

Databases

Alignment & structure classification

GOALS

1. Known structures
2. Structure comparison
3. Structure classification
4. Number of folds in nature
5. Sequences VS fold structures

1. Known structures

PDB

RCSB Protein Data Bank - RCSB PDB

www.pdb.org/pdb/home/home.do

An Information Portal to Biological Macromolecular Structures

As of Tuesday Feb 19, 2013 at 4 PM PST there are 68325 Structures | PDB Statistics

Search | All Categories: e.g., PDB ID, molecule name, author

Customize This Page

Available on the App Store

PDB-101

Structural View of Biology

Molecule of the Month

Proton-Gated Urea Channel

The acid in your stomach helps to digest food, but it also helps protect you from bacterial infection. However, one type of bacteria, *Helicobacter pylori*, is able to live in the acidic environment of the stomach. It is one of the most common bacterial infections, found worldwide in half of the population. It causes a continued inflammation of the stomach, which leads in some cases to stomach ulcers and stomach cancer.

Full Article

Protein Structure Initiative Featured System

Designer Proteins

The engineering of new proteins with novel structures and functions is one of the grand challenges facing the scientific community. This goal is particularly tempting, because we can look to nature to see thousands of working examples of proteins that spontaneously fold and perform diverse functions. By looking at natural proteins, scientists have discovered many of the features that are required to create a functional protein, and now, researchers at PSI have proven that these rules may be used for design.

Full Article | Archive | PSI Structural Biology Knowledgebase

Explore Archive

Organism

Exp. Method

Release Date

Enzyme Classification

Taxonomy

X-ray Resolution

Polymer Type

SCOP Classification

Show all

Latest Structures

4G4R : MutM containing F114A mutation bound to oxoG-containing DNA

Structural and biochemical analysis of DNA helix-invasion by the bacterial 8-oxoguanine DNA glycosylase MutM.

View in 3D (3mol)

check out all of the latest structures released.

New Structures

Latest Release

New Structure Papers

Search Unreleased Entries

New Features

Macromolecule Names: Synonyms

Latest features released:

Website Release Archive:

RCSB PDB News

Weekly | Quarterly | Yearly

2013-02-19

Build complex queries with Advanced Search

Advanced Search

Combine different searches to find structures and refine search results. New options include quick searches by experimental and/or molecule type, structure determination method, and intermediate connectivity (LINK records). more

Poster Prize Awarded at AsC

Author Profiles: Timeline

Display of a Researcher's Structures

News Reader Survey

Visit the Biomolecular Discovery Dome at Biophysical Society

Join the RCSB PDB Team

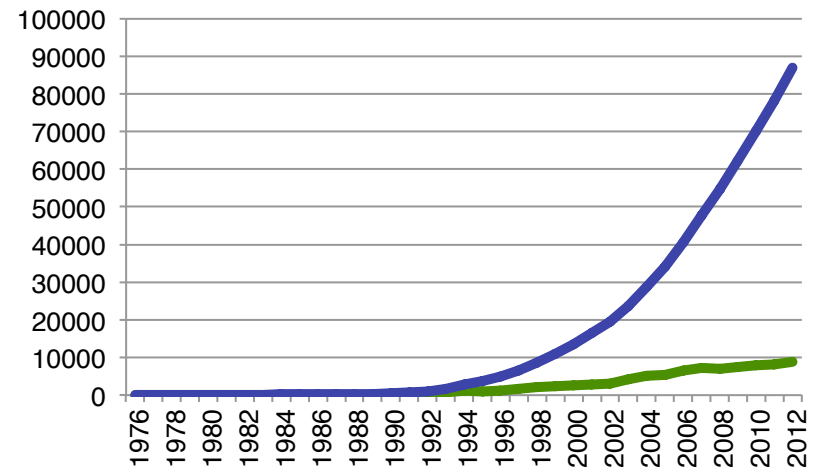
wwPDB News

X-ray Validation Task Force Report Published more

2012-12-21

Announcement: Remediation of Structure

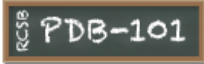

Yearly and total PDB structures per year



Yearly

Total

PDB search



A MEMBER OF

An Information Portal to Biological Information


As of **Tuesday Mar 05, 2013 at 4 PM PST** there are **88714** Structures

Search
Advanced
Browse

Everything Author Macromolecule Sequence Ligand ?

Search History (1), Previous Results (2)

Customize This Page

 Available on the App Store

PDB-101 Hide
Structural View of Biology
Understanding PDB Data
Molecule of the Month
Educational Resources
Author Profiles

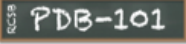
MyPDB Hide
Login to your Account
Register a New Account
Query Results (2)
Query History (1)

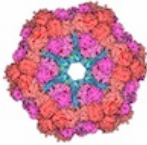
Biological Macromolecular Resource

Full Description

Learn: Featured Molecules

Structural View of Biology












Molecule of the Month
Erythrocrutorin

Hemoglobin comes in many shapes and sizes. In our blood, a **hemoglobin** with four chains carries oxygen from the lungs to cells throughout the body. Some plants build a single-chain hemoglobin to help protect sensitive nitrogen-fixing bacteria from oxygen, similar to the single chain **myoglobin** that stores oxygen in our muscle cells. Some bacteria also make simple forms of hemoglobin to help manage oxygen and other small molecules. Earthworms, however, are the champions when it comes to building huge hemoglobins. They, and a few other types of invertebrate animals, build enormous complexes of hemoglobin to carry their oxygen, termed erythrocrutors.

List View of Archive By: **Title** | **Date** | **Category**



Biological Energy

5

PDB search

Showing 1 - 2 of 2 Results

Results : 25 Page: 1 of 1

Filter: Check All

View: Detailed

Reports: Select one...

Sort: Relevance

☒ 2RO2



Solution structure of domain I of the negative polarity CChMVd hammerhead ribozyme

Authors: Gallego, J., Dufour, D., Gago, S., de la Pena, M., Flores, R.

Release: 2008-12-30

Classification: RNA

Experiment: SOLUTION NMR

Residue Count: 23

Compound: 1 Polymer [[Display Full Polymer Details](#) | [Display for All Results](#)]

Citation: **Structure-function analysis of the ribozymes of chrysanthemum chlorotic mottle viroid: a loop-loop interaction motif conserved in most natural hammerheads**
(2009) Nucleic Acids Res. **37**: 368-381 [[Display Full Abstract](#) | [Display for All Results](#)]

Search Hit: Title: Solution structure of domain I of the negative polarity CChMVd hammerhead ribozyme

☒ 2RPK



Solution Structure of Domain II of the Positive Polarity CChMVd Hammerhead Ribozyme

Authors: Gallego, J., Dufour, D., de la Pena, M., Gago, S., Flores, R.

Release: 2008-12-30

Classification: RNA

Experiment: SOLUTION NMR

Residue Count: 20

Compound: 1 Polymer [[Display Full Polymer Details](#) | [Display for All Results](#)]

Citation: **Structure-function analysis of the ribozymes of chrysanthemum chlorotic mottle viroid: a loop-loop interaction motif conserved in most natural hammerheads**
(2009) Nucleic Acids Res. **37**: 368-381 [[Display Full Abstract](#) | [Display for All Results](#)]

Search Hit: Title: Solution Structure of Domain II of the Positive Polarity CChMVd Hammerhead Ribozyme

Advanced search

Search

Advanced

Browse

Everything

Author

Macromolecule

Sequence

Ligand

?

e.g., PDB ID, molecule name, author

Search History , Previous Results

↑ PDB-101

Hide

Structural View of Biology

Understanding PDB Data

Molecule of the Month

Educational Resources

Author Profiles

↑ MyPDB

Hide

Login to your Account

Register a New Account

↑ Home

Hide

News & Publications

Usage/Reference Policies

Deposition Policies

Website FAQ

Deposition FAQ

Contact Us

About Us

Careers

External Links

Sitemap

New Website Features

↑ Deposition

Hide

All Deposit Services

Advanced Search Interface

Macromolecule Type

?

Search based on whether the structure contains chains of certain molecule types (e.g. protein vs. DNA)

Contains Protein

Yes

Contains DNA

No

Contains RNA

Yes

Contains DNA/RNA Hybrid

No

Result Count

AND

X-ray Resolution

?

Search by X-ray resolution (mmCIF item _refine.ls_d_res_high)

Between

0

and

3

Result Count

Add Search Criteria +

PDB comparison tool

RCSB PDB Protein Comparison Tool

Calculate pairwise sequence or structure alignments.

Compare the following two proteins ⓘ

ID 1:

Cytoplasmic dynein 1 heavy chain 1, seryl t-RNA synthetase chimera
KQQEVIADKQMSVKEDLDK...

ID 2:

Cytoplasmic dynein 1 heavy chain 1, seryl t-RNA synthetase chimera
KQQEVIADKQMSVKEDLDK...

--- Select Comparison Method --- ▾

--- Select Comparison Method ---

Pairwise Sequence Alignment

blast2seq
Smith-Waterman
Needleman-Wunsch

Pairwise Structure Alignment

jFATCAT - rigid
jFATCAT - flexible
jCE algorithm
jCE Circular Permutation
external server: FATCAT
external server: TM-Align
external server: TopMatch
external server: Dali

Compare

you can use the auto-suggest feature. It supports searching by

)

ptions)

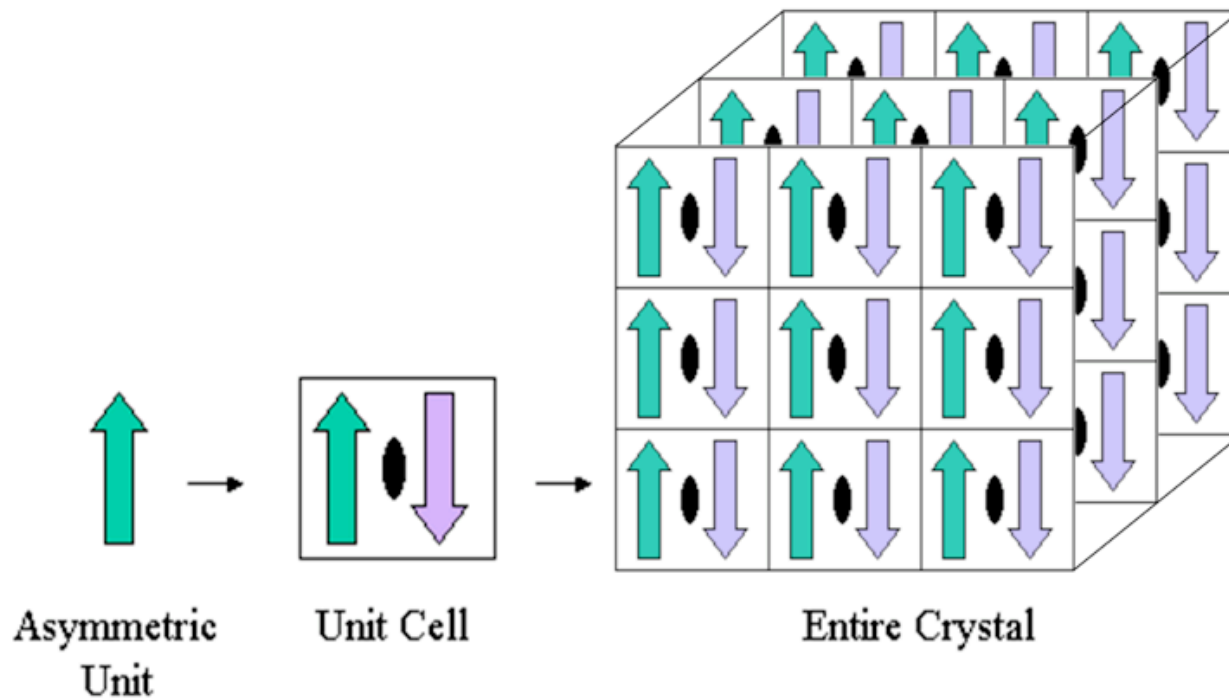
If you are having trouble using the tool, or if you are using the tool for Java Web Start applications, view our [troubleshooting Java Web Start page](#) for more help.

