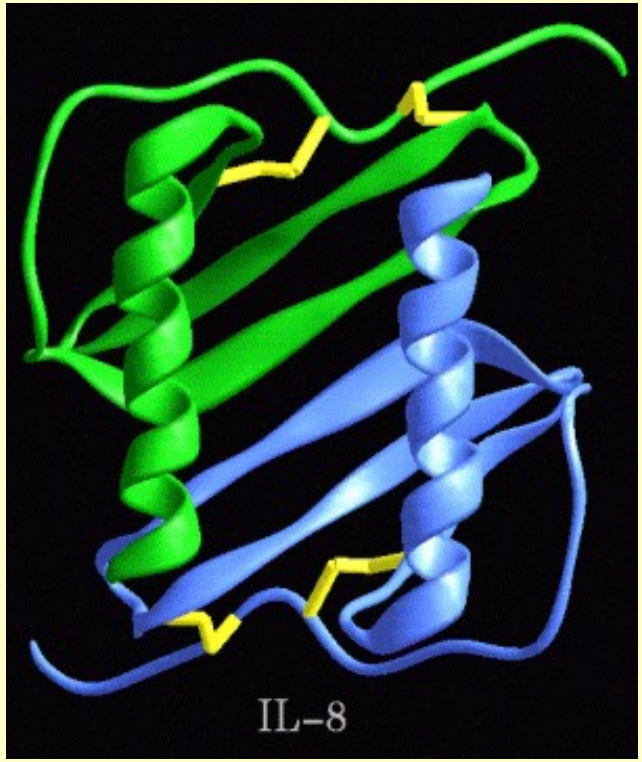
Computational Molecular Biology
Biochem 218 – BioMedical Informatics 231

http://biochem218.stanford.edu/

Protein Structural Motifs



Doug Brutlag
Professor Emeritus
Biochemistry & Medicine (by courtesy)

Homework 5: Phylogenies

- For this homework assignment take 20 to 30 protein sequences which are at least 30% similar or better and:
  - 1) make a multiple sequence alignment with them using ClustalW and
  - 2) make two phylogenies, one using UPGMA method and the other using the Neighbor Joining method
- Describe the resulting alignments and include graphic images of the phylogenies in a message to homework218@cmgm.stanford.edu
- Mention if the trees seem reasonable biologically or taxonomically by comparison with standard taxonomies
- Do the two trees have the same topology?
- Do the trees have the same branch lengths?
- If the two trees do not have the same topology or branch lengths, describe the differences and indicate why you think the two trees differ. Are the differences significant?
- Do the trees show evidence of paralogous evolution? Which nodes are orthologous and which are paralogous bifurcations?
- Do the trees show evidence of either gene conversion or horizontal gene transfer?

Final Projects Due March 12

- Examples of Previous Final Projects
  - http://biochem218.stanford.edu/Projects.html
- Critical review of any area of computational molecular biology.
  - Area from the lectures but in more depth
  - Any other area of bioinformatics or genomics focused on computational approaches
- Proposed improvement or novel approach
- Can be a combined experimental/computational method.
- Could be an implementation or just pseudocode.
- Please do a MeSH literature search for Reviews on your topic.
  Some useful MeSH terms include:
  - Algorithms
  - Statistics
  - Molecular Sequence Data
  - Molecular Structure etc.
- Please send a proposed final project topic to homework218@cmgm.stanford.edu by next Friday

