by wwPDB partners since 2003
- Collaboration on all aspects of the PDB archive
  - Policies, procedures, formats, validation standards, ligands, journals, etc.
- Friendly competition on "data-out"
  - Serving PDB data with added-value; PDB-based services; other services, resources and activities

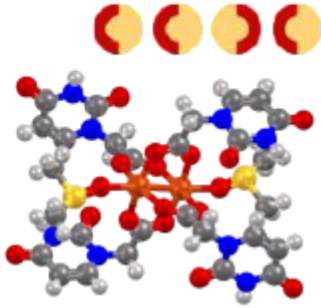


Other experimental 3D model archives

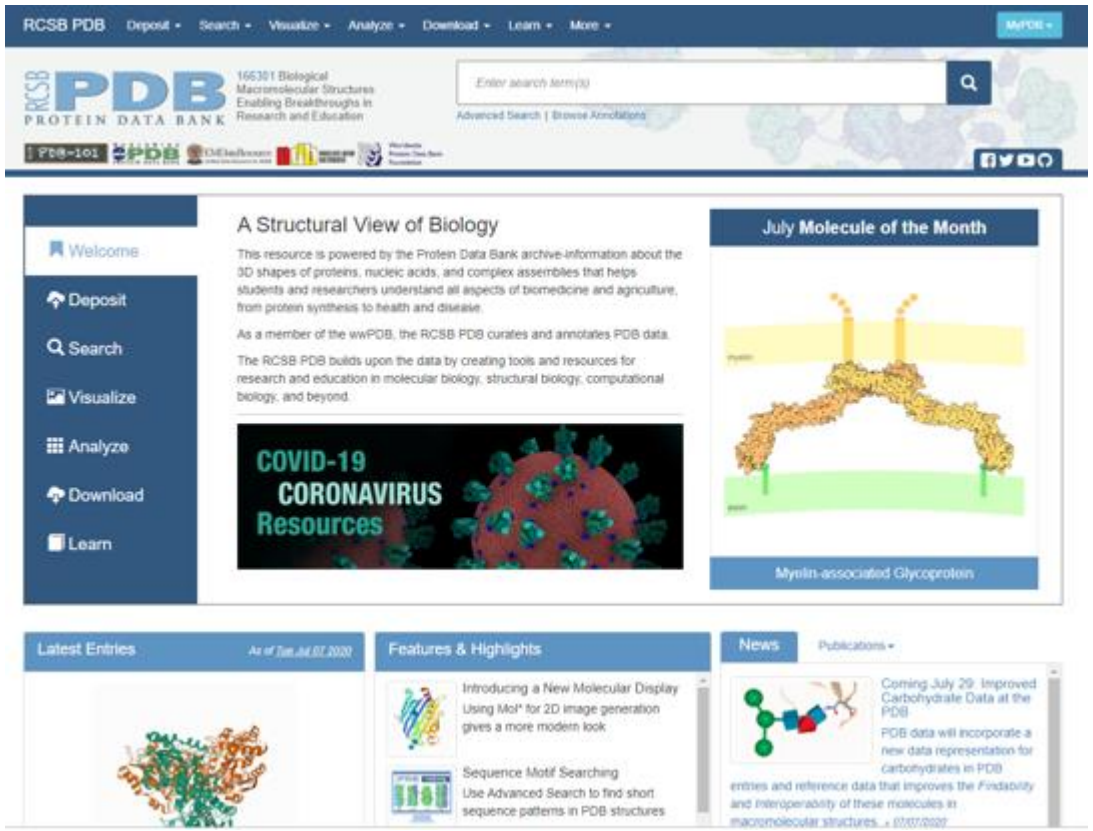
- NDB - Nucleic acid Database
  - Operated by RCSB (1995)
  - >9100 structures (Oct-17)
  - Most structures also in PDB
  - http://ndbserver.rutgers.edu/
- CSD – Cambridge Structural Database
  - Operated by CCDC (1965)
  - Crystal structures of "small molecules"
  - >875,000 structures (Oct-17)
  - http://www.ccdc.cam.ac.uk/

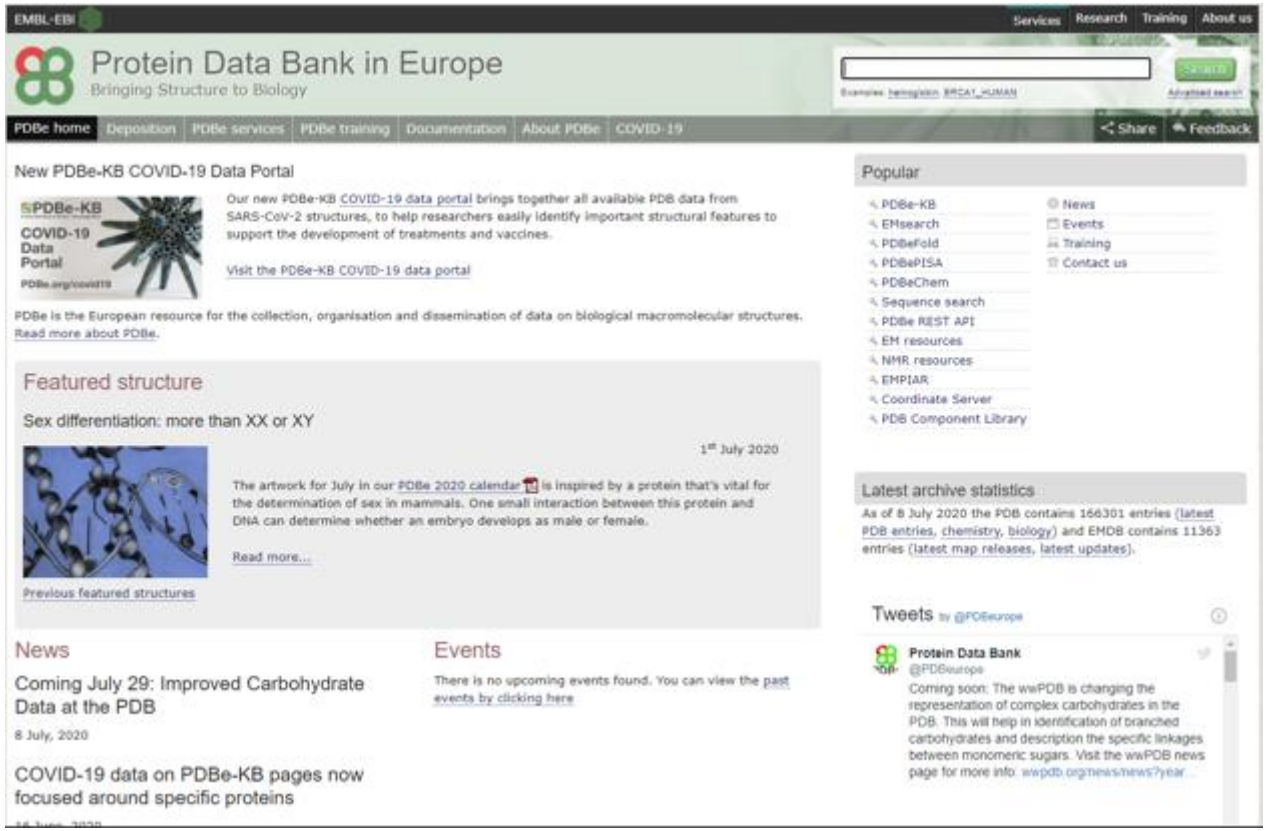
ndb NUCLEIC ACID DATABASE





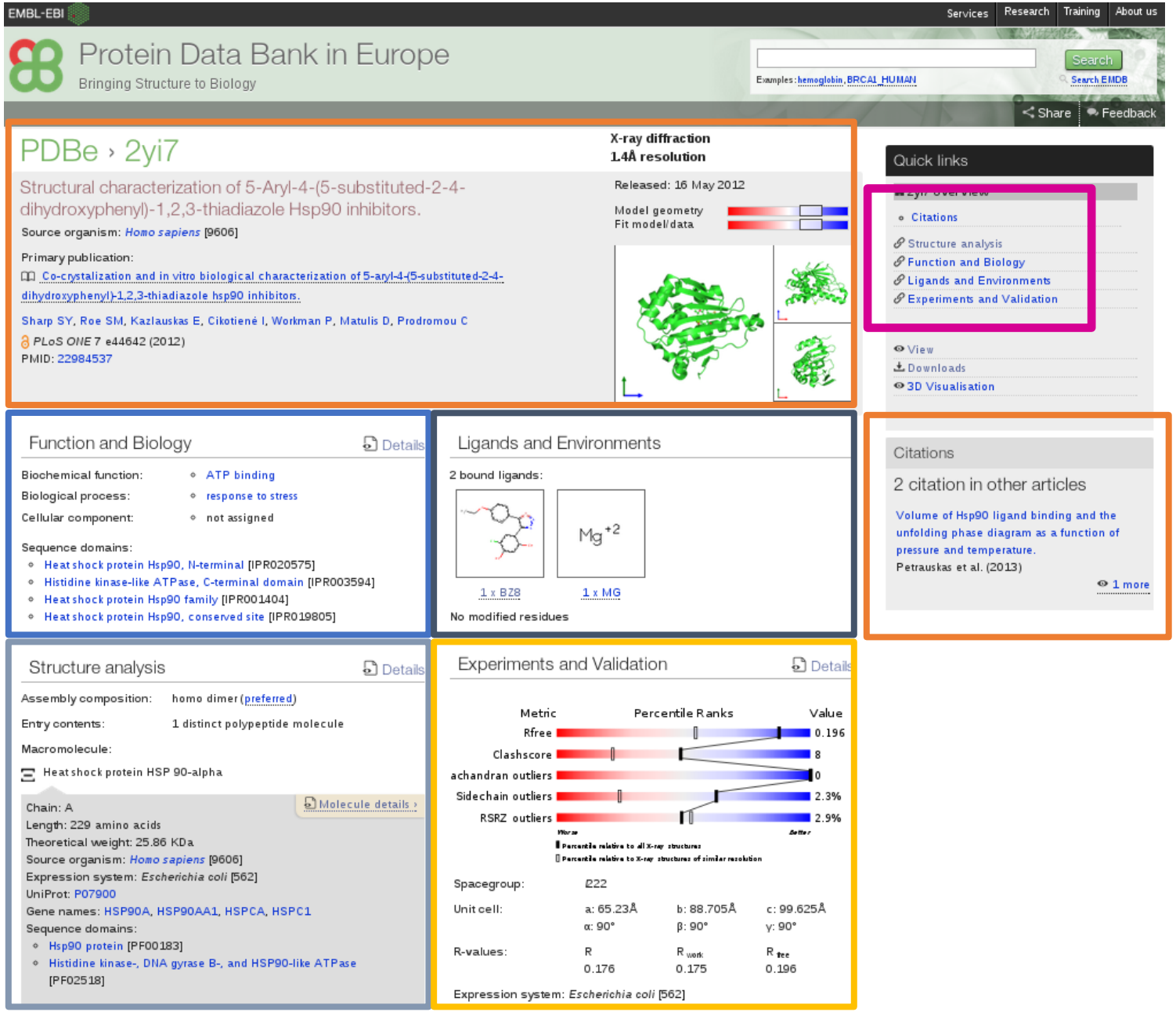
PDB – PROTEIN DATA BANK





https://www.rcsb.org/

https://www.ebi.ac.uk/pdbe/node/1



pdbe.org/2yi7

Some FAQs:

- Where is the structure published?
- Is it any good?
- How do I find out who has cited the structure?
- What is the assembly?
- What sequence / structural domains are present?
- Where is the ligand binding?

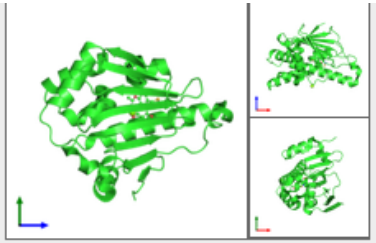
Primary publication:

Co-crystalization and in vitro biological characterization of 5-aryl-4-(5-substituted-2-4-dihydroxyphenyl)-1,2,3-thiadiazole hsp90 inhibitors.

Sharp SY, Roe SM, Kazlauskas E, Cikotienė I, Workman P, Matulis D, Prodromou C

PLoS ONE **7** e44642 (2012)

PMID: 22984537



- Function and Biology
- Ligands and Environments
- Experiments and Validation

- View
- Downloads
- 3D Visualisation

Function and Biology

Details

Biochemical function: ATP binding

Biological process: response to stress

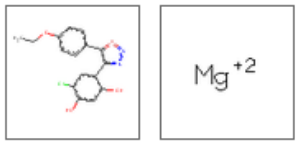
Cellular component: not assigned

Sequence domains:

- Heat shock protein Hsp90, N-terminal
- Histidine kinase-like ATPase, C-terminal domain
- Heat shock protein Hsp90 family
- Heat shock protein Hsp90, conserved site

Ligands and Environments

2 bound ligands:



1 x BZ8 1 x MG

No modified residues

Structure analysis

Details

Assembly composition: homo dimer (preferred)

Entry contents: 1 distinct polypeptide molecule

Macromolecule:

Heat shock protein HSP 90-alpha

Molecule details ›

Chain: A
Length: 229 amino acids
Theoretical weight: 25.86 KDa
Source organism: *Homo sapiens*
Expression system: *Escherichia coli*
UniProt:

- P07900 (Residues: 1-229; Coverage: 31%)

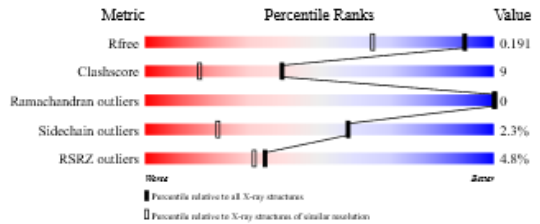
Gene names: HSP90A, HSP90AA1, HSPCA, HSPC1
Sequence domains:

- Hsp90 protein
- Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase

Structure domains: Heat Shock Protein 90

Experiments and Validation

Details



X-ray source: DIAMOND BEAMLINE I02

Spacegroup: *I*222

Unit cell: a: 65.23Å b: 88.705Å c: 99.625Å α: 90° β: 90° γ: 90°

R-values:

R	R work	R free
0.176	0.175	0.196

Expression system: *Escherichia coli*

Citations

2 citation in other articles

Volume of Hsp90 ligand binding and the unfolding phase diagram as a function of pressure and temperature.
Petrauskas et al. (2013)

1 more

PDB_REDO

The sliders below show the change in model quality between original PDB entry and the PDB_REDO entry

PDB_REDO

EMBL-EBI

Services | Research | Training | About us

Protein Data Bank in Europe
Bringing Structure to Biology

Search

Examples: hemoglobin, BRCA1_HUMAN

Search EMDB

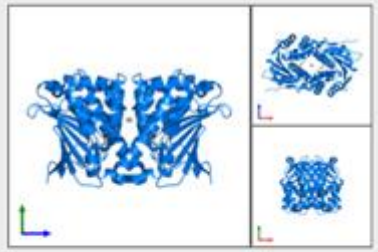
Share | Feedback

2yi7 › Heat shock protein HSP 90-alpha

Chain: A
Length: 229 amino acids
Theoretical weight: 25.86 KDa
Source organism: *Homo sapiens*
Expression system: *Escherichia coli*
UniProt:
- P07900 (Residues: 1-229; Coverage: 31%)

Gene names: HSP90A, HSP90AA1, HSPCA, HSPC1
FASTA Sequence

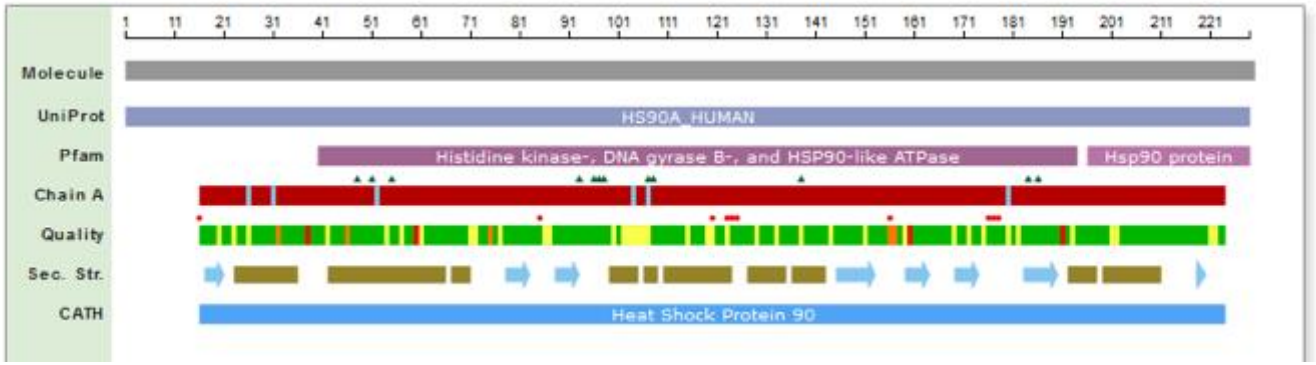
```
>pdb|2yi7|A
MPEETQTQDQPMEEEEVETFAFQAEIAQLMSLIINTFYSNKEIFLRELISNSSDALDKIRYESLTDPSKLDSGKELHINLIPNK
QDRTLTIVDTGIGMTKADLINNLGTIAKSGTKAFMEALQAGADISMIGQFGVGFYSAYLVAEKVTVITKHNDDEQYAWESSAGG
SFTVRTDTGEPMGRGTKVILHLKEDQTEYLEERRIKEIVKKHSQFIGYPITLFVEKERDKE
```



Visualisation

PDB Sequence Viewer | UniProt Coverage Viewer

Compact | Expanded



Quick links

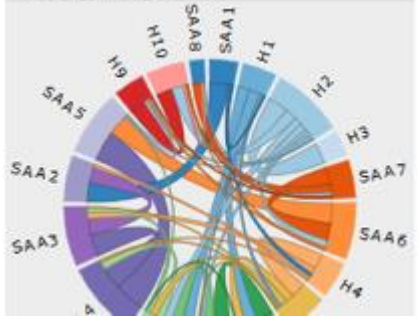
- 2yi7 overview
- Citations
- Structure analysis
- Function and Biology
- Ligands and Environments
- Experiments and Validation

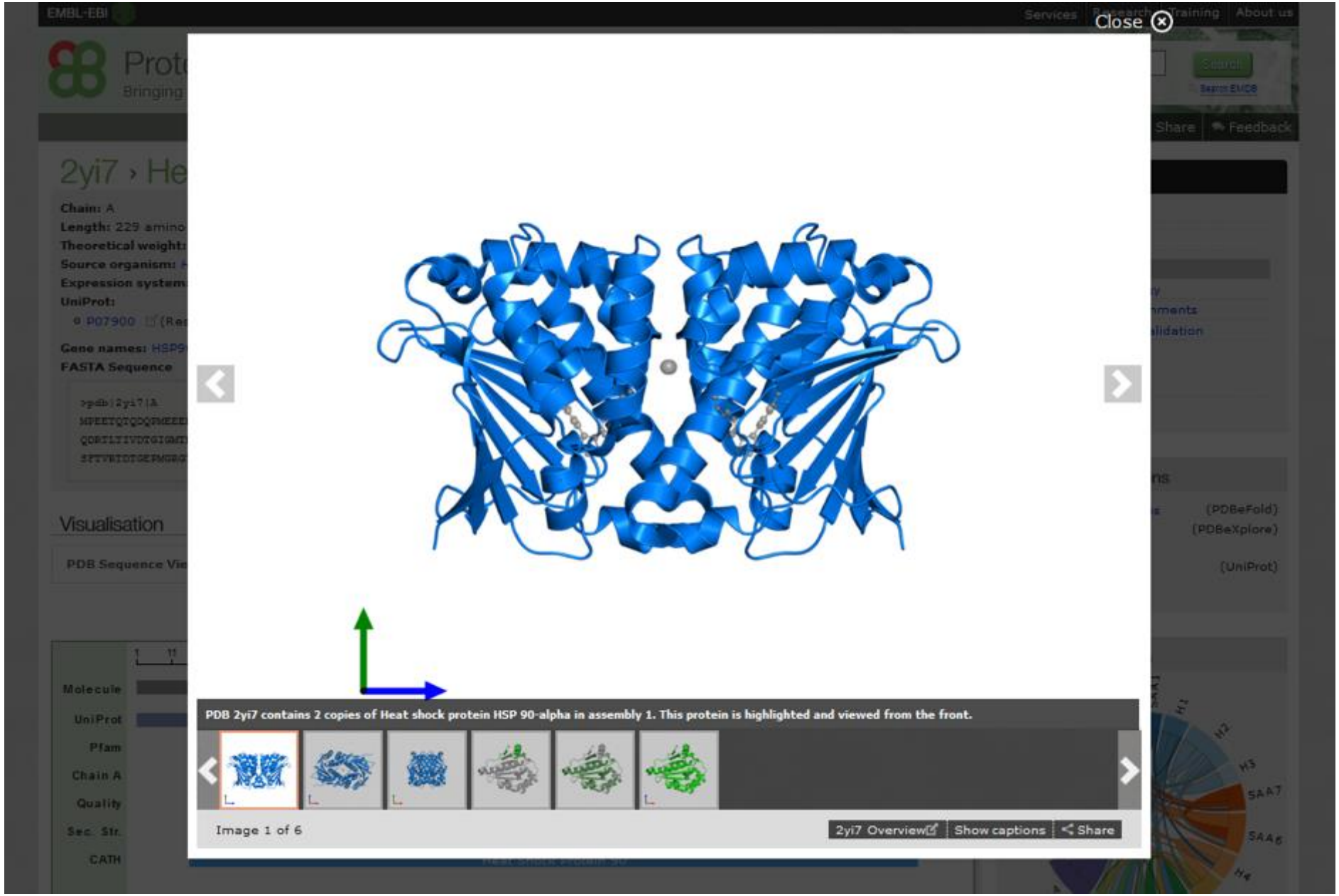
Downloads

Search similar proteins

- Similar 3D structures (PDBeFold)
- Similar sequences (PDBeXplore)
- BLAST P07900 (UniProt)

Residue interactions



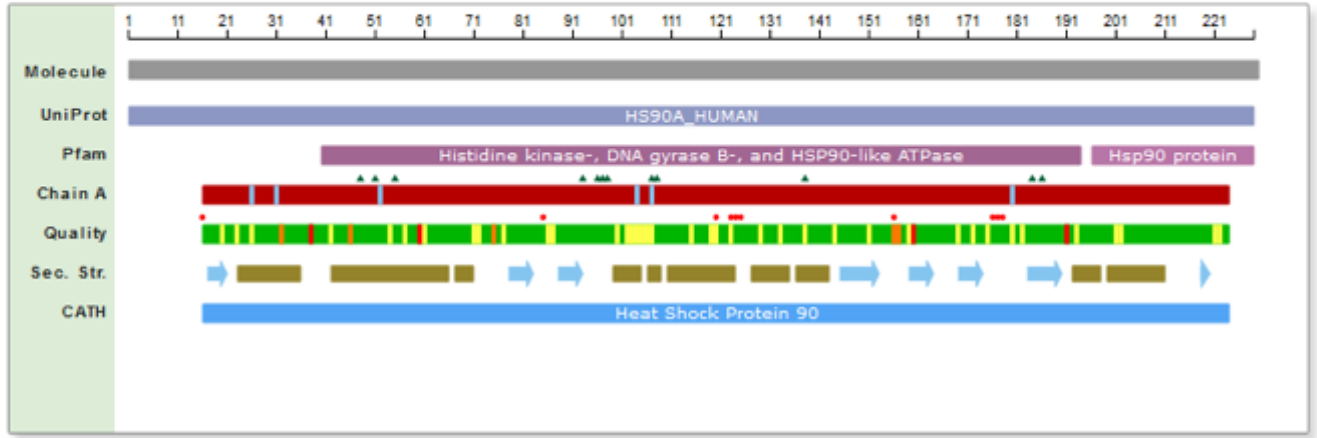

Close
PDB 2yi7 contains 2 copies of Heat shock protein HSP 90-alpha in assembly 1. This protein is highlighted and viewed from the front.
Image 1 of 6
2yi7 Overview
Show captions
Share

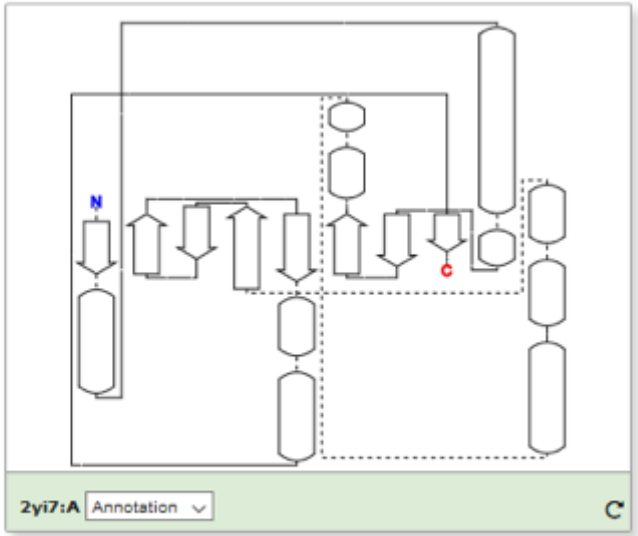

Close
1 copy of Pfam domain PF02518 (Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase) in Heat shock protein HSP 90-alpha in PDB 2yi7.
Image 5 of 6
2yi7 Overview
Show captions
Share
Visualisation
Molecule
UniProt
Pfam
Chain A
Quality
Sec. Str
CATH

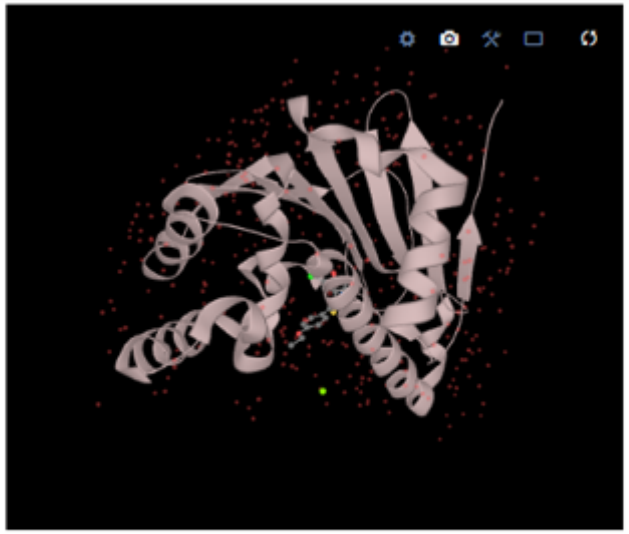
Visualisation

PDB Sequence Viewer | UniProt Coverage Viewer

Compact | Expanded



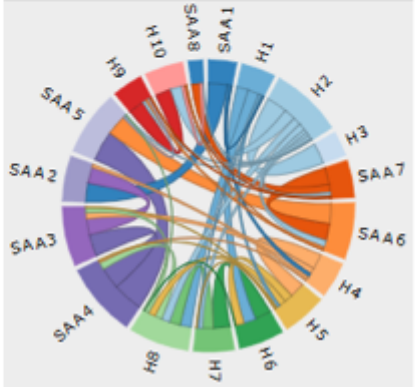




Similar 3D structures (PDBeFold)
Similar sequences (PDBeXplore)

BLAST P07900 (UniProt)

Residue interactions

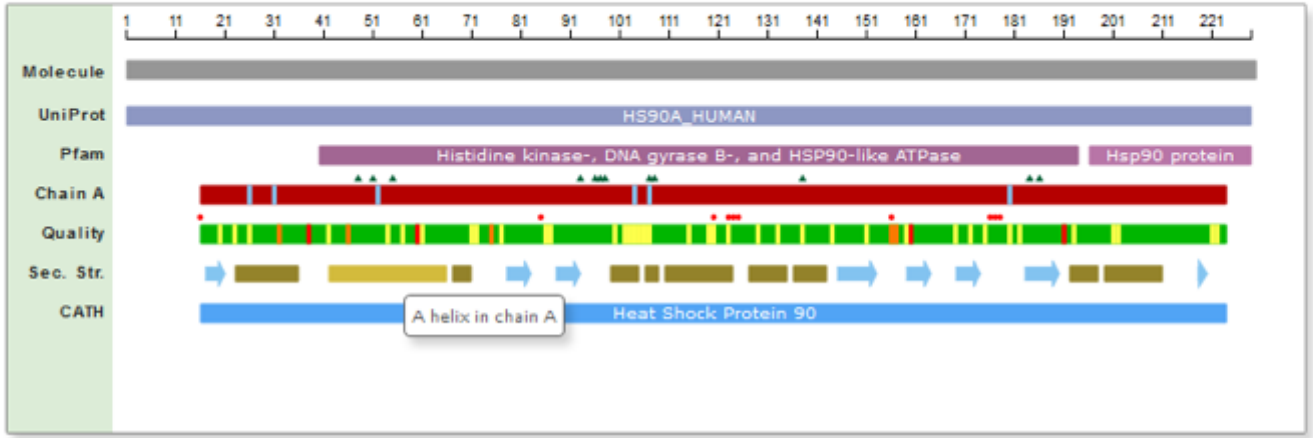


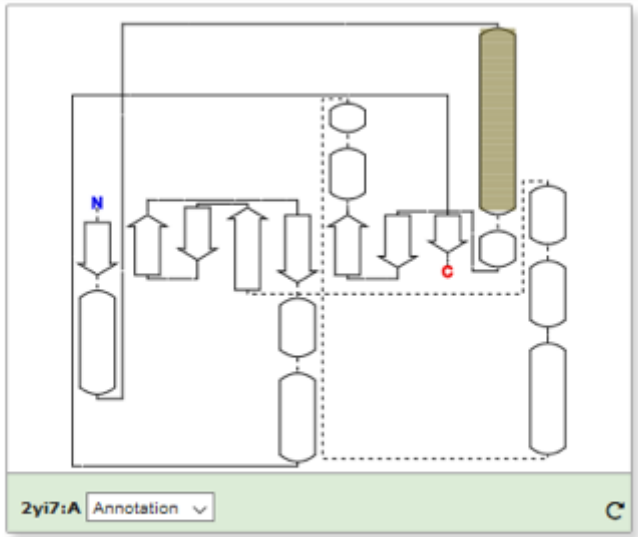
View residue interactions in Rajini

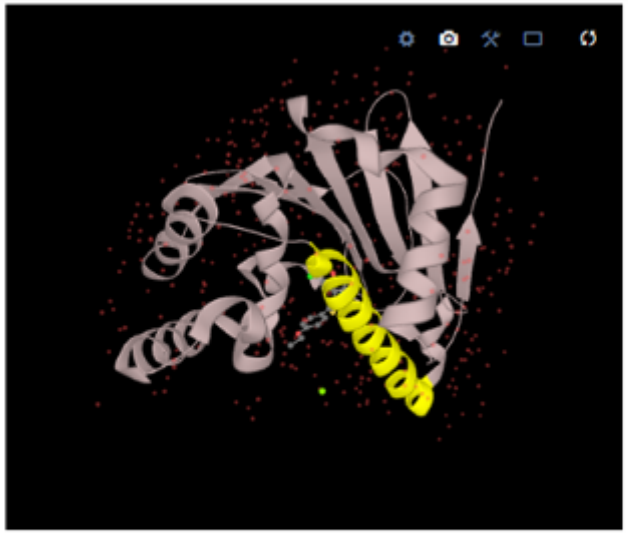
Visualisation

PDB Sequence Viewer | UniProt Coverage Viewer

Compact | Expanded



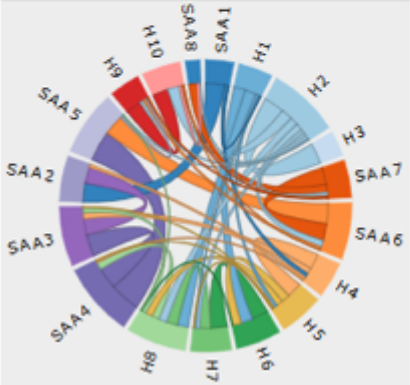




Similar 3D structures (PDBeFold)
Similar sequences (PDBeXplore)