PDB format

<http://www.wwpdb.org/documentation/format33/v3.3.html>

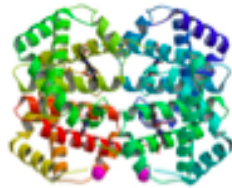
```
HEADER      MOTOR PROTEIN/STRUCTURAL PROTEIN          25-JUN-12   3J1U
TITLE       LOW AFFINITY DYNEIN MICROTUBULE BINDING DOMAIN - TUBULIN COMPLEX
COMPND      MOL_ID: 1;
COMPND      2 MOLECULE: CYTOPLASMIC DYNEIN 1 HEAVY CHAIN 1, SERYL T-RNA SYNTHETASE
COMPND      3 CHIMERA;
COMPND      4 CHAIN: A;
COMPND      5 FRAGMENT: SEE REMARK 999;
COMPND      6 SYNONYM: CYTOPLASMIC DYNEIN HEAVY CHAIN 1, DYNEIN HEAVY CHAIN,
COMPND      7 CYTOSOLIC;
COMPND      8 ENGINEERED: YES;
COMPND      9 MOL_ID: 2;
COMPND     10 MOLECULE: TUBULIN ALPHA-1B CHAIN;
COMPND     11 CHAIN: B;
COMPND     12 SYNONYM: ALPHA-TUBULIN UBIQUITOUS, TUBULIN K-ALPHA-1, TUBULIN ALPHA-
COMPND     13 UBIQUITOUS CHAIN;
COMPND     14 MOL_ID: 3;
COMPND     15 MOLECULE: TUBULIN BETA-2B CHAIN;
COMPND     16 CHAIN: C;
COMPND     17 SYNONYM: BETA TUBULIN
SOURCE      MOL_ID: 1;
SOURCE      2 ORGANISM_SCIENTIFIC: MUS MUSCULUS;
SOURCE      3 ORGANISM_COMMON: MOUSE;
SOURCE      4 ORGANISM_TAXID: 10090;
SOURCE      5 GENE: DYNC1H1, DHCL, DNCH1, DNCHCL, DYHC;
SOURCE      6 EXPRESSION_SYSTEM: ESCHERICHIA COLI;
SOURCE      7 EXPRESSION_SYSTEM_TAXID: 562;
SOURCE      8 MOL_ID: 2;
SOURCE      9 ORGANISM_SCIENTIFIC: BOS TAURUS;
SOURCE     10 ORGANISM_COMMON: BOVINE;
SOURCE     11 ORGANISM_TAXID: 9913;
SOURCE     12 MOL_ID: 3;
SOURCE     13 ORGANISM_SCIENTIFIC: BOS TAURUS;
SOURCE     14 ORGANISM_COMMON: BOVINE;
SOURCE     15 ORGANISM_TAXID: 9913
KEYWDS      MOTOR PROTEIN-STRUCTURAL PROTEIN COMPLEX
EXPDTA      ELECTRON MICROSCOPY
AUTHOR      W.B.REDWINE,R.HERNANDEZ-LOPEZ,S.ZOU,J.HUANG,S.L.RECK-PETERSON,
AUTHOR      2 A.E.LESCHZINER
```

Assymmetric Unit VS Biological Assembly



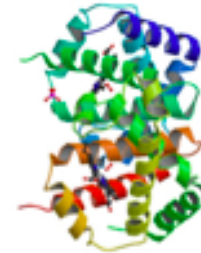
Assymmetric Unit VS Biological Assembly

Asymmetric unit with one biological assembly



Entry **2hhb** contains **one** hemoglobin molecule (**4 chains**) in the asymmetric unit.

Asymmetric unit with a portion of a biological assembly



Entry **1hho** contains **half** a hemoglobin molecule (**2 chains**) in the asymmetric unit. A crystallographic two-fold axis generates the other 2 chains of the hemoglobin molecule.

2. Structure comparison

Structure-Structure alignments

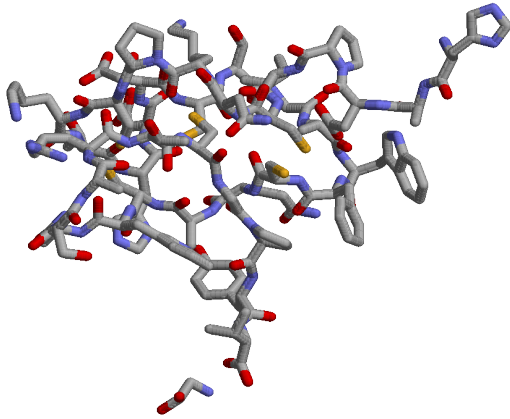
General steps in a bioinformatics procedure:

Representation

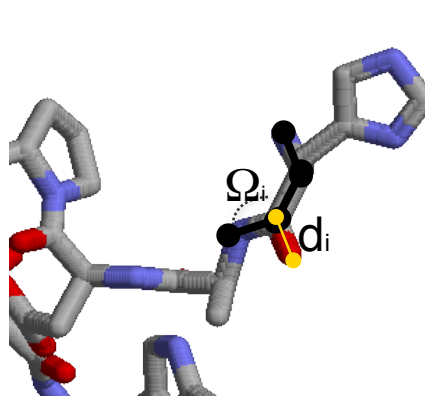
Scoring

Optimizer

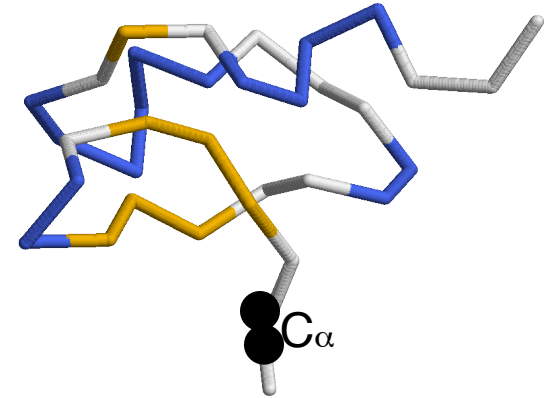
Representation Structures



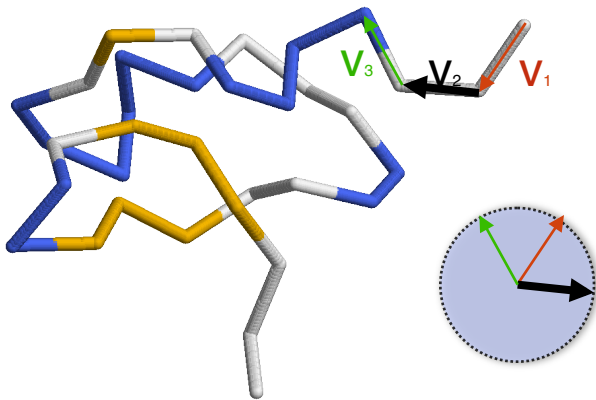
All atoms and coordinates



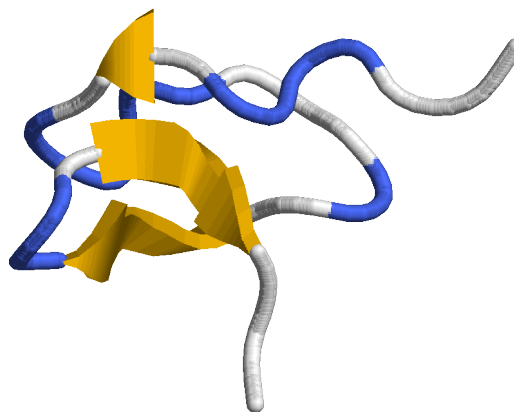
Dihedral space or distance space



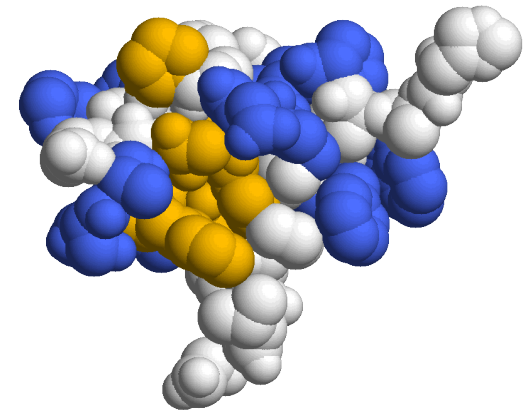
Reduced atom representation



Vector representation



Secondary Structure



Accessible surface (and others)

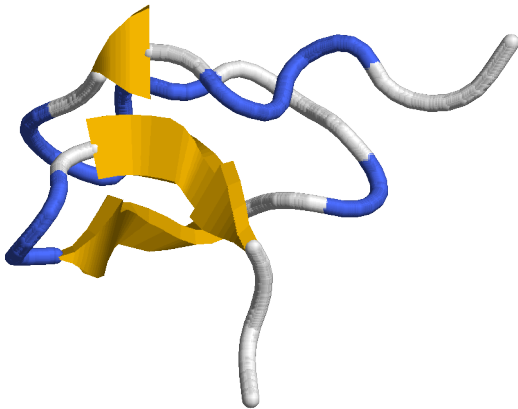
Raw scores

	C	S	T	P	A	G	N	D	E	Q	H	R	K	M	I	L	V	F	Y	W
C	9	-1	-1	-3	0	-3	-3	-3	-4	-3	-3	-3	-3	-1	-1	-1	-1	-2	-2	-2
S	-1	4	1	-1	1	0	1	0	0	-1	-1	0	-1	-2	-2	-2	-2	-2	-2	-3
T	-1	1	4	1	-1	1	0	1	0	0	-1	0	-1	-2	-2	-2	-2	-2	-2	-3
P	-3	-1	1	7	-1	-2	-1	-1	-1	-1	-2	-2	-1	-2	-3	-3	-2	-4	-3	-4
A	0	1	-1	-1	4	0	-1	-2	-1	-1	-2	-1	-1	-1	-1	-1	-2	-2	-2	-3
G	-3	0	1	-2	0	6	-2	-1	-2	-2	-2	-2	-3	-4	-4	0	-3	-3	-2	-2
N	-3	1	0	-2	-2	0	6	1	0	0	-1	0	0	-2	-3	-3	-3	-2	-4	-4
D	-3	0	1	-1	-2	-1	1	6	2	0	-1	-2	-1	-3	-3	-4	-3	-3	-3	-4
E	-4	0	0	-1	-1	-2	0	2	5	2	0	0	1	-2	-3	-3	-3	-2	-3	-3
Q	-3	0	0	-1	-1	-2	0	0	2	5	0	1	1	0	-3	-2	-2	-3	-1	-2
H	-3	-1	0	-2	-2	-2	1	1	0	0	8	0	-1	-2	-3	-3	-2	-1	2	-2
R	-3	-1	-1	-2	-1	-2	0	-2	0	1	0	5	2	-1	-3	-2	-3	-3	-2	-3
K	-3	0	0	-1	-1	-2	0	-1	1	1	-1	2	5	-1	-3	-2	-3	-3	-2	-3
M	-1	-1	-1	-2	-1	-3	-2	-3	-2	0	-2	-1	-1	5	1	2	-2	0	-1	-1
I	-1	-2	-2	-3	-1	-4	-3	-3	-3	-3	-3	-3	-3	1	4	2	1	0	-1	-3
L	-1	-2	-2	-3	-1	-4	-3	-4	-3	-2	-3	-2	-2	2	2	4	3	0	-1	-2
V	-1	-2	-2	-2	0	-3	-3	-3	-2	-2	-3	-3	-2	1	3	1	4	-1	-1	-3
F	-2	-2	-2	-4	-2	-3	-3	-3	-3	-3	-1	-3	-3	0	0	0	-1	6	3	1
Y	-2	-2	-2	-3	-2	-3	-2	-3	-2	-1	2	-2	-2	-1	-1	-1	-1	3	7	2
W	-2	-3	-3	-4	-3	-2	-4	-4	-3	-2	-2	-3	-3	-1	-3	-2	-3	1	2	11