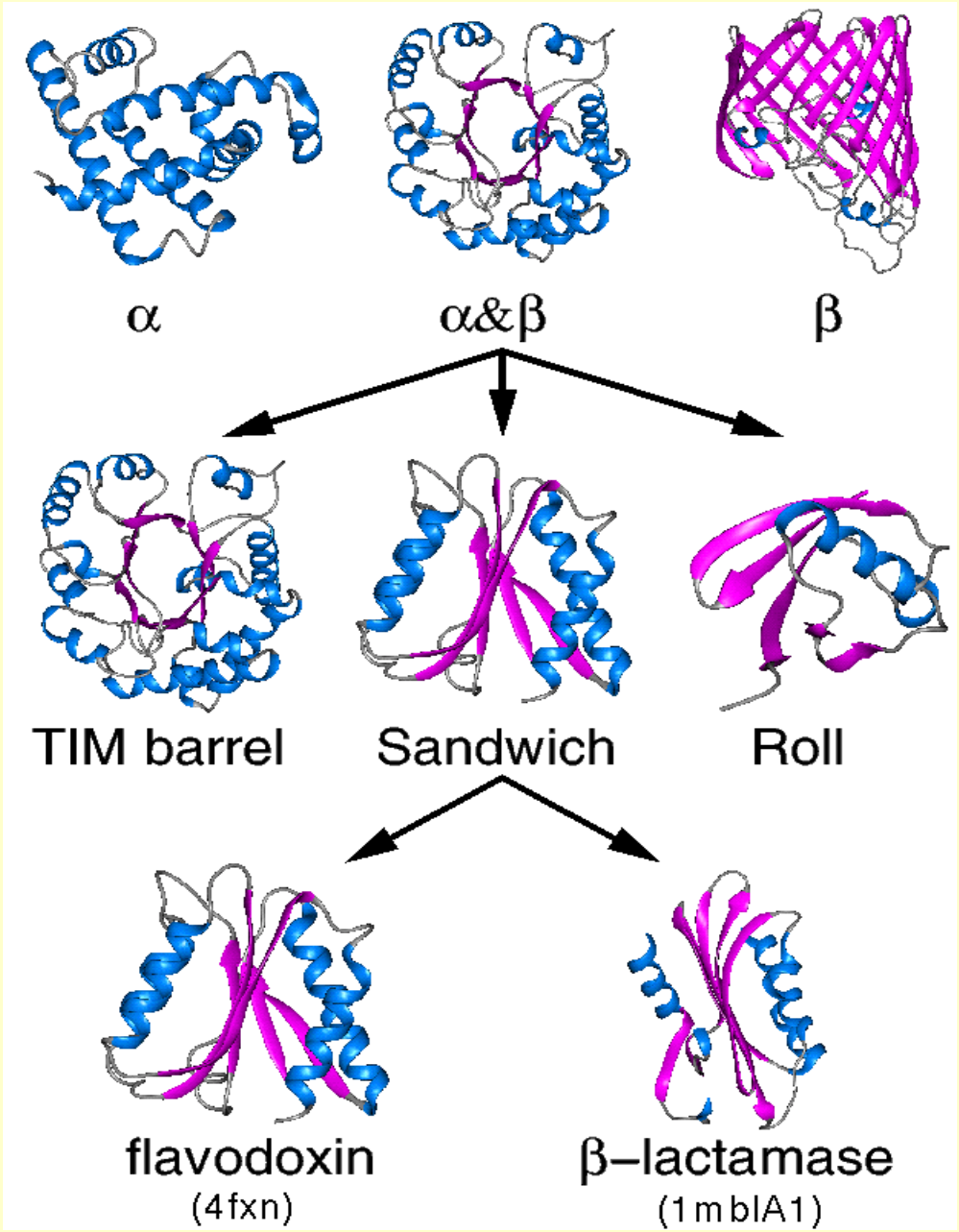
Protein Structure Computational Goals

- Compare all known structures to each other
- Compute distances between protein structures
- Classify and organize all structures in a biologically meaningful way
- Discover conserved substructure domain
- Discover conserved substructural motifs
- Find common folding patterns and structural/functional motifs
- Discover relationship between structure and function.
- Study interactions between proteins and other proteins, ligands and DNA (Protein Docking)
- Use known structures and folds to infer structure from sequence (Protein Threading)
- Use known structural motifs to infer function from structure
- Many more…

Structural Classification of Proteins (SCOP)

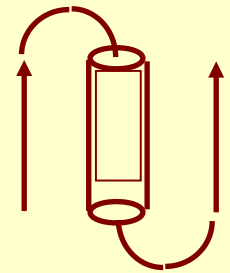
http://scop.berkeley.edu/

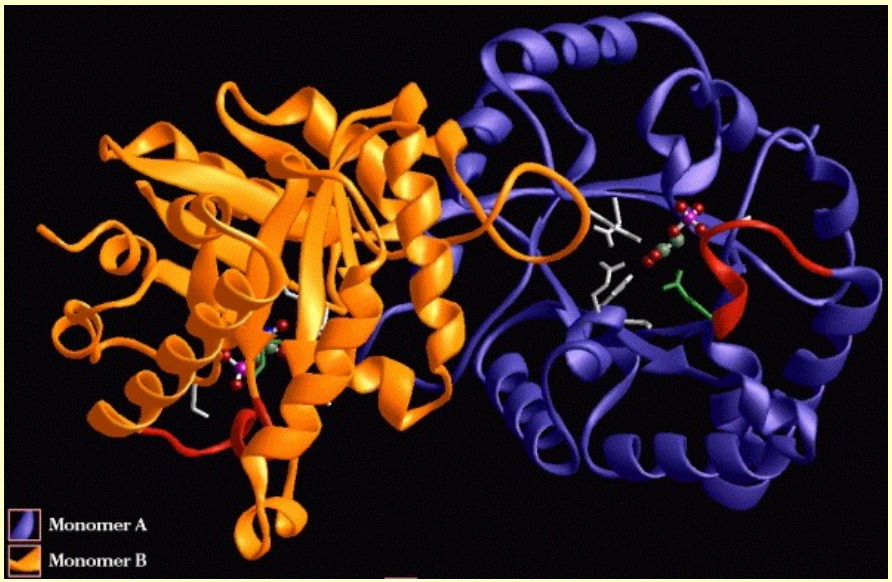
- Class
  - Similar secondary structure content
  - All α, all β, alternating α/β etc
- Fold (Architecture)
  - Major structural similarity
  - SSE's in similar arrangement
- Superfamily (Topology)
  - Probable common ancestry
  - HMM family membership
- Family
  - Clear evolutionary relationship
  - Pairwise sequence similarity > 25%