BLAST P07900 (UniProt)

Residue interactions

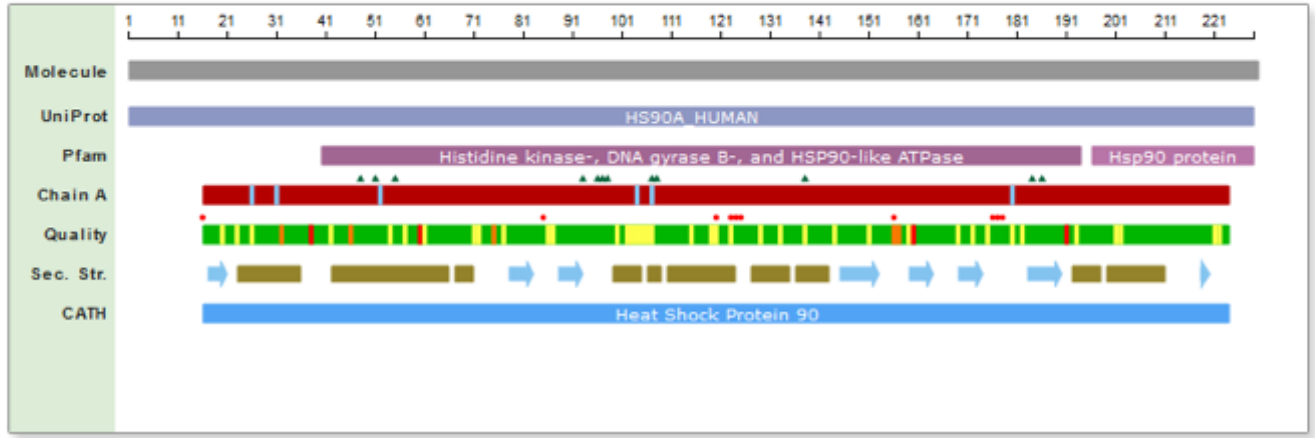


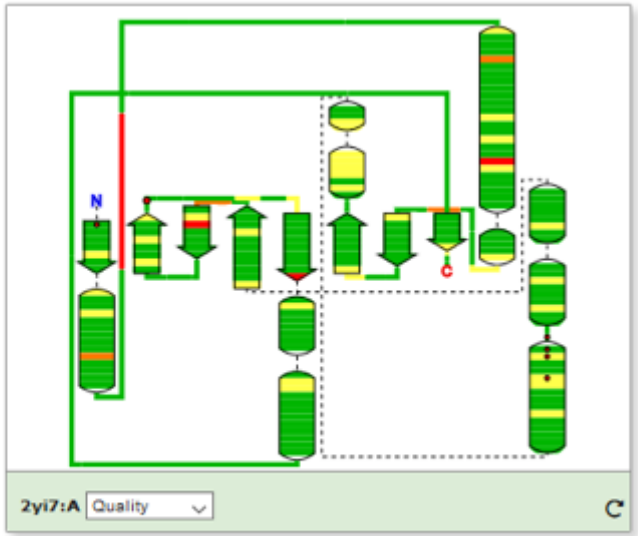
View residue interactions in Rajini

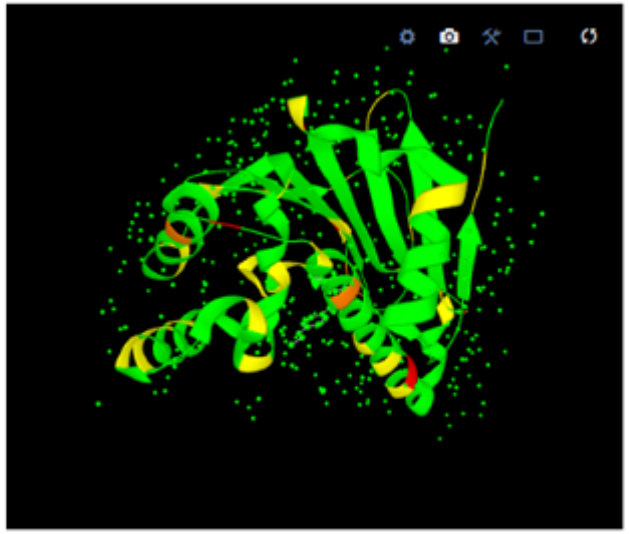
Visualisation

PDB Sequence Viewer | UniProt Coverage Viewer

Compact | Expanded



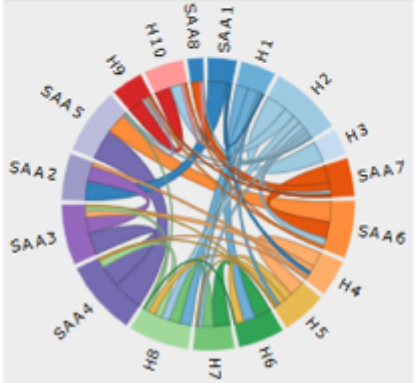




Similar 3D structures (PDBeFold)
Similar sequences (PDBeXplore)

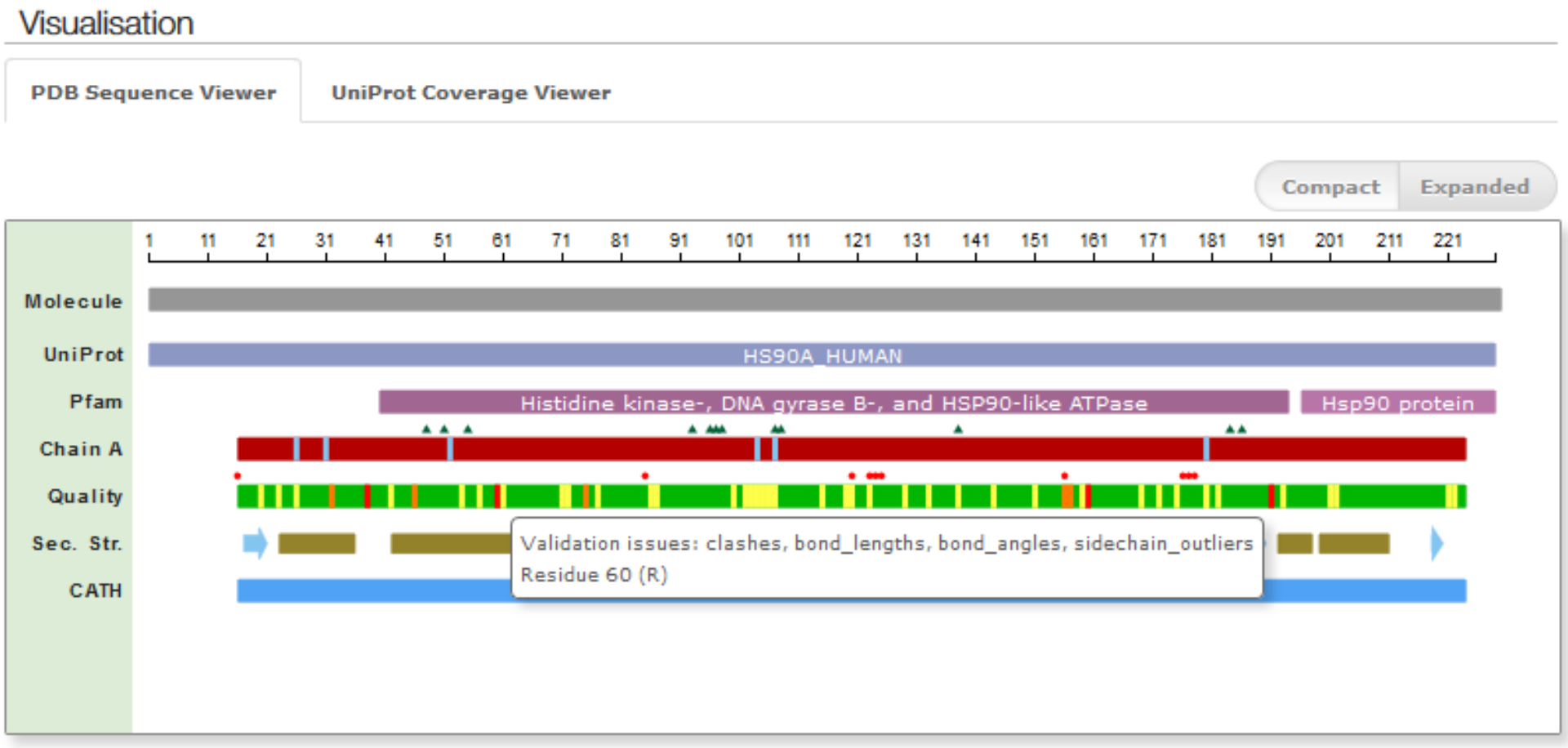
BLAST P07900 (UniProt)

Residue interactions

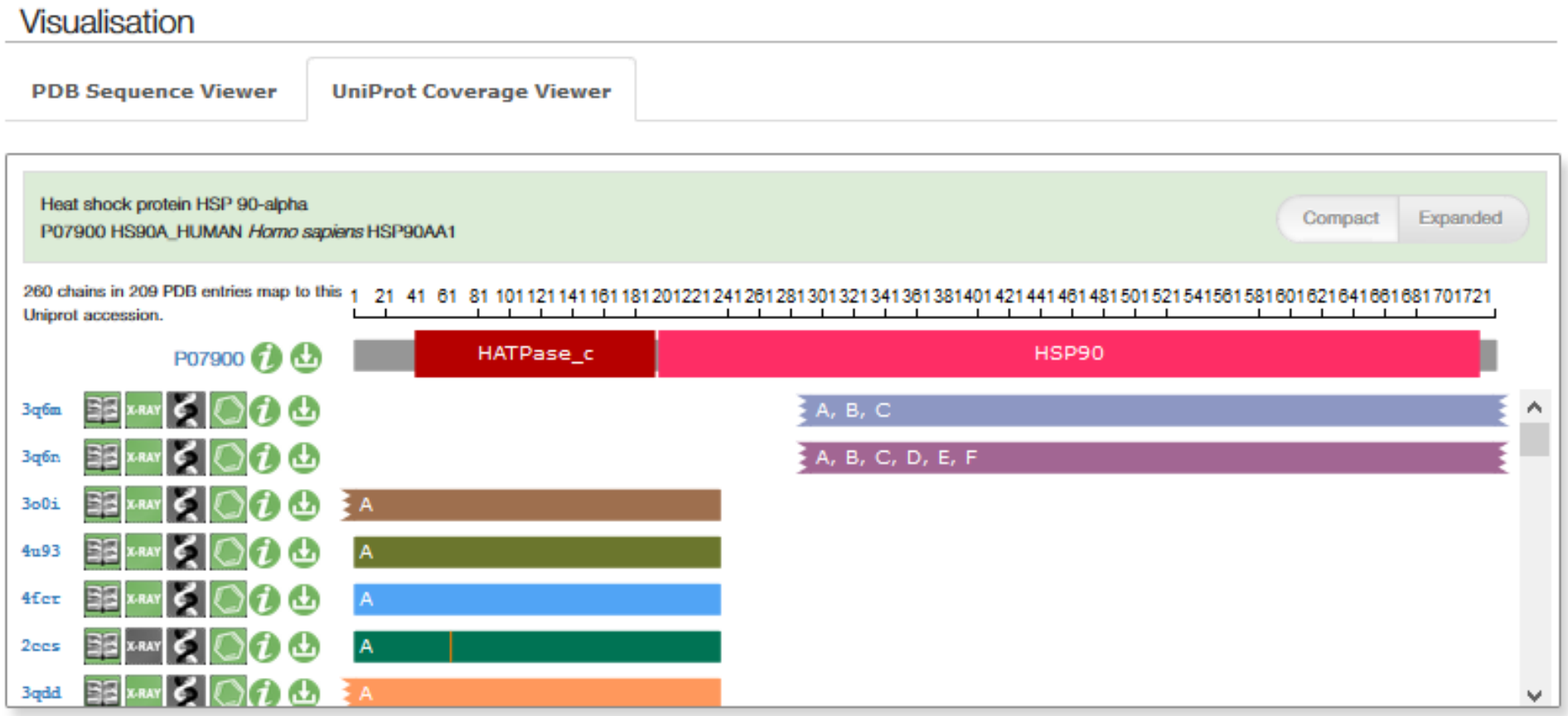


View residue interactions in Rajini

Validation issues detailed on sequence



UniProt coverage viewer



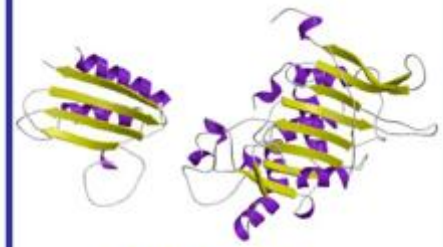
Protein Folds and Families, CATH, SCOP and Pfam

Proteins are generally composed of one or more functional regions, commonly termed **domains**. Different combinations of domains give rise to the diverse range of proteins found in nature. The identification of domains that occur within proteins can therefore provide insights into their function.

CATH — Domain structures

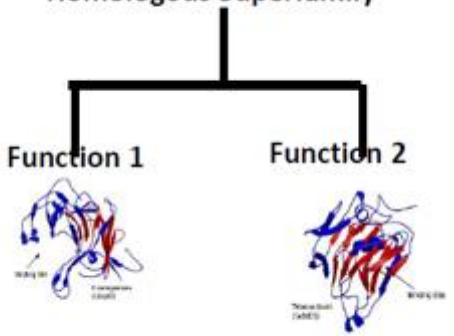
Gene3D — Domain structure predictions

Domain Architecture Classification



SCOP/CATH - Classification and evolution of domain structure superfamilies

SUPERFAMILY/Gene3D – assigning sequences to SCOP/CATH
Pfam – domain sequence families

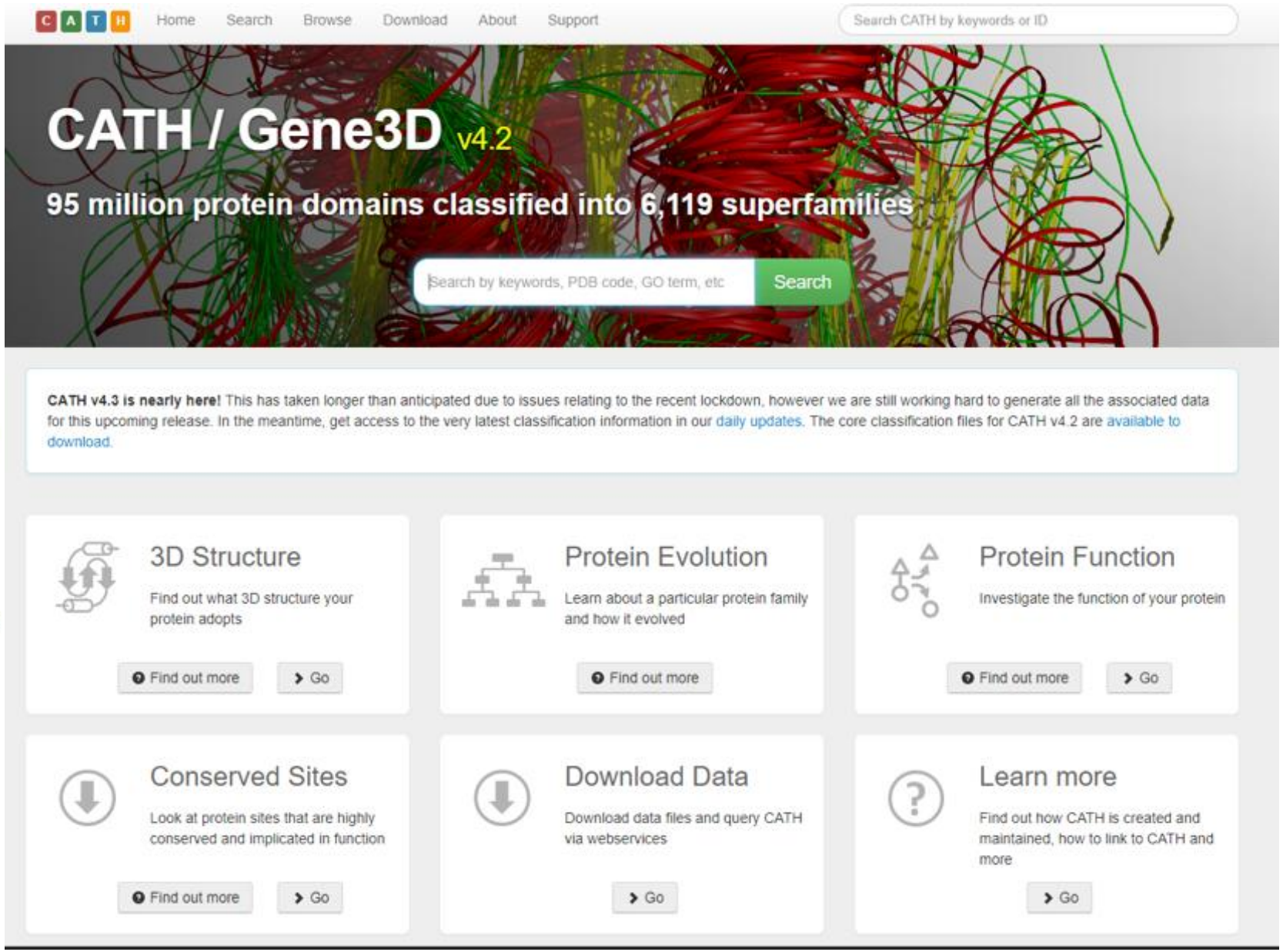


Functional subfamilies within superfamilies

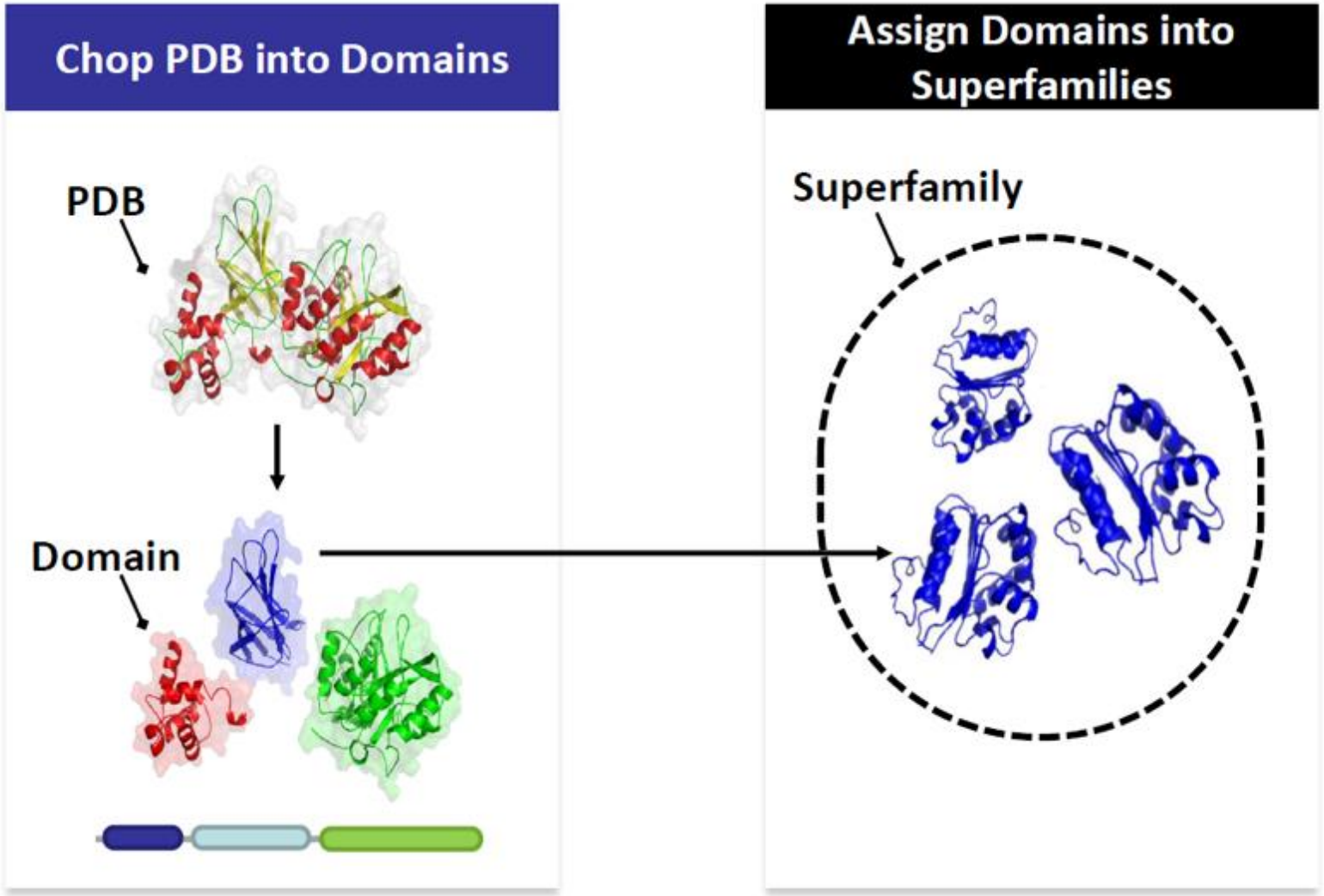
cath

~450,000 domain structures

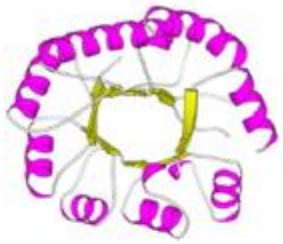
~6000 superfamilies



CATH: Protein Structure Classification

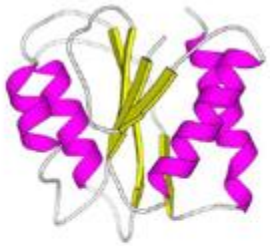


Cath architectures

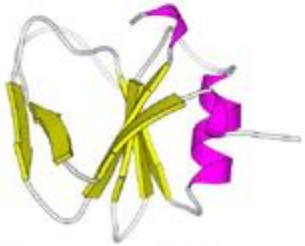


αβ-barrel

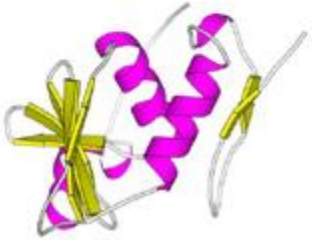
2-layer (αβ) sandwich



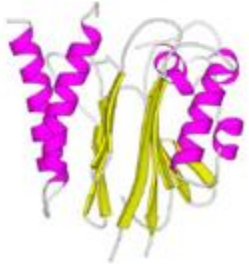
3-layer (αβα) sandwich



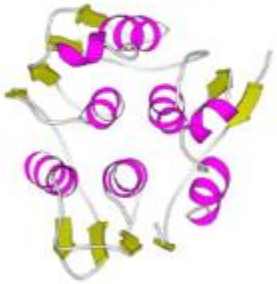
3-layer (ββα) sandwich



3-layer (βαβ) sandwich

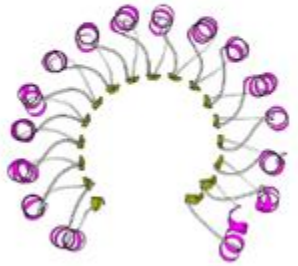


4-layer (αββα) sandwich



αβ-prism

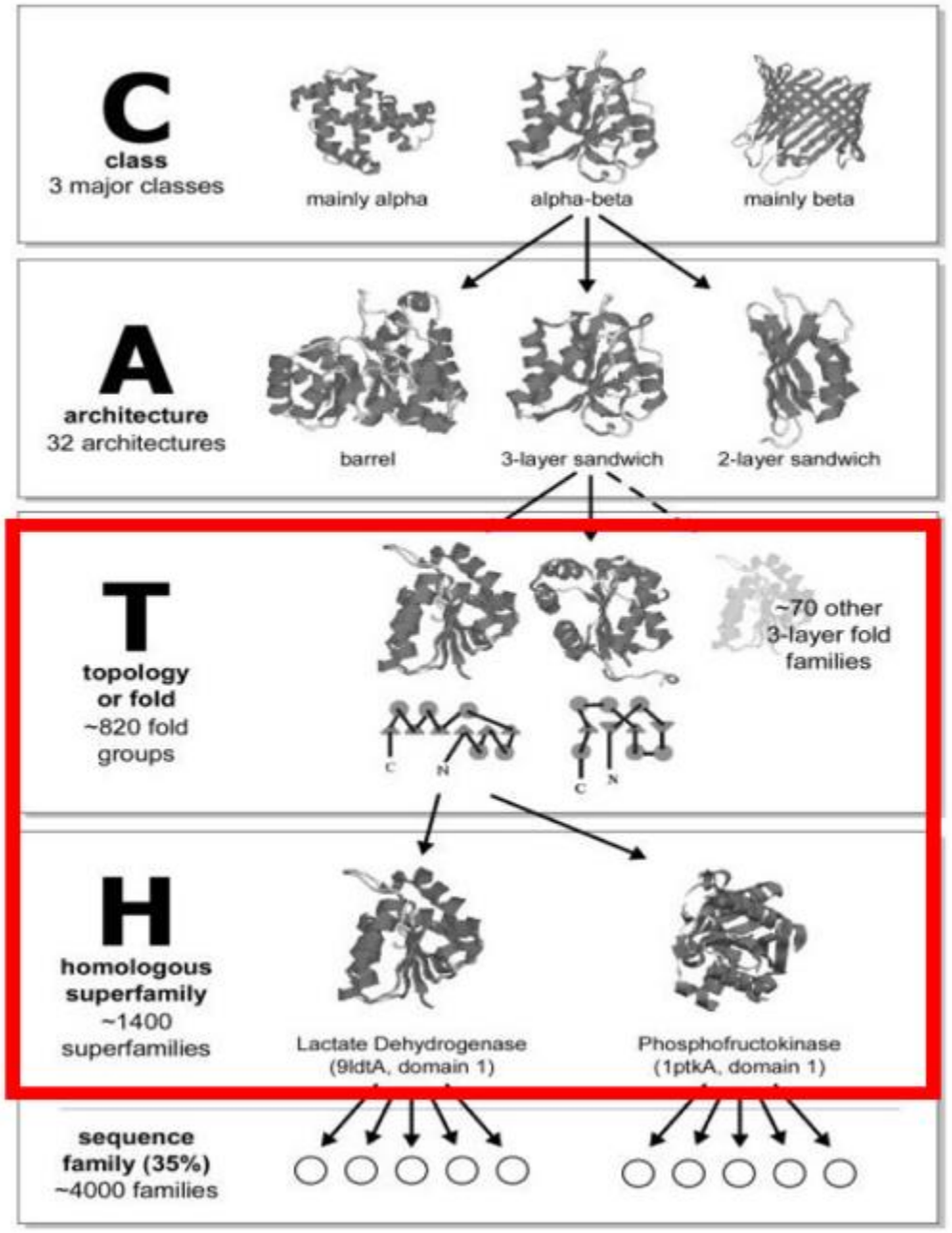
αβ-box



αβ-horseshoe

Domain Structure Classification

Structure based methods
for recognising fold similarities

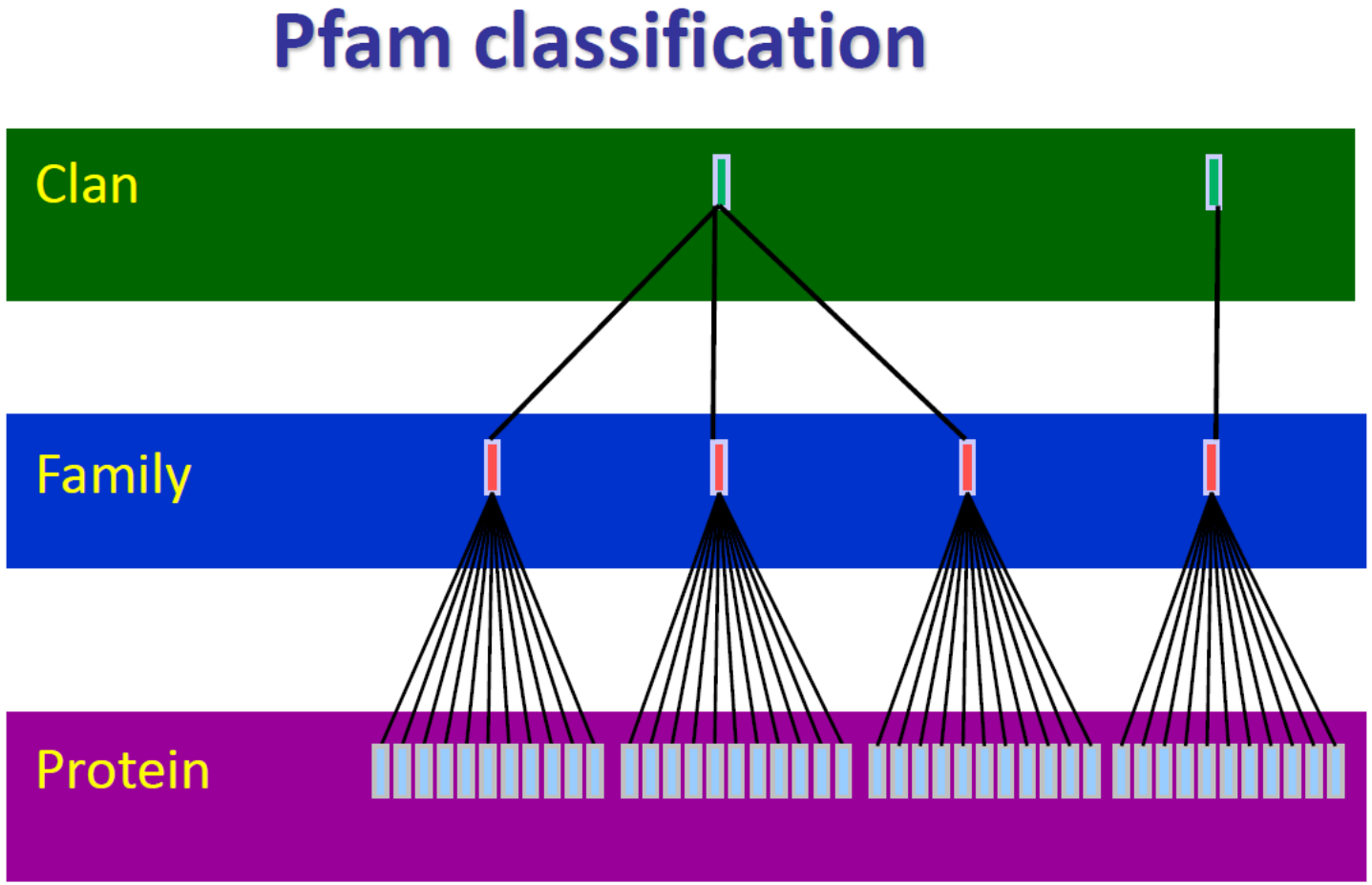


Pfam database

The Pfam database is a large collection of protein families, each represented by **multiple sequence alignments** and **hidden Markov models (HMMs)**.