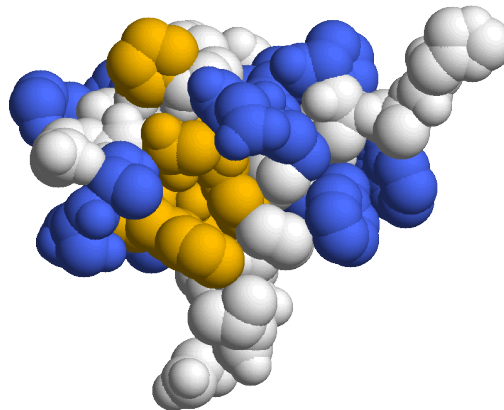
Aminoacid substitutions

$$RMSD(x, y) = \sqrt{\left(\frac{1}{N}\right) \sum_{i=1}^N (\|x(i) - y(i)\|^2)}$$

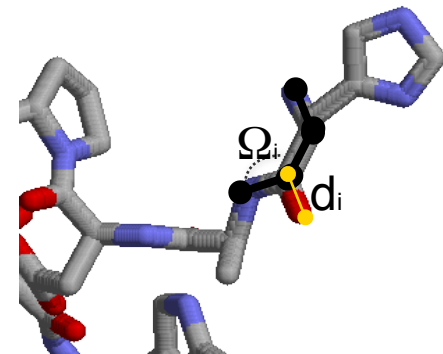
Root Mean Square Deviation



Secondary Structure (H,B,C)



Accessible surface (B,A [%])

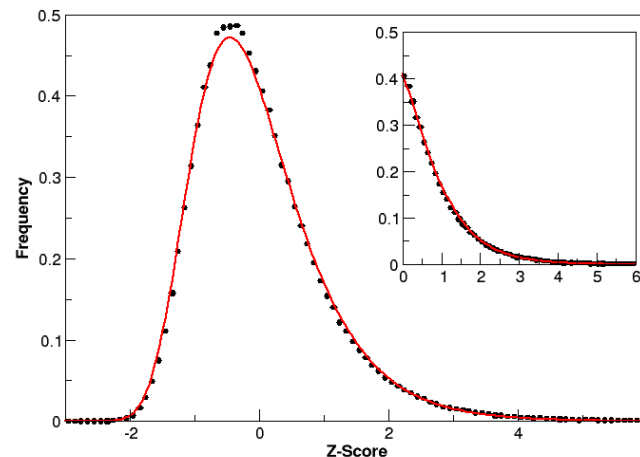
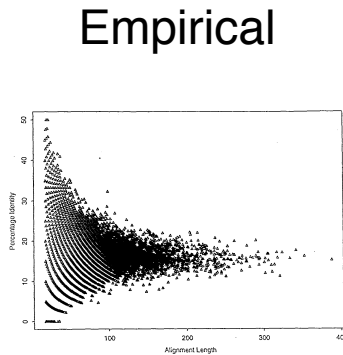


Angles or distances

Scoring

Significance of an alignment (score)

Probability that the optimal alignment of two random sequences/structures of the same length and composition as the aligned sequences/structures have at least as good a score as the evaluated alignment.



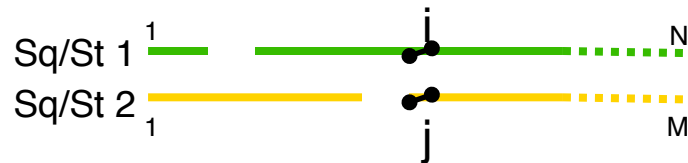
Sometimes approximated by Z-score (normal distribution).

Analytic

$$P(s) = e^{-\lambda (s-\mu)}$$

$$P(s \geq x) = 1 - \exp\left(-e^{-\lambda (s-\mu)}\right)$$

Global dynamic programming alignment



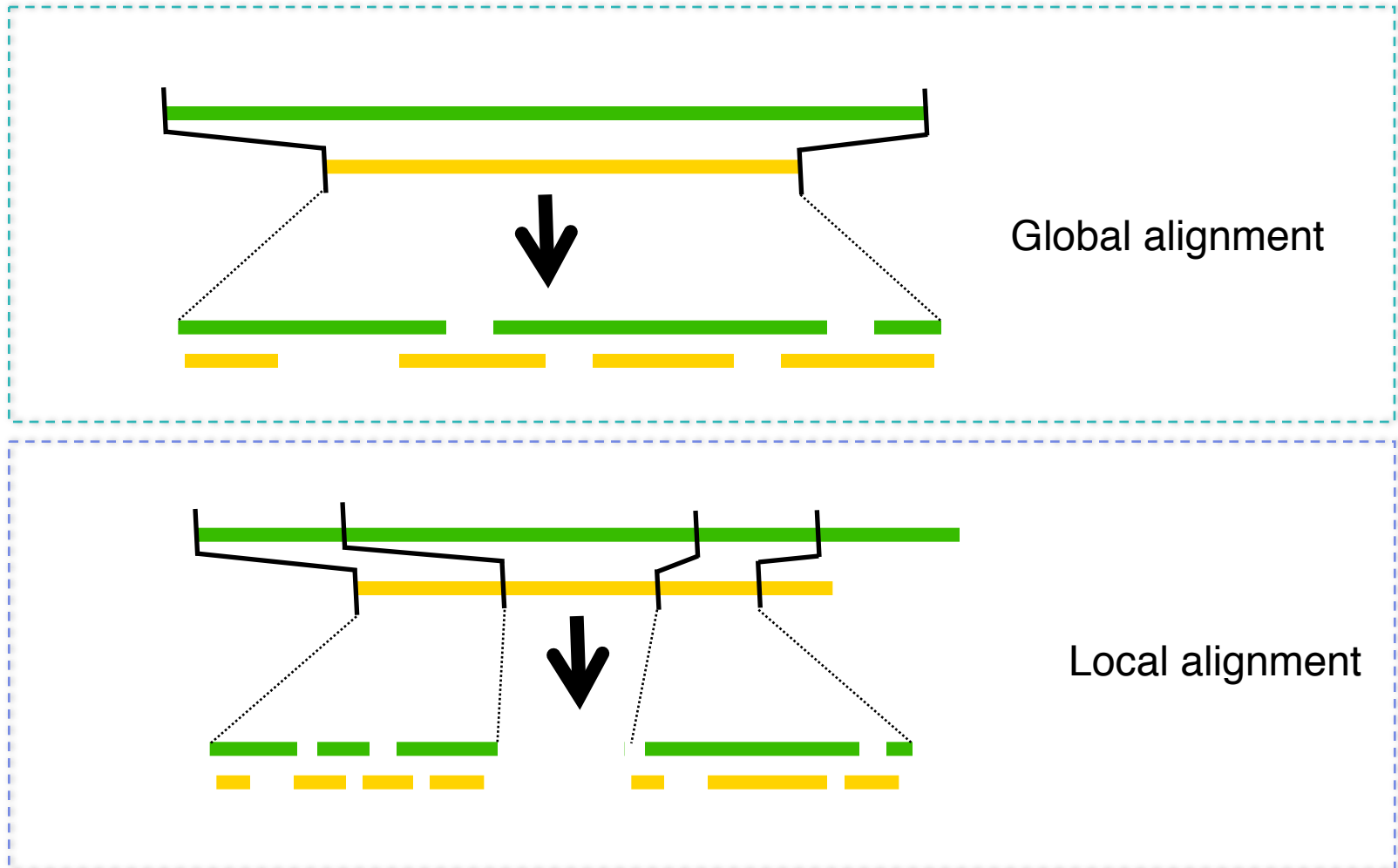
	1	2	3	...	N
1	*	*	*	*	*
2	*	*	*	*	*
3	*	*	*	*	*
...					
M					*

$$D_{i,j} = \min \begin{cases} D_{i,j-1} + \text{Score}_{(\Delta, \eta)} \\ D_{i-1,j-1} + \text{Score}_{(\eta, \eta)} \\ D_{i-1,j} + \text{Score}_{(\eta, \Delta)} \end{cases}$$

Best alignment score

Backtracking to get the best alignment

Global .vs. local alignment



Multiple alignment

Pairwise alignments

Example – 4 sequences A, B, C, D.



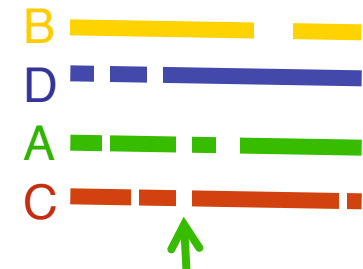
6 pairwise comparisons
then cluster analysis

Multiple alignments

Following the tree from step 1



Align B-D with A-C

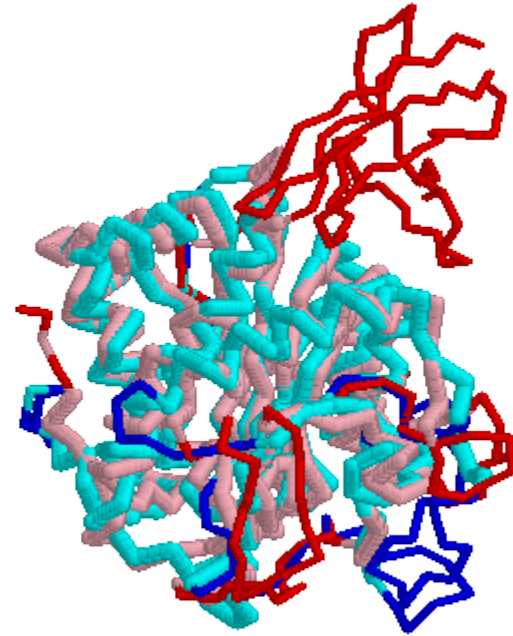


New gap in A-C to optimize
its alignment with B-D

Coverage .vs. Accuracy



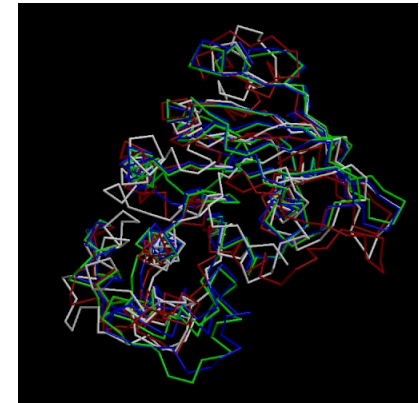
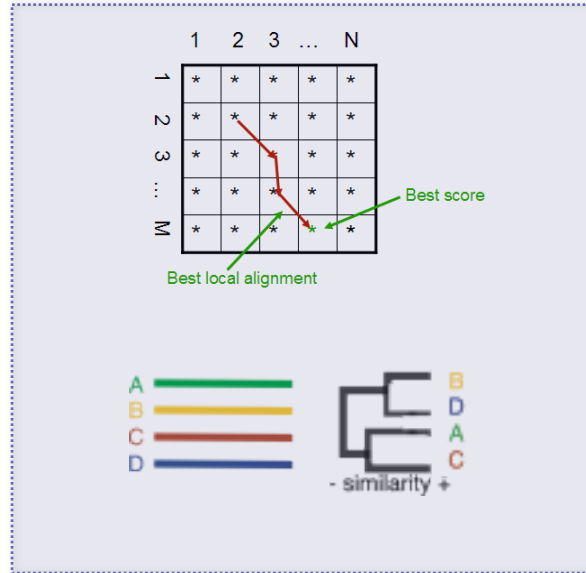
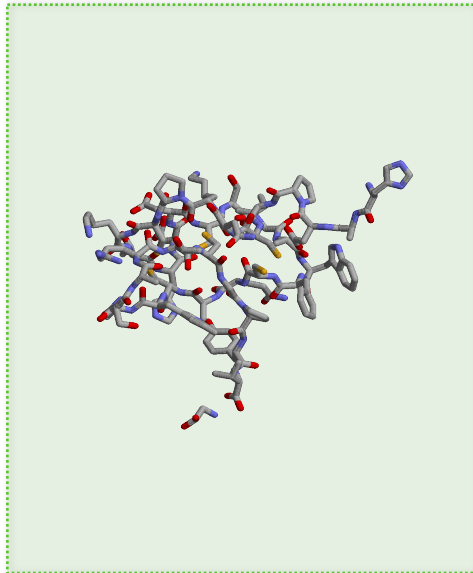
Coverage ~90% C α