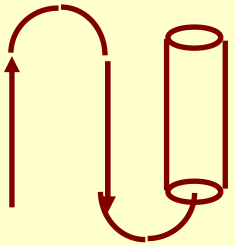
Classes of Protein Structures

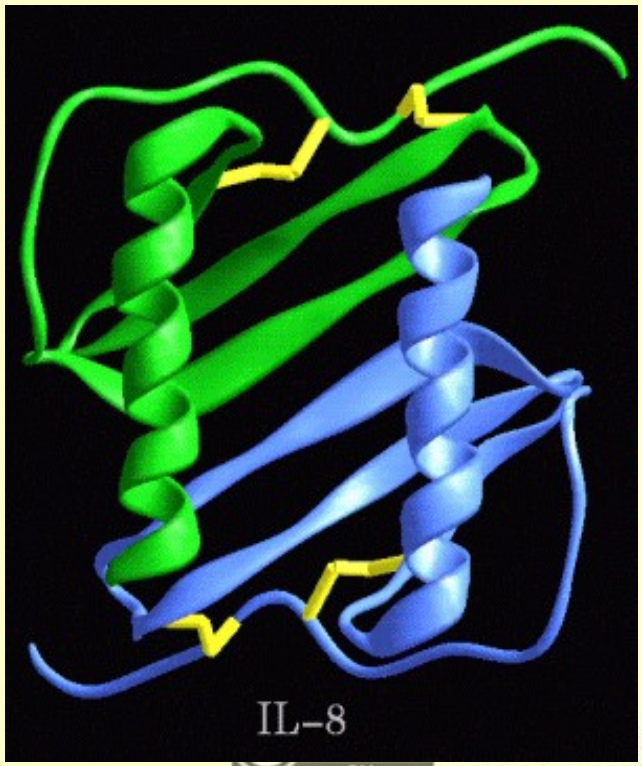
- Mainly α
- Mainly β
- α/β alternating
  - Parallel β sheets, β-α-β units



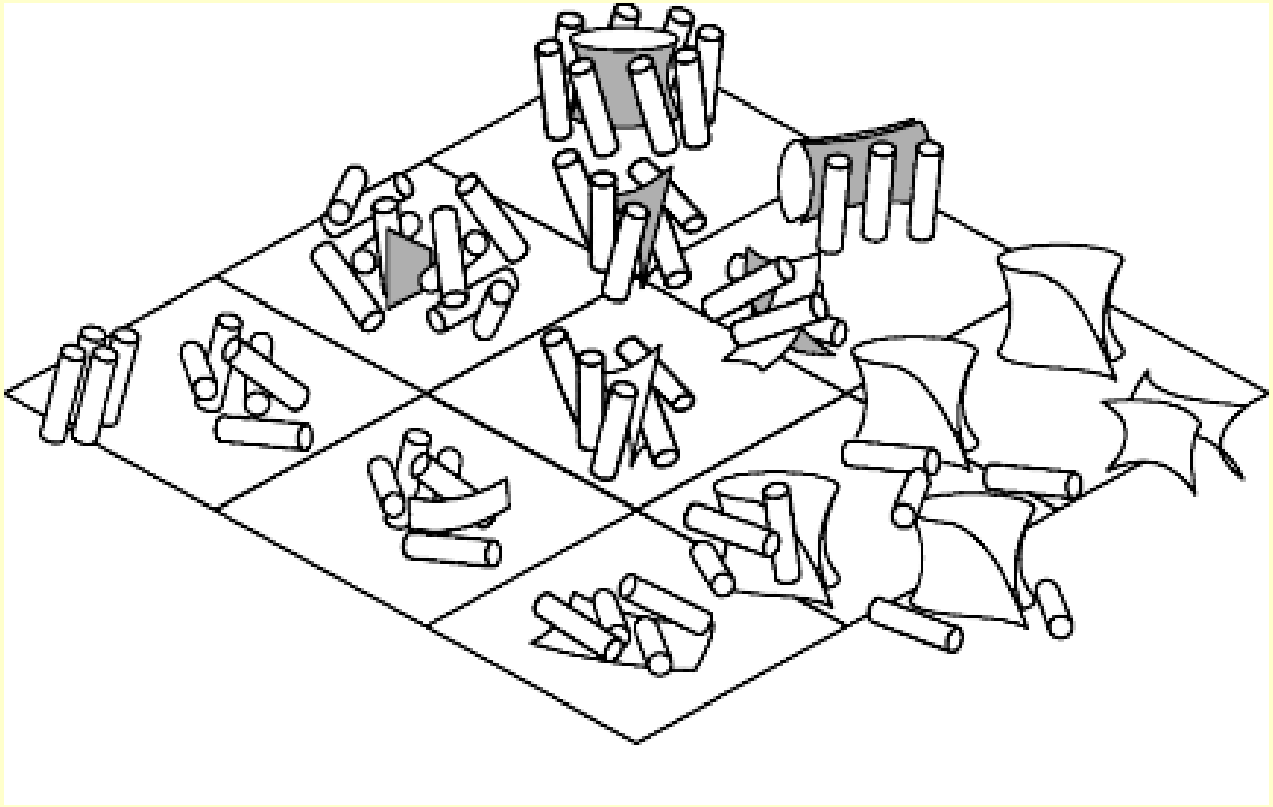


- α +β
  - Anti-parallel β sheets, segregated α and β regions
  - helices mostly on one side of sheet





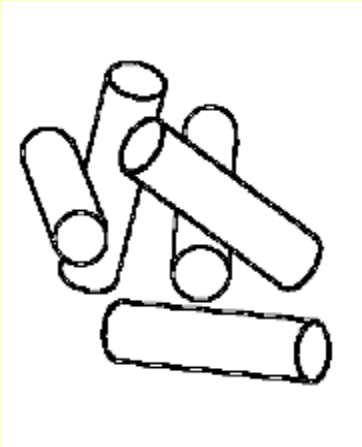
Classes of Protein Structures

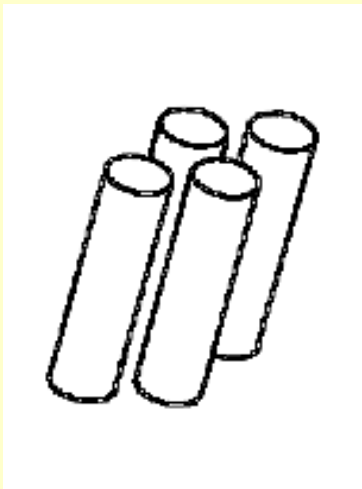


- Others
  - Multi-domain, membrane and cell surface, small proteins, peptides and fragments, designed proteins