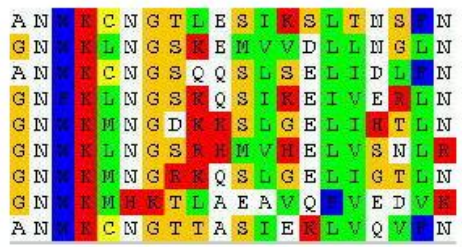
Pfam also generates higher-level groupings of related entries, known as **clans**. A clan is a collection of Pfam entries which are related by similarity of sequence, structure or profile-HMM.

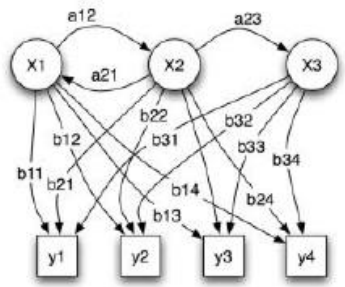
The data presented for each entry is based on the UniProt Reference Proteomes but information on individual UniProtKB sequences can still be found by entering the protein accession. Pfam *full* alignments are available from searching a variety of databases, either to provide different accessions (e.g. all UniProt and NCBI GI) or different levels of redundancy.



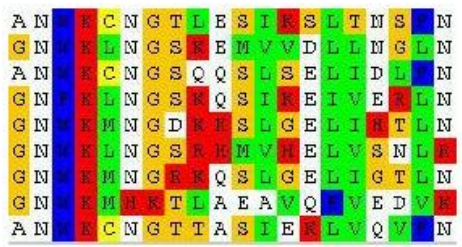
Sequence based methods for detecting protein homologues

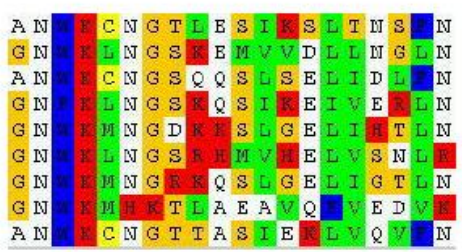
sequence-sequence comparison
BLAST, NW, SSEARCH





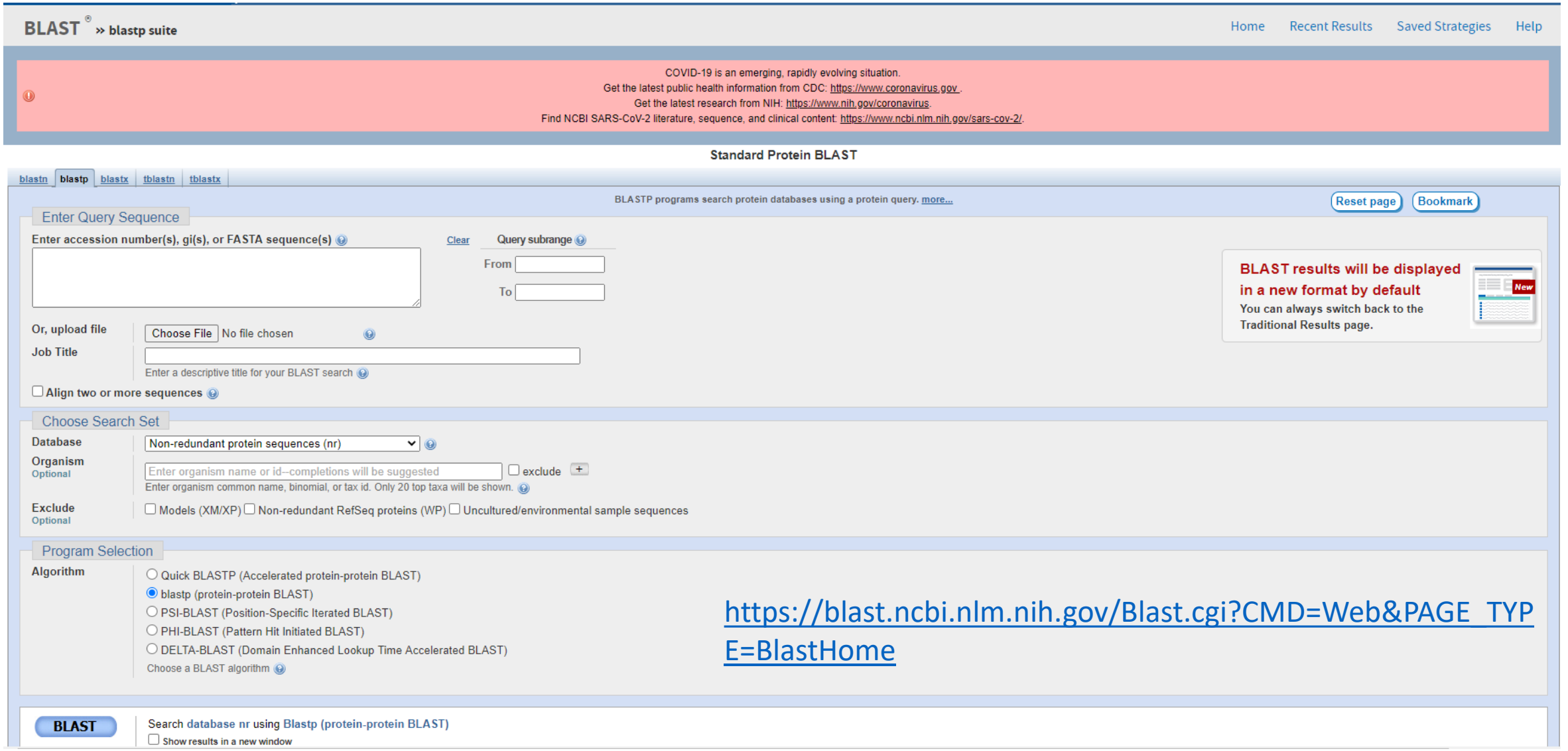
sequence-profile comparison
PSI-BLAST, HMMer, SAM-T





profile-profile comparison,
HMM-HMM comparison
PRC, COMPASS, HHpred

Blast



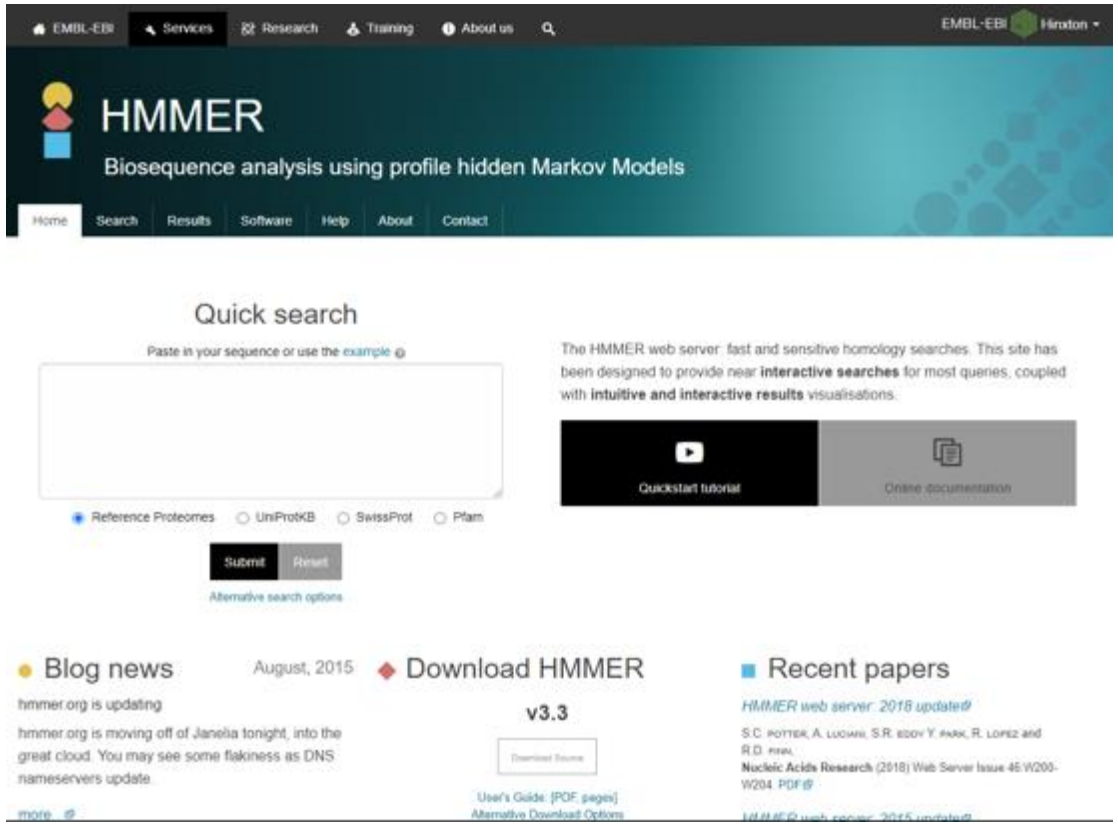
https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE_TYPE=BlastHome

HMMER: biosequence analysis using profile hidden Markov models

HMMER is used for searching sequence databases for sequence homologs, and for making sequence alignments. It implements methods using probabilistic models called profile hidden Markov models (profile HMMs).

HMMER is often used together with a profile database, such as Pfam. But HMMER can also work with query *sequences*, not just profiles, just like BLAST. For example, you can search a protein query sequence against a database with **phmmer**, or do an iterative search with **jackhmmer**.

HMMER is designed to detect remote homologs as sensitively as possible, relying on the strength of its underlying probability models. In the past, this strength came at significant computational expense, but as of the new HMMER3 project, HMMER is now essentially as fast as BLAST.



VALIDATION OF STRUCTURAL SIMILARITY

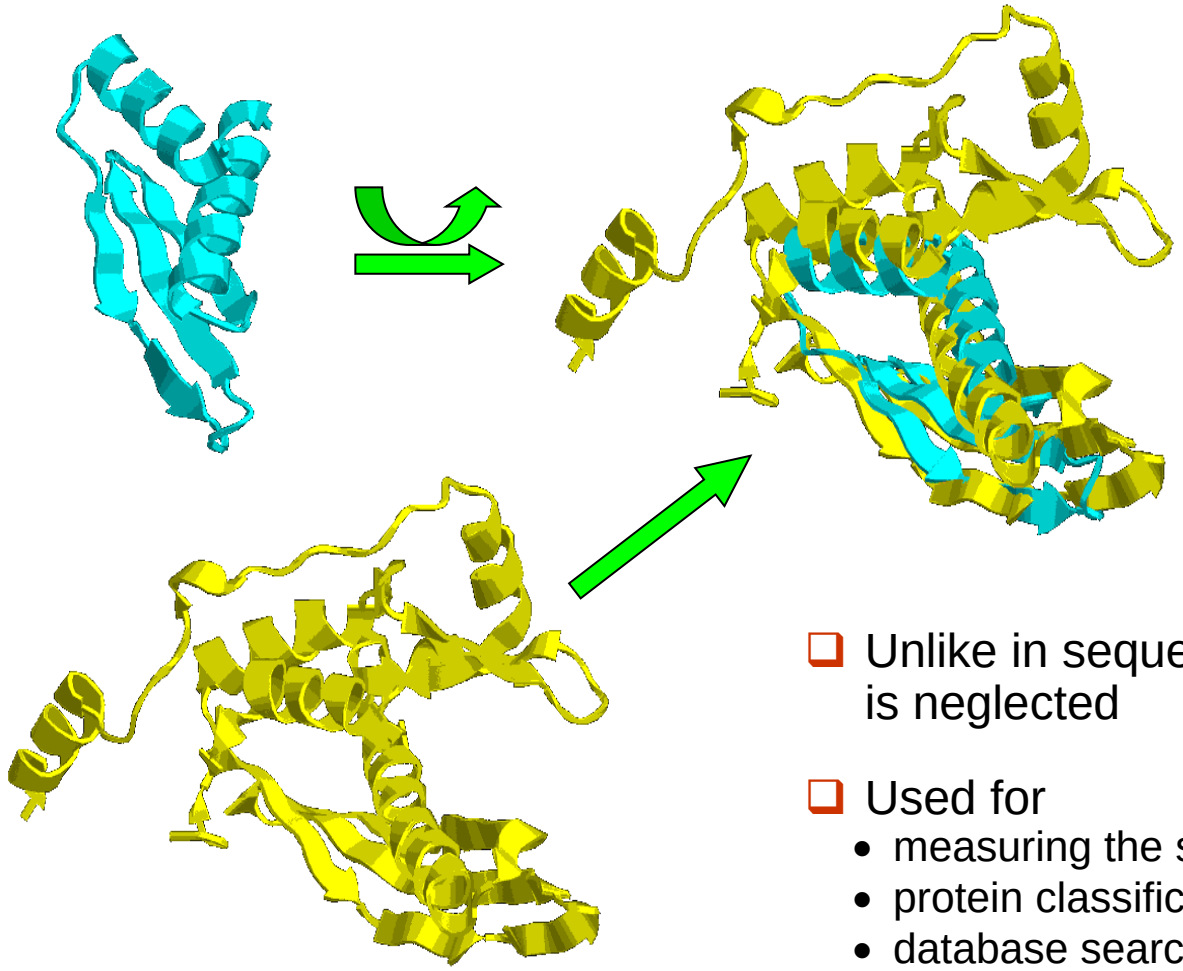
Based on CATH (Protein Structure Classification database) to Validate homologues at least 2 out of 3 of the following conditions should be met:

Significant structural similarity

Significant sequence similarity

Functional similarity

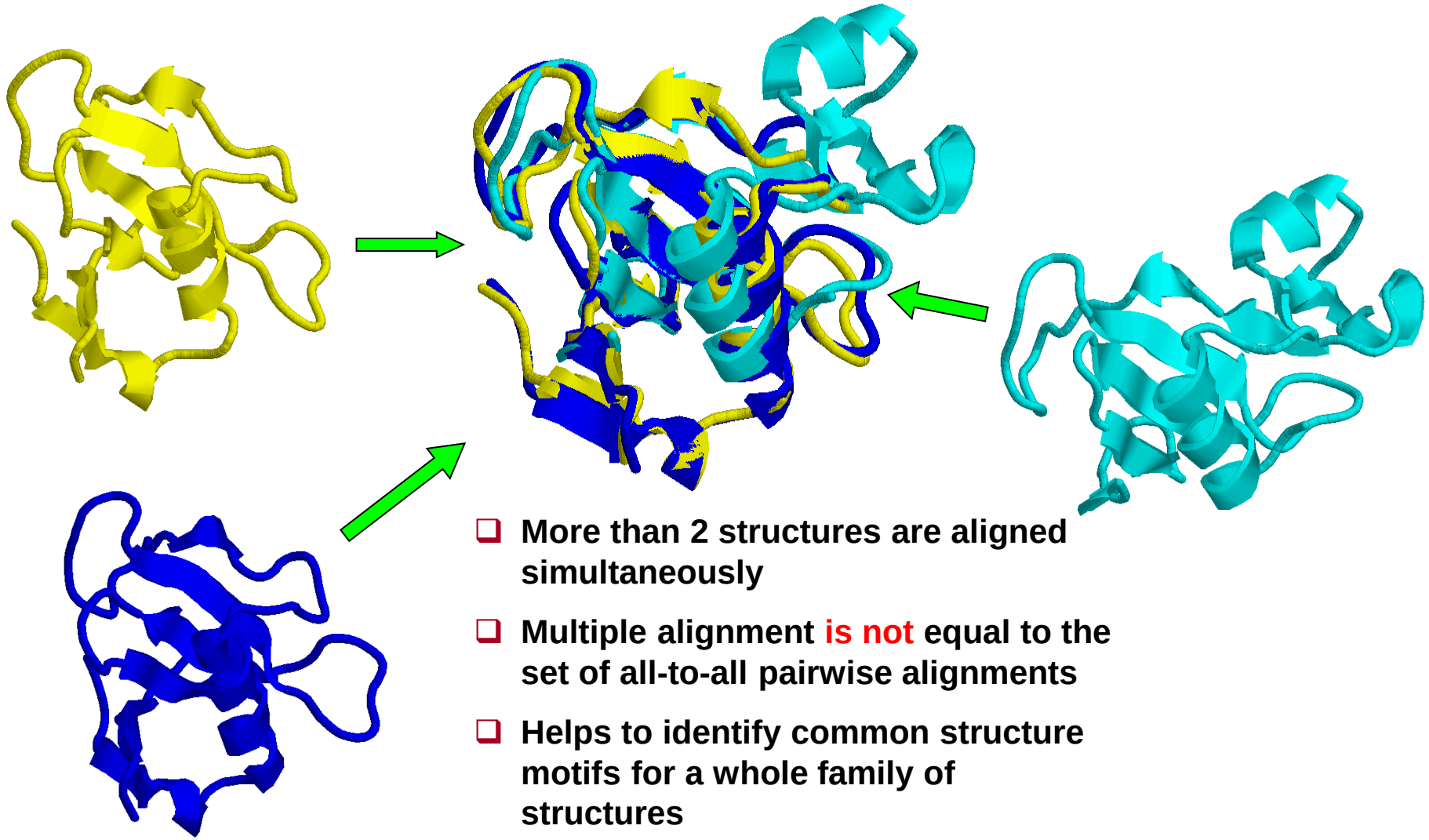
Structure alignment



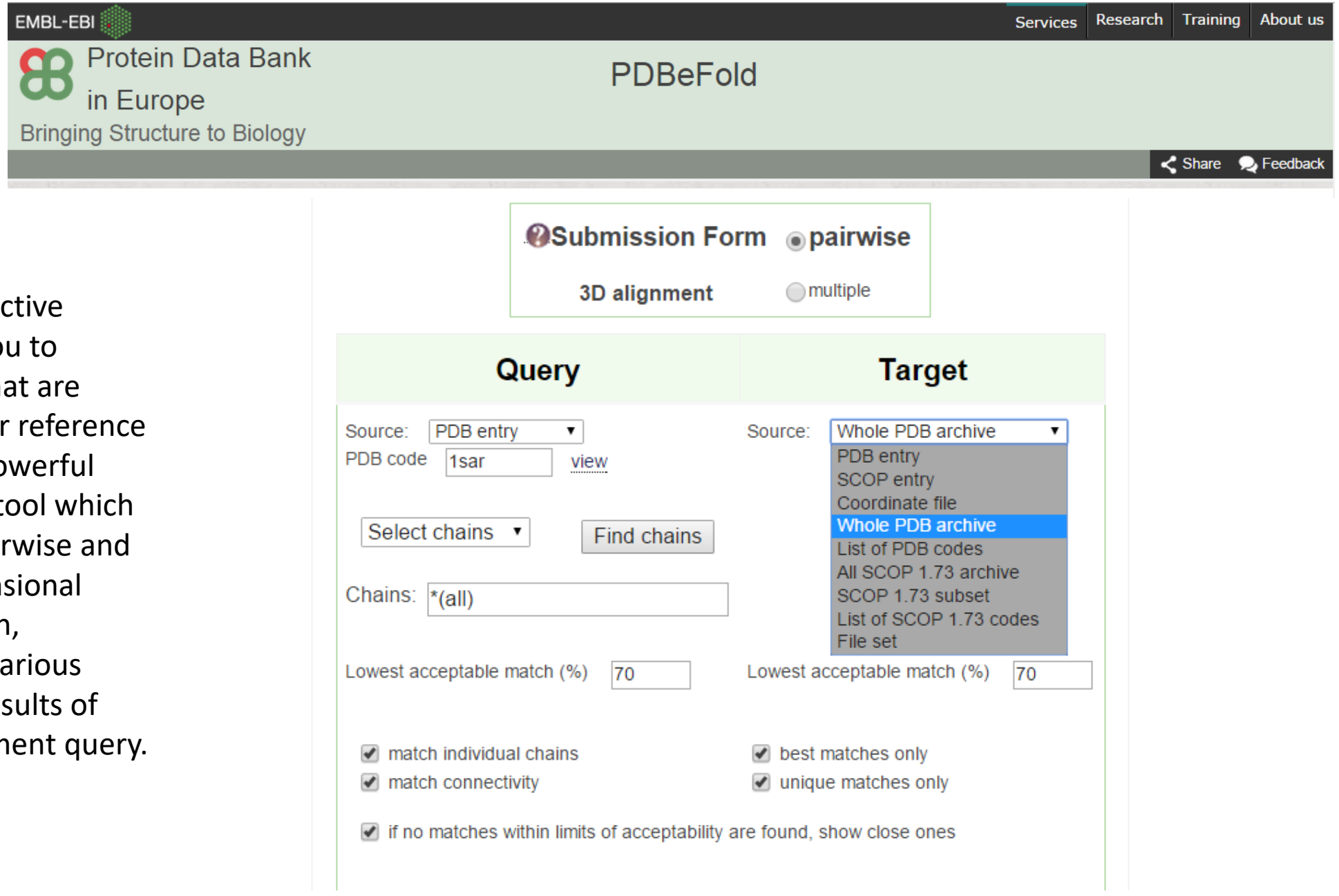
Structure alignment may be defined as identification of residues occupying "equivalent" geometrical positions

- Unlike in sequence alignment, residue type is neglected
- Used for
  - measuring the structural similarity
  - protein classification and functional analysis
  - database searches

Multiple structure alignment



- More than 2 structures are aligned simultaneously
- Multiple alignment is not equal to the set of all-to-all pairwise alignments
- Helps to identify common structure motifs for a whole family of structures



PDBeFold is an interactive service that allows you to **identify structures** that are **similar** to that of your reference protein. It is a very powerful **structure alignment** tool which can perform both pairwise and multiple three-dimensional alignment. In addition, PDBeFold gives you various options to sort the results of your structural alignment query.

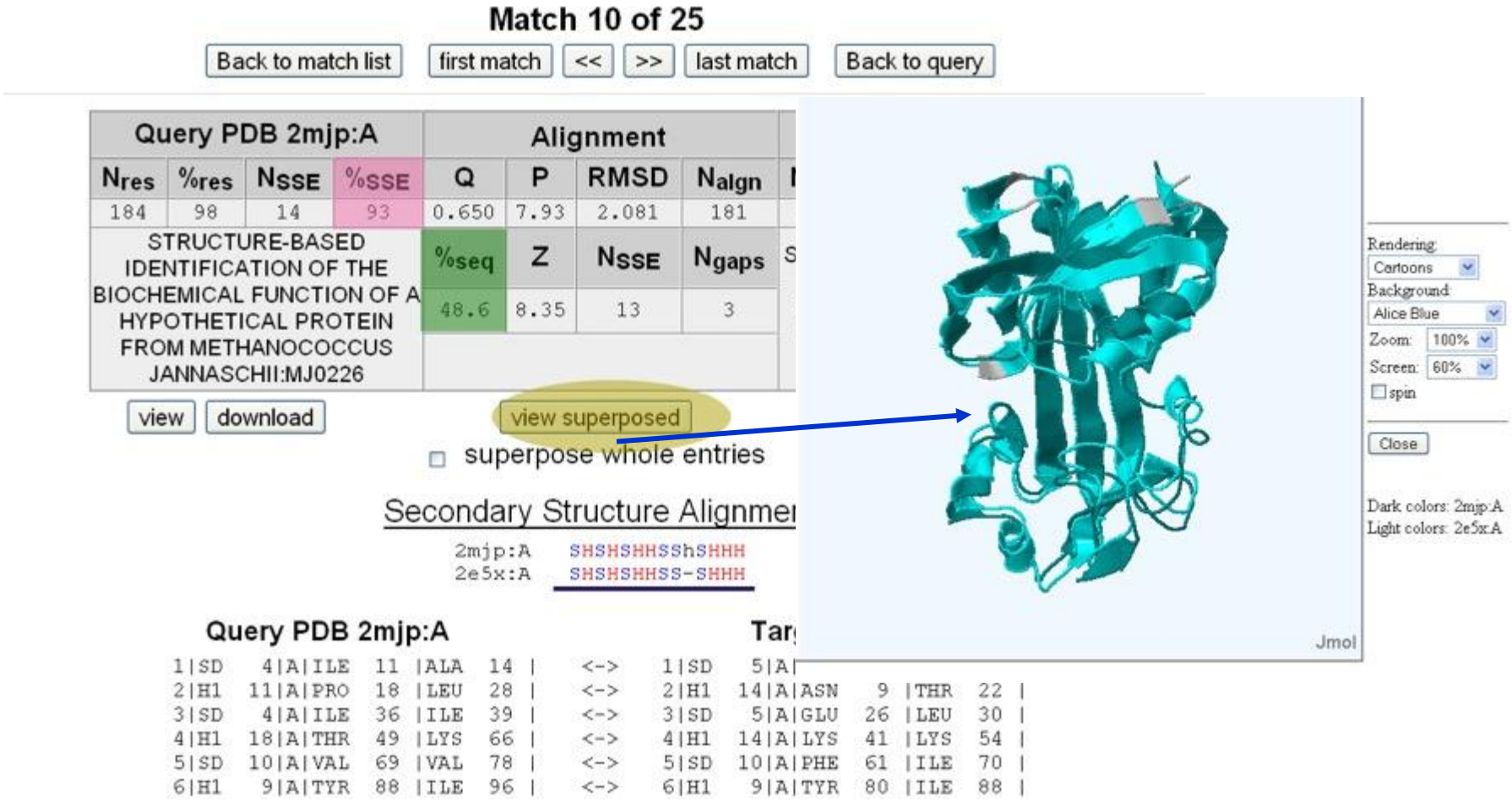
PDBeFold output

- Table of matched Secondary Structure Elements
- Table of matched backbone C_α-atoms with distances between them at best structure superposition
- Rotation-translation matrix of best structure superposition
- Visualisation in Jmol and Rasmol
- *r.m.s.d.* of C_α-alignment
- Length of C_α-alignment N_{align}
- Number of gaps in C_α-alignment
- Quality score Q
- Statistical significance scores $P(S)$, Z
- Sequence identity

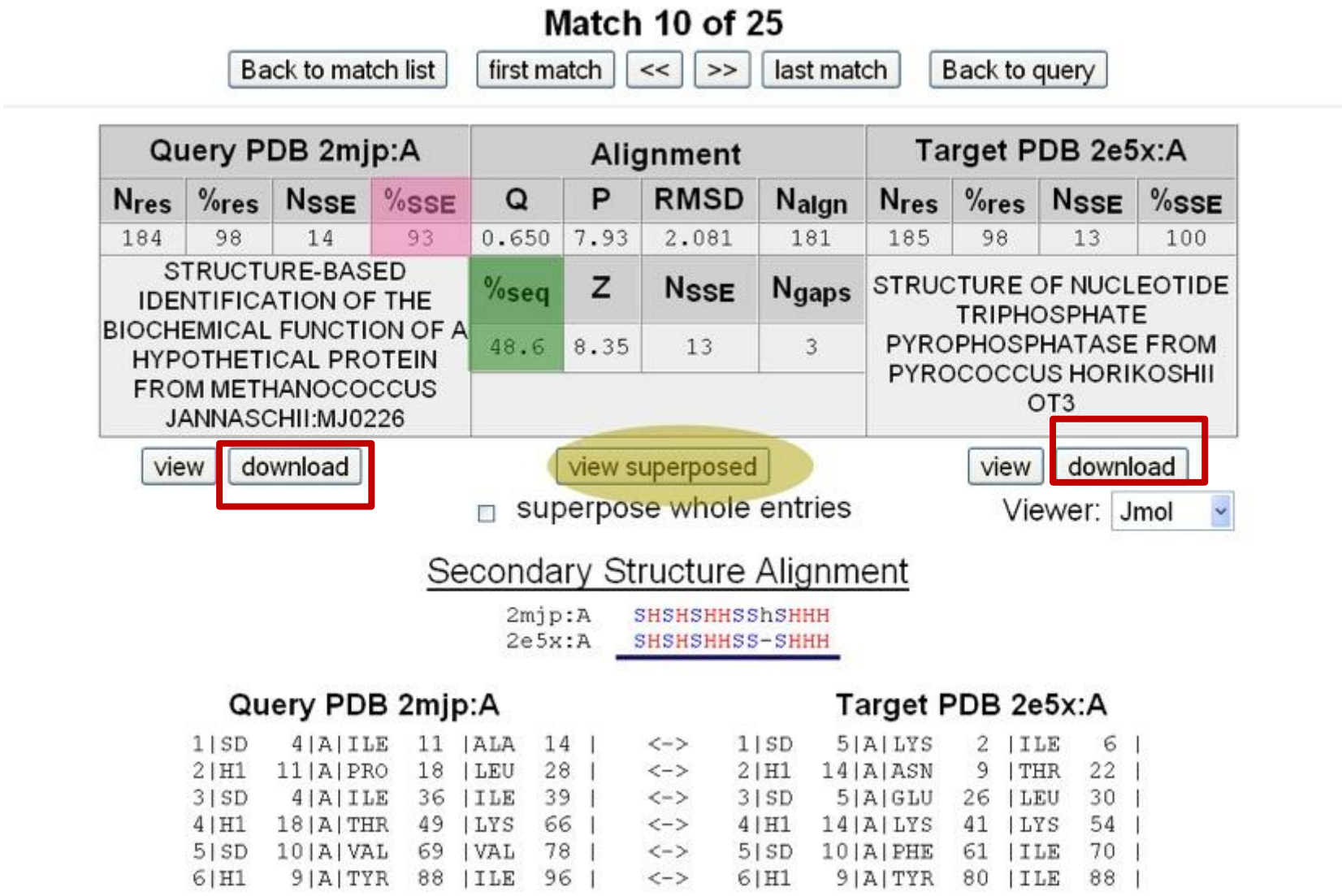
The Results Page For Pairwise Alignment

##	Q	P	Z	Rmsd	N_{algn}	N_g	%seq	%sse	Match	%sse	N_{res}	×	Title
1	1.00	26.8	15.5	0.00	184	0	100	100	2mjp:A	100	184	☐	STRUCTURE-BASED IDENTIFICATION OF THE BIOCHEMICAL FUNCTION OF A HYPOTHETICAL PROTEIN FROM METHANOCOCCUS JANNASCHII:MJ0226
2	1.00	25.4	15.0	0.20	184	0	100	100	1b78:A	100	184	☐	STRUCTURE-BASED IDENTIFICATION OF THE BIOCHEMICAL FUNCTION OF A HYPOTHETICAL PROTEIN FROM METHANOCOCCUS JANNASCHII:MJ0226
3	0.95	22.3	14.1	0.49	182	1	99	93	1b78:B	93	184	☐	STRUCTURE-BASED IDENTIFICATION OF THE BIOCHEMICAL FUNCTION OF A HYPOTHETICAL PROTEIN FROM METHANOCOCCUS JANNASCHII:MJ0226
4	0.95	20.0	13.3	0.55	182	1	99	86	2mjp:B	100	184	☐	STRUCTURE-BASED IDENTIFICATION OF THE BIOCHEMICAL FUNCTION OF A HYPOTHETICAL PROTEIN FROM METHANOCOCCUS JANNASCHII:MJ0226
5	0.85	15.3	11.6	1.14	182	2	50	93	2dvn:B	100	186	☐	STRUCTURE OF PH1917 PROTEIN WITH THE COMPLEX OF IMP FROM PYROCOCCUS HORIKOSHII
6	0.68	9.1	8.9	1.92	181	3	49	93	1v7r:A	100	186	☐	STRUCTURE OF NUCLEOTIDE TRIPHOSPHATE PYROPHOSPHATASE FROM PYROCOCCUS HORIKOSHII OT3
7	0.67	8.5	8.6	1.96	181	3	49	93	2dvo:A	100	185	☐	STRUCTURE OF PH1917 PROTEIN WITH THE COMPLEX OF ITP FROM PYROCOCCUS HORIKOSHII
8	0.67	9.0	8.9	1.95	181	3	49	93	2dvn:A	100	186	☐	STRUCTURE OF PH1917 PROTEIN WITH THE COMPLEX OF IMP FROM PYROCOCCUS HORIKOSHII
9	0.66	6.7	7.7	1.93	177	5	50	79	2dvp:A	92	184	☐	STRUCTURE OF NTPASE FROM PYROCCOUS HORIKOSHII
10	0.65	7.9	8.3	2.08	181	3	49	93	2e5x:A	100	185	☐	STRUCTURE OF NUCLEOTIDE TRIPHOSPHATE PYROPHOSPHATASE FROM PYROCOCCUS HORIKOSHII OT3
11	0.65	8.7	8.8	1.42	168	7	35	71	2car:B	83	194	☐	CRYSTAL STRUCTURE OF HUMAN INOSINE TRIPHOSPHATASE
12	0.64	8.6	8.8	1.48	170	7	35	71	2car:A	83	196	☐	CRYSTAL STRUCTURE OF HUMAN INOSINE TRIPHOSPHATASE
13	0.64	8.8	8.8	1.41	168	7	35	71	2i5d:A	83	195	☐	CRYSTAL STRUCTURE OF HUMAN INOSINE TRIPHOSPHATE PYROPHOSPHATASE
14	0.63	5.4	7.0	1.81	173	6	35	71	1vp2:A	91	189	☐	CRYSTAL STRUCTURE OF PUTATIVE XANTHOSINE TRIPHOSPHATE PYROPHOSPHATASE/HAM1 PROTEIN HOMOLOG (TM0159) FROM THERMOTOGA MARITIMA AT 1.78 A RESOLUTION