Coverage ~75% C α

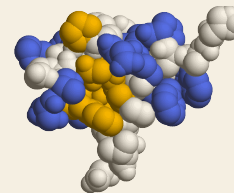
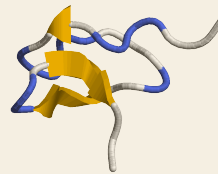
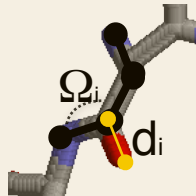
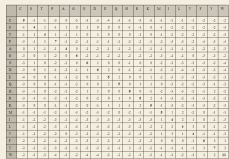
Same RMSD ~ 2.5Å

Structural alignment by properties conservation (SALIGN-MODELLER)



- ✓ Uses all available structural information
- ✓ Provides the optimal alignment

Computationally expensive



$$RMSD(x, y) = \sqrt{\left(\frac{1}{N}\right) \sum_{i=1}^N (\|x(i) - y(i)\|^2)}$$

$R_{i,j}$

$D_{i(3),j(3)}$


$S_{i,j}$

$B_{i,j}$

$I_{i,j}$

Structural alignment by properties conservation (SALIGN-MODELLER)

<http://salilab.org/salign>

 **SALIGN Webserver**

[Sali Lab Home](#) • [ModWeb](#) • [ModBase](#) • [ModEval](#) • [PCSS](#) • [FoXS](#) • [IMP](#) • [MultiFit](#) • [ModPipe](#) •

[SALIGN Home](#) • [SALIGN Help](#) • [SALIGN Examples](#) • [SALIGN Contact](#)

SALIGN: A multiple protein sequence/structure alignment server.

Developers:
Hannes Braberg
M.S. Madhusudhan
Ursula Pieper
Ben Webb
Elina Tjioe
Andrej Sali

SALIGN is a general alignment module of the modeling program [MODELLER](#). The alignments are computed using dynamic programming, making use of several features of the protein sequences and structures. SALIGN benchmarks from published papers are [also available](#).

General information

Email address 

Input alignment information

Users can either upload their own sequences/structures to align or choose structures from the PDB.

Paste one sequence at a time, without header 

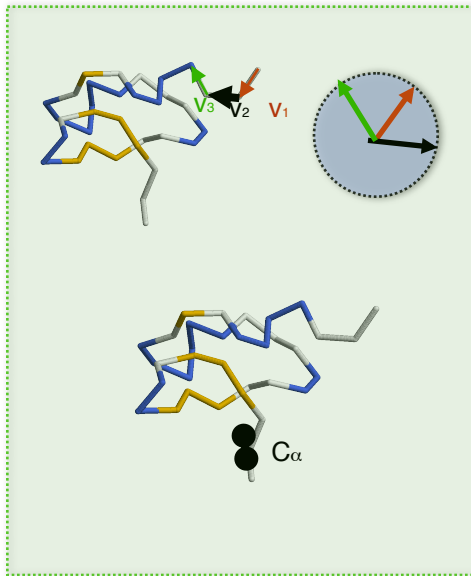
Upload 

Upload sequence/PDB file(s) 

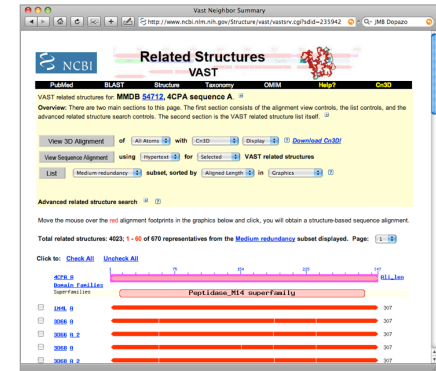
 Choose File No file chosen

Upload 

Vector Alignment Search Tool (VAST)

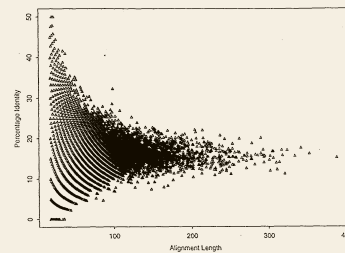
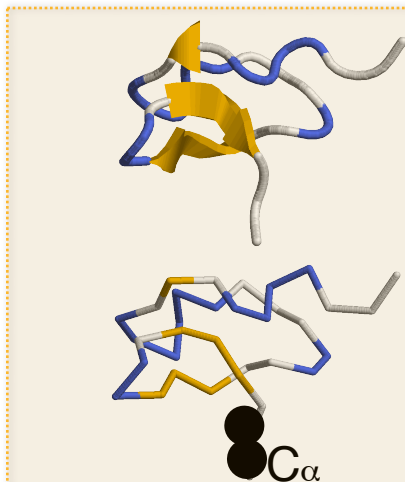


Graph theory search
of similar SSE
Refining by Monte Carlo
at all atom resolution

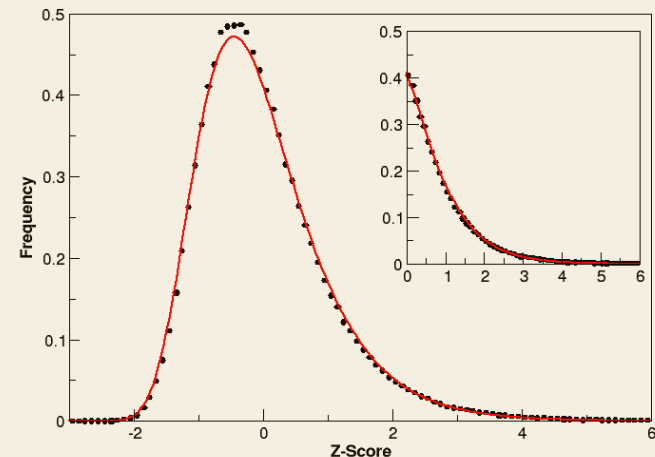


✓ Good scoring system with significance

Reduces the protein representation




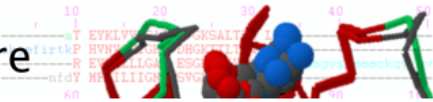
$$RMSD(x, y) = \sqrt{\left(\frac{1}{N}\right) \sum_{i=1}^N (\|x(i) - y(i)\|^2)}$$



Vector Alignment Search Tool (VAST)

<http://www.ncbi.nlm.nih.gov/Structure/VAST/vast.shtml>

NCBI > Structure Home > 3D Macromolecular Structures > VAST

 **Structure** 

HOME | SEARCH | GUIDE | Structure Home | 3D Macromolecular Structures | Conserved Domains | PubChem | BioSystems

VAST: Vector Alignment Search Tool ABOUT | SEARCH | HELP | nr-PDB | PUBLICATIONS | RESOURCES | NEWS

About VAST

VAST, short for **Vector Alignment Search Tool**, is a computer algorithm developed at NCBI and used to identify similar protein 3-dimensional structures by purely geometric criteria, and to identify distant homologs that cannot be recognized by sequence comparison.

VAST is applied on every protein in the [Molecular Modeling Database \(MMDB\)](#) during [MMDB data processing](#) in order to identify similar 3D structures. The pre-computed results are accessible from a [structure's summary page](#); to retrieve them, you can either:

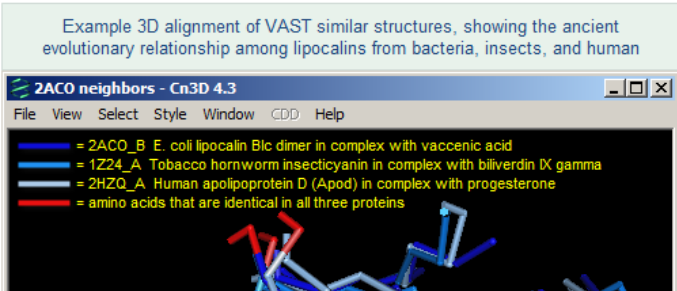
1. view the "[show annotation](#)" graphic for any protein molecule of interest on a structure summary page, then click on the bar graphic for the overall protein molecule or for any **3D domain** it contains in order to view a list of structures that are similar in shape to the molecule or 3D domain you selected. The [VAST Help](#) document provides additional details and illustrated examples.
2. follow the link for "[Similar Structures: VAST](#)" in the upper right corner of a structure summary page to open a tabular list of the protein molecules and **3D domains** in the structure. Then select the protein or 3D domain of interest to view a list of structures that are similar in shape to the region you selected.

Show "Similar Structures" for PDB ID or MMDB ID:

If you have a newly determined protein structure that is not yet in MMDB, then you can use the [VAST Search](#) service to input your data in [PDB file format](#) and compare your structure against all those in MMDB. The [VAST Search Help](#) document provides additional information about using the VAST Search page.

Whether you retrieve similar structures from the summary page of a publicly available structure or through a VAST search, you can then use the free **Cn3D** structure viewing program to view a superposition of the query structure and similar structures, as shown in the illustration below. The [Cn3D Tutorial](#) provides additional details about [viewing structure alignments in Cn3D](#).

Example 3D alignment of VAST similar structures, showing the ancient evolutionary relationship among lipocalins from bacteria, insects, and human

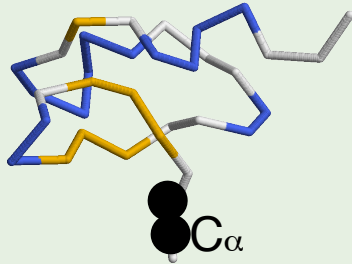


2ACO neighbors - Cn3D 4.3

File View Select Style Window CDD Help

- = 2ACO_B E. coli lipocalin B1c dimer in complex with vaccenic acid
- = 1Z24_A Tobacco hornworm insecticyanin in complex with biliverdin IX gamma
- = 2HZQ_A Human apolipoprotein D (ApoD) in complex with progesterone
- = amino acids that are identical in all three proteins

Incremental combinatorial extension (CE)



Exhaustive combination
of fragments

Longest combination of
AFPs

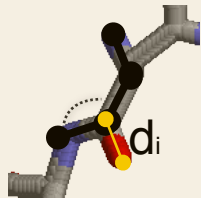
Heuristic similar to
PSI-BLAST



✓ FAST!

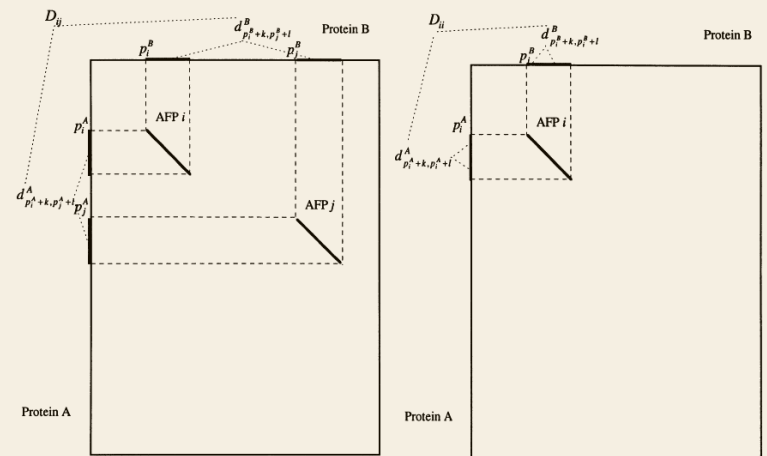
✓ Good quality of local alignments

Complicated scoring and heuristics




8 residues peptides

$$RMSD(x, y) = \sqrt{\left(\frac{1}{N}\right) \sum_{i=1}^N (\|x(i) - y(i)\|^2)}$$



Incremental combinatorial extension (CE)

<http://source.rcsb.org/jfatcatserver/ceHome.jsp>



Combinatorial Extension (CE)
A method for comparing and aligning protein structures

Help

Calculate

Two Chains

Symmetry

DB Search

All vs All

Download

Download JCE/jFatCat

Documentation

History

Release History

CE History

Links

Multiple Structure

Alignment


Results

Subdomains

(PDF)

CruiseControl

CruiseControl



Combinatorial Extension (CE)

A method for comparing and aligning protein structures

This page is intended as a pointer to get you to the most recent information on CE and to enable you to perform the calculations you need. CE is now an integral part of the [RCSB Protein Data Bank](http://www.rcsb.org/pdb) (PDB) and continues to be developed in the [Bourne laboratory](http://source.rcsb.org/jfatcatserver/) as needed.