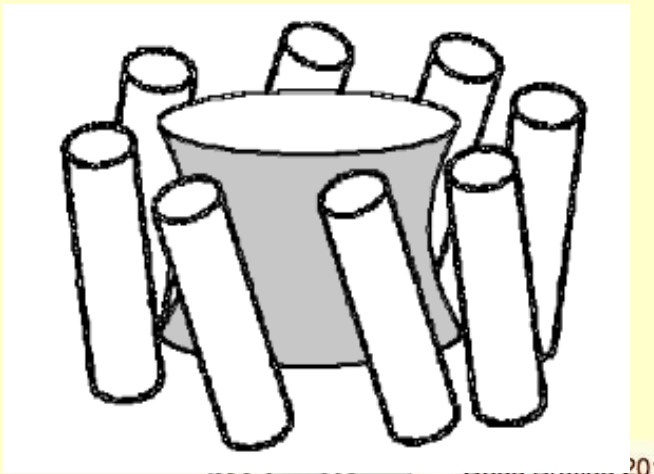
Folds / Architectures

- Mainly α
  - Bundle
  - Non-Bundle
- Mainly β
  - Single sheet
  - Roll
  - Barrel
  - Clam
  - Sandwich
  - Prism
  - 4/6/7/8 Propeller
  - Solenoid

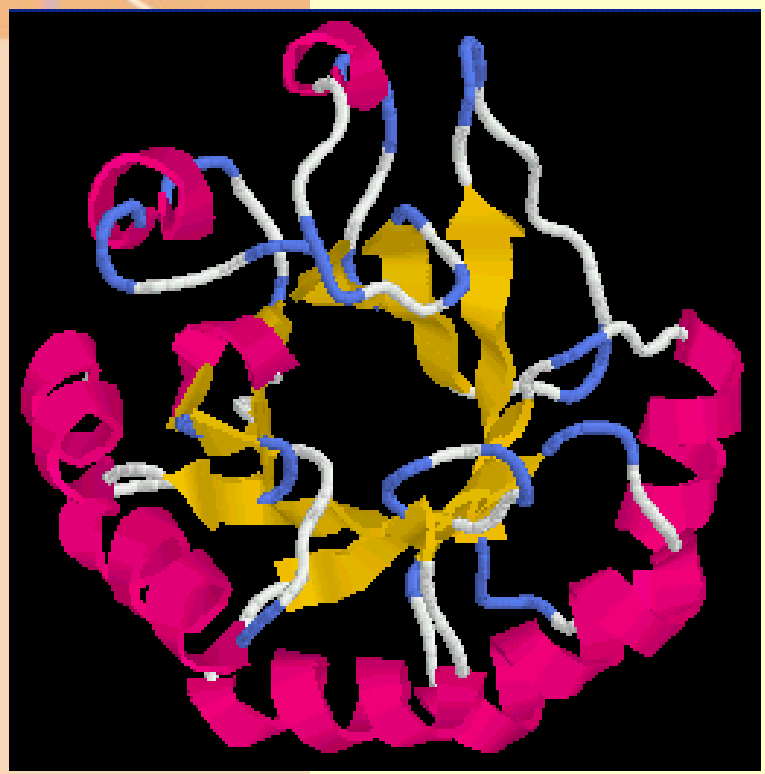


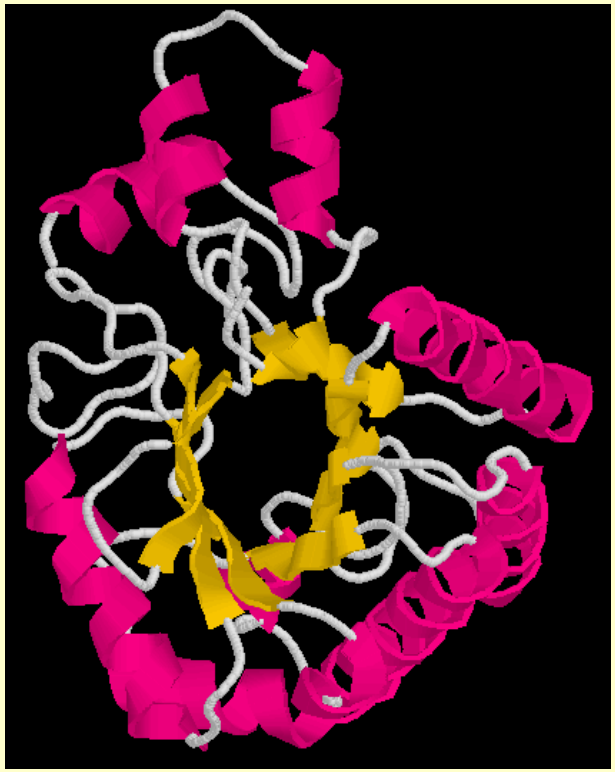


- **α/ β**and α+β
  - Closed
    - Barrel
    - Roll, ...
  - Open
    - Sandwich
    - Clam, ...