Analyzing the result from a particular pairwise alignment



Analyzing the result from a particular pairwise alignment



Match 10 of 25

Back to match list | first match | << | >> | last match | Back to query

Query PDB 2mjp:A				Alignment				Target PDB 2e5x:A			
Nres	%res	Nsse	%sse	Q	P	RMSD	Nalgn	Nres	%res	Nsse	%sse
184	98	14	93	0.650	7.93	2.081	181	185	98	13	100
STRUCTURE-BASED IDENTIFICATION OF THE BIOCHEMICAL FUNCTION OF A HYPOTHETICAL PROTEIN FROM METHANOCOCCUS JANNASCHII:MJ0226				%seq	Z	Nsse	Ngaps	STRUCTURE OF NUCLEOTIDE TRIPHOSPHATE PYROPHOSPHATASE FROM PYROCOCCUS HORIKOSHII OT3			
				48.6	8.35	13	3				

view | download | view superposed | view | download

☐ superpose whole entries

Viewer: Jmol

Secondary Structure Alignment

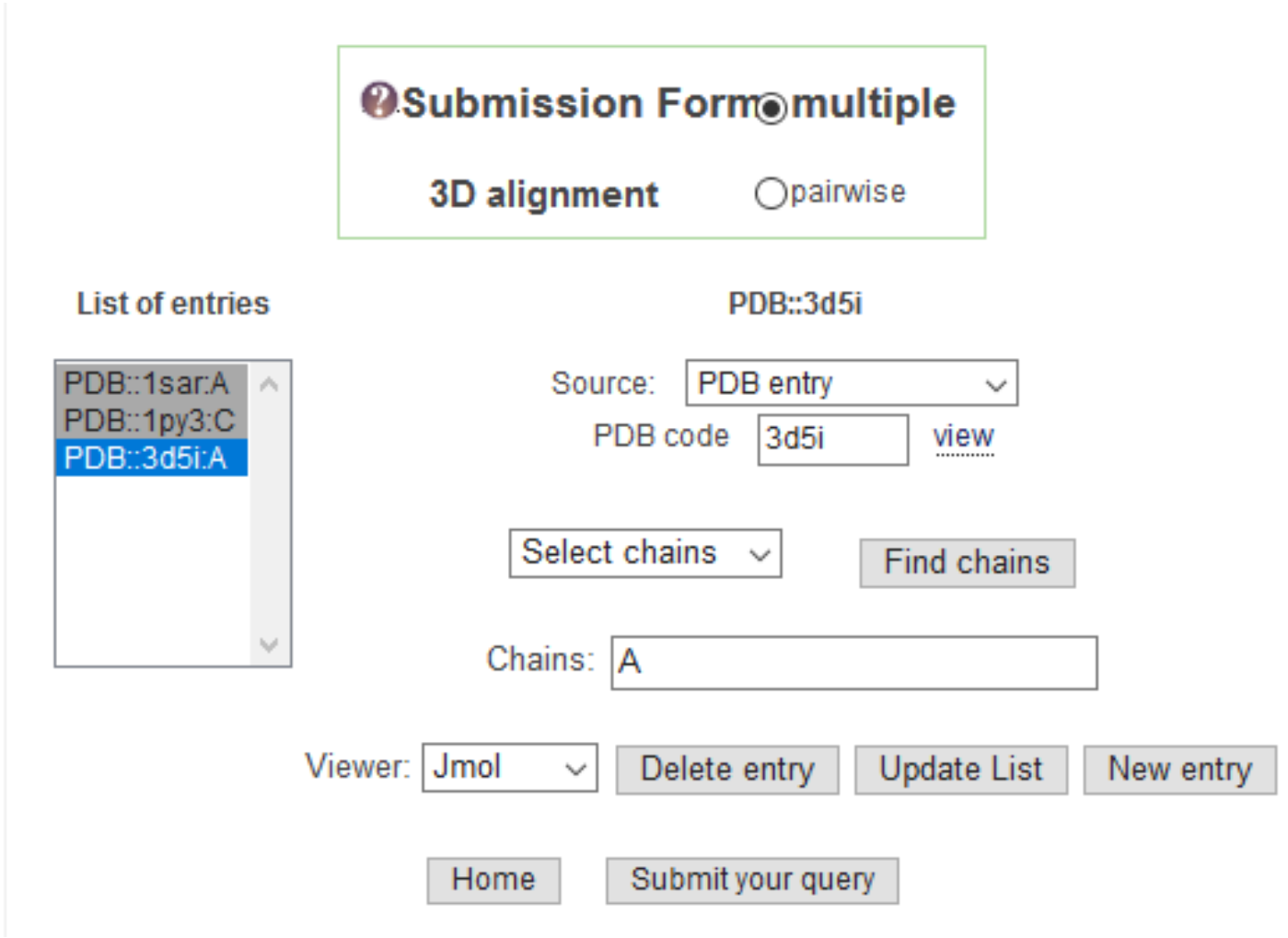
```
2mjp:A   SHSHSHHSShSHHH
2e5x:A   SHSHSHHSS-SHHH
```

Query PDB 2mjp:A							Target PDB 2e5x:A						
1	SD	4	A	ILE	11	ALA 14	<->	1	SD	5	A	LYS 2	ILE 6
2	H1	11	A	PRO	18	LEU 28	<->	2	H1	14	A	ASN 9	THR 22
3	SD	4	A	ILE	36	ILE 39	<->	3	SD	5	A	GLU 26	LEU 30
4	H1	18	A	THR	49	LYS 66	<->	4	H1	14	A	LYS 41	LYS 54
5	SD	10	A	VAL	69	VAL 78	<->	5	SD	10	A	PHE 61	ILE 70
6	H1	9	A	TYR	88	ILE 96	<->	6	H1	9	A	TYR 80	ILE 88

Residue-by-Residue Structure alignment result

- Structure alignment includes both secondary structure assignment and residue identity:
- For aligned residues:
  - **Exact match**
  - **No match**

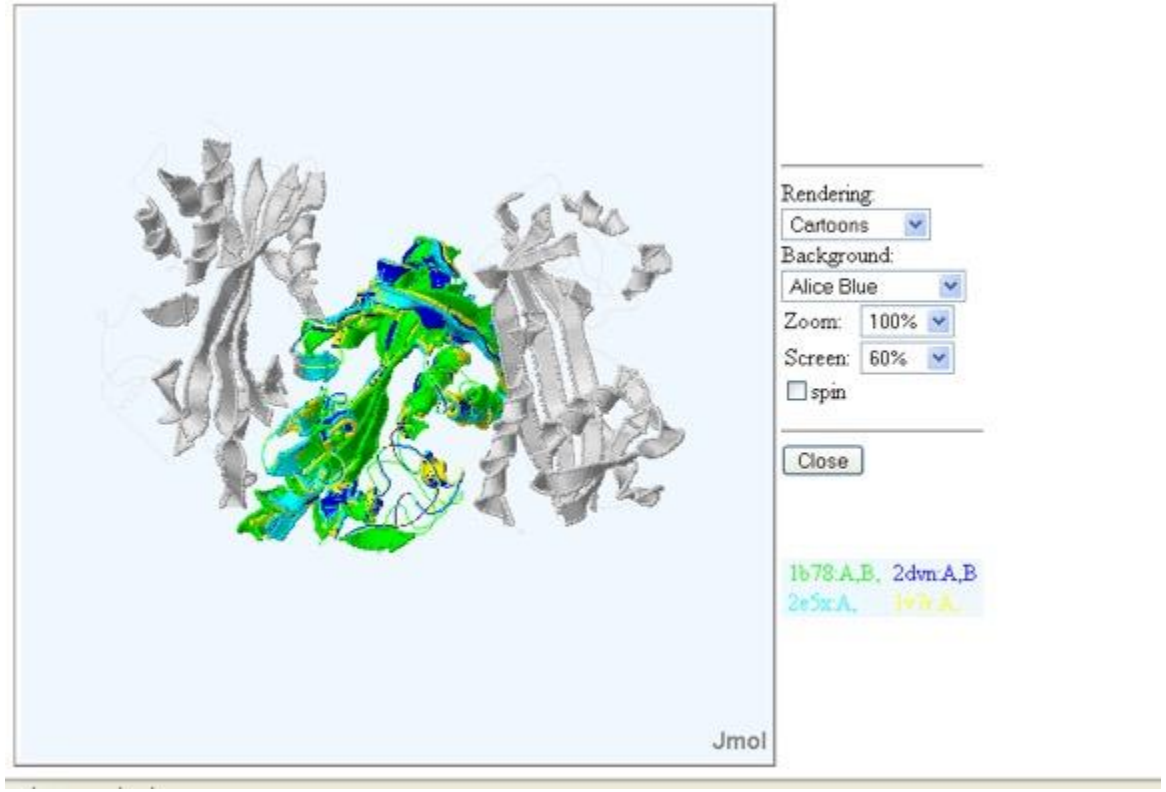
Multiple 3D alignment using PDBefold



Results from multiple 3D alignment

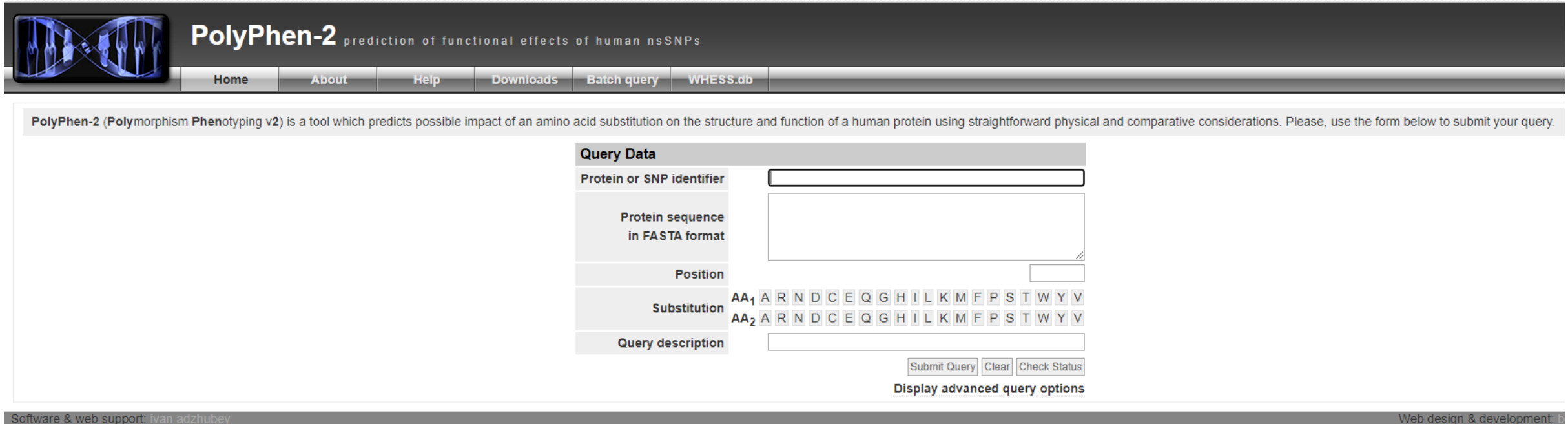
3D Structural alignment

1b78			2dvn			2e5x:A			1v7r:A		
				A:MET	1		MET	1		MET	1
	A:LYS	10	S	A:LYS	2	S	LYS	2	S	LYS	2
S	A:ILE	11	S	A:ILE	3	S	ILE	3	S	ILE	3
S	A:TYR	12	S	A:PHE	4	S	PHE	4	S	PHE	4
S	A:PHE	13	S	A:PHE	5	S	PHE	5	S	PHE	5
S	A:ALA	14	S	A:ILE	6	S	ILE	6	S	ILE	6
	A:THR	15		A:THR	7		THR	7		THR	7
	A:GLY	16		A:SER	8		SER	8		SER	8
H	A:ASN	17	H	A:ASN	9	H	ASN	9	H	ASN	9
H	A:PRO	18	H	A:PRO	10	H	PRO	10	H	PRO	10
H	A:ASN	19	H	A:GLY	11	H	GLY	11	H	GLY	11
H	A:LYS	20	H	A:LYS	12	H	LYS	12	H	LYS	12
H	A:ILE	21	H	A:VAL	13	H	VAL	13	H	VAL	13
H	A:LYS	22	H	A:ARG	14	H	ARG	14	H	ARG	14
H	A:GLU	23	H	A:GLU	15	H	GLU	15	H	GLU	15
H	A:ALA	24	H	A:VAL	16	H	VAL	16	H	VAL	16
H	A:ASN	25	H	A:ALA	17	H	ALA	17	H	ALA	17
H	A:ILE	26	H	A:ASN	18	H	ASN	18	H	ASN	18
H	A:ILE	27	H	A:PHE	19	H	PHE	19	H	PHE	19
H	A:LEU	28	H	A:LEU	20	H	LEU	20	H	LEU	20
	A:LYS	29	H	A:GLY	21	H	GLY	21	H	GLY	21

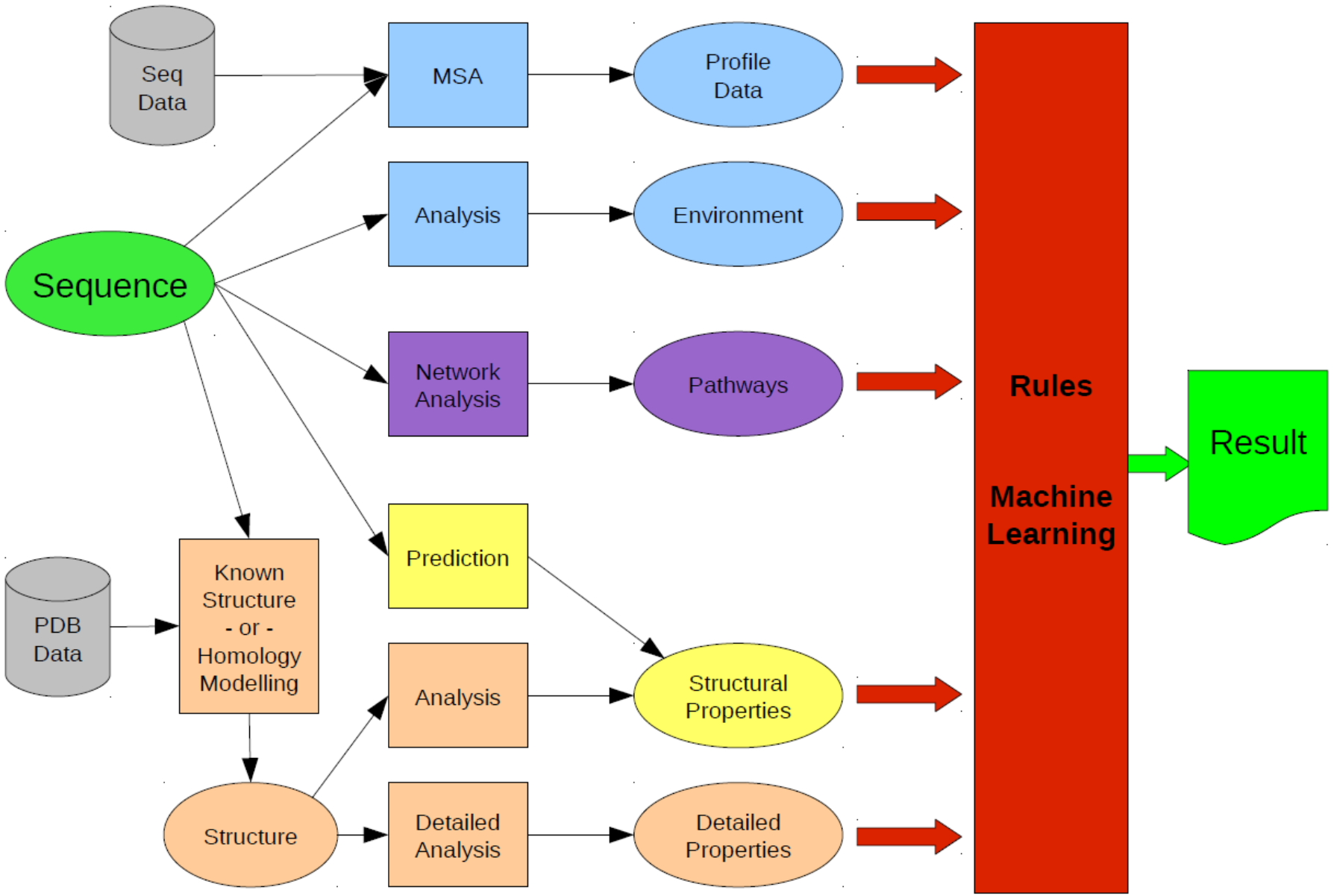


PolyPhen-2 prediction of functional effects of human SNPs

PolyPhen-2 (**Poly**morphism **Phen**otyping v**2**) is a tool which predicts possible impact of an amino acid substitution on the structure and function of a human protein using straightforward physical and comparative considerations.



Διαδικασία πρόβλεψης μία νέας δομής



3D structure prediction tools

Tertiary (or 3-D) structure prediction tools fall into two main methods: Ab initio, and comparative protein modeling.

Ab initio (or de novo) protein structure prediction methods attempt to predict tertiary structures from sequences based on general principles that govern protein folding energetics and/or statistical tendencies of conformational features that native structures acquire, without the use of explicit templates.
Ab initio protein structure prediction thus requires vast amount of computational power and time to solve the native conformation of a protein, and remains one of the top challenges for modern science.

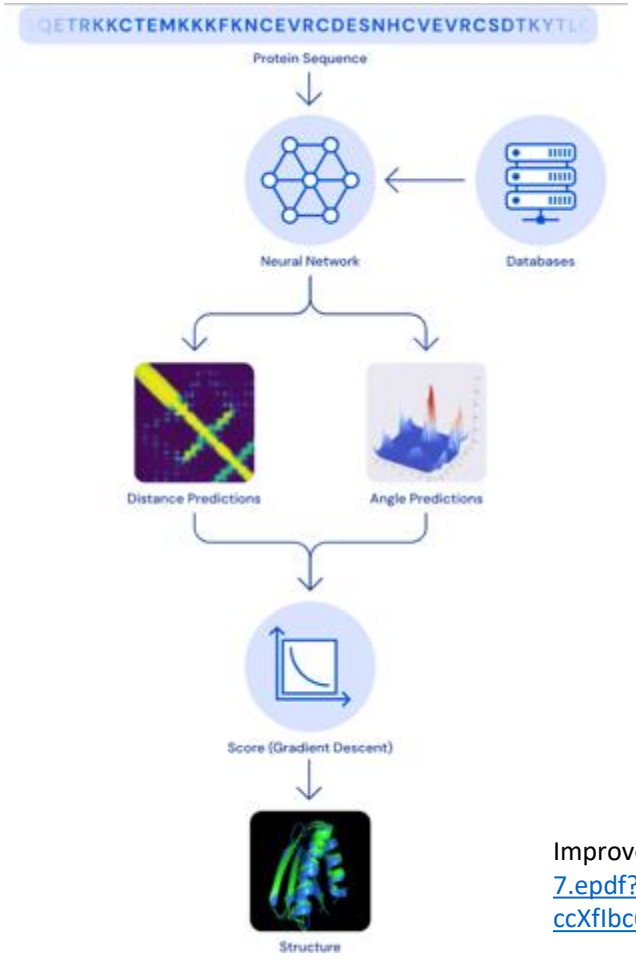
Most popular servers include Robetta (using the Rosetta software package), SWISS-MODEL, PEPstr, QUARK.

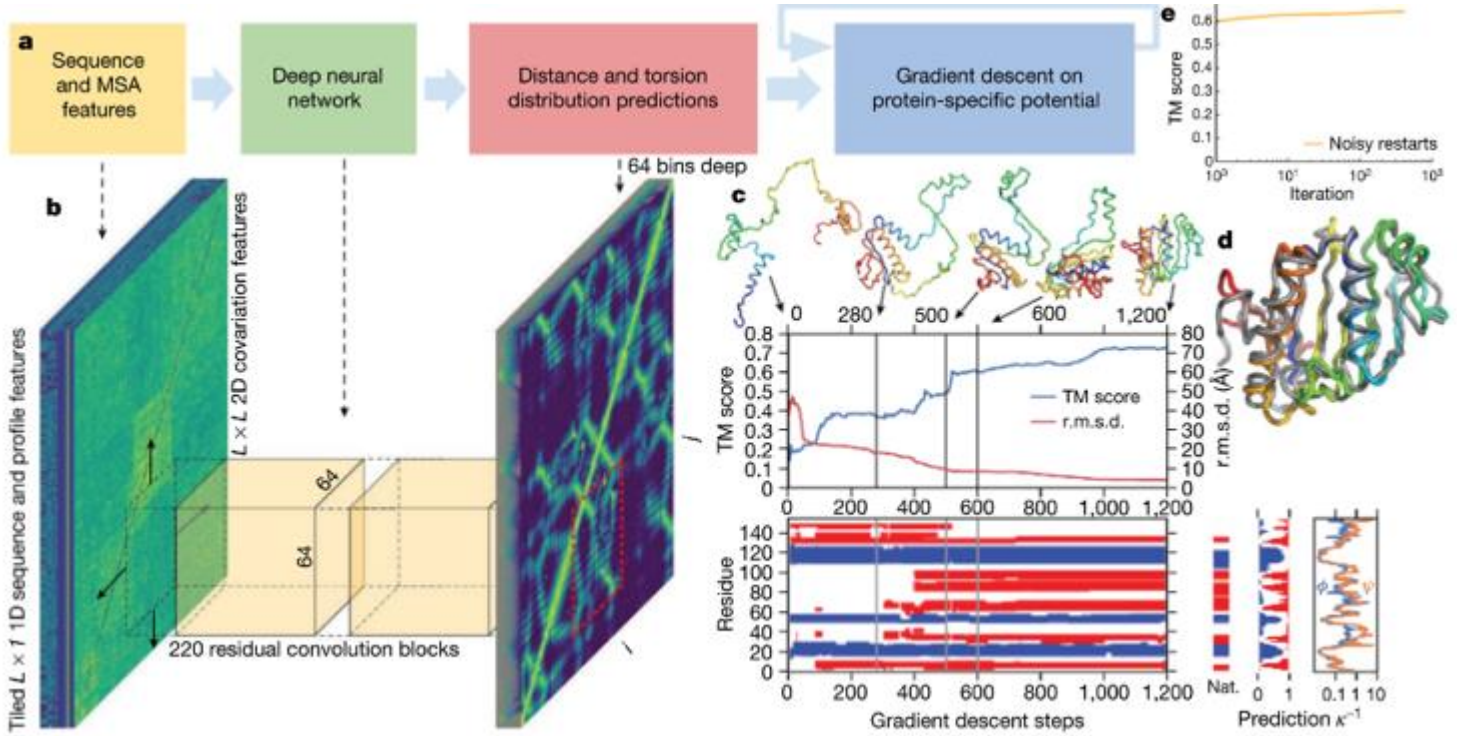
If a protein of known tertiary structure shares at least 30% of its sequence with a potential homolog of undetermined structure, comparative methods that overlay the putative unknown structure with the known can be utilized to predict the likely structure of the unknown. Homology modeling and protein threading are two main strategies that use prior information on other similar protein to propose a prediction of an unknown protein, based on its sequence.

Homology modeling and protein threading software include RaptorX, FoldX, HHpred, I-TASSER, and more.

ALPHA FOLD (https://deepmind.com/blog/article/AlphaFold-Using-AI-for-scientific-discovery)

AlphaFold placed first in the overall rankings of the 13th CASP
In November 2020, DeepMind's new version, AlphaFold 2, won CASP14





Improved protein structure prediction using potentials from deep learning. https://www.nature.com/articles/s41586-019-1923-7.epdf?author_access_token=Z_KaZKDqtKzbE7Wd5HtwI9RgN0jAjWel9jnR3ZoTv0MCcgAwHMgRx9mvLjNQdB2TlQQaa7l420UCtGo8vYQ39gg8lFWR9mAZtvsN_1PrccXfIbc6e-tGSgazNL_XdtQzn1PHfy21qdcxV7Pw-k3htw%3D%3D

https://github.com/deepmind/deepmind-research/tree/master/alphafold_casp13

I-TASSER

Protein Structure & Function Predictions

(The server completed predictions for 555532 proteins submitted by 132543 users from 148 countries or regions)
(The template library was updated on 2020/07/11)

I-TASSER (Iterative Threading ASSEmbly Refinement) is a hierarchical approach to protein structure and function prediction. It first identifies structural templates from the PDB by multiple threading approach LOMETS, with full-length atomic models constructed by iterative template-based fragment assembly simulations. Function insights of the target are then derived by re-threading the 3D models through protein function database BioLiP. I-TASSER (as 'Zhang-Server') was ranked as the No 1 server for protein structure prediction in recent community-wide CASP7, CASP8, CASP9, CASP10, CASP11, CASP12, and CASP13 experiments. It was also ranked as the best for function prediction in CASP9. The server is in active development with the goal to provide the most accurate protein structure and function predictions using state-of-the-art algorithms. Please report problems and questions at I-TASSER message board and our developers will study and answer the questions accordingly. (>> More about the server ...)

Structure models for the 2019-nCov Coronavirus genome by C-I-TASSER NEW

I-TASSER On-line Server (View an example of I-TASSER output):

Copy and paste your sequence below ([10, 1500] residues in FASTA format). Click here for a sample input:

Or upload the sequence from your local computer:

Choose File No file chosen

Email: (mandatory, where results will be sent to)

Password: (mandatory, please click here if you do not have a password)

ID: (optional, your given name of the protein)

▶ **Option I:** Assign additional restraints & templates to guide I-TASSER modeling.

▶ **Option II:** Exclude some templates from I-TASSER template library.

▶ **Option III:** Specify secondary structure for specific residues.

I-TASSER results for job id S555535

[Click on S555535_results.tar.bz2 to download the tarball file including all modeling results listed on this page]

(Click on Annotation of I-TASSER Output to read the instructions for how to interpret the results on this page)

Submitted Sequence in FASTA format

```
>seq
MAKSSFKISNPLEARMSESSRIREKYPDRIPVIVEKAGQSDVPDIDKKKYLVPADLTVGQ
FVYVVRKRIKLGAEKAIFVFVKNTLPPTAALMSAIYEEHKDEDGFLYMTYSGENTFGSLT
VA
```

Predicted Secondary Structure

```
                  20                  40                  60                  80                 100                 120
                   |                   |                   |                   |                   |                   |
Sequence    MAKSSFKISNPLEARMSESSRIREKYPDRIPVIVEKAGQSDVPDIDKKKYLVPADLTVGQFVYVVRKRIKLGAEKAIFVFVKNTLPPTAALMSAIYEEHKDEDGFLYMTYSGENTFGSLTVA
Prediction  CCCCCCCCCCCHHHHHHHHHHHHHCCCCCSSSSSSCCCCCCCCCCCCCSSSCCCCCCHHHHHHHHHHCCCCCCCSSSSSSCCCCCCCCCCHHHHHHHHCCCCCCSSSSSSCCCCCCCCCCC
Conf.Score  986532331899999999999987876515788763625785224523760699708899999764325698861999989851665475999999834877877999617865688565 9
            H:Helix; S:Strand; C:Coil
```

Predicted Solvent Accessibility

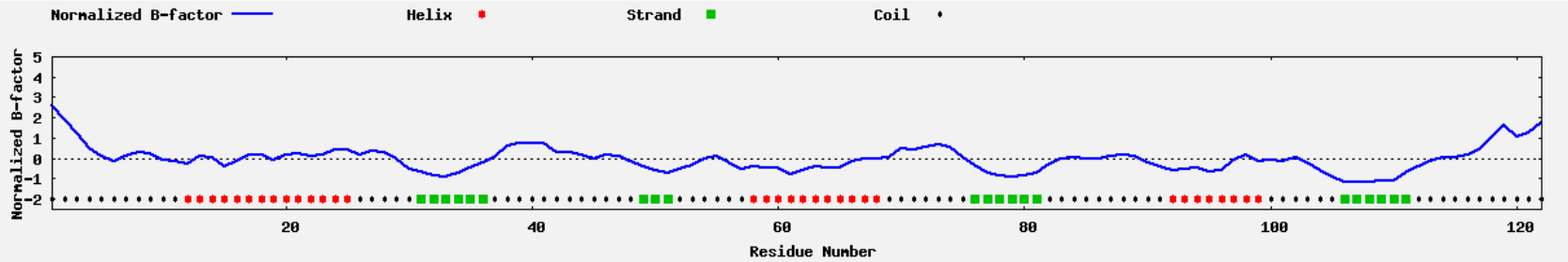
```
                  20                  40                  60                  80                 100                 120
                   |                   |                   |                   |                   |                   |
Sequence    MAKSSFKISNPLEARMSESSRIREKYPDRIPVIVEKAGQSDVPDIDKKKYLVPADLTVGQFVYVVRKRIKLGAEKAIFVFVKNTLPPTAALMSAIYEEHKDEDGFLYMTYSGENTFGSLTVA
Prediction  876451476331550364054127414730301033267561451534321134723233023104431714663101010243224643303401563444210000001324122336 48
            Values range from 0 (buried residue) to 9 (highly exposed residue)
```

Predicted normalized B-factor

(B-factor is a value to indicate the extent of the inherent thermal mobility of residues/atoms in proteins. In I-TASSER, this value is deduced from threading template proteins from the PDB in combination with the sequence profiles derived from sequence databases. The reported B-factor profile in the figure below corresponds to the normalized B-factor of the target protein, defined by B=(B'-u)/s, where B' is the raw B-factor value, u and s are respectively the mean and standard deviation of the raw B-factors along the sequence. Click here to read more about predicted normalized B-factor)

Predicted normalized B-factor

(B-factor is a value to indicate the extent of the inherent thermal mobility of residues/atoms in proteins. In I-TASSER, this value is deduced from threading template proteins from the PDB in combination with the sequence profiles derived from sequence databases. The reported B-factor profile in the figure below corresponds to the normalized B-factor of the target protein, defined by B=(B'-u)/s, where B' is the raw B-factor value, u and s are respectively the mean and standard deviation of the raw B-factors along the sequence. Click here to read more about predicted normalized B-factor)



Top 10 threading templates used by I-TASSER

(I-TASSER modeling starts from the structure templates identified by LOMETS from the PDB library. LOMETS is a meta-server threading approach containing multiple threading programs, where each threading program can generate tens of thousands of template alignments. I-TASSER only uses the templates of the highest significance in the threading alignments, the significance of which are measured by the Z-score, i.e. the difference between the raw and average scores in the unit of standard deviation. The templates in this section are the 10 best templates selected from the LOMETS threading programs. Usually, one template of the highest Z-score is selected from each threading program, where the threading programs are sorted by the average performance in the large-scale benchmark test experiments.)