Key Pointers

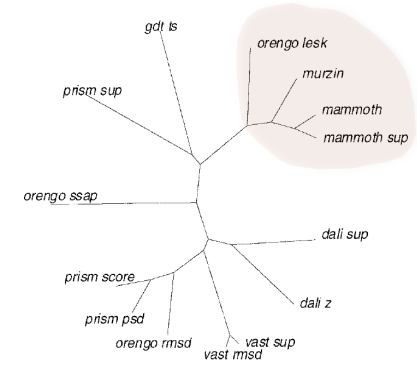
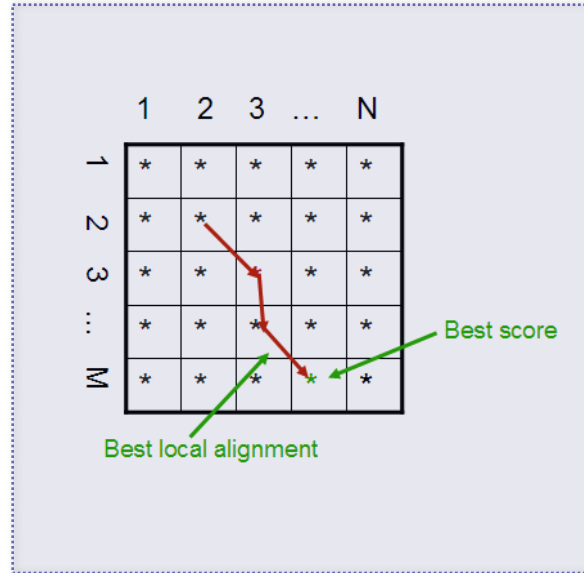
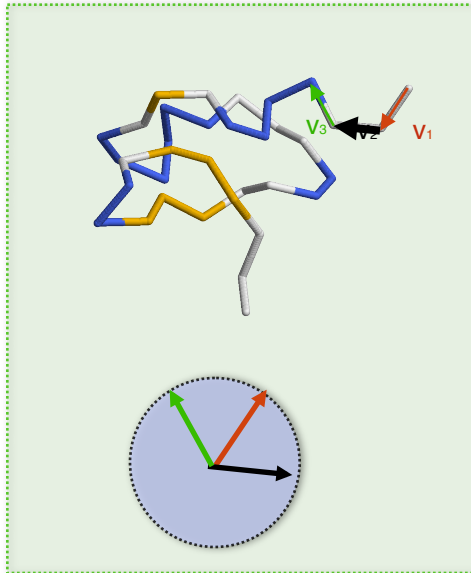
- Access to CE from the RCSB PDB <http://www.rcsb.org/pdb/workbench/workbench.do>
- Standalone server <http://source.rcsb.org/jfatcatserver/>
- Access to the CE code in Java (JCE) and the original source <http://source.rcsb.org/jfatcatserver/download.jsp>
- [Legacy](#) CE - web site

What follows is a brief description of the history of CE and some additional references and pointers.

Chronology

- 1998 - CE method released and original paper published [1]
- 2000 - CE used to map existing protein fold space [2]
- 2001 - Pairwise alignment database made available [3]

Matching molecular models obtained from theory (MAMMOTH)

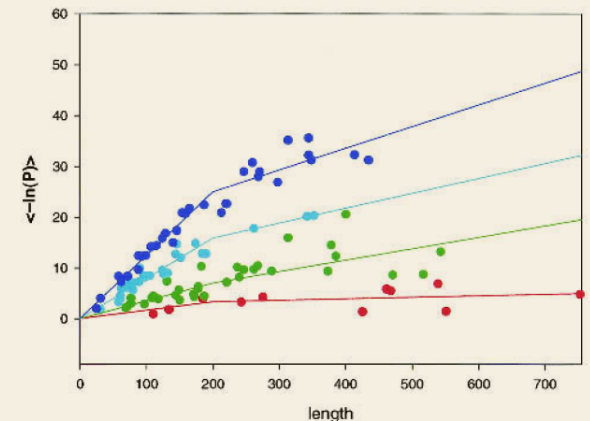
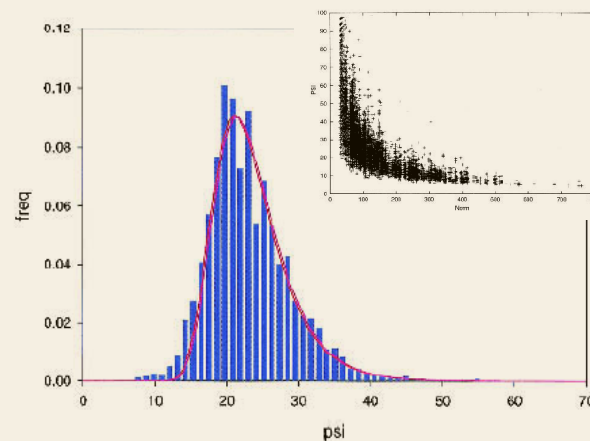


- ✓ VERY FAST!
- ✓ Good scoring system with significance

Reduces the protein representation

$$URMS^R = \sqrt{2.0 - \frac{2.84}{\sqrt{n}}}$$

$$S_{AB} = \frac{(URMS^R - URMS^{AB})D}{URMS^R}$$

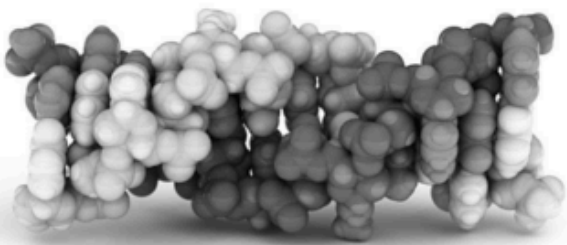


Matching molecular models obtained from theory (MAMMOTH)

<http://ub.cbm.uam.es/software/online/mammoth.php>

Bioinformatics Unit

Centro de Biología Molecular Severo Ochoa



[Home](#)[Publications](#)[Research](#)[Services](#)[Software](#)

Currently we are under testing, so this service may not be functioning properly. Sorry for the inconvenience



MAMMOTH

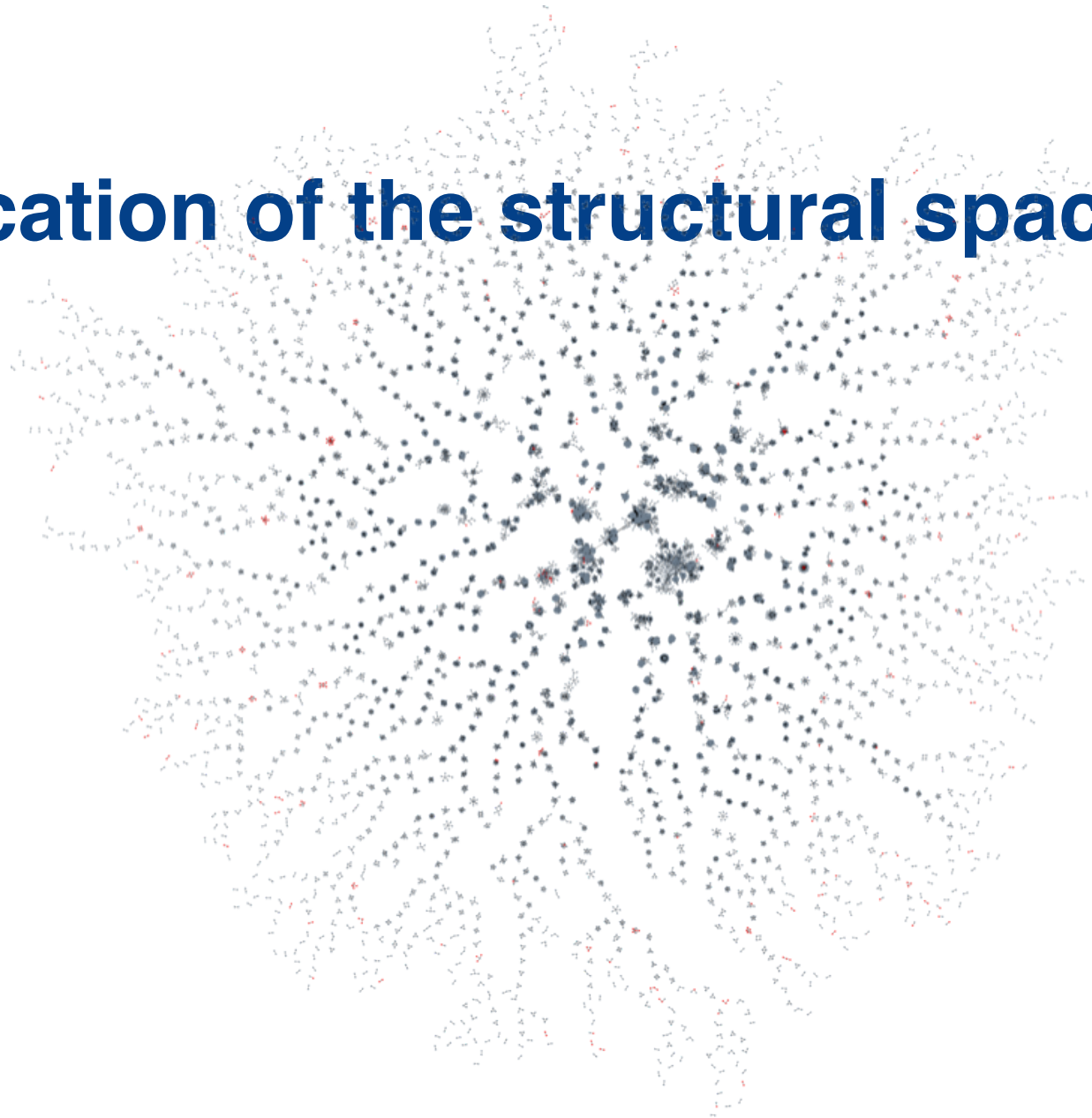
- MAMMOTH (Matching Molecular Models Obtained from Theory) es un método de alineamiento estructural de proteínas independiente de su secuencia. Esto permite la comparación de la estructura experimental de una proteína con un modelo arbitrario de baja resolución. También permite la comparación de dos estructuras experimentales, así como la búsqueda de estructuras similares en una base de datos.
- Versión: 1.0
- Uso gratuito para fines educativos y de investigación.
- Referencia: Ortiz AR, Strauss CE, Olmea O (2002) *Protein Sci.* 11:2606-21.

Alinea tus propias proteínas.

- Sube el archivo de coordenadas (PDB) de tu **primera proteína**:
- Sube el archivo de coordenadas (PDB) de tu **segunda proteína**:
- Tu **correo electrónico** para el envío de los resultados:

3. Structure classification

Classification of the structural space



SCOP_{1.75} database

<http://scop.berkeley.edu/>

Murzin A. G., et al. (1995). *J. Mol. Biol.* **247**, 536-540.

- ✓ Largely recognized as “standard of gold”
- ✓ Manually classification
- ✓ Clear classification of structures in:
CLASS
FOLD
SUPER-FAMILY
FAMILY
- ✓ Some large number of tools already available

Structural Classification of Proteins and ASTRAL release 1.75B (January 2013)

[Browse](#) [Stats & History](#) [ASTRAL Subsets](#) [Downloads](#) [Related Resources](#) [References](#) [Help](#) [About](#)












Welcome to the new SCOP+ASTRAL website!

This release is part of a series of planned releases based on SCOP 1.75 before the advent of a major reclassification, **SCOP 2.0**.