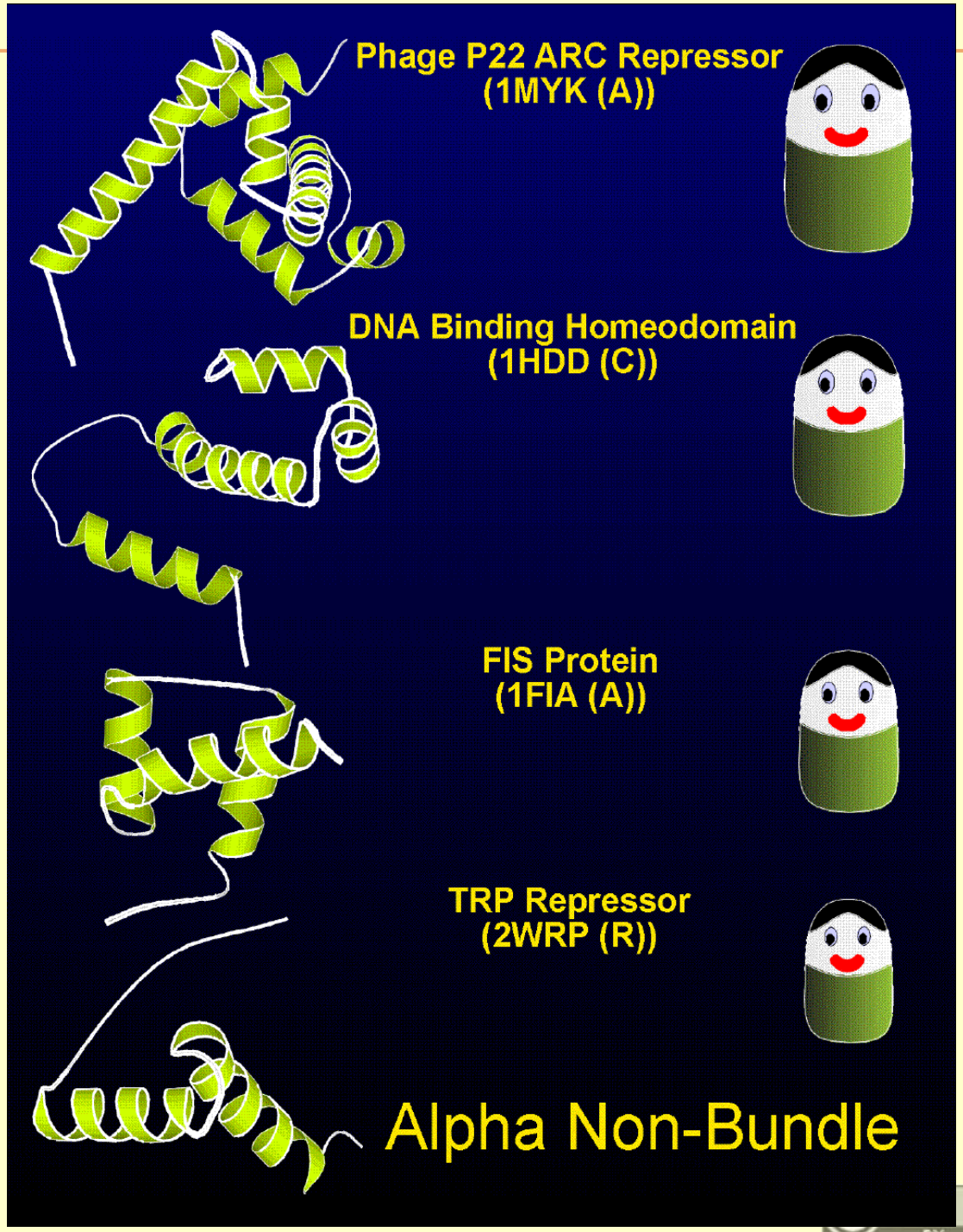


The TIM Barrel Fold

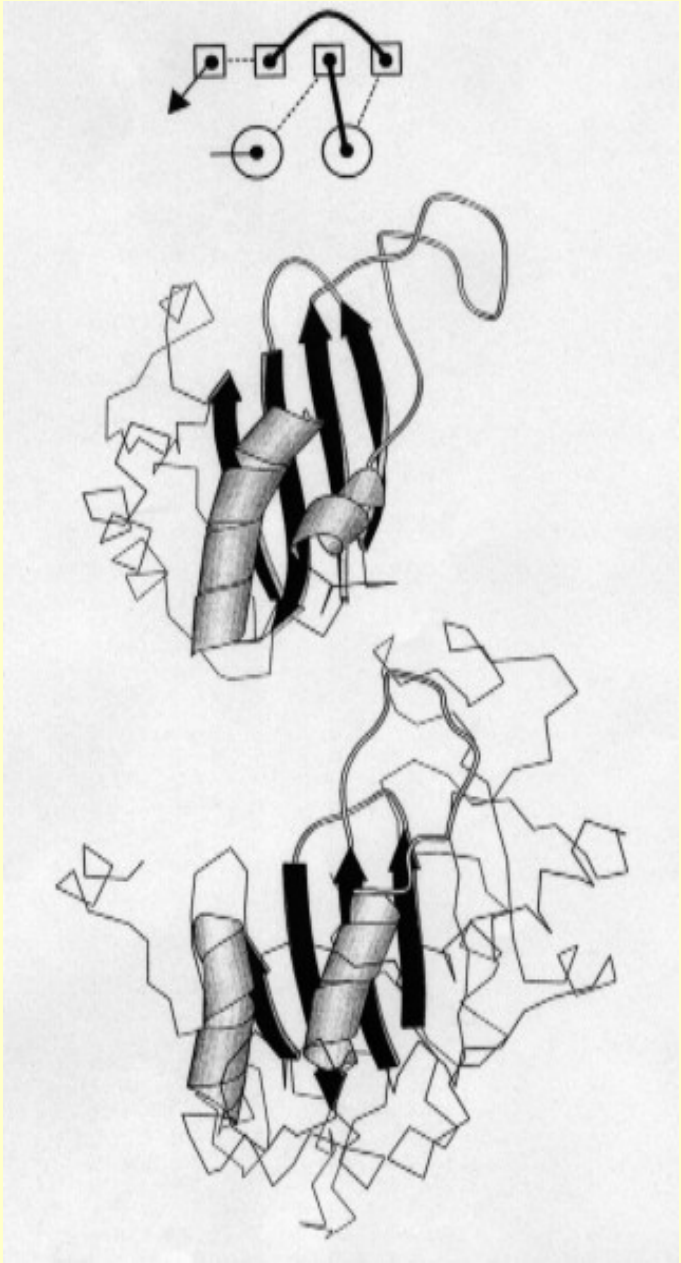




A Conceptual Problem ...



Fold versus Topology



Another example:
Globin
vs.
Colicin

PDB Protein Database
http://www.rcsb.org/pdb/

- Protein DataBase
  - Multiple Structure Viewers
  - Sequence & Structure Comparison Tools
  - Derived Data
    - SCOP
    - CATH
    - pFAM
    - Go Terms
  - Education on Protein Structure
  - Download Structures and Entire Database

PDB Protein Database

http://www.rcsb.org/pdb/

RCSB PDB PROTEIN DATA BANK

MyPDB Login

An Information Portal to Biological Macr

As of **Tuesday Feb 02, 2010 at 4 PM PST** there are 63093 St

WHAT'S NEW | **HELP** | **PRINT**

PDB ID or keyword: P0ACF0

- Home Hide
  - News & Publications
  - Policies
  - FAQ
  - Contact
  - Feedback
  - About Us
- Deposition Hide
  - All Deposit Services
  - Electron Microscopy
  - NMR
  - Validation Server
  - BioSync Beamline
  - Related Tools
- Search Hide
  - Advanced Search
  - Latest Release
  - Latest Publications
  - Sequence Search
  - Ligand Search
  - Unreleased Entries
  - Browse Database
  - Histograms
- Tools Hide
  - File Downloads
  - FTP Services
  - File Formats
  - Services: RESTful | SOAP
  - Widgets
  - Compare Structures
- Education Hide
  - Looking at Structures
  - Molecule of the Month
  - Educational Resources

A Resource for Studying Biological Macromolecules

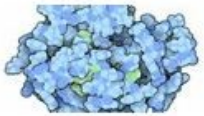
The PDB archive contains information about experimentally-determined structures of proteins, nucleic acids, and complex assemblies. As a member of the **wwPDB**, the RCSB PDB curates and annotates PDB data according to agreed upon standards.

The RCSB PDB also provides a variety of tools and resources. Users can perform simple and advanced searches based on annotations relating to sequence, structure and function. These molecules are visualized, downloaded, and analyzed by users who range from students to specialized scientists.

Molecule of the Month: Enhanceosome

Take a moment to ponder the form of your body: the shape of your face, the color of your eyes, the length of your fingers, the perfect articulation of your bones and muscles, the way your hair grows curly or straight. Now let your imagination travel inward, and think of the complex shapes and functions of your different cells, and the teeming molecular world inside each one. Remarkably, this amazing structure and form and function is specified by information in the genome, which encodes a mere 20,000-25,000 protein-coding genes. One of the great puzzles being pieced together by scientists is the mechanism by which these genes, and the methods used to control their expression, specify all of these different aspects of life. ■ **Read more ...** ■ **Previous Features**