Rank	PDB Hit	Iden1	Iden2	Cov	Norm. Z-score	Download Align.	20 | 40 | 60 | 80 | 100 | 120
							Sec.Str CCCCCCCCCCCCHHHHHHHHHHHHHHCCCCCSSSSSSCCCCCCCCCCCCCSSSCCCCCCHHHHHHHHHHCCCCCCCSSSSSSCCCCCCCCCCHHHHHHHHCCCCCCSSSSSSCCCCCCCCCCC
							Seq MAKSSFKISNPLEARMSESSRIREKYPDRIPVIVEKAGQSDVPDIDKKKYLVPADLTVGQFVYVVRKRIKLGAEKAIFVFVKNTLPPTAALMSAIYEEHKDEDGFLYMTYSGENTFGSLTVA
1	1eo6B	0.60	0.57	0.96	3.40	Download	M-KWMFKEDHSLEHRCVESAKIRAKYPDRVPVIVEKVSGSQIVDIDKRKYLVPSDITVAQFMWIIRKRIQLPSEKAIFLFVDKTVPQSSLTMGQLYEKEKDEDGFLYVAYSGENTFGF----
2	3h9dA	0.53	0.51	0.95	2.81	Download	-KDSKYKMSHTFESRQSDAAKVRERHPDRLPIICEKVYNSDIGELDRCKFLVPSDLTVGQFVSVLRKRVQLEAESALFVYTNDTVLPSSAQMADIYSKYKDEDGFLYMKYSGEATFG-----
3	4co7A	0.58	0.57	0.97	3.22	Download	SMKWMFKEDHSLEHRCVESAKIRAKYPDRVPVIVEKVSGSQIVDIDKRKYLVPSDITVAQFMWIIRKRIQLPSEKAIFLFVDKTVPQSSLTMGQLYEKEKDEDGFLYVAYSGENTFGF----
4	5lxiD	0.56	0.53	0.96	3.11	Download	-MKFQYKEDHPFEYRKKEGEKIRKKYPDRVPVIVEKAPKARVPDLDKRKYLVPSDLTVGQFYFLIRKRIHLRPEDALFFFVNNTIPPTSATMGQLYEDNHEEDYFLYVAYSDESVYGK----
5	5lxiD	0.56	0.53	0.96	2.28	Download	-MKFQYKEDHPFEYRKKEGEKIRKKYPDRVPVIVEKAPKARVPDLDKRKYLVPSDLTVGQFYFLIRKRIHLRPEDALFFFVNNTIPPTSATMGQLYEDNHEEDYFLYVAYSDESVYGK----
6	3h9dA	0.53	0.51	0.95	3.88	Download	-KDSKYKMSHTFESRQSDAAKVRERHPDRLPIICEKVYNSDIGELDRCKFLVPSDLTVGQFVSVLRKRVQLEAESALFVYTNDTVLPSSAQMADIYSKYKDEDGFLYMKYSGEATFG-----
7	5lxiD	0.56	0.53	0.96	3.45	Download	-MKFQYKEDHPFEYRKKEGEKIRKKYPDRVPVIVEKAPKARVPDLDKRKYLVPSDLTVGQFYFLIRKRIHLRPEDALFFFVNNTIPPTSATMGQLYEDNHEEDYFLYVAYSDESVYGK----
8	1eo6A	0.60	0.57	0.95	2.98	Download	M-KWMFKEDHSLEHRCVESAKIRAKYPDRVPVIVEKVSGSQIVDIDKRKYLVPSDITVAQFMWIIRKRIQLPSEKAIFLFVDKTVPQSSLTMGQLYEKEKDEDGFLYVAYSGENTFG-----
9	3h9dA	0.53	0.51	0.95	3.24	Download	-KDSKYKMSHTFESRQSDAAKVRERHPDRLPIICEKVYNSDIGELDRCKFLVPSDLTVGQFVSVLRKRVQLEAESALFVYTNDTVLPSSAQMADIYSKYKDEDGFLYMKYSGEATFG-----
10	1gnuA	0.00	0.54	0.96	3.21	Download	-----

(a) All the residues are colored in black; however, those residues in template which are identical to the residue in the query sequence are highlighted in color. Coloring scheme is based on the property of amino acids, where polar are brightly coloured while non-polar residues are colored in dark shade. (more about the colors used)

(b) Rank of templates represents the top ten threading templates used by I-TASSER.

(c) Ident1 is the percentage sequence identity of the templates in the threading aligned region with the query sequence.

(d) Ident2 is the percentage sequence identity of the whole template chains with query sequence.

(e) Cov represents the coverage of the threading alignment and is equal to the number of aligned residues divided by the length of query protein.

(f) Norm. Z-score is the normalized Z-score of the threading alignments. Alignment with a Normalized Z-score >1 mean a good alignment and vice versa.

(g) Download Align. provides the 3D structure of the aligned regions of the threading templates.

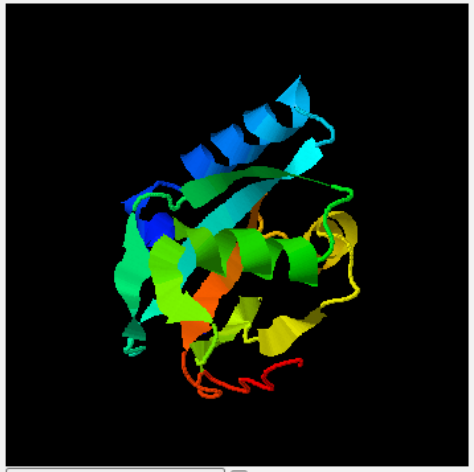
(h) The top 10 alignments reported above (in order of their ranking) are from the following threading programs:
1: MUSTER 2: FFAS-3D 3: SPARKS-X 4: HHSEARCH2 5: HHSEARCH I 6: Neff-PPAS 7: HHSEARCH 8: pGenTHREADER 9: wdPPAS 10: cdPPAS

Top 5 final models predicted by I-TASSER

(For each target, I-TASSER simulations generate a large ensemble of structural conformations, called decoys. To select the final models, I-TASSER uses the SPICKER program to cluster all the decoys based on the pair-wise structure similarity, and reports up to five models which corresponds to the five largest structure clusters. The confidence of each model is quantitatively measured by C-score that is calculated based on the significance of threading template alignments and the convergence parameters of the structure assembly simulations. C-score is typically in the range of [-5, 2], where a C-score of a higher value signifies a model with a higher confidence and vice-versa. TM-score and RMSD are estimated based on C-score and protein length following the correlation observed between these qualities. Since the top 5 models are ranked by the cluster size, it is possible that the lower-rank models have a higher C-score in rare cases. Although the first model has a better quality in most cases, it is also possible that the lower-rank models have a better quality than the higher-rank models as seen in our benchmark tests. If the I-TASSER simulations converge, it is possible to have less than 5 clusters generated; this is usually an indication that the models have a good quality because of the converged simulations.)

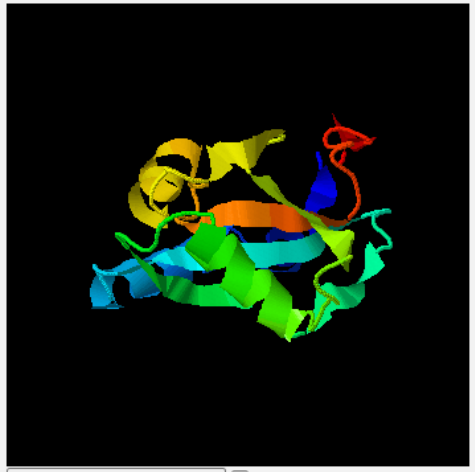
- More about C-score
- Local structure accuracy profile of the top five models

(By right-click on the images, you can export image file or change the configurations, e.g. modifying the background color or stopping the spin of your models)



Reset to initial orientation ☐ Spin On/Off

- Download Model 1
- C-score=0.51 (Read more about C-score)
- Estimated TM-score = 0.78±0.10
- Estimated RMSD = 3.4±2.3Å



Reset to initial orientation ☐ Spin On/Off

- Download Model 2
- C-score = -0.65

Reset to initial orientation ☐ Spin On/Off

- Download Model 3
- C-score = 0.47

Reset to initial orientation ☐ Spin On/Off

- Download Model 4
- C-score = -2.03

Proteins structurally close to the target in the PDB (as identified by TM-align)

(After the structure assembly simulation, I-TASSER uses the TM-align structural alignment program to match the first I-TASSER model to all structures in the PDB library. This section reports the top 10 proteins from the PDB that have the closest structural similarity, i.e. the highest TM-score, to the predicted I-TASSER model. Due to the structural similarity, these proteins often have similar function to the target. However, users are encouraged to use the data in the next section 'Predicted function using COACH' to infer the function of the target protein, since COACH has been extensively trained to derive biological functions from multi-source of sequence and structure features which has on average a higher accuracy than the function annotations derived only from the global structure comparison.)

Top 10 Identified stuctural analogs in PDB

Click to view	Rank	PDB Hit	TM-score	RMSD[a]	IDEN[a]	Cov	Alignment
○	1	4co7A	0.946	0.66	0.585	0.967	Download
○	2	5lxiD	0.924	1.13	0.556	0.959	Download
○	3	3dowA	0.914	1.44	0.568	0.967	Download
○	4	2xidC	0.891	2.07	0.361	0.967	Download

TM-align

Quick & Accurate Structural Alignment

TM-align is an algorithm for sequence independent protein structure comparisons. For two protein structures of unknown equivalence, TM-align first generates optimized residue-to-residue alignment based on structural similarity using heuristic dynamic programming iterations. An optimal superposition of the two structures built on the detected alignment, as well as the TM-score value which scales the structural similarity, will be returned. TM-score has the value in (0,1], where 1 indicates a perfect match between two structures. Following strict statistics of structures in the PDB, scores below 0.2 correspond to randomly chosen unrelated proteins while those higher than 0.5 assume generally the same fold in SCOP/CATH.

News: NEW TM-align now allows for input structure with either PDB or PDBx/mmCIF format. Meanwhile C++ version of TM-align is now officially released.

TM-align on-line (view an example of output)

Note: This server is only for pair-wise structure comparison. If you want to match one protein structure with all proteins in the PDB library, you can do it in COFACTOR Server.

- Input Structure 1 in PDB format or PDBx/mmCIF format (mandatory):
  Please copy and paste your structure file here. Sample input

```
HEADER    OXYGEN TRANSPORT                        13-DEC-97   101M
TITLE     SPERM WHALE MYOGLOBIN F46V N-BUTYL ISOCYANIDE AT PH 9.0
COMPND    MOL_ID: 1;
COMPND   2 MOLECULE: MYOGLOBIN;
COMPND   3 CHAIN: A;
COMPND   4 ENGINEERED: YES;
COMPND   5 MUTATION: YES
SOURCE    MOL_ID: 1;
SOURCE   2 ORGANISM_SCIENTIFIC: PHYSETER CATODON;
SOURCE   3 ORGANISM COMMON: SPERM WHALE;
```

  Or upload the stucture file:
  Choose File No file chosen

- Input Structure 2 in PDB format or PDBx/mmCIF format (mandatory):
  Please copy and paste your structure file here. Sample input

```
HEADER    OXYGEN STORAGE                          22-FEB-89   1MBA
TITLE     APLYSIA LIMACINA MYOGLOBIN. CRYSTALLOGRAPHIC ANALYSIS AT
TITLE    2 1.6 ANGSTROMS RESOLUTION
COMPND    MOL_ID: 1;
COMPND   2 MOLECULE: MYOGLOBIN;
COMPND   3 CHAIN: A;
COMPND   4 ENGINEERED: YES
SOURCE    MOL_ID: 1;
SOURCE   2 ORGANISM_SCIENTIFIC: APLYSIA LIMACINA;
SOURCE   3 ORGANISM COMMON: SLUG SEA HARE;
```

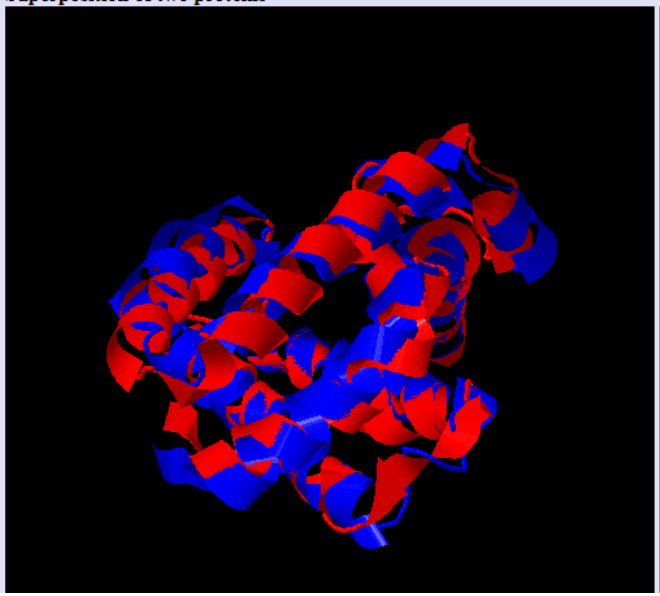
  Or upload the stucture file:
  Choose File No file chosen

- Input Email: (optional, where results will be sent to)

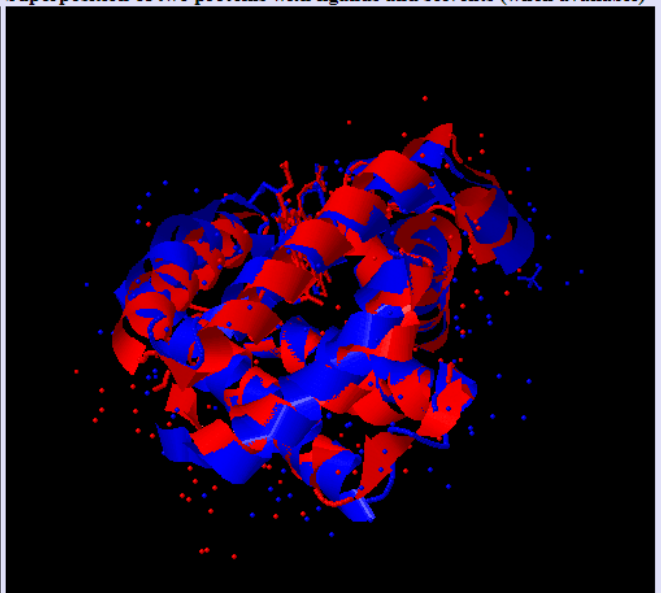
```
 ****************************************************************************
 *                      TM-align (Version 20190822)                         *
 * An algorithm for protein structure alignment and comparison              *
 * Based on statistics:                                                     *
 *       0.0 < TM-score < 0.30, random structural similarity                *
 *       0.5 < TM-score < 1.00, in about the same fold                      *
 * Reference: Y Zhang and J Skolnick, Nucl Acids Res 33, 2302-9 (2005)      *
 * Please email your comments and suggestions to: zhng@umich.edu            *
 ****************************************************************************

Name of Chain_1: A835798
Name of Chain_2: B835798
Length of Chain_1:  154 residues
Length of Chain_2:  146 residues

Aligned length=  143, RMSD=   1.83, Seq_ID=n_identical/n_aligned= 0.245
TM-score= 0.81453 (if normalized by length of Chain_1)
TM-score= 0.85377 (if normalized by length of Chain_2)
(You should use TM-score normalized by length of the reference protein)

(":" denotes aligned residue pairs of d < 5.0 A, "." denotes other aligned residues)
MVLSEGEWQLVLHVWAKVEADVAGHGQDILIRLFKSHPETLEKFDRVKHLKTEAEMKASEDLKKHGVTVLTALGAILKKK--G-HHEAELKPLAQSHATKHKIPIKYLEFISEAIIHVLHSRHPGNFGADAQGAMNKALELFRKDIAAKYKELGYQG
 :::::::::::::::::::::::::::::::::::::::::::::::::::: :::::::::::::::::::::::::::::  : :::::::::::::: :::::::::::::::::::::::::::  ::::::::::::::::::::::: :     ::
-SLSAAEADLAGKSWAPVFANKNANGLDFLVALFEKFPDSANFFADFKGKS-VADIKASPKLRDVSSRIFTRLNEFVNNAANAGKMSAMLSQFAKEHV-GFGVGSAQFENVRSMFPGFVASVAA--PPAGADAAWTKLFGLIIDALKA-A-----GA
```

Visualization (Protein-1 in blue and Protein-2 in red)

Superposition of two proteins



Superposition of two proteins with ligands and solvents (when available)



- **TMScore**: TM-score is intended as a more accurate measure of the quality of full-length protein structures than the often used RMSD measure
- **RMSD** — a different structure comparison measure
- **GDT** — a different structure comparison measure
- Longest continuous segment (**LCS**)
- Global distance calculation (**GDC**_sc, GDC_all) — Structure comparison measures that use full-model information (not just α-carbon) to assess similarity
- Local global alignment (**LGA**) — Protein structure alignment program and structure comparison measure

USE CASES

Chronic Lymphocytic Leukemia CLL

Somatic hypermutation (SHM) analysis

- 19,635 IGHV-IGHD-IGHJ gene rearrangement sequences from CLL patients were examined.
  Based on IGHV gene usage, all cases were subgrouped into 3 phylogenetic clans namely
  - Clan I | IGHV1/5/7), n=5439;
  - Clan II | IGHV2/4/6, n=5092 and
  - Clan III | IGHV3, n=9104.
- A region weight factor was set based on the physicochemical properties of the amino acids and the entropy of the SHMs in the complementarity determining regions (CDRs) and the framework regions (FRs).
- The Euclidean distance between the sequences was defined as a combination of the Blosum Matrix and the region weight factor,
- DBSCAN clustering algorithm
- A force-directed graph drawing visualization the associations between SHMs and CLL groups was underlined using the VH CDR3 patterns as similarity metric.

Somatic hypermutation (SHM) analysis

Distance between sequences:

- Position Entropy
- the Blosum Matrix
- The physicochemical similarity was defined according to the 11 “IMGT Physicochemical” classes, for the 20 amino acids”

	Ala	Arg	Asn	Asp	Cys	Gln	Glu	Gly	His	Ile	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val
Ala	4																			
Arg	−1	5																		
Asn	−2	0	6																	
Asp	−2	−2	1	6																
Cys	0	−3	−3	−3	9															
Gln	−1	1	0	0	−3	5														
Glu	−1	0	0	2	−4	2	5													
Gly	0	−2	0	−1	−3	−2	−2	6												
His	−2	0	1	−1	−3	0	0	−2	8											
Ile	−1	−3	−3	−3	−1	−3	−3	−4	−3	4										
Leu	−1	−2	−3	−4	−1	−2	−3	−4	−3	2	4									
Lys	−1	2	0	−1	−3	1	1	−2	−1	−3	−2	5								
Met	−1	−1	−2	−3	−1	0	−2	−3	−2	1	2	−1	5							
Phe	−2	−3	−3	−3	−2	−3	−3	−3	−1	0	0	−3	0	6						
Pro	−1	−2	−2	−1	−3	−1	−1	−2	−2	−3	−3	−1	−2	−4	7					
Ser	1	−1	1	0	−1	0	0	0	−1	−2	−2	0	−1	−2	−1	4				
Thr	0	−1	0	−1	−1	−1	−1	−2	−2	−1	−1	−1	−1	−2	−1	1	5			
Trp	−3	−3	−4	−4	−2	−2	−3	−2	−2	−3	−2	−3	−1	1	−4	−3	−2	11		
Tyr	−2	−2	−2	−3	−2	−1	−2	−3	2	−1	−1	−2	−1	3	−3	−2	−2	2	7	
Val	0	−3	−3	−3	−1	−2	−2	−3	−3	3	1	−2	1	−1	−2	−2	0	−3	−1	4

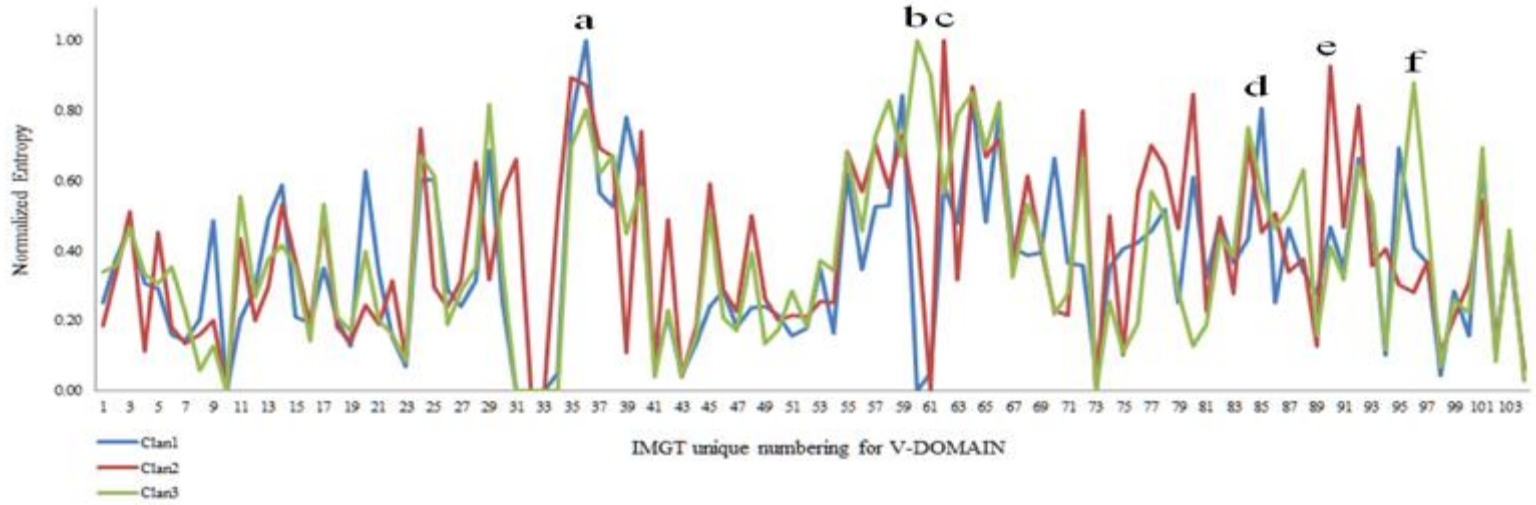
Volume' classes		Hydropathy' classes											
	in A³	Hydrophobic				Neutral				Hydrophilic			
Very large	189-228	F		W		Y							
Large	162-174	I	L	M					K	R			
Medium	138-154	V						H			E		Q
Small	108-117			C	P	T					D		N
Very small	60-90	A			G	S							
		Alphatic		Sulfur		Hydroxil		Basic		Acidic		Amide	
		Uncharged				Uncharged		Charged				Uncharged	
		Nonpolar				Polar							

Somatic hypermutation (SHM) analysis

- Region weight factor was set based on the physicochemical properties of the amino acids and the entropy of the SHMs in the complementarity determining regions (CDRs) and the framework regions (FRs).

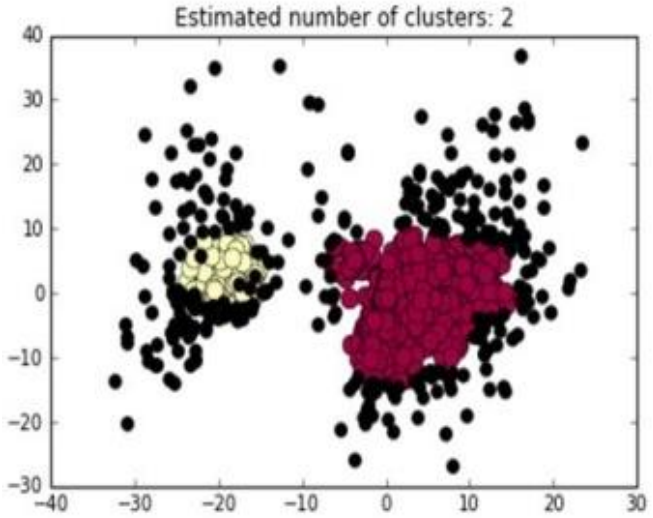
- The SHM variance of each position was calculated through the position Shannon entropy $H_2(X) = -\sum_{i=1}^{n} \frac{m_i}{N} \log_2 \frac{m_i}{N}$

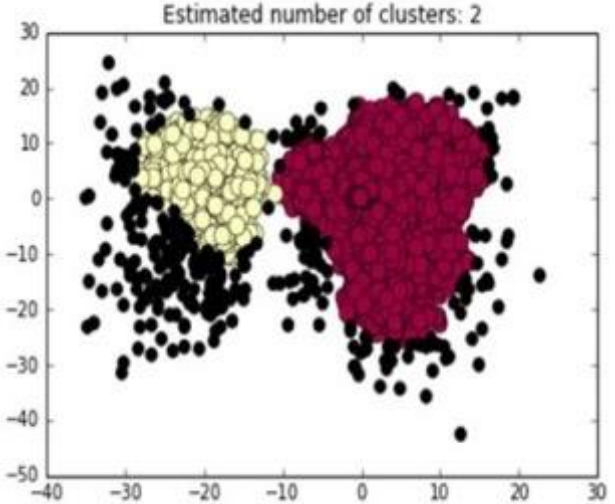
where m_i / N is the frequency of the SHM i in every codon of the IGHV genes sequence.

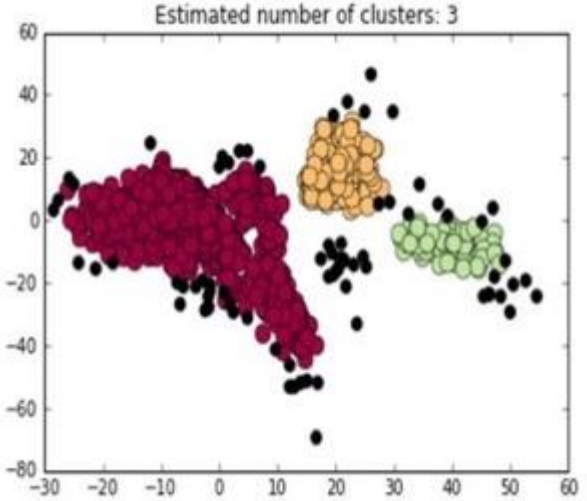


The points a,b,c underline the maximum entropy in CDRs while d,e,f in FRs for ClanI, Clan II, ClanIII respectively.