This website now provides integrated access to data previously found in the [SCOP](#) and [ASTRAL](#) databases. For prior releases of SCOP and ASTRAL, click on the [Stats & History](#) tab above. For more info, click on the [About](#) tab above.

Search SCOP (example):

Classes in SCOP 1.75B:

1.  [a: All alpha proteins](#) [46456] (284 folds)
2.  [b: All beta proteins](#) [48724] (174 folds)
3.  [c: Alpha and beta proteins \(a/b\)](#) [51349] (147 folds)
4.  [d: Alpha and beta proteins \(a+b\)](#) [53931] (376 folds)
5.  [e: Multi-domain proteins \(alpha and beta\)](#) [56572] (66 folds)
6.  [f: Membrane and cell surface proteins and peptides](#) [56835] (57 folds)
7.  [g: Small proteins](#) [56992] (90 folds)
8.  [h: Coiled coil proteins](#) [57942] (7 folds)
9.  [i: Low resolution protein structures](#) [58117] (25 folds)
10.  [j: Peptides](#) [58231] (120 folds)
11.  [k: Designed proteins](#) [58788] (44 folds)

Manually classification
Not 100% up-to-date
Domain boundaries definition

Class	Number of folds	Number of superfamilies	Number of families
All alpha proteins	284	507	928
All beta proteins	174	354	815
Alpha and beta proteins (a/b)	147	244	902
Alpha and beta proteins (a+b)	376	552	1170
Multi-domain proteins	66	66	100
Membrane and cell surface proteins	57	109	127
Small proteins	90	129	230
Total	1194	1961	4272



Copyright © 1994-2013 The SCOP and ASTRAL authors
scop@mrc-lmb.cam.ac.uk and astral@compbio.berkeley.edu





[a: All alpha proteins](#) -> [a.3: Cytochrome c](#) -> [a.3.1: Cytochrome c](#) ->
(class) (fold) (superfamily)
[a.3.1.4: Two-domain cytochrome c](#)
(family)

Structural Classification of Proteins and ASTRAL release 1.75B (January 2013)






Browse	Stats & History	ASTRAL Subsets	Downloads	Related Resources	References	Help	About
---------------	-----------------	----------------	-----------	-------------------	------------	------	-------

Search SCOP ([example](#)):

Lineage for Family **a.3.1.4: Two-domain cytochrome c**

1. Root: [SCOP 1.75B](#)
2.  Class [a: All alpha proteins](#) [46456] (284 folds)
3.  Fold [a.3: Cytochrome c](#) [46625] (1 superfamily)
core: 3 helices; folded leaf, opened
4.  Superfamily [a.3.1: Cytochrome c](#) [46626] (9 families) \mathcal{S}
covalently-bound heme completes the core
5.  Family **a.3.1.4: Two-domain cytochrome c** [46680] (2 protein domains)
duplication: consists of two cytochrome c type domains

Protein Domains:







1.  [Cytochrome c4](#) [46681] (2 species)
 1.  Species [Pseudomonas stutzeri](#) [[TaxId:316](#)] [46682] (3 PDB entries)
 2.  Species [Thiobacillus ferrooxidans](#) [[TaxId:920](#)] [88972] (1 PDB entry)
2.  [Flavocytochrome c sulfide dehydrogenase, FCSD, cytochrome subunit](#) [46683] (1 species)
 1.  Species [Chromatium vinosum](#) [[TaxId:1049](#)] [46684] (1 PDB entry)


More info for Family **a.3.1.4: Two-domain cytochrome c**

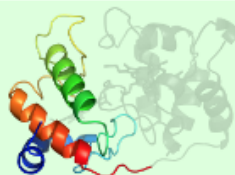
Timeline for Family a.3.1.4: Two-domain cytochrome c:

Family a.3.1.4: Two-domain cytochrome c [appears in SCOP 1.75A](#)

Lineage for d1etpa1 (1etp A:1-92)

1. Root: [SCOP 1.75B](#)
2.  Class [a: All alpha proteins](#) [46456] (284 folds)
3.  Fold [a.3: Cytochrome c](#) [46625] (1 superfamily)
core: 3 helices; folded leaf, opened
4.  Superfamily [a.3.1: Cytochrome c](#) [46626] (9 families) *S*
covalently-bound heme completes the core
5.  Family [a.3.1.4: Two-domain cytochrome c](#) [46680] (2 protein domains)
duplication: consists of two cytochrome c type domains
6.  Protein [Cytochrome c4](#) [46681] (2 species)
7.  Species [Pseudomonas stutzeri](#) [[TaxId:316](#)] [46682] (3 PDB entries)

8.  Domain d1etpa1: 1etp A:1-92 [15962]
complexed with hem



Details for d1etpa1

PDB Entry: [1etp \(more details\)](#)

PDB Description: crystal structure of cytochrome c4 from pseudomonas stutzeri

PDB Compounds: (A:) cytochrome c4

SCOP Domain Sequences for d1etpa1:

Sequence; same for both **SEQRES** and **ATOM** records: ([download](#))

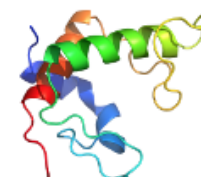
```
>d1etpa1 a.3.1.4 (A:1-92) Cytochrome c4 {Pseudomonas stutzeri [TaxId: 316]}  
agdaeaagqgkvavcgachgvdgnsapnfpklaggqeryllkqlqdkagstpgapevg  
rkvlemtgmldplsdqdlediaayfssqkgs
```

SCOP Domain Coordinates for d1etpa1:

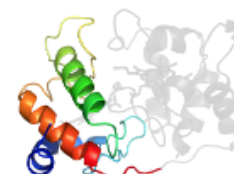
Click to download the [PDB-style file with coordinates for d1etpa1](#).

(The format of our PDB-style files is described [here](#).)

Timeline for d1etpa1:



[d1etpa1](#)
([click for larger image](#))



[d1etpa1 in context of chain](#)
Domains from same chain:
[d1etpa2](#)

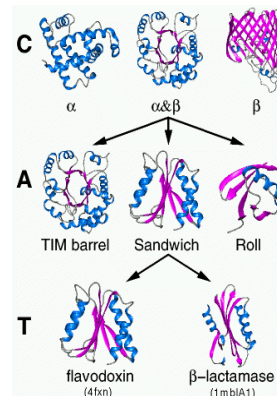
CATH_{3.5} database

<http://www.cathdb.info>

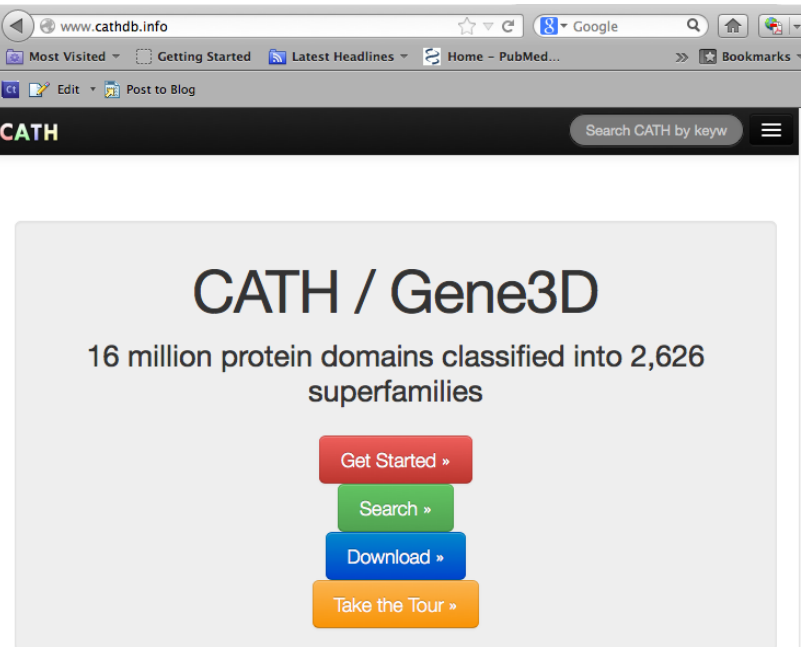
Uses FSSP for superimposition

- ✓ Recognized as “standard of gold”
- ✓ Semi-automatic classification
- ✓ Clear classification of structures in:
CLASS
ARCHITECTURE
TOPOLOGY
HOMOLOGOUS SUPERFAMILIES
- ✓ Some large number of tools already available
- ✓ Easy to navigate

Semi-automatic classification
Domain boundaries definition



173,536 CATH Domains
2,626 CATH Superfamilies
51,334 PDBs



What's New?

The CATH website has recently undergone a big overhaul. We really hope you find the new pages more useful, easier to use and quicker to load. Please [get in touch](#) and let us know what you think.

Searching CATH

- [Search by ID / keyword](#)
- [Search by FASTA sequence](#)
- [Search by PDB structure](#)

Example pages

- [PDB "1dan"](#)
- [Domain "1cukA01"](#)
- [Relatives of "1cukA01"](#)
- [Superfamily "HUPs"](#)
- [Functional Family](#)

Browse - tree

Details on the currently selected CATH node are displayed in the panel below

H Nop10-like SnoRNP

[View Superfamily](#)

CATH ID **2.20.28.40**

Non-redundant Sequences (<35% seq id) **1**

Domains **3**

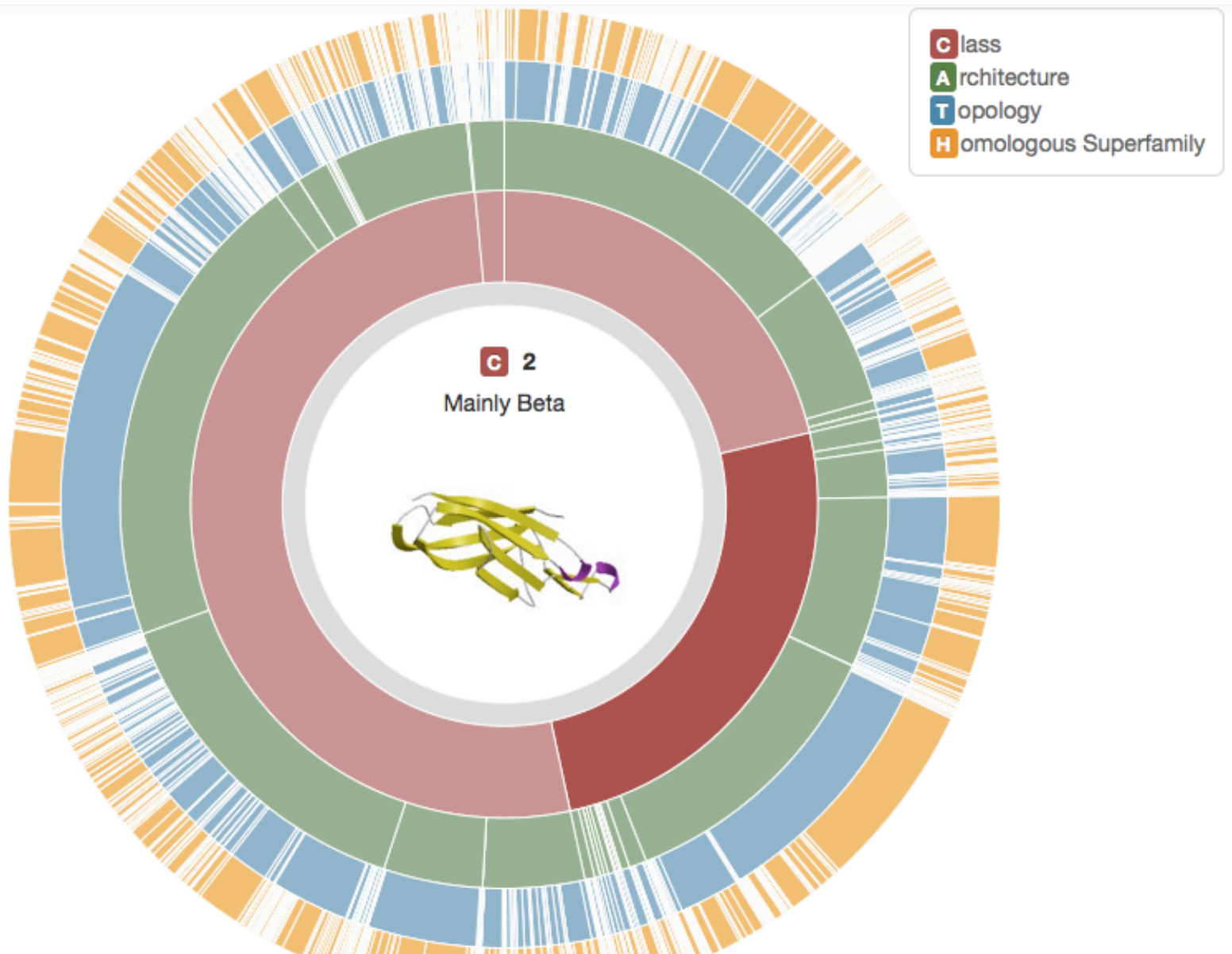
Example Domain **2apoB00**
[\[PDB\]](#)