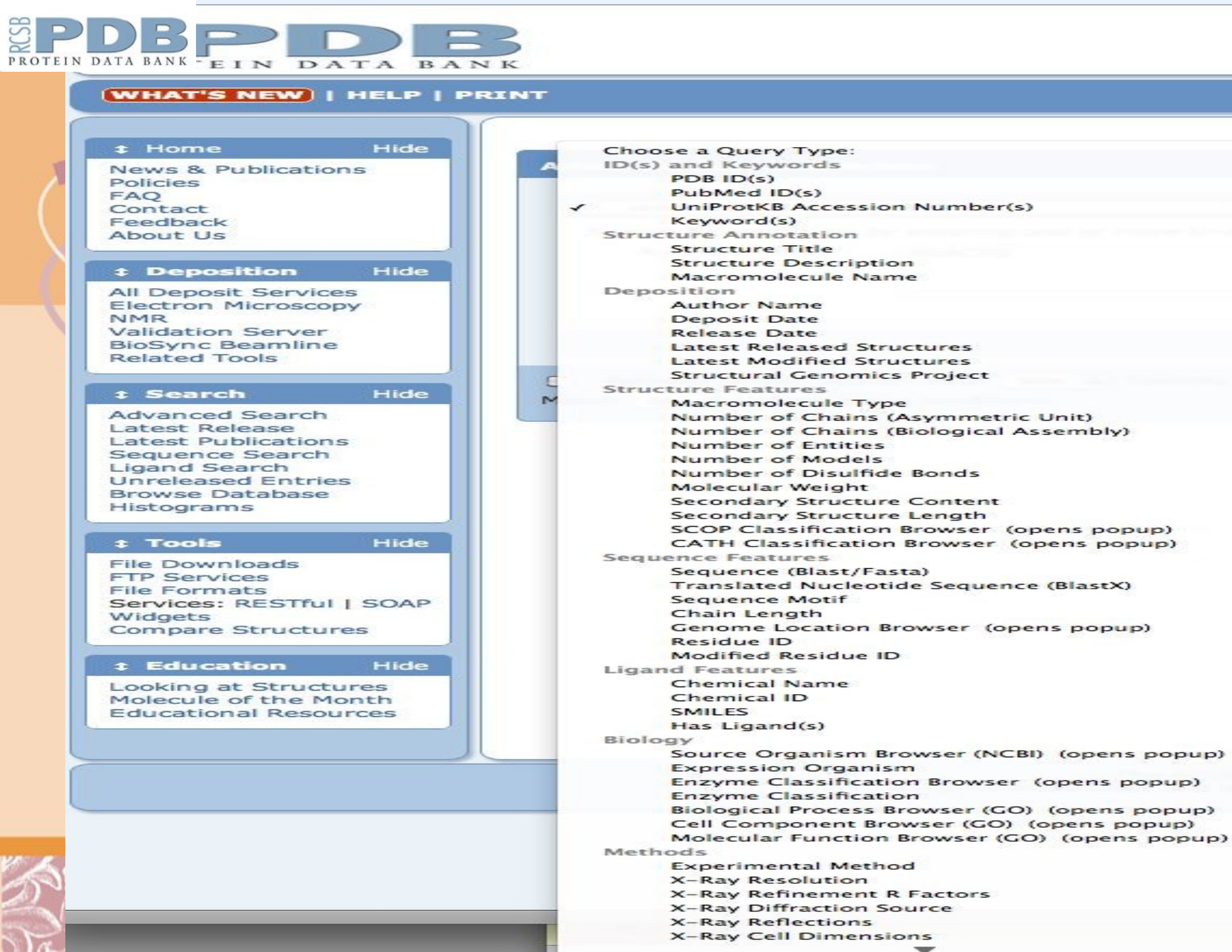
PSI Featured Molecule: Sugarcoating the surface: yeast Alg13



Many proteins in our cells are decorated with carbohydrate chains, which make the proteins more stable and assist with their function. Using NMR, PSI researchers now understand how this enzyme builds these essential carbohydrates.

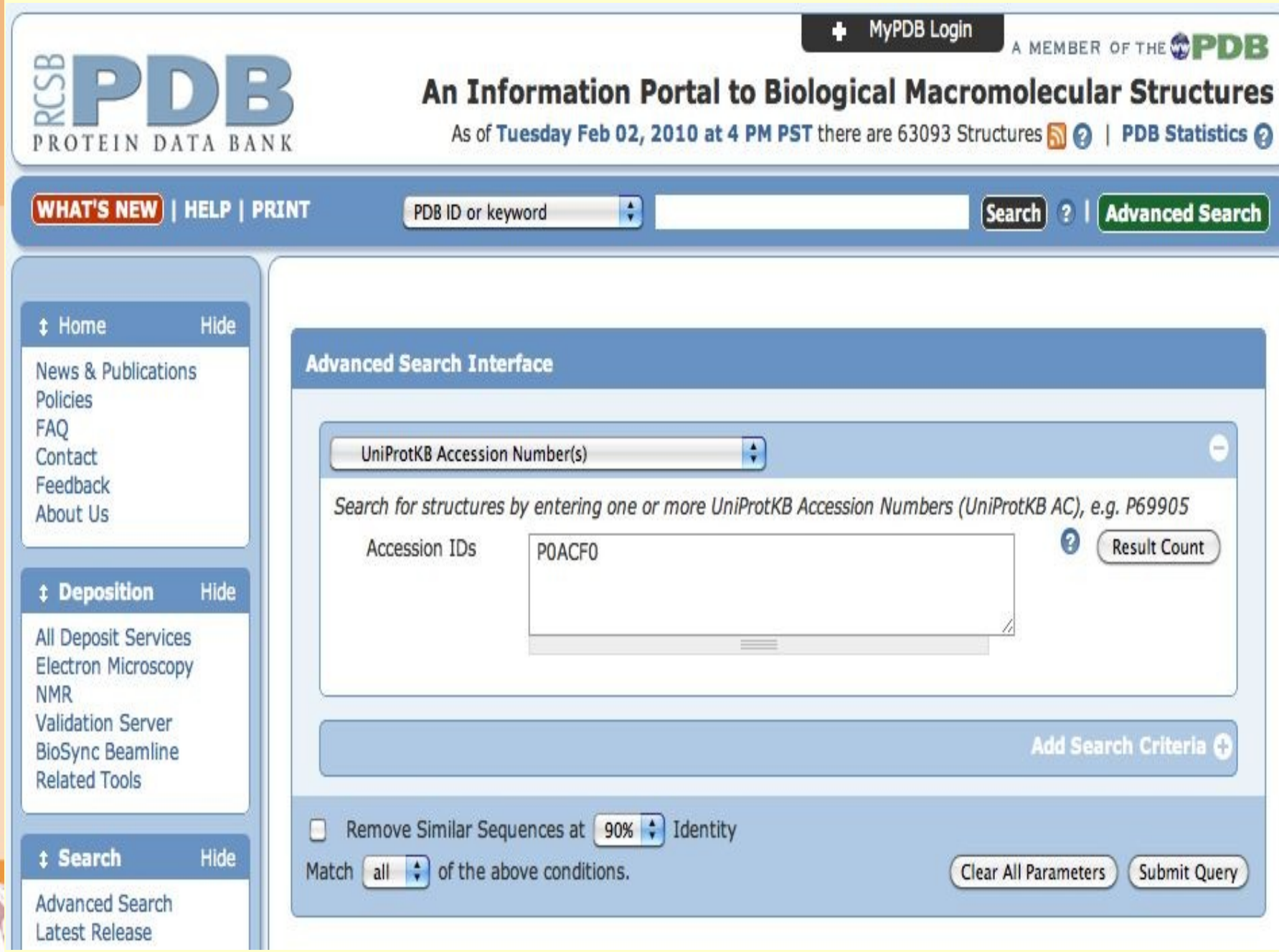
■ **Read more from the Structural Genomics Knowledgebase** ■ **Previous Features**