Somatic hypermutation (SHM) analysis



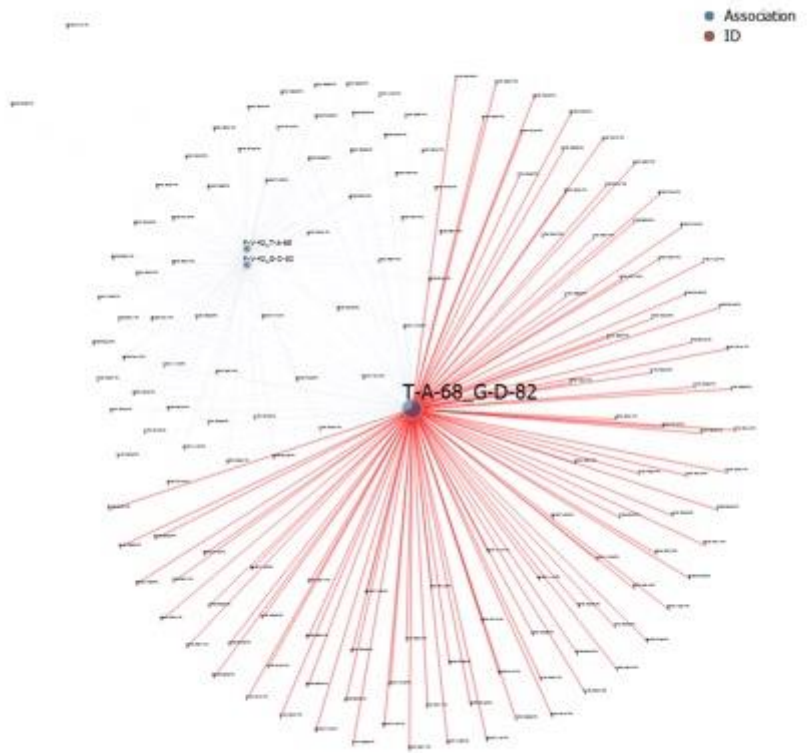


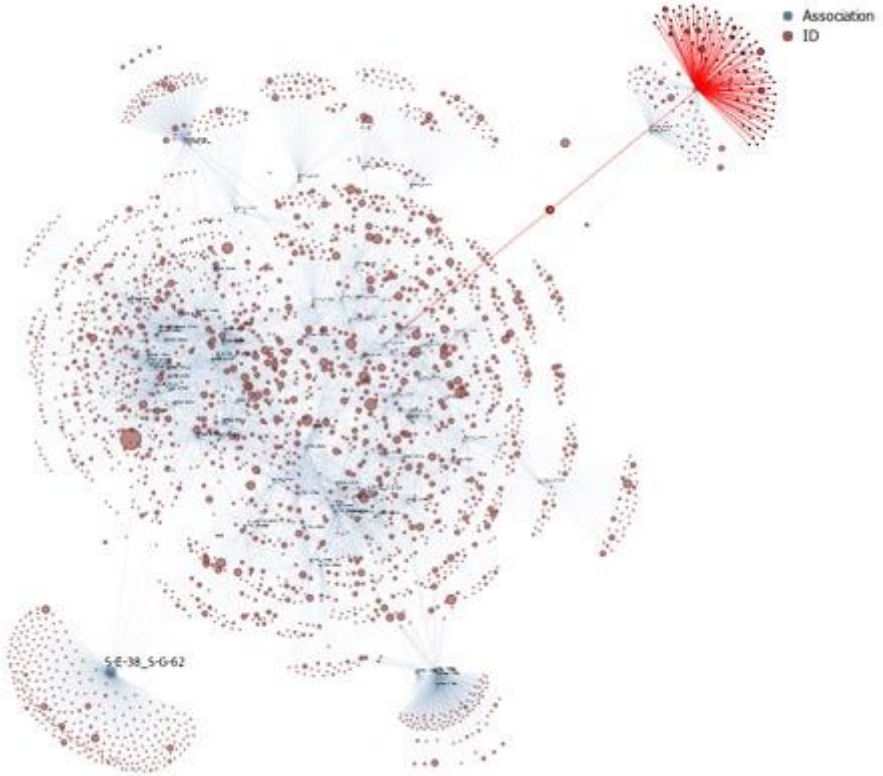


DBScan Clusters:

	Cluster 1	Cluster 2	Cluster 3	Outliers
Gr. clan I	4496	766	n/a	218
Gr. clan II	3992	824	171	107
Gr. clan III	7671	1092	n/a	342

Somatic hypermutation (SHM) analysis





Somatic hypermutation (SHM) analysis

Group No	Jaccard Index	Subset	Associations	VH CDR3 pattern	Sequence Logo
Group A-Clan1_Cluster1	0.841	#28A	S-M-78_V-A-100	ARxxxGxxtYYYYYGMDx	
GroupB-Clan1_Cluster1	0.317	#1	R-W-55_V-A-100, G-D-36_V-A-100, G-D-36_R-W-55, M-I-39_R-W-55, R-W-55_S-M-78	ARx[NQ]W[AVLI]xxxxFDx	
Group B-Clan1_Cluster1	0.493	#28A	R-W-55_V-A-100, G-D-36_V-A-100, G-D-36_R-W-55, M-I-39_R-W-55, R-W-55_S-M-79	ARxxxGxxtYYYYYGMDx	
Group C-Clan1_Cluster1	0.384	#59	A-T-38_G-R-55, F-L-62_T-I-64	AxxxDFWSGxxx	
Group D-Clan2_Cluster1	0.067	#4	S-T-40_P-S-45	[AVLI]RGxxxxxxx[KRH]RYYYYGx[DE]x	

Group E-Clan2_Cluster1	0.330	#148B	S-T-31_N-D-59	*(not officially defined)*	
Group F-Clan2_Cluster1	0.250	#14	P-A-46_Y-T-58	x[KRH]GGxWxFDx	
Group G-Clan2_Cluster1	0.240	#14	S-T-40_I-M-78	x[KRH]GGxWxFDx	
Group H-Clan2_Cluster1	0.400	#77	P-S-16_C-Y-103	[AVLI]RGxxx[ST]GWxxxxx	
Group I-Clan3_Cluster1	0.292	#31	S-E-38_S-G-62	ARxxxxxxxxxxxYYYxMDx	

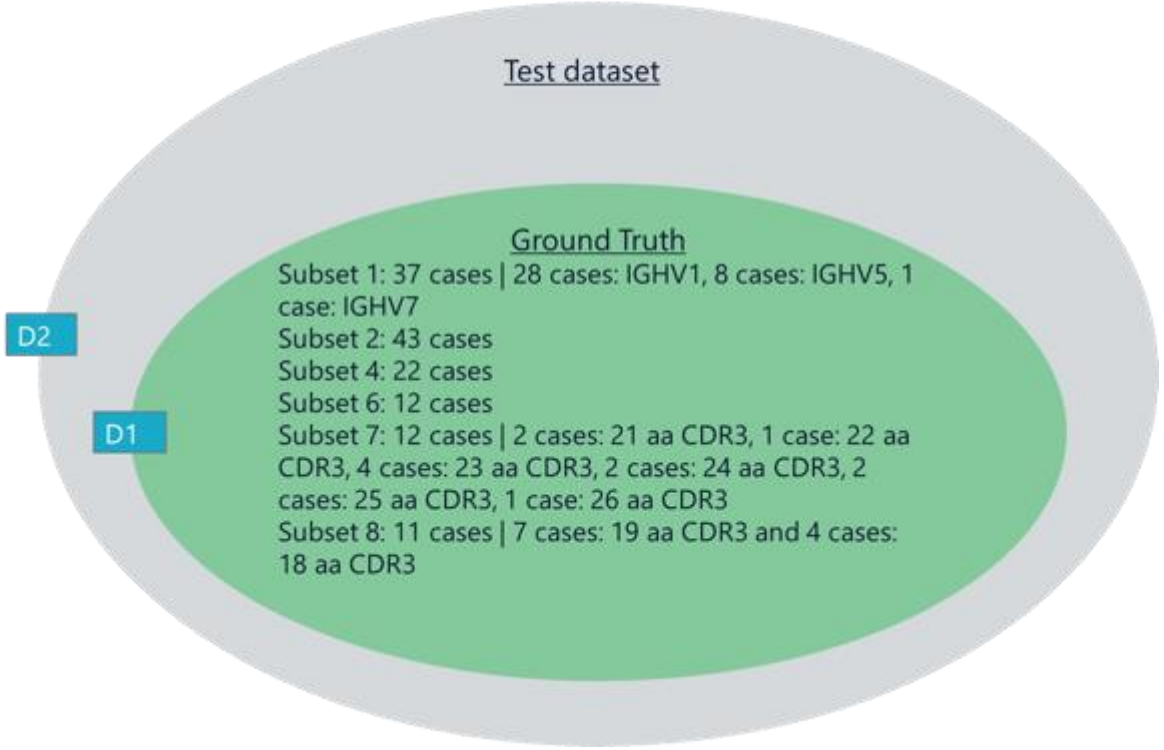
results

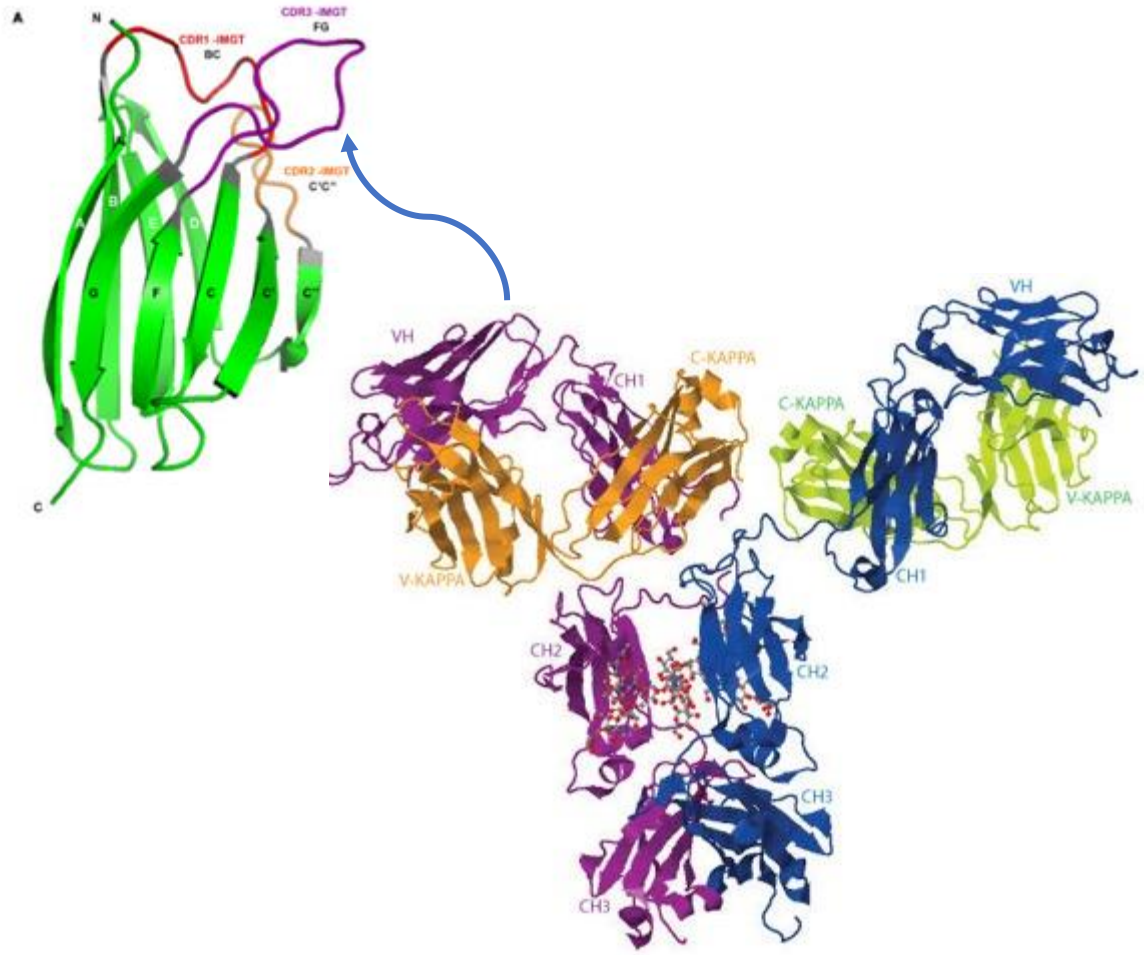
- Clan based patterns in terms of both quantity and quality of mutations.
- Some frequent SHMs were linked to specific subsets. And are now considered subset specific.
- Next level analysis will include associations between more than two SHMs.
- SHM patterns in the VH region will then be associated with CDR3 patterns leading to the identification of yet uncharacterized group of patients.

STRUCTURAL SIMILARITY

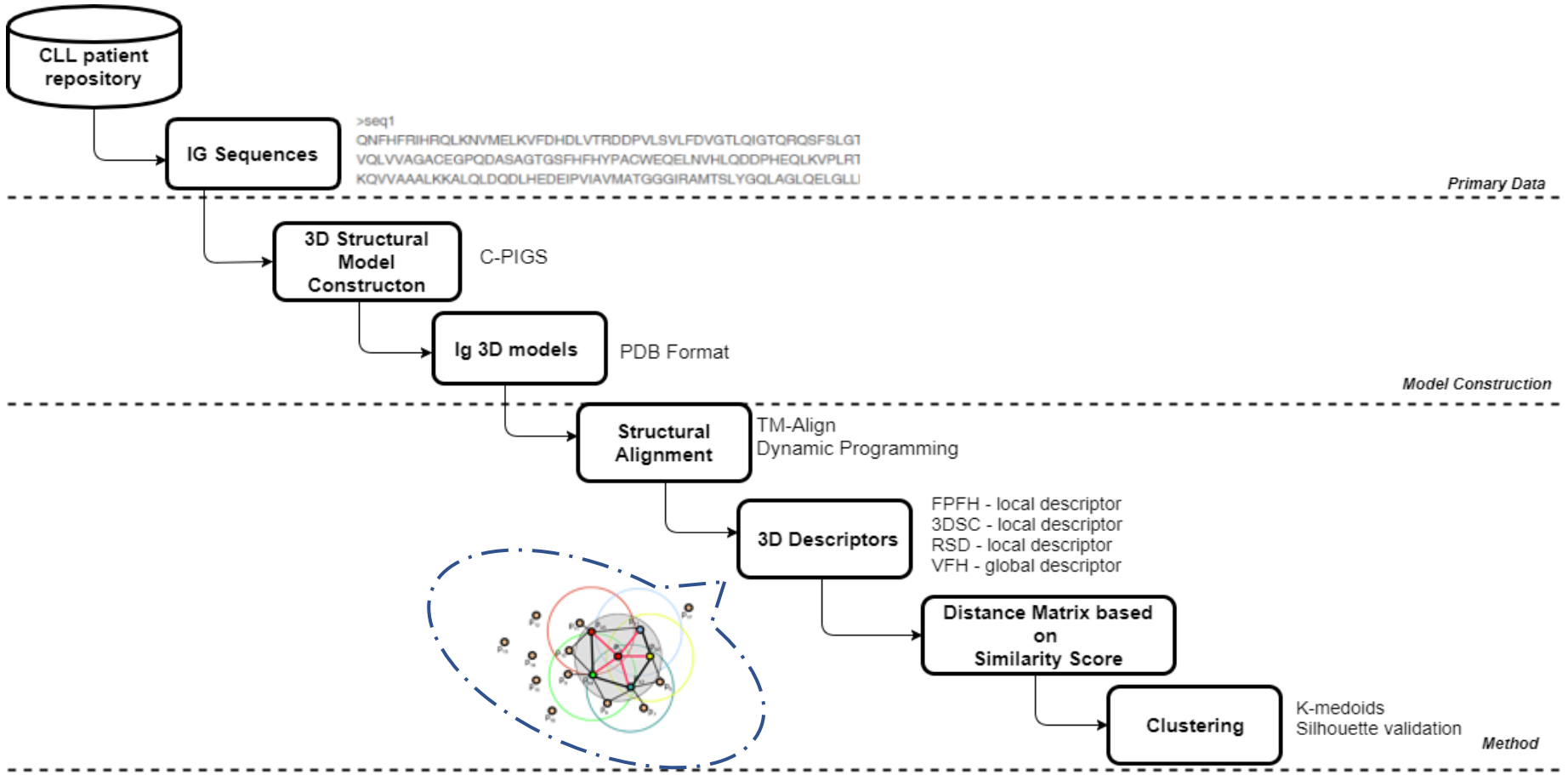
Datasets Description

Dataset	Protein structures	Predefined Subsets
D1	137	6 (D1.A ~ D1.F)
D2	925	6 (D1.A ~ D1.F) + Unknown





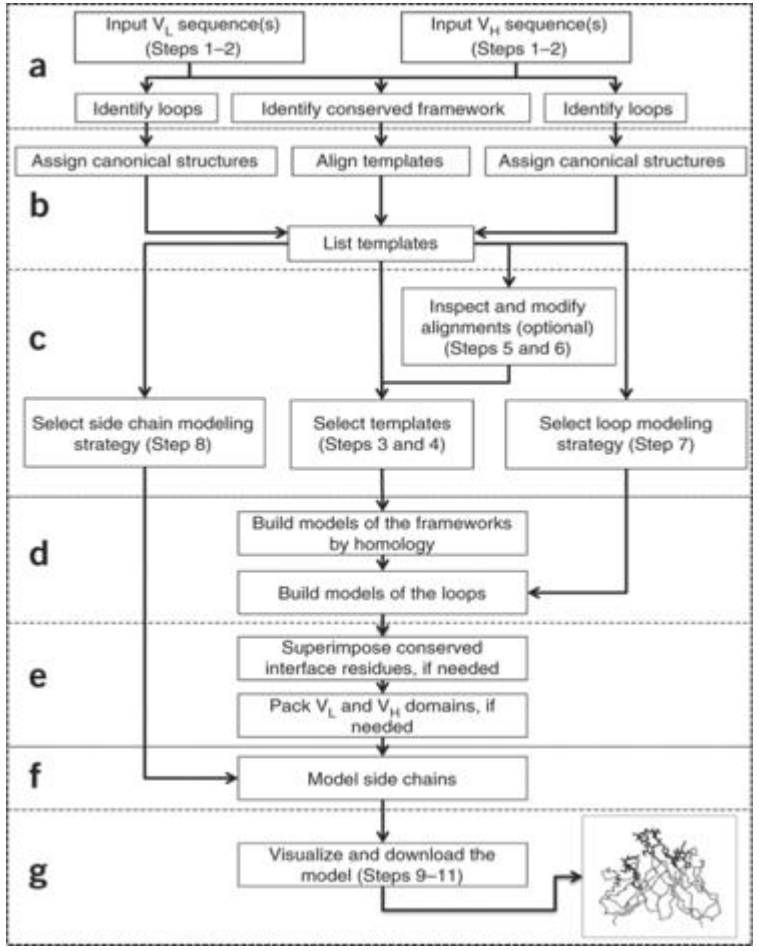
STRUCTURAL SIMILARITY



STRUCTURAL SIMILARITY

Marcatili, P., Rosi, A., & Tramontano, A. (2008). PIGS: automatic prediction of antibody structures. Bioinformatics, 24(17), 1953-1954.

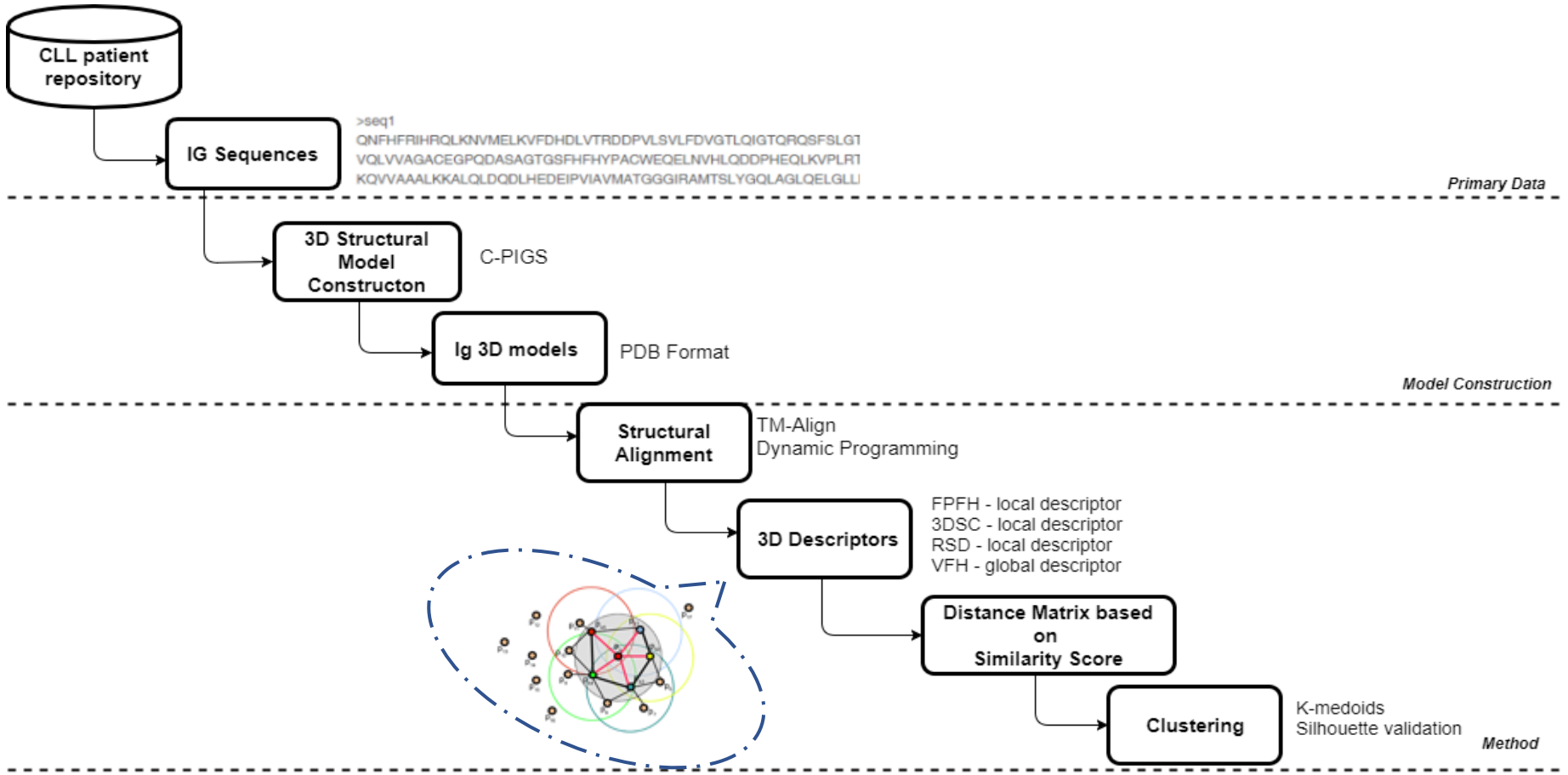
C-PIGS structural prediction protocol



Dataset	Patients	Predefined Subsets
D1	106	6 (D1.A ~ D1.F)
D2	894	N/A

```
EXPDTA    THEORETICAL MODEL, MODELLER 9.14 2016/01/20 18:57:51
REMARK   6 MODELLER OBJECTIVE FUNCTION:       297.9526
REMARK   6 MODELLER BEST TEMPLATE % SEQ ID: 100.000
ATOM      1  N   VAL H   2     -10.312   3.544  19.814  1.00  0.00           N
ATOM      2  CA  VAL H   2     -11.387   2.542  19.698  1.00  0.00           C
ATOM      3  CB  VAL H   2     -11.473   1.658  20.956  1.00  0.00           C
ATOM      4  CG1 VAL H   2     -12.684   0.741  20.881  1.00  0.00           C
ATOM      5  CG2 VAL H   2     -10.196   0.849  21.127  1.00  0.00           C
ATOM      6  C   VAL H   2     -12.770   3.081  19.445  1.00  0.00           C
ATOM      7  O   VAL H   2     -13.074   3.395  18.296  1.00  0.00           O
ATOM      8  N   GLN H   3     -13.662   3.226  20.453  1.00  0.00           N
ATOM      9  CA  GLN H   3     -14.986   3.636  20.044  1.00  0.00           C
ATOM     10  CB  GLN H   3     -15.890   2.416  19.853  1.00  0.00           C
ATOM     11  CG  GLN H   3     -15.420   1.460  18.769  1.00  0.00           C
ATOM     12  CD  GLN H   3     -15.622   2.016  17.374  1.00  0.00           C
ATOM     13  OE1 GLN H   3     -16.525   2.819  17.140  1.00  0.00           O
ATOM     14  NE2 GLN H   3     -14.778   1.590  16.441  1.00  0.00           N
ATOM     15  C   GLN H   3     -15.613   4.581  21.036  1.00  0.00           C
ATOM     16  O   GLN H   3     -15.496   4.408  22.247  1.00  0.00           O
ATOM     17  N   LEU H   4     -16.308   5.623  20.523  1.00  0.00           N
ATOM     18  CA  LEU H   4     -17.040   6.541  21.354  1.00  0.00           C
ATOM     19  CB  LEU H   4     -16.344   7.903  21.390  1.00  0.00           C
ATOM     20  CG  LEU H   4     -17.047   9.003  22.188  1.00  0.00           C
ATOM     21  CD1 LEU H   4     -17.113   8.639  23.663  1.00  0.00           C
ATOM     22  CD2 LEU H   4     -16.340  10.337  22.003  1.00  0.00           C
ATOM     23  C   LEU H   4     -18.425   6.616  20.781  1.00  0.00           C
ATOM     24  O   LEU H   4     -18.597   6.919  19.600  1.00  0.00           O
ATOM     25  N   VAL H   5     -19.460   6.339  21.604  1.00  0.00           N
ATOM     26  CA  VAL H   5     -20.814   6.394  21.116  1.00  0.00           C
ATOM     27  CB  VAL H   5     -21.490   5.011  21.169  1.00  0.00           C
ATOM     28  CG1 VAL H   5     -22.939   5.109  20.719  1.00  0.00           C
ATOM     29  CG2 VAL H   5     -20.728   4.013  20.310  1.00  0.00           C
ATOM     30  C   VAL H   5     -21.547   7.404  21.945  1.00  0.00           C
ATOM     31  O   VAL H   5     -21.410   7.435  23.167  1.00  0.00           O
ATOM     32  N   GLN H   6     -22.369   8.249  21.291  1.00  0.00           N
ATOM     33  CA  GLN H   6     -23.047   9.315  21.972  1.00  0.00           C
ATOM     34  CB  GLN H   6     -22.757  10.655  21.293  1.00  0.00           C
ATOM     35  CG  GLN H   6     -21.294  11.067  21.331  1.00  0.00           C
ATOM     36  CD  GLN H   6     -21.049  12.407  20.666  1.00  0.00           C
ATOM     37  OE1 GLN H   6     -19.919  12.731  20.301  1.00  0.00           O
ATOM     38  NE2 GLN H   6     -22.109  13.190  20.507  1.00  0.00           N
ATOM     39  C   GLN H   6     -24.522   9.037  22.000  1.00  0.00           C
ATOM     40  O   GLN H   6     -25.031   8.210  21.246  1.00  0.00           O
ATOM     41  N   SER H   7     -25.246   9.732  22.904  1.00  0.00           N
```

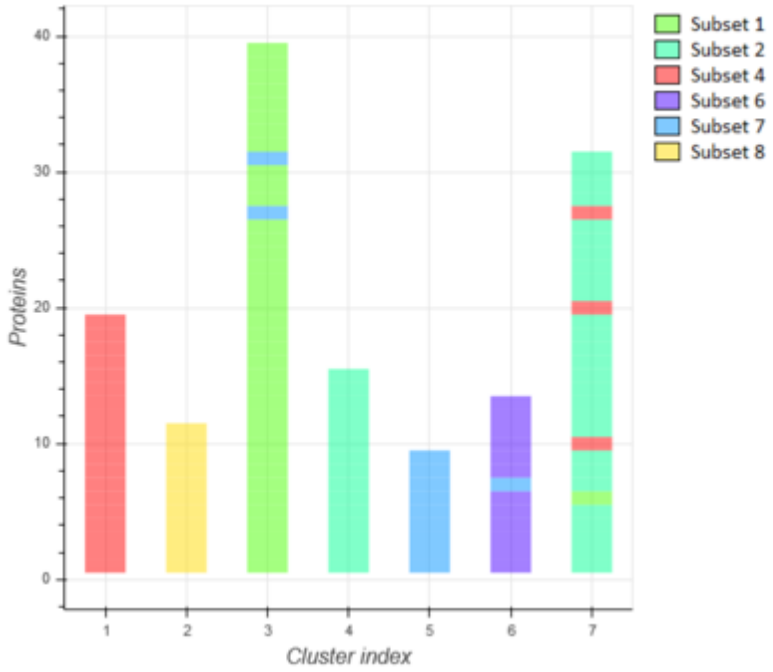
STRUCTURAL SIMILARITY



STRUCTURAL SIMILARITY

1st Approach : Clustering D1 using Single descriptors

A high degree of similarity was observed in all cases (85%-95% depending on the structure).



Descriptor	Rand Index	
TM-SCORE	0.854	Common approach in structural comparison
3DSC	0.879	Object recognition approach
FPFH	0.891	
RSD	0.895	
Combined	0.938	Combination of object recognition methodologies

STRUCTURAL SIMILARITY

Based on CATH (Protein Structure Classification database) to Validate homologues at least 2 out of 3 of the following conditions should be met:

Significant structural similarity

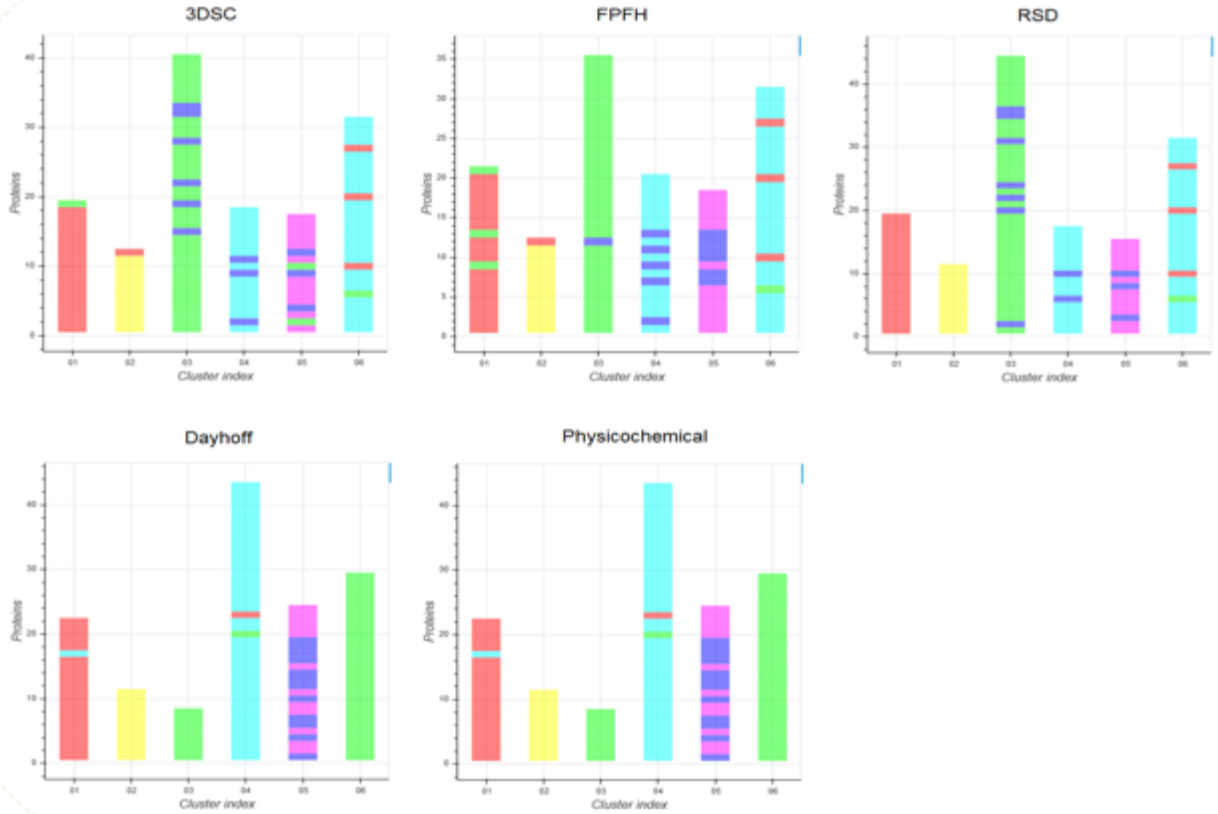
Significant sequence similarity

Functional similarity

To examine sequence similarity two new descriptors: Dayhoff & Physicochmical Similarity were examined

STRUCTURAL SIMILARITY

2nd Approach : Clustering based on 3D descriptors and Sequence Similarity



Descriptor	Rand Index
3DSC	0.879669
FPFH	0.890833
RSD	0.895234
Dayhoff	0.937849
Physicochemical	0.937849

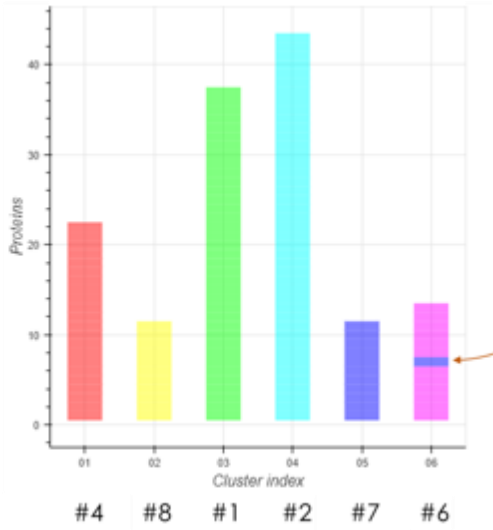
STRUCTURAL SIMILARITY

2nd Approach : Clustering based on 3D descriptors and Sequence Similarity

Clustering D1 using Descriptors Combination

B

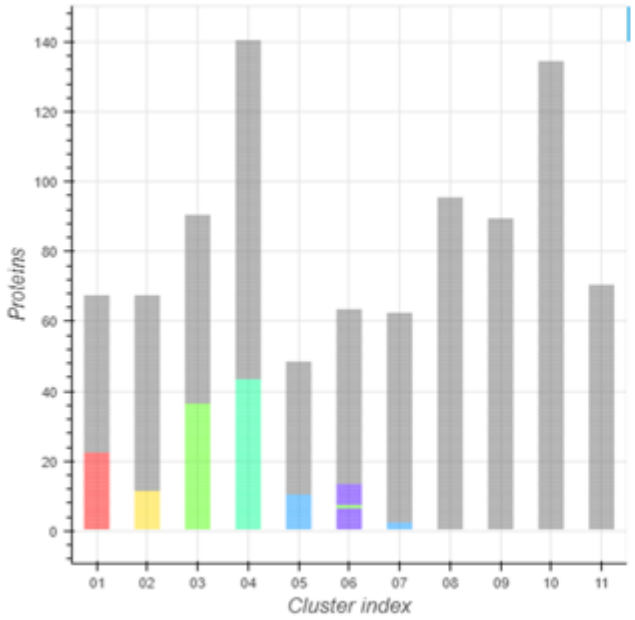
	3DSC	FPFH	RSD	Dayhoff	Physicochemical	Rand Index
6 clusters	0.3	0.5	0.1	0	0.1	0.99753
6 clusters	0.3	0.4	0.3	0	0	0.987334



- All models were grouped into clusters corresponding to individual stereotyped subsets
- The only exception was a single #7 case (GR-02-0494) that was grouped into the cluster of #6 cases

STRUCTURAL SIMILARITY

2nd Approach : Clustering based on 3D descriptors and Sequence Similarity



Clustering D2 using Descriptors Combination

	3dsc	fpfh	rsd	Dayhoff	Physicochemical	Rand Index
11 clusters	0.5	0.2	0.1	0.2	0	0.993

RESULTS

- Clusters were distinct ang highly homogeneous.
- This finding could be interpreted as important evidence that this novel approach enables robust clustering of patients with CLL.
- IG 3D models from cases with stereotyped subsets were clustered together strengthening the relevance of our findings.
- Altogether, clustering of CLL patients into discrete subgroups based on feature extraction from IG 3D models may have highly relevant implications for refined patient sub-classification.