Top of CATH Hierarchy (4 Classes)

▷ C 1 Mainly Alpha	5 Architectures, 386 Folds, 875 Superfamilies, 37038 Domains
◄ C 2 Mainly Beta	20 Architectures, 229 Folds, 520 Superfamilies, 43881 Domains
▷ A 2.10 Ribbon	25 Folds, 31 Superfamilies, 1733 Domains
◄ A 2.20 Single Sheet	20 Folds, 31 Superfamilies, 683 Domains
▷ T 2.20.20 Anthopleurin-A	3 Superfamilies, 21 Domains
▷ T 2.20.25 N-terminal domain of TfIIb	6 Superfamilies, 342 Domains
◄ T 2.20.28 Rubrerythrin, domain 2	5 Superfamilies, 209 Domains
H 2.20.28.10 Not yet named	131 Domains
H 2.20.28.20 Not yet named	11 Domains
H 2.20.28.30 RNA polymerase ii, chain L	61 Domains
H 2.20.28.40 Nop10-like SnoRNP	3 Domains
H 2.20.28.50 degv family protein	3 Domains
▷ T 2.20.50 Outer Surface Protein A; domain 2	1 Superfamilies, 1 Domains
▷ T 2.20.60 Heparin-binding Growth Factor, Midkine; Chain A	1 Superfamilies, 1 Domains
▷ T 2.20.70 Ubiquitin Ligase Nedd4; Chain: W;	1 Superfamilies, 60 Domains
▷ T 2.20.80 Lipovitellin-phosvitin complex, chain A, domain 4	1 Superfamilies, 1 Domains
▷ T 2.20.90 Lipovitellin-phosvitin complex; beta-sheet shell regions	1 Superfamilies, 1 Domains
▷ T 2.20.100 TSP-1 type 1 repeat	1 Superfamilies, 3 Domains
▷ T 2.20.110 Histone H3 K4-specific methyltransferase SET7/9 N-terminal domain	1 Superfamilies, 3 Domains
▷ T 2.20.120 Multimodular pneumococcal cell wall endolysin, domain 3	1 Superfamilies, 6 Domains
▷ T 2.20.130 S-adenosyl-L-methionine-dependent methyltransferases	1 Superfamilies, 0 Domains
▷ T 2.20.140 α64vE2_hafr protein fold	

Browse - sunburst



SUPERFAMILY LINKS

Summary

[Superfamily Superposition](#)[Classification / Domains](#)[Alignments](#)[Structural Neighbourhood](#)[Functional Annotations](#)[Taxonomy](#)[Multi-Domain Organisation](#)

Functional Families

Overview of the Structural Clusters (SC) and Functional Families (FF) within this CATH Superfamily

SC:1 — TATA-box-b

GO Diversity

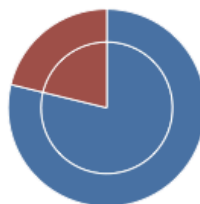
Unique GO annotations



93 Unique GO terms >

EC Diversity

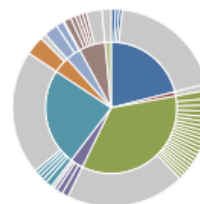
Unique EC annotations



2 Unique EC terms >

Species Diversity

Unique species annotations



433 Unique species >

Superfamily Summary

A general summary of information for this superfamily.

Structures

Domains: 140

Domains (< 95% seq id): 15

Domains (< 35% seq id): 4

Unique PDBs: 41

Alignments

Structural Clusters: 2

FunFam Clusters: 2

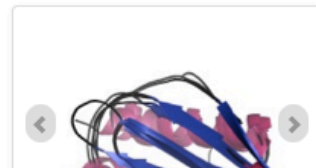
Function

Unique EC: 2

Unique GO: 93

Structural Diversity

Structural domains within this superfamily



Domain Organisation

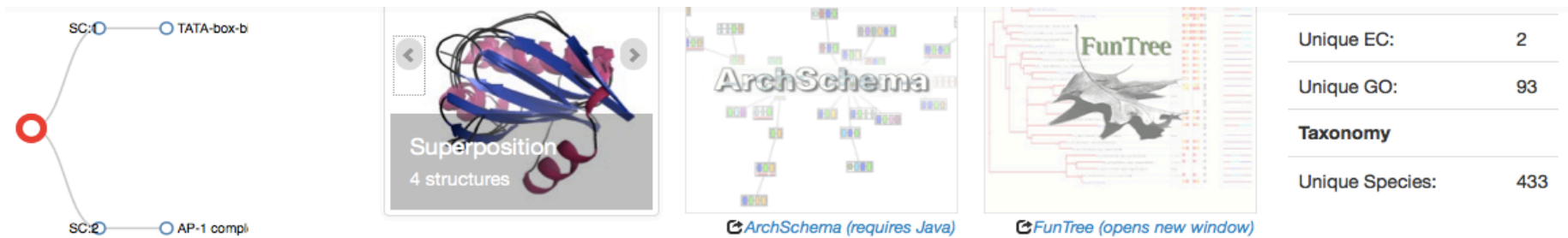
View multi-domain architectures via ArchSchema (Laskowski/EBI)



Enzyme Function

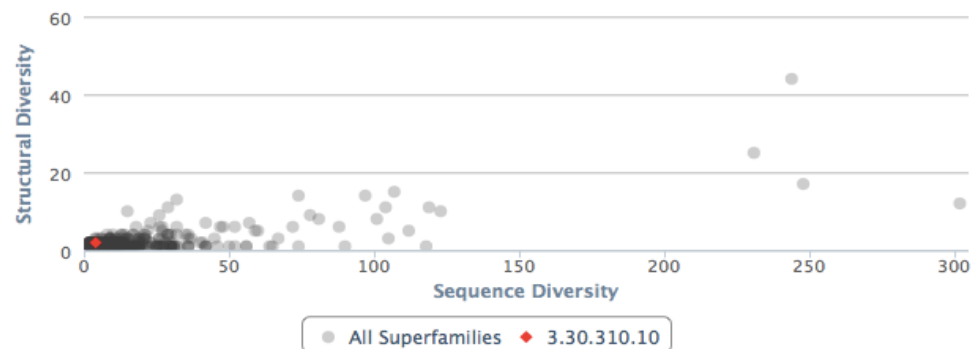
Evolution of Enzyme Function via FunTree (Furnham/EBI)





Sequence/Structure Diversity

Overview of the sequence / structure diversity of this superfamily compared to other superfamilies in CATH. Click on the chart to view the data in more detail.



Classification of the structural space

Not an easy task!

Domain definition AND domain classification

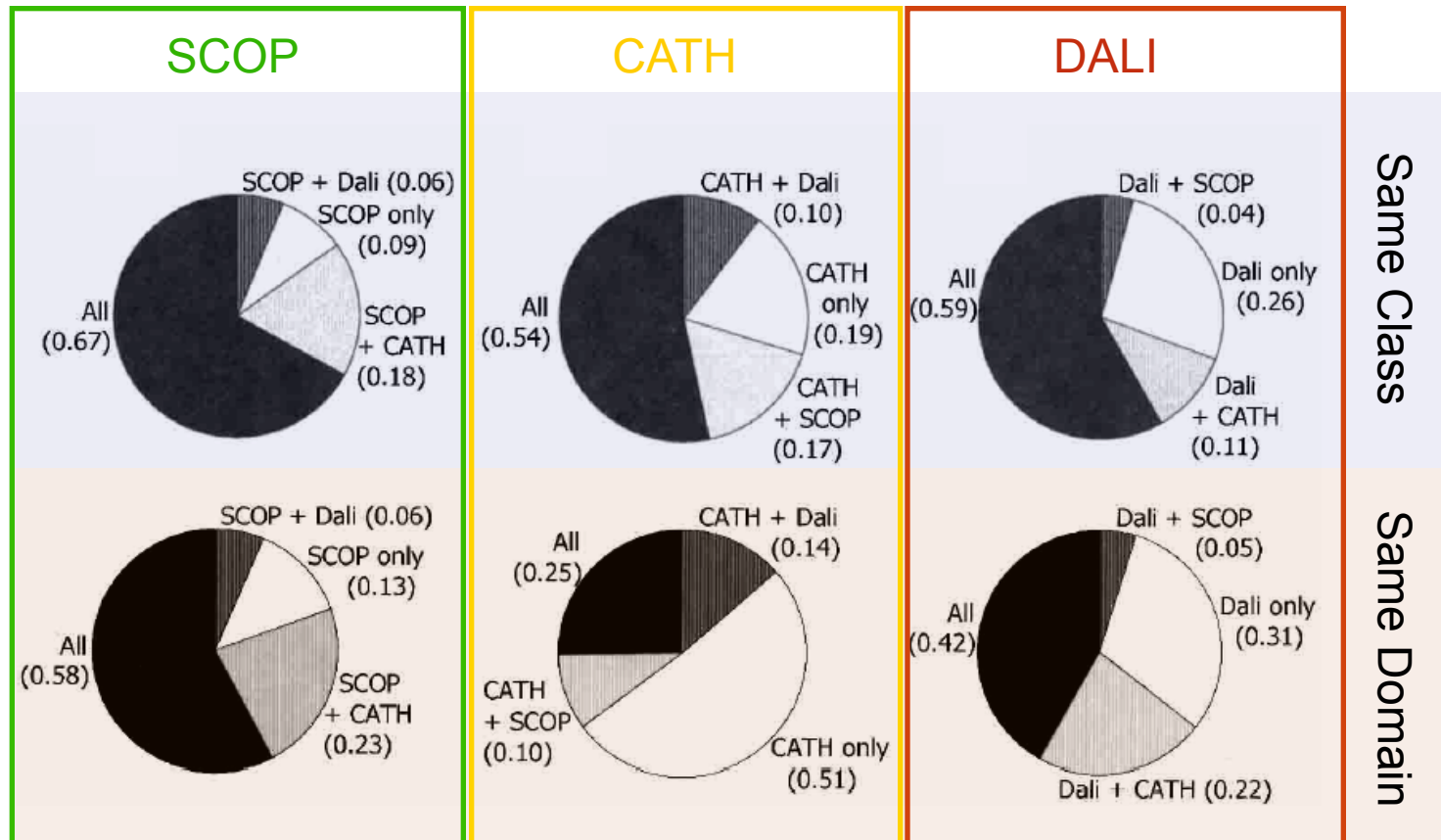
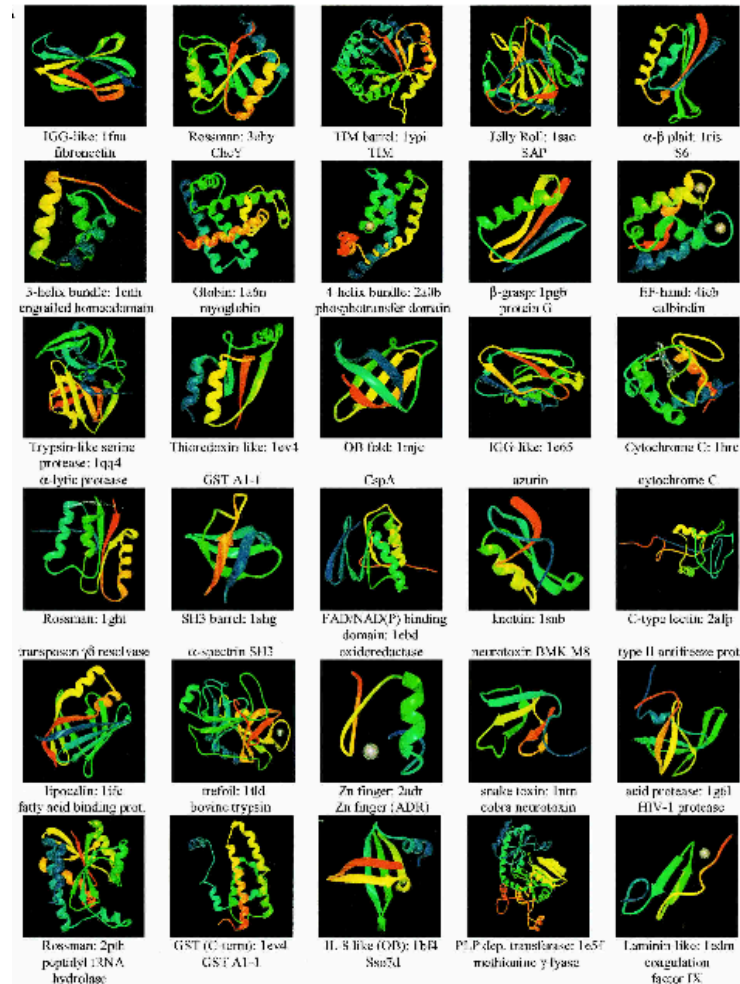
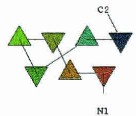
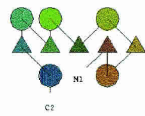
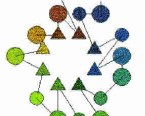
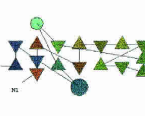
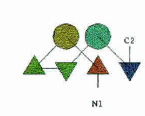
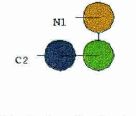
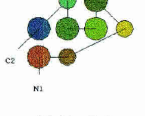
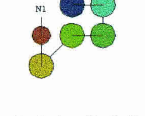
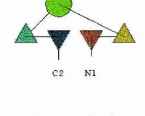
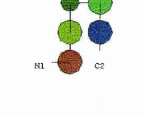
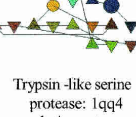
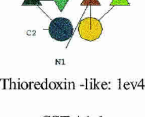
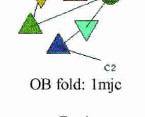