New user? Try the browser **compatibility check** and information on **Getting Started**.


RCSB PDB
PROTEIN DATA BANK
PDB
EIN DATA BANK
WHAT'S NEW | HELP | PRINT
Home Hide
News & Publications
Policies
FAQ
Contact
Feedback
About Us
Deposition Hide
All Deposit Services
Electron Microscopy
NMR
Validation Server
BioSync Beamline
Related Tools
Search Hide
Advanced Search
Latest Release
Latest Publications
Sequence Search
Ligand Search
Unreleased Entries
Browse Database
Histograms
Tools Hide
File Downloads
FTP Services
File Formats
Services: RESTful | SOAP
Widgets
Compare Structures
Education Hide
Looking at Structures
Molecule of the Month
Educational Resources
Choose a Query Type:
ID(s) and Keywords
PDB ID(s)
PubMed ID(s)
UniProtKB Accession Number(s)
Keyword(s)
Structure Annotation
Structure Title
Structure Description
Macromolecule Name
Deposition
Author Name
Deposit Date
Release Date
Latest Released Structures
Latest Modified Structures
Structural Genomics Project
Structure Features
Macromolecule Type
Number of Chains (Asymmetric Unit)
Number of Chains (Biological Assembly)
Number of Entities
Number of Models
Number of Disulfide Bonds
Molecular Weight
Secondary Structure Content
Secondary Structure Length
SCOP Classification Browser (opens popup)
CATH Classification Browser (opens popup)
Sequence Features
Sequence (Blast/Fasta)
Translated Nucleotide Sequence (BlastX)
Sequence Motif
Chain Length
Genome Location Browser (opens popup)
Residue ID
Modified Residue ID
Ligand Features
Chemical Name
Chemical ID
SMILES
Has Ligand(s)
Biology
Source Organism Browser (NCBI) (opens popup)
Expression Organism
Enzyme Classification Browser (opens popup)
Enzyme Classification
Biological Process Browser (GO) (opens popup)
Cell Component Browser (GO) (opens popup)
Molecular Function Browser (GO) (opens popup)
Methods
Experimental Method
X-Ray Resolution
X-Ray Refinement R Factors
X-Ray Diffraction Source
X-Ray Reflections
X-Ray Cell Dimensions

PDB Advanced Search for UniProt Entry

http://www.rcsb.org/pdb/



PDB Search Results

http://www.rcsb.org/pdb/

RCSB PDB PROTEIN DATA BANK

MyPDB Login

A MEMBER OF THE PD

An Information Portal to Biological Macromolecular Structur

As of **Tuesday Feb 02, 2010 at 4 PM PST** there are 63093 Structures | PDB Statistics

WHAT'S NEW | HELP | PRINT | PDB ID or keyword | Search | Advanced Sear

- Home Hide
  - News & Publications
  - Policies
  - FAQ
  - Contact
  - Feedback
  - About Us
- Deposition Hide
  - All Deposit Services
  - Electron Microscopy
  - NMR
  - Validation Server
  - BioSync Beamline
  - Related Tools
- Search Hide
  - Advanced Search
  - Latest Release
  - Latest Publications
  - Sequence Search
  - Ligand Search
  - Unreleased Entries
  - Browse Database
  - Histograms

Results:

Query (2 hits):

- Query Results
- Query Details
- Query History
- Save Query to MyPDB

- Tools Hide
  - File Downloads
  - FTP Services
  - File Formats

2 Structure Hits | 2 Citations | 2 Ligand Hits | GO Hits | SCOP Hits | CATH Hits

Simple query for a list of Uniprot Accession IDs: P0ACF0 | Query Options:

Display/Download: | Generate Reports: | Sort by: | Results per Page:

Displaying results 1 - 2 of 2 total | Page 1 c

2097

Crystal Structure of E. coli HU heterodimer

Chains	A
Characteristics	Release Date: 06-Mar-2007 Exp. Method: X-RAY DIFFRACTION Resolution: 2.45 Å
Classification	**DNA