Polychronidou *et al. BMC Bioinformatics* 2018, **19**(Suppl 14):414
https://doi.org/10.1186/s12859-018-2381-1

BMC Bioinformatics

METHODOLOGY Open Access

Automated shape-based clustering of 3D immunoglobulin protein structures in chronic lymphocytic leukemia

Eleftheria Polychronidou[1*†], Ilias Kalamaras[1†], Andreas Agathangelidis[2], Lesley-Ann Sutton[5], Xiao-Jie Yan[6], Vasilis Bikos[7], Anna Vardi[8], Konstantinos Mochament[1], Nicholas Chiorazzi[6], Chrysoula Belessi[9], Richard Rosenquist[5], Paolo Ghia[11], Kostas Stamatopoulos[2], Panayiotis Vlamos[10], Anna Chailyan[3], Nanna Overby[4], Paolo Marcatili[4], Anastasia Hatzidimitriou[2] and Dimitrios Tzovaras[1]

From 5th International Work-Conference on Bioinformatics and Biomedical Engineering
Granada, Spain. 26-28 April 2017

Polychronidou E, et. al.Automated shape-based clustering of 3D immunoglobulin protein structures in chronic lymphocytic leukemia. BMC Bioinformatics. 2018 Nov 20;19(Suppl 14):414. doi: 10.1186/s12859-018-2381-1. PMID: 30453883; PMCID: PMC6245605.

Alzheimer’s Disease

Alzheimer’s Disease is a folding disorder

- Irreversible progressive dementia with long prodromal stages and up to now there is a lack of effective pharmacotherapy options evaluation of the prediction algorithms
- While the clinical symptoms of the disease are defined by cognitive impairment, the causes leading to memory decline are strongly tied to deposits of misfolded protein aggregates.

OPEN ACCESS Freely available online PLoS COMPUTATIONAL BIOLOGY

In Silico Analysis of the Apolipoprotein E and the Amyloid β Peptide Interaction: Misfolding Induced by Frustration of the Salt Bridge Network

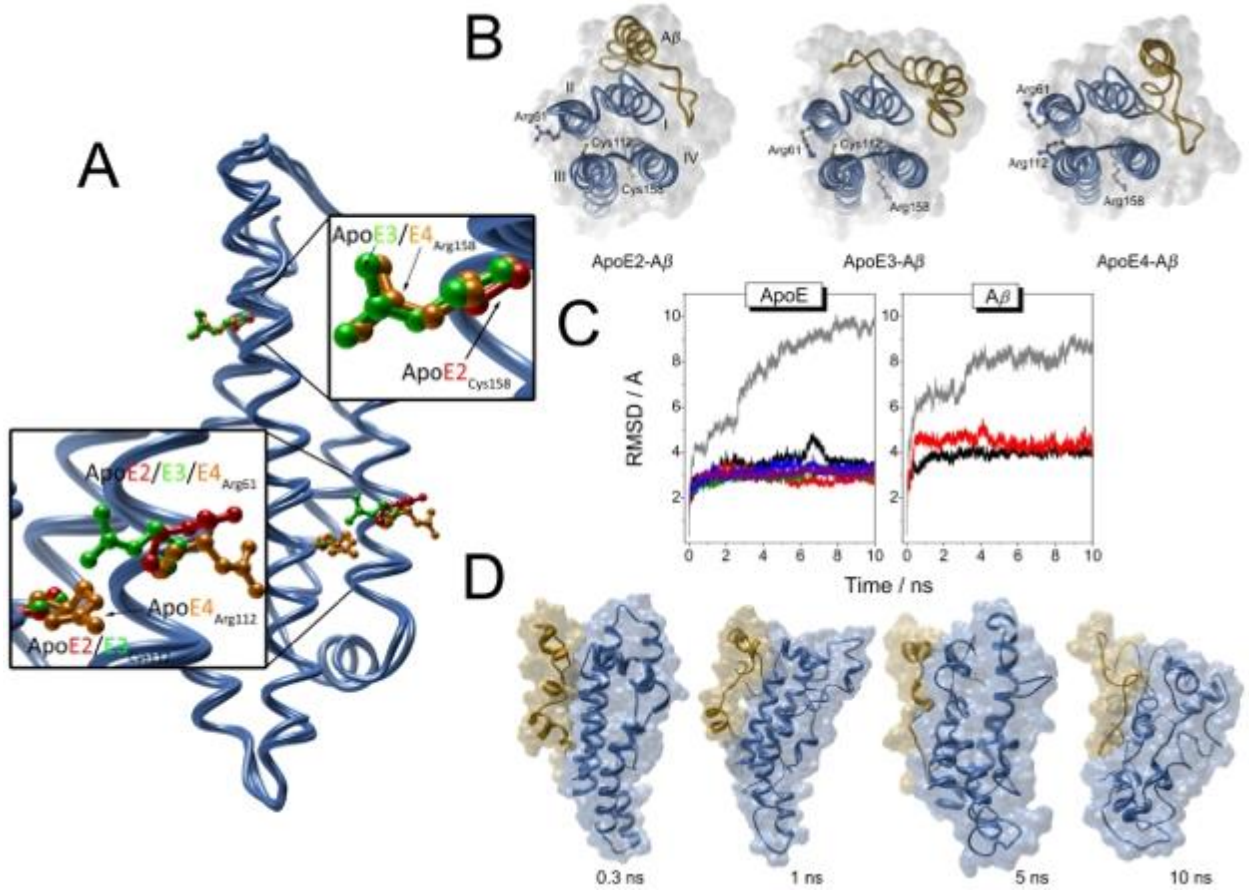
Jinghui Luo[1], Jean-Didier Maréchal[2], Sebastian Wärmländer[1], Astrid Gräslund[1], Alex Perálvarez-Marín[1,3*¤]

1 Department of Biochemistry and Biophysics, Stockholm University, Stockholm, Sweden, **2** Unitat de Química Física, Departament de Química, Universitat Autònoma de Barcelona, Bellaterra, Spain, **3** Unitat de Biofísica, Departament de Bioquímica i de Biologia Molecular i Centre d'Estudis en Biofísica, Universitat Autònoma de Barcelona, Bellaterra, Spain

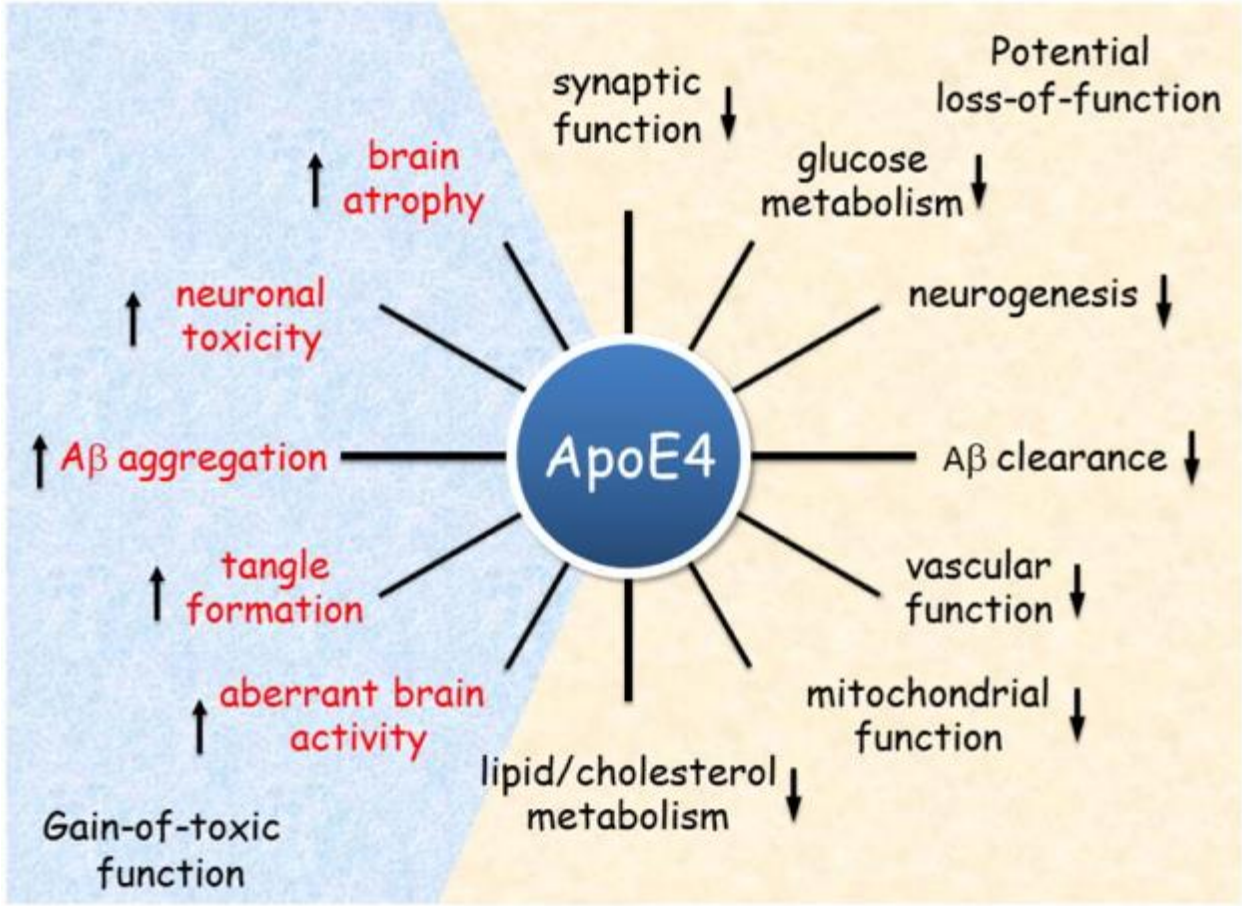
Abstract

The relationship between Apolipoprotein E (ApoE) and the aggregation processes of the amyloid β (Aβ) peptide has been shown to be crucial for Alzheimer's disease (AD). The presence of the ApoE4 isoform is considered to be a contributing risk factor for AD. However, the detailed molecular properties of ApoE4 interacting with the Aβ peptide are unknown, although various mechanisms have been proposed to explain the physiological and pathological role of this relationship. Here, computer simulations have been used to investigate the process of Aβ interaction with the N-terminal domain of the human ApoE isoforms (ApoE2, ApoE3 and ApoE4). Molecular docking combined with molecular dynamics simulations have been undertaken to determine the Aβ peptide binding sites and the relative stability of binding to each of the ApoE isoforms. Our results show that from the several ApoE isoforms investigated, only ApoE4 presents a misfolded intermediate when bound to Aβ. Moreover, the initial α-helix used as the Aβ peptide model structure also becomes unstructured due to the interaction with ApoE4. These structural changes appear to be related to a rearrangement of the salt bridge network in ApoE4, for which we propose a model. It seems plausible that ApoE4 in its partially unfolded state is incapable of performing the clearance of Aβ, thereby promoting amyloid forming processes. Hence, the proposed model can be used to identify potential drug binding sites in the ApoE4-Aβ complex, where the interaction between the two molecules can be inhibited.

Apolipoprotein E (ApoE) and the aggregation processes of the amyloid *β* (A*β*) peptide



Apolipoprotein E (ApoE) and the aggregation processes of the amyloid *β* (A*β*) peptide



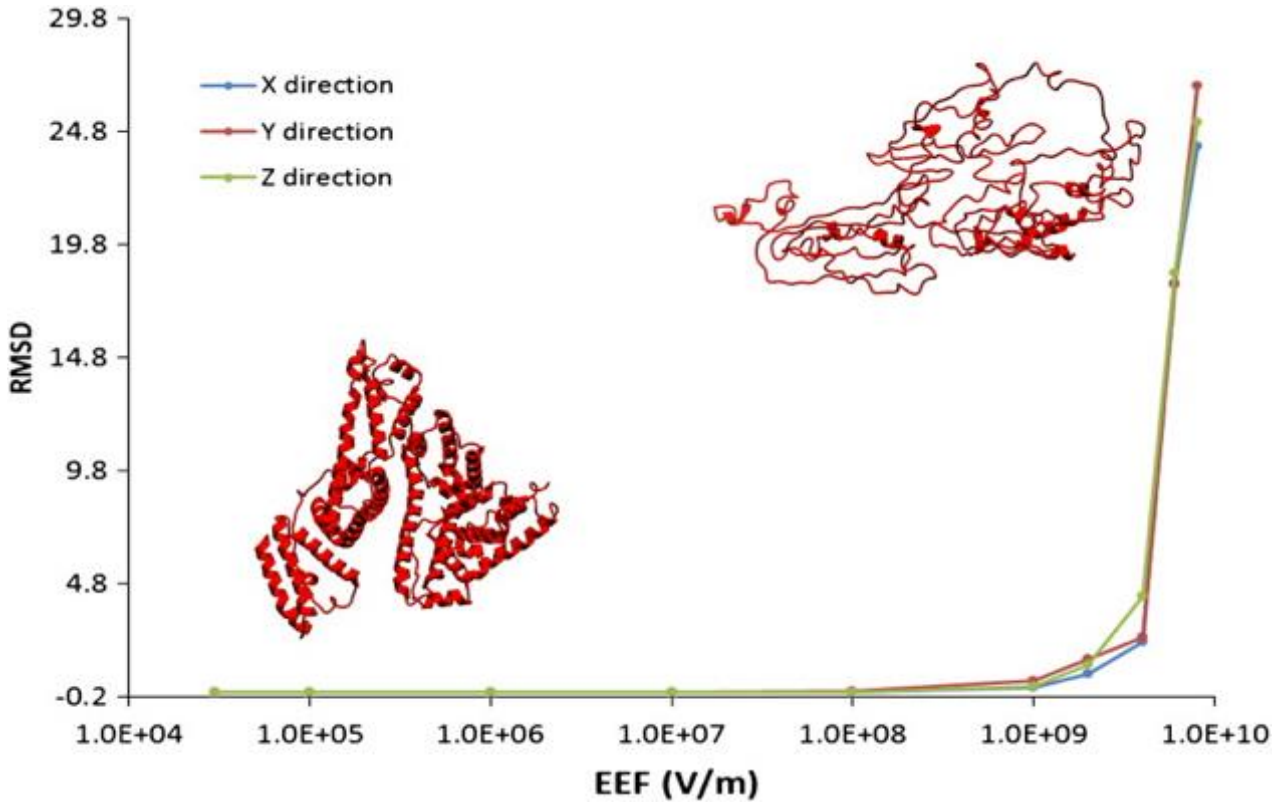
Nat Rev Neurol. 2013 Feb; 9(2): 106–118.

Hypothesis

What is the connection between the different isoforms of well known proteins implicated in the pathogenesis and progression of Alzheimer's Disease?

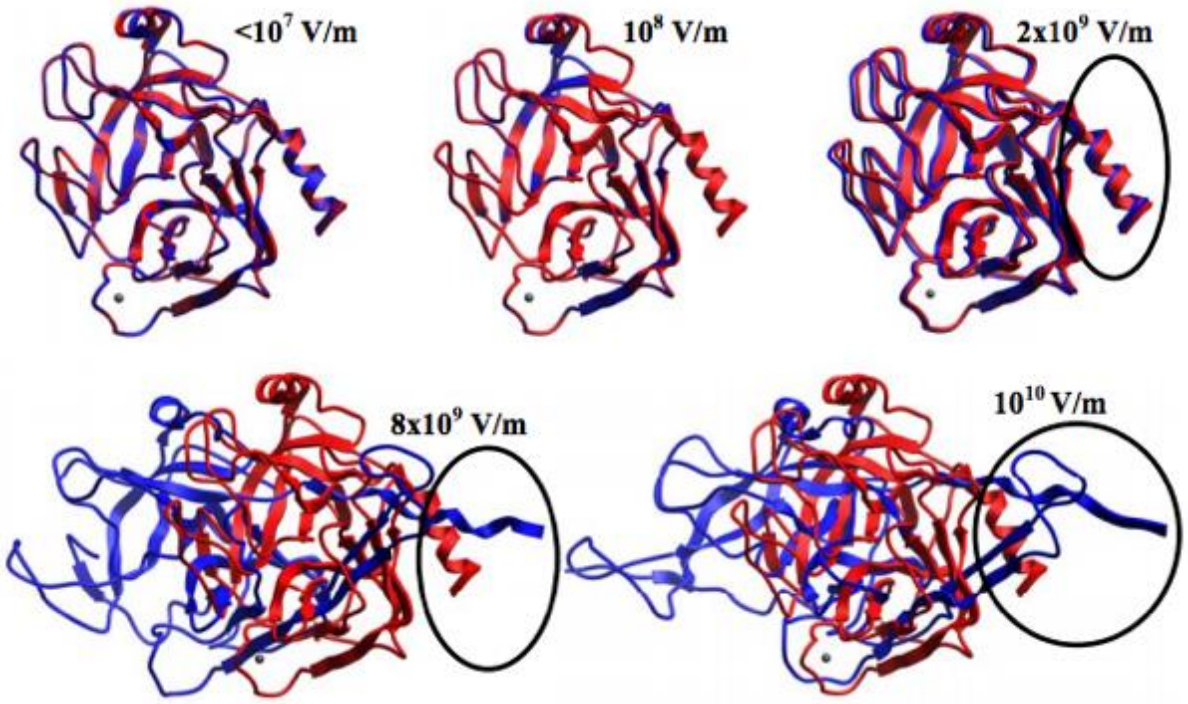
Could we find a correlation between their misfolded forms and the events occurring at the molecular level?

- **Computational simulation of the effect of an increasing electric field strength on the protein structure**



J Proteomics. 2012 Oct 22; 75(18-2): 5533–5543.

- **Computational simulation of the effect of an increasing electric field strength on the protein structure**



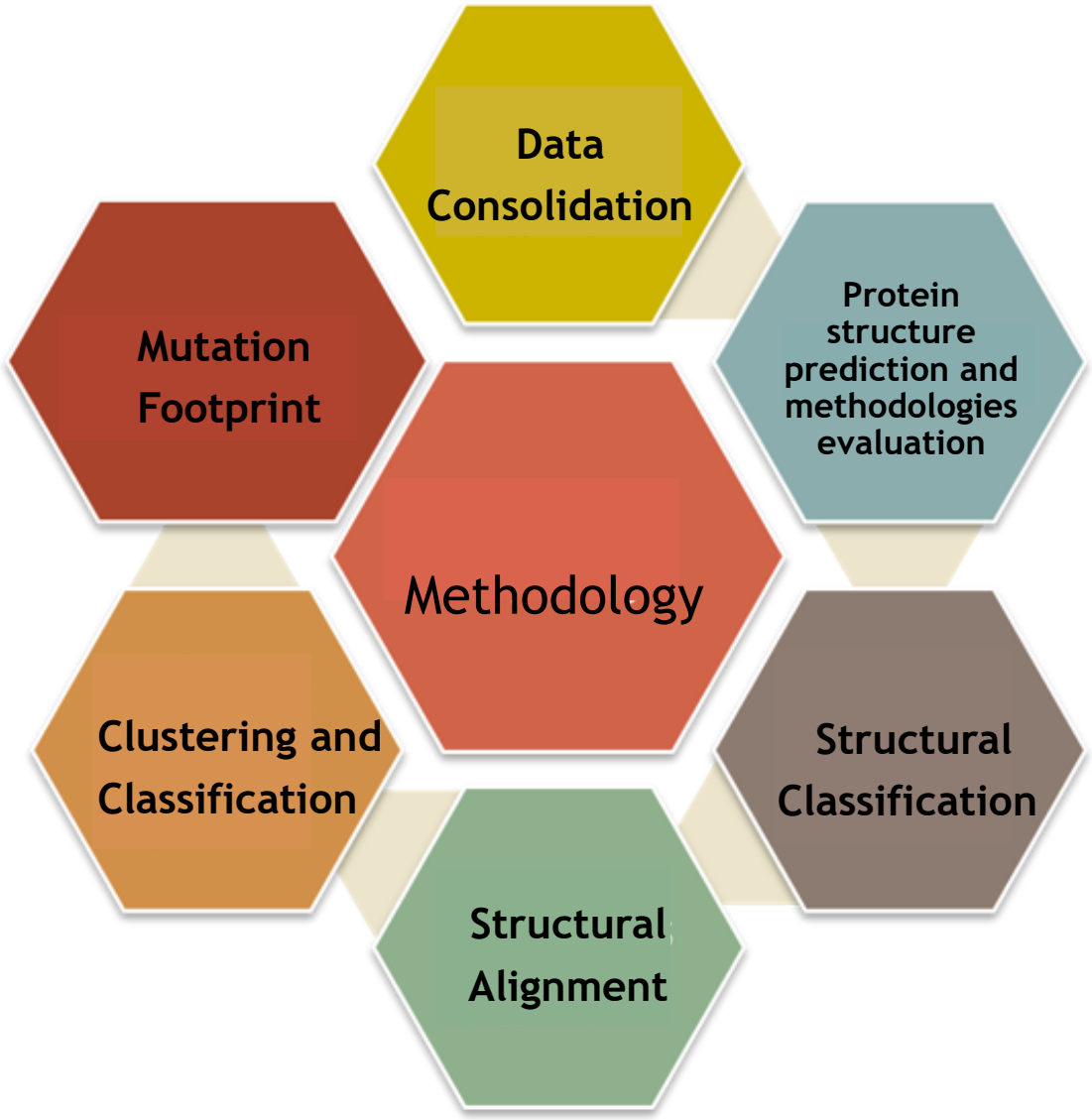
J Proteomics. 2012 Oct 22; 75(18-2): 5533–5543.

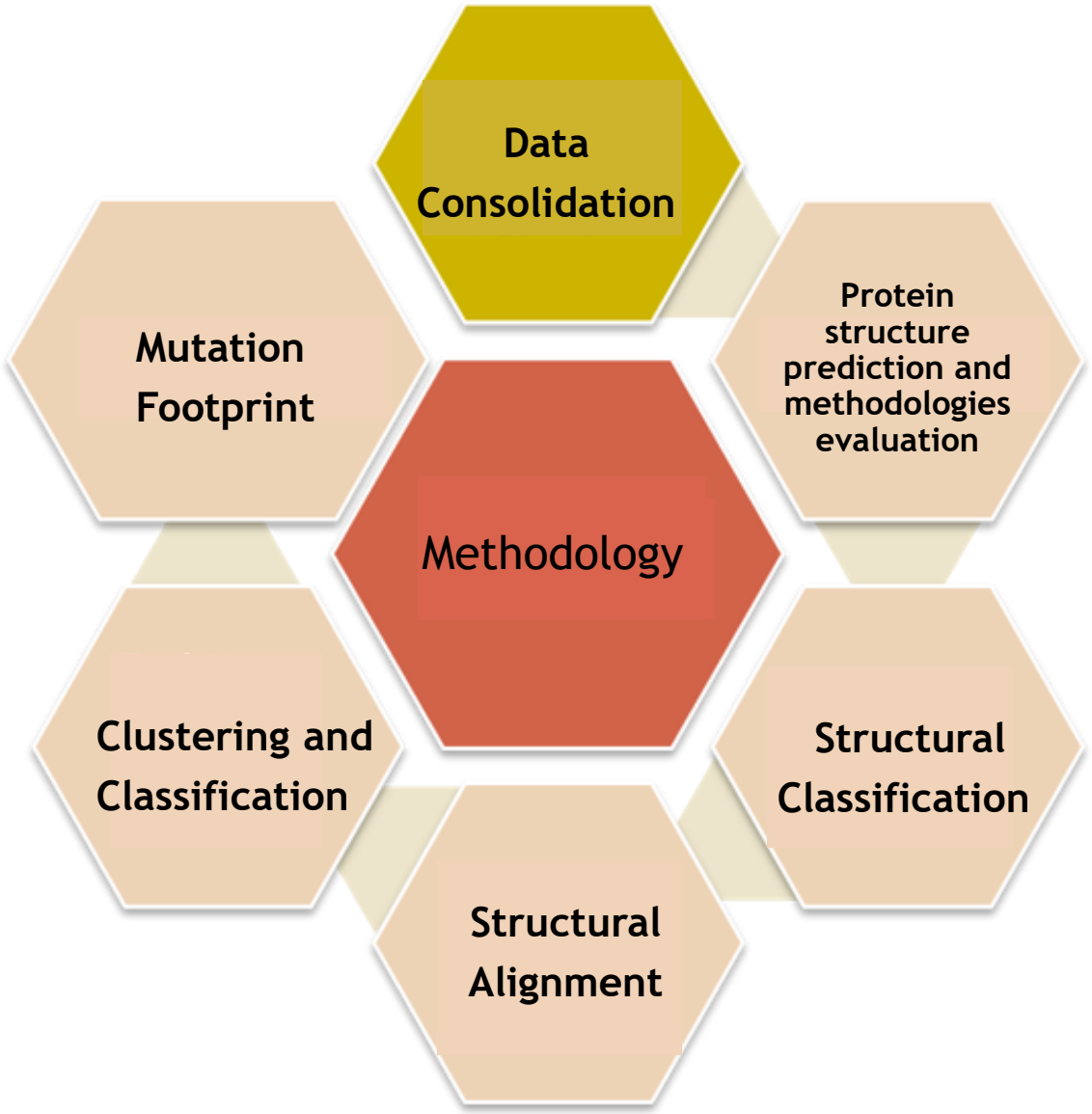
Methodology

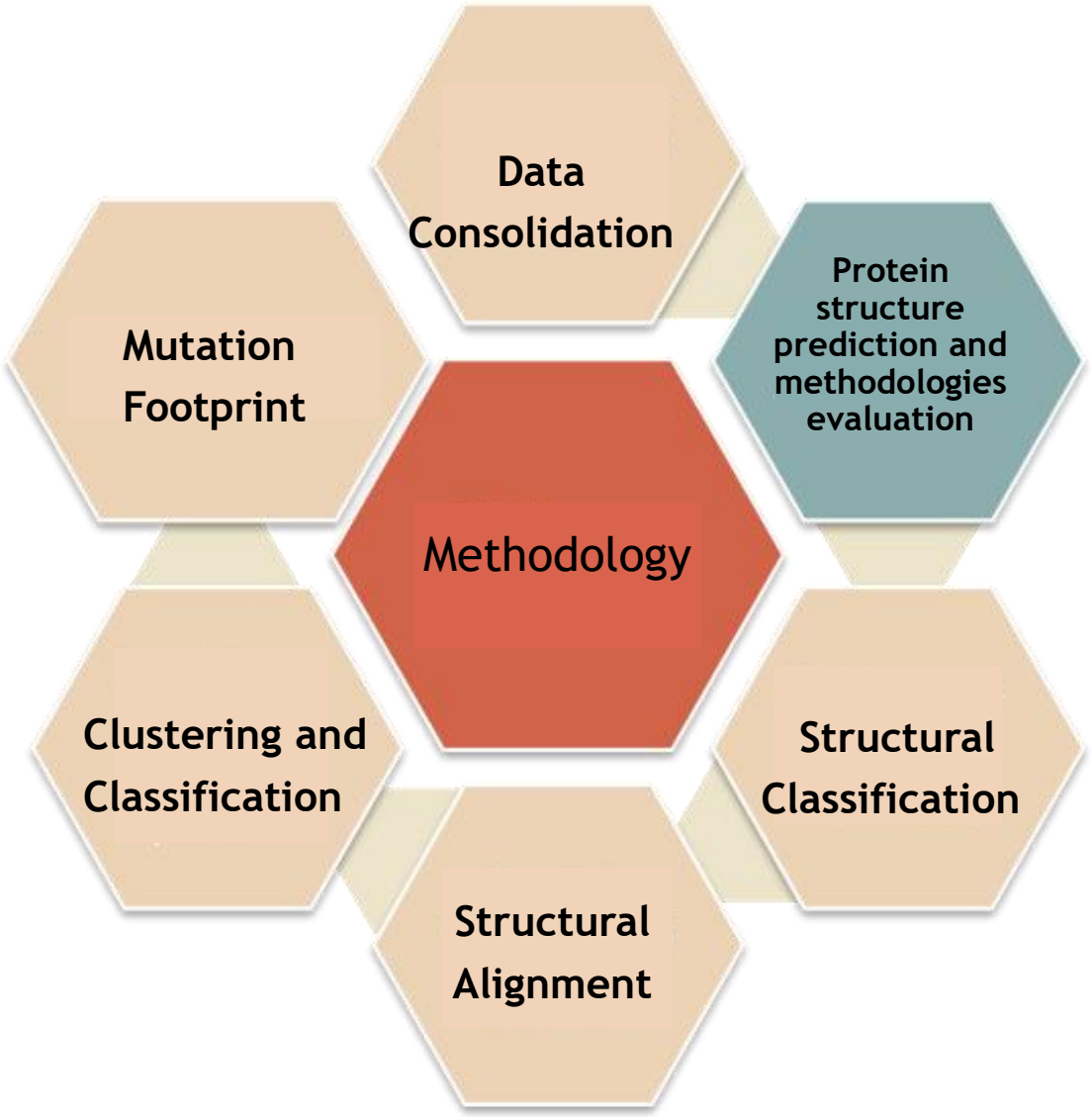
- study of the experimentally determined protein structures,
- evaluation of the prediction algorithms
- prediction of the AD related proteins with unknown structures
- evaluation of the mutation footprint of the established mutations in the tertiary structure.

Innovation

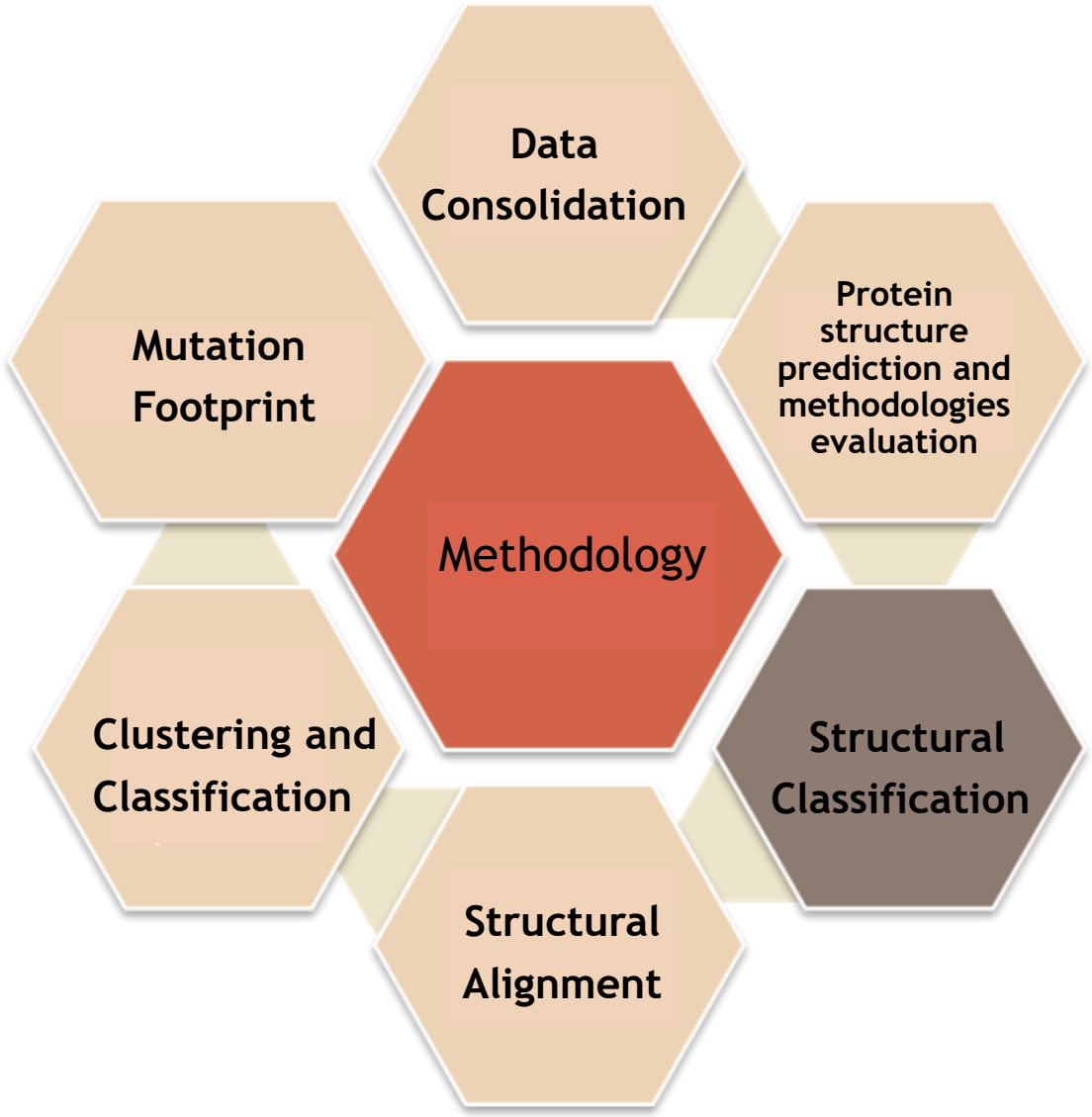
- ability to highlight broad ensembles of genes that drive key disease mechanisms
- These genes may hold strong potential for the identification of promising therapeutic targets.

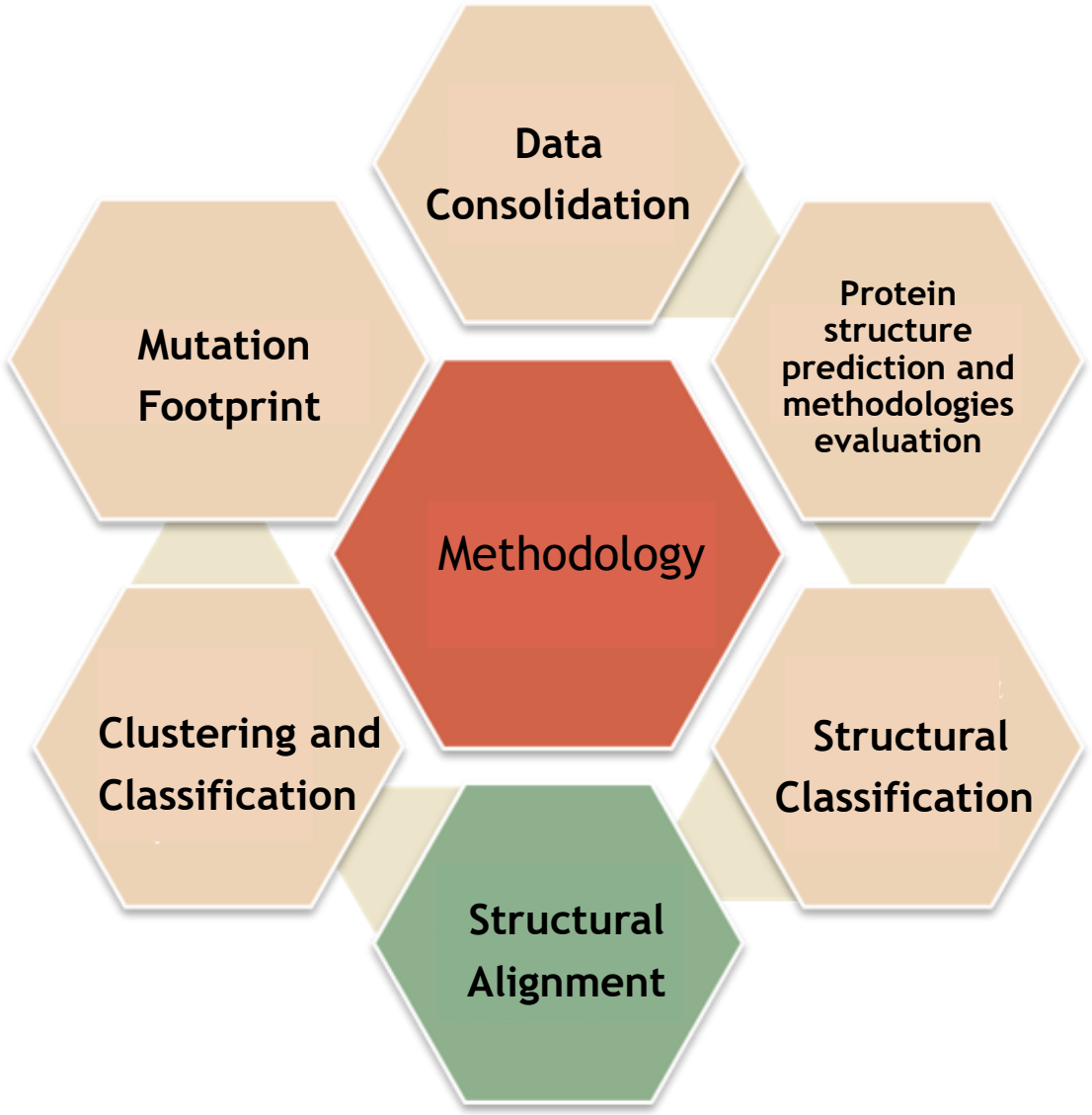

Data Consolidation
Protein structure prediction and methodologies evaluation
Mutation Footprint
Methodology
Clustering and Classification
Structural Classification
Structural Alignment

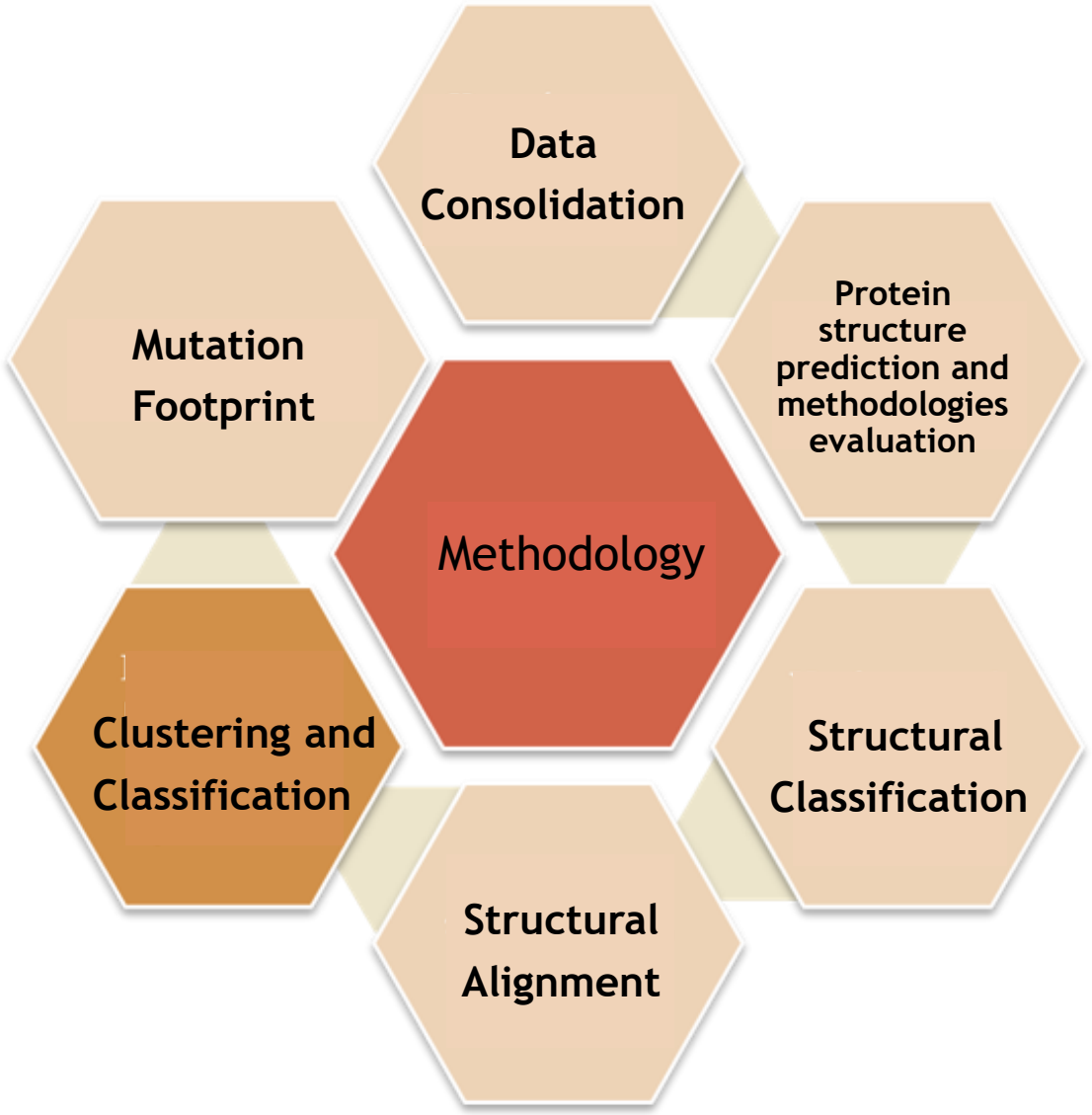

Data Consolidation
Protein structure prediction and methodologies evaluation
Mutation Footprint
Methodology
Clustering and Classification
Structural Classification
Structural Alignment

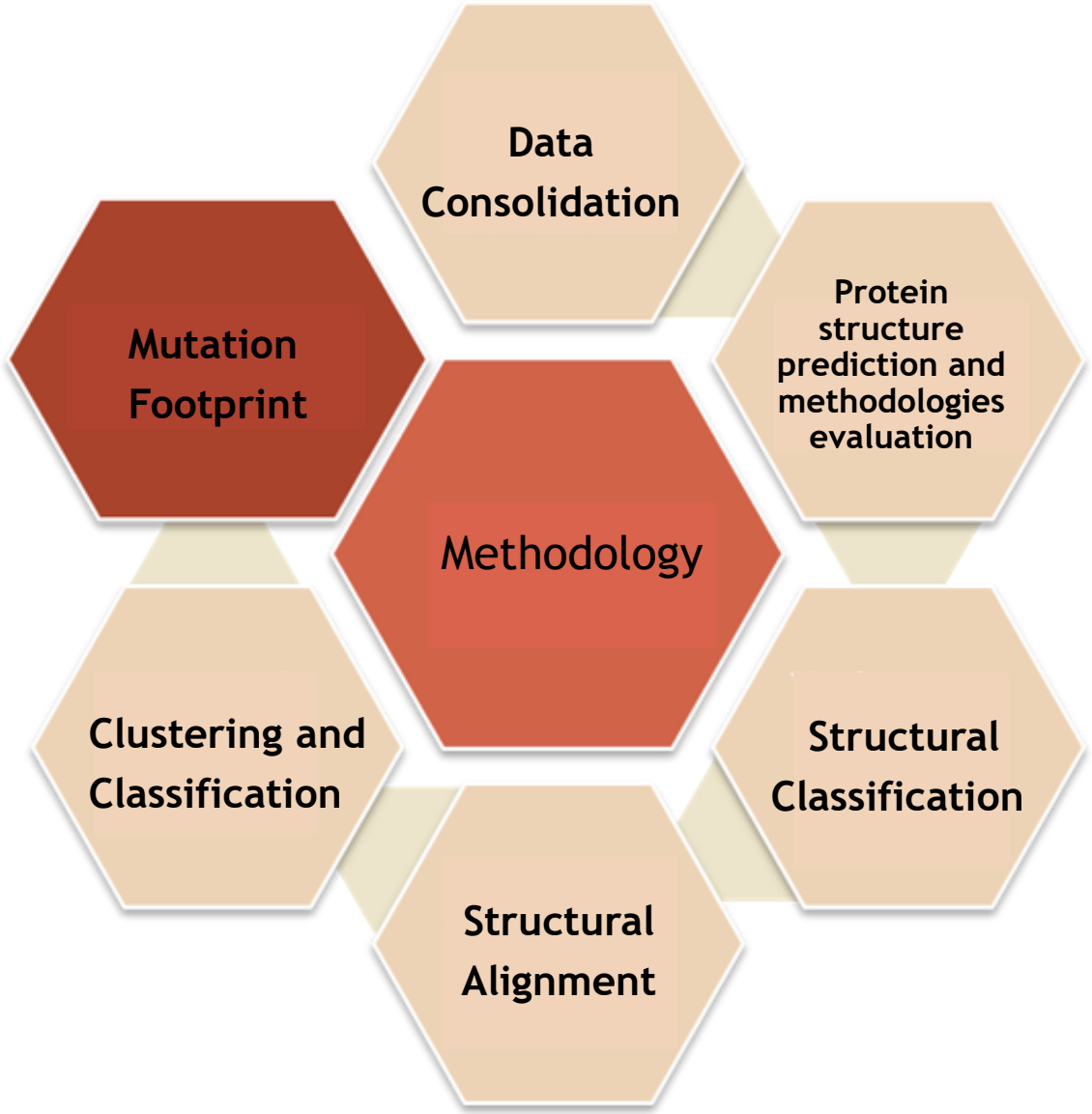

Data Consolidation
Protein structure prediction and methodologies evaluation
Mutation Footprint
Methodology
Clustering and Classification
Structural Classification
Structural Alignment

Software	Method	Description	Category
MODELLER	Satisfaction of spatial restraints	Standalone program mainly in Fortran and Python	homology
SWISS-MODEL	Local similarity/fragment assembly	Automated webserver (based on ProModII)	homology
HHpred	Template detection, alignment, 3D modeling	Interactive webserver with help facility	threading
I-TASSER	Combination of ab initio folding and threading methods	Structural and function predictions	Combination
ROBETTA	Rosetta homology modeling and ab initio fragment assembly with Ginzu domain prediction	Webserver	Ab initio
QUARK	Monte Carlo fragment assembly	On-line server for protein modeling (best for ab initio folding in CASP9)	Ab initio


Data Consolidation
Mutation Footprint
Protein structure prediction and methodologies evaluation
Methodology
Clustering and Classification
Structural Classification
Structural Alignment


Data Consolidation
Protein structure prediction and methodologies evaluation
Mutation Footprint
Methodology
Clustering and Classification
Structural Classification
Structural Alignment


Data Consolidation
Protein structure prediction and methodologies evaluation
Mutation Footprint
Methodology
Clustering and Classification
Structural Classification
Structural Alignment


Data Consolidation
Mutation Footprint
Protein structure prediction and methodologies evaluation
Methodology
Clustering and Classification
Structural Classification
Structural Alignment

STRUCTURAL SIMILARITY and MUTATION analysis IN Alzheimer’s disease

Presenilin 1 is a large multi-pass transmembrane protein which participates in the formation of γ-secretase complex and has a strong impact in the clinical course of the disease.

Summary of the 246 mutations of the Presenilin-1

>pdb|5a63|B

sp|P49768|PSN1_HUMAN Presenilin-1 OS=Homo sapiens OX=9606 GN=PSEN1 PE=1 SV=1

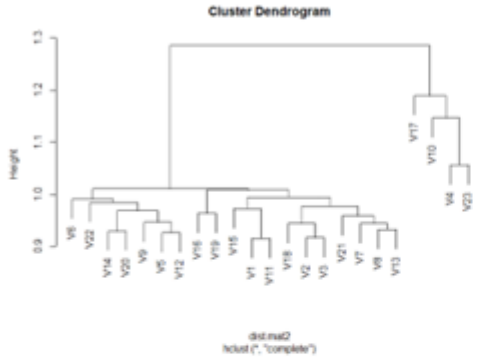
MTELPAPLSYFQNAQMSEDNHLSNTVRSQNDNRERQEHNDRRSLGHPEPLSNGRPQGN
SRQVVEQDEEEDEELTLKYGAKHVIMLFVPVTLCMVVVVATIKSVSFYTRKDGQLIYTPFTE
DTETVGQRALHSILNAAIMISVIVVMTILLVVLYKYRCYKVIHAWLIISSLLLLFFFSFIYLGEVF
KTYNVAVDYITVALLIWNFGVVGMISIHWKGPLRLQQAYLIMISALMALVFIKYLPEWTAWLI
LAVISVYDLVAVLCPKGPLRMLVETAQERNETLFPALIYSSTMVWLVNMAEGDPEAQRRVS
KNSKYNAESTERESQDTVAENDDGGFSEEWEAQRDSHLGPHRSTPESRAAVQELSSSILA
GEDPEERGVKLGLGDFIFYSVLVGKASATASGDWNTTIACFVAILIGLCLTLLLLAIFKKALP
ALPISITFGLVFYFATDYLVQPFMDQLAFHQFYI

*Blue : AD - Unclear Pathogenicity, *Green : AD - Not Pathogenic, *Red: AD - Pathogenic

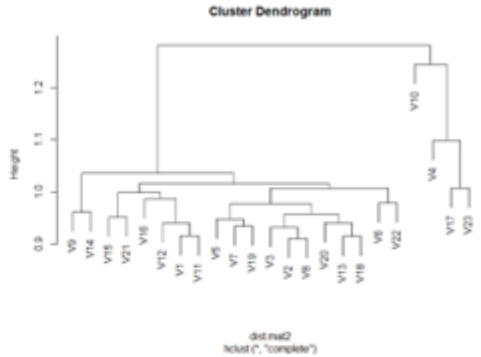
Mutations that are reported in PSEN1 protein. The mutations are grouped into three categories.

STRUCTURAL SIMILARITY and MUTATION analysis IN Alzheimer's disease

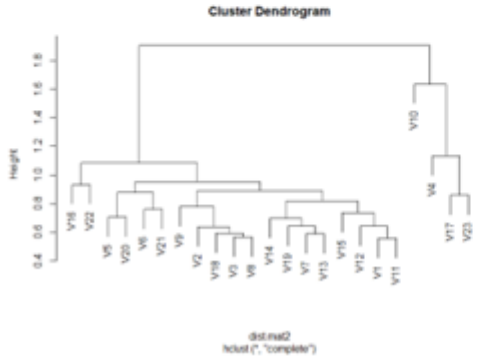
Dendrogram based on 3DSC Descriptors



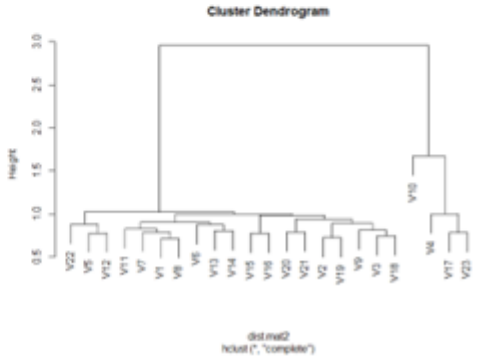
Dendrogram based on FPFH Descriptors

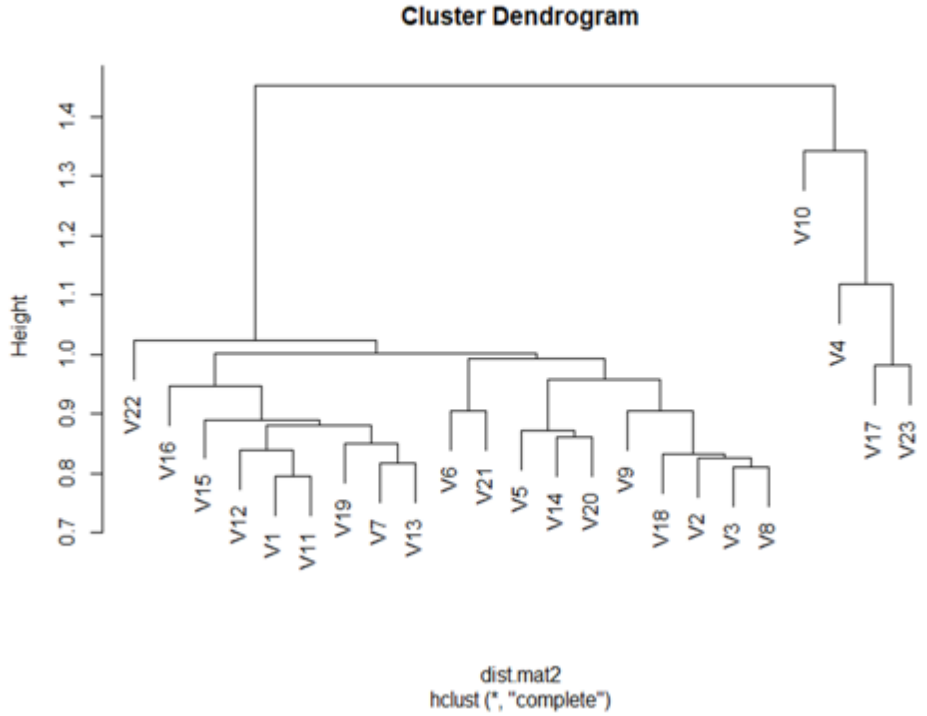


Dendrogram based on RSD Descriptors

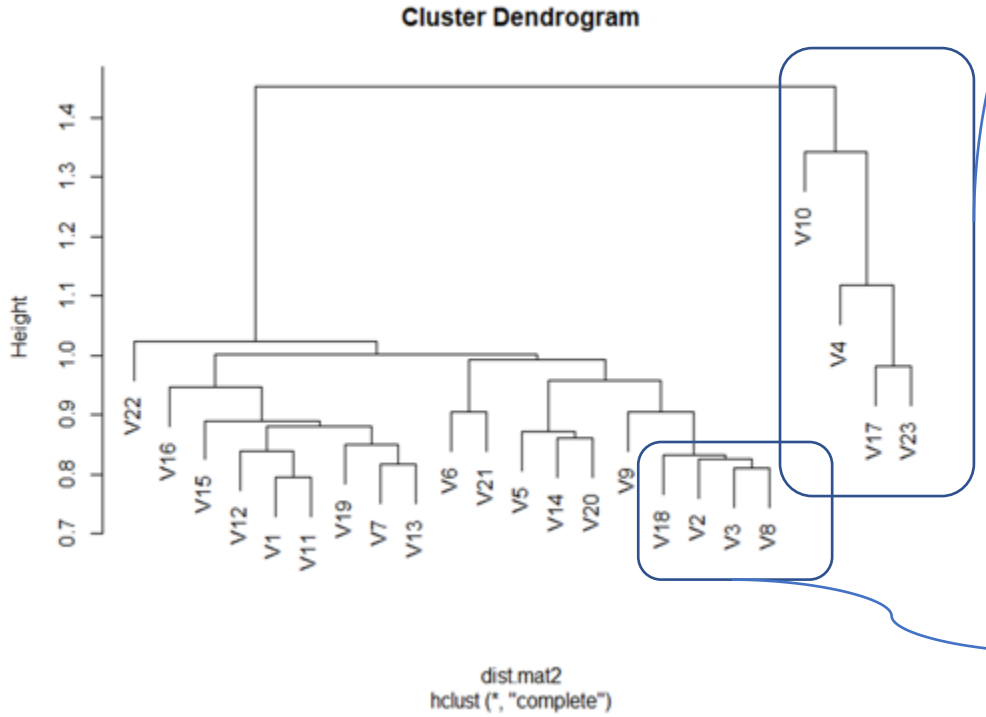


Dendrogram based on VFH Descriptors





STRUCTURAL SIMILARITY and MUTATION analysis IN Alzheimer's disease



Higher Percentage of structure alteration due to mutation

Mapping	Mutation	Pathogenicity
V17	G378V	AD : Pathogenic
V23	S390N	AD : Pathogenic
V4	I83_M84del	AD : Pathogenic

Mapping	Mutation	Pathogenicity
V8	E120K	AD : Pathogenic
V2	R35Q	AD : Unclear Pathogenicity
V3	I83T	AD : Pathogenic
V18	F386I	AD : Pathogenic

Insights for the pathogenicity of the R35Q mutation

Summary

- Risk prediction standpoint,
- networks exhibiting coexistent genetic variation and biological perturbation would represent prime targets in the development of personalized, burden-based genetic susceptibility tests.
- Therapeutic strategies development
- pathways and networks displaying multi-omics relationships in AD would reduce the search space for rational drug design and may highlight “hub” genes for therapeutic cocktail approaches, such as in the polypharmacy strategies successfully employed for AIDS and various cancers.

> Adv Exp Med Biol. 2020;1195:227-236. doi: 10.1007/978-3-030-32633-3_31.

Alzheimer's Disease: The Role of Mutations in Protein Folding

Eleftheria Polychronidou [1], Antigoni Avramouli [2], Panayiotis Vlamos [2]

Affiliations + expand

PMID: 32468481 DOI: 10.1007/978-3-030-32633-3_31

Abstract

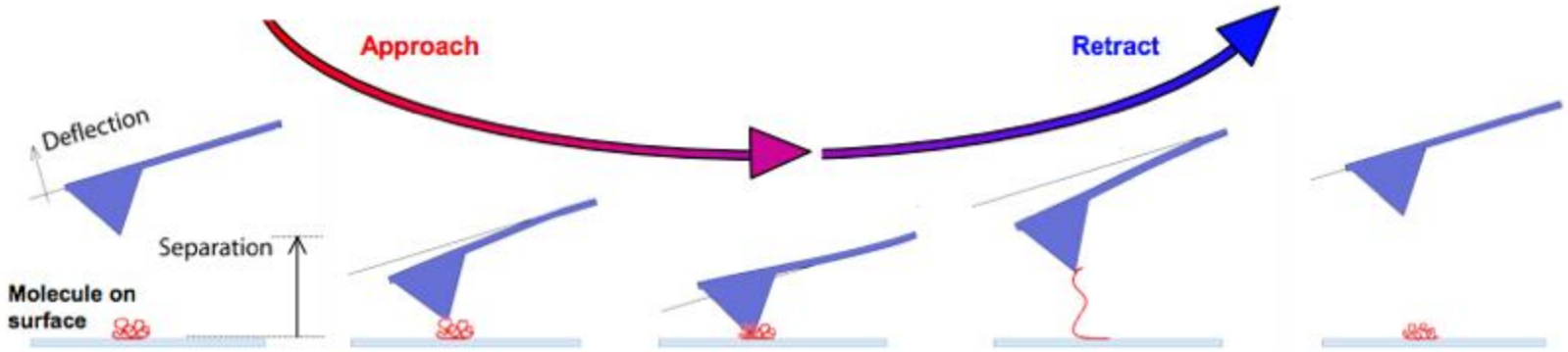
Misfolded proteins result when a protein follows the wrong folding pathway. Accumulation of misfolded proteins can cause disorders, known as amyloid diseases. Unfortunately, some of them are very common. The most prevalent one is Alzheimer's disease. Alzheimer's disease is a neurodegenerative disorder and the commonest form of dementia. The current study aims to assess the impact of somatic mutations in PSEN1 gene. The said mutations are the most common cause of familial Alzheimer's disease. As protein functionality can be affected by mutations, the study of possible alterations in the tertiary structure of proteins may reveal new insights related to the relationship between mutations and protein functions. To examine the effect of mutations, the

Polychronidou E, Avramouli A, Vlamos P. Alzheimer's Disease: The Role of Mutations in Protein Folding. Adv Exp Med Biol. 2020;1195:227-236. doi: 10.1007/978-3-030-32633-3_31. PMID: 32468481.

5. Experimental design for studying protein misfolding

JPK Force Robot

- Atomic force microscope is a very sensitive tool for quantitatively measuring forces on a single molecule level
    - Molecular stretching experiments



JPK Force Robot

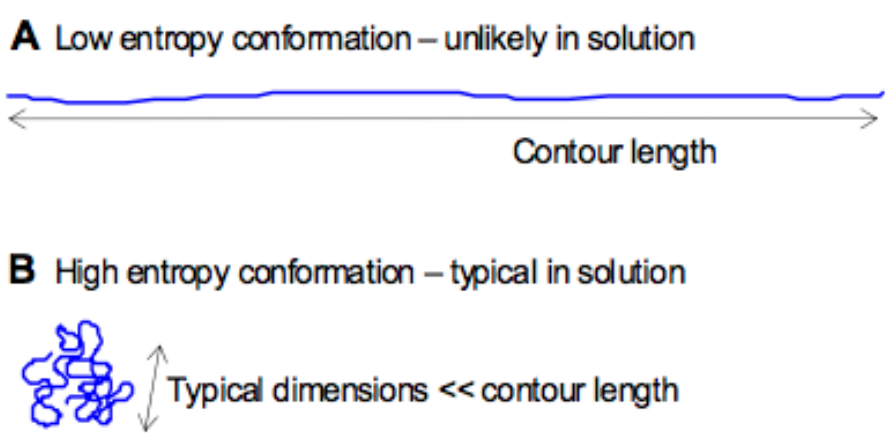


Fig. 2 Schematic diagram of two conformations of a long molecule.

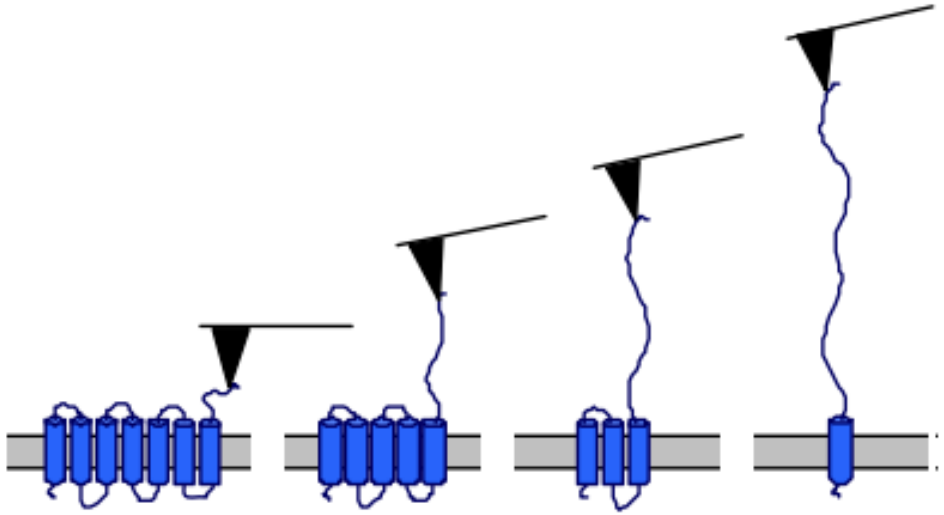


Fig. 7 Schematic diagram of bacteriorhodopsin unfolding, with intermediate states as the alpha-helices are pulled out in pairs.

Thank you!

Computer Science is no more about computers as astronomy is about telescopes.

There should be no such thing as boring Mathematics!

Edsger Dijkstra