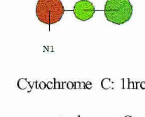
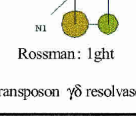
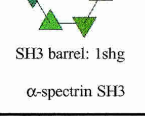
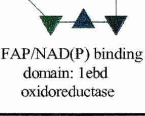
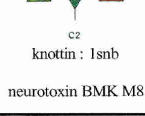
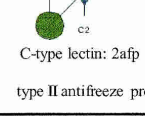
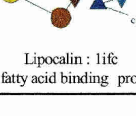
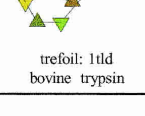
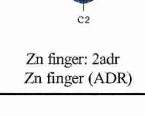
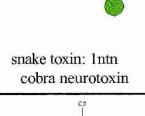
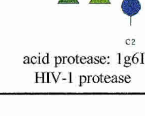
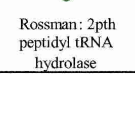
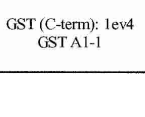
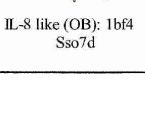
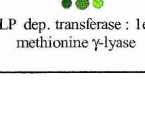
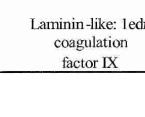


Table 2. SCOP, CATH, and Dali codes associated with the 30 most populated metafolds

Fold	SCOP codes	CATH codes	Dali codes	Population
Immunoglobulin-like	b.1, b.2, b.3, b.5, b.7	2.60.40, 2.60.97, 2.70.50, 6.1.250, 6.1.251	2.27, 2.28, 2.29, 2.30, 2.31, 2.32, 2.33, 2.35, 2.43	642
Rossmann fold	c.2, c.16, c.23, c.32, c.33, c.37, c.41, c.44, c.62, c.65, c.66, c.69, c.78	3.40.50	1.1, 1.3, 1.15, 1.17	424
TIM barrel	c.1, c.6	3.20.20, 6.1.22	1.35	205
Jelly roll	b.10, b.13, b.18, b.19, b.22, b.23, b.29	2.60.120	2.43, 2.45	162
α - β plait	d.41, d.51, d.58, d.89, d.94	3.30.70, 3.40.462	5.3, 5.8, 5.12, 6.23, 6.26, 6.28, 6.45, 6.60	121
DNA/RNA binding 3-helix bundle	a.4	1.10.10	3.153, 3.154, 3.155, 3.156, 3.157, 3.160, 3.162, 3.164, 5.17, 6.157, 6.158, 6.160	105
Globin	a.1, f.1	1.10.490	3.62, 3.159	84
Four-helix bundle	a.24, a.25, a.26, a.63	1.20.120	3.35	78
β grasp (ubiquitin-like)	d.15	3.10.20	2.3, 2.4, 2.5, 6.62, 6.63	70
EF-hand	a.39	1.10.238	3.136, 3.146, 3.147, 6.156	67
Trypsin-like serine protease	b.47	2.40.10	6.108	66
Thioredoxin-like	c.47	3.40.30	1.51, 1.52	64
OB fold	b.40	2.40.50	4.10, 4.11, 4.16, 7.266	64
More IG-like	b.6	2.60.40	2.34	55
Cytochrome C	a.3	1.10.1250	3.232	48
More Rossmann folds	c.5, c.25, c.26, c.29, c.53	3.40.50	1.7, 1.8	45
SH3 barrel	b.34	2.30.30	4.23, 4.26, 4.27, 6.111, 6.112, 6.117	45
FAD/NAD(P) binding domain	c.3	3.50.50	1.6	42
Knottins	g.3	3.30.30	5.17, 6.196, 6.197	39
C-type lectin-like	d.169	3.10.40, 3.10.100, 3.90.6	6.195	36
Lipocalin	b.60	2.40.128	6.71	36
Trefoil	b.42	2.80.10	6.81	32
C2H2 & C2H2 Zinc finger	g.37	3.30.160	6.206, 6.207	30
snake toxin-like	g.7	2.10.60	6.198	29
Acid proteases	b.50	2.40.70	2.21	29
more Rossmann folds	c.48, c.54, c.56	3.40.50, 3.40.630	1.2	29
Glutathione S-transferase	a.45	1.20.1050	3.84	28
IL-8 like (OB fold)	d.9	2.40.50	4.19, 6.103	28
PLP dependent transferases	c.67		1.23	27
Laminin-like (knottins)	g.3	2.10.25	7.244, 7.246	26

4. Number of folds in nature



				
IGG-like: 1fna fibronectin	Rossman: 3chy CheY	TIM barrel: 1typi TIM	Jelly roll: 1sac SAP	a-b plait: 1ris S6
				
3-helix bundle: 1enh Engrailed Homeodomain	Globin: 1a6n myoglobin	4-helix bundle: 2a0b phosphotransfer domain	b-grasp: 1pgh Protein G	EF-hand: 4icb calbindin
				
Trypsin-like serine protease: 1qq4 α -lytic protease	Thioredoxin-like: 1ev4 GST A1-1	OB fold: 1mj2 CspA	IGG-like: 1e65 azurin	Cytochrome C: 1hrc cytochrome C
				
Rossman: 1ght transposon γ resolvase	SH3 barrel: 1shg α -spectrin SH3	FAP/NAD(P) binding domain: 1ebd oxidoreductase	knottin: 1snb neurotoxin BMK M8	C-type lectin: 2afp type II antifreeze prot.
				
Lipocalin: 1lfc fatty acid binding prot.	trefoil: 1td bovine trypsin	Zn finger: 2adr Zn finger (ADR)	snake toxin: 1ntn cobra neurotoxin	acid protease: 1g61 HIV-1 protease
				
Rossman: 2pth peptidyl tRNA hydrolase	GST (C-term): 1ev4 GST A1-1	IL-8 like (OB): 1b4 Sso7d	PLP dep. transferase : 1e5f methionine γ -lyase	Laminin-like: 1edm coagulation factor IX

5. Sequences VS fold structures

Structure is three to ten times more conserved than sequence--a study of structural response in protein cores.

Illergård K, Ardell DH, Elofsson A.

Center for Biomembrane Research, Department of Biochemistry and Biophysics, Stockholm University, SE-106 91 Stockholm, Sweden.



Structure is three to ten times more conserved than sequence—A study of structural response in protein cores

Kristoffer Illergård,¹ David H. Ardell,^{2,3} and Arne Elofsson^{1*}

¹ Center for Biomembrane Research and Stockholm Bioinformatics Center, Department of Biochemistry and Biophysics, Stockholm University, SE-106 91 Stockholm, Sweden

² Department of Natural Sciences, School of Natural Sciences, University of California, Merced, California 95344

³ Linnaeus Centre for Bioinformatics, Uppsala University, SE-751 24 Uppsala, Sweden

ABSTRACT

Protein structures change during evolution in response to mutations. Here, we analyze the mapping between sequence and structure in a set of structurally aligned protein domains. To avoid artifacts, we restricted our attention only to the core components of these structures. We found that on average, using different measures of structural change, protein cores evolve linearly with evolutionary distance (amino acid substitutions per site). This is true irrespective of

INTRODUCTION

Evolutionary changes of individual protein domain primary structures that become fixed in populations are mainly replacements of single amino acid residues and short insertions or deletions. Since most three-dimensional structures of proteins are determined by their sequences¹ and solvent interactions, higher-order structure will also change in response to these changes. The extent of higher-order structural perturbation in response to sequence evolution will depend on the type and location of sequence changes. Some single mutations will completely disrupt structure, while others that conserve the physicochemical properties of the sequence will barely affect structure at all.²

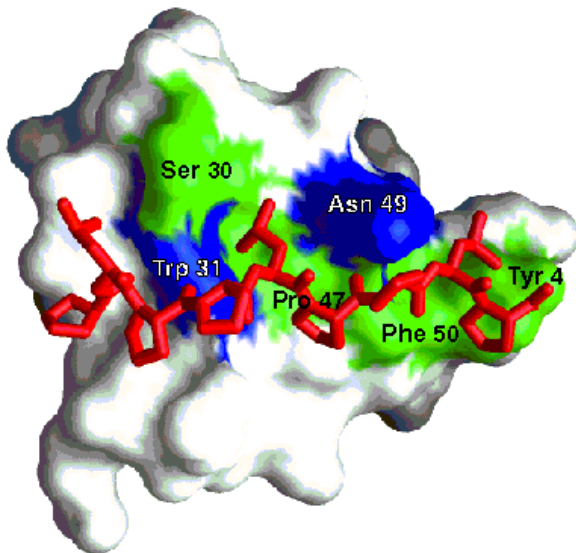
Why is it useful to know the **structure** of a protein, not only its sequence?

- ◆ The biochemical function (activity) of a protein is defined by its interactions with other molecules.
- ◆ The biological function is in large part a consequence of these interactions.
- ◆ The 3D structure is more informative than sequence because interactions are determined by residues that are close in space but are frequently distant in sequence.

YDL117W
(15-64)

10 20 30 40 50

K A R Y G W S G Q T K G D L G F L E G D I M E V T R I A G S W F Y G K L L R N K K C S G Y F P H N F



In addition, since evolution tends to conserve function and function depends more directly on structure than on sequence, **structure is more conserved in evolution than sequence.**

The net result is that **patterns in space are frequently more recognizable than patterns in sequence.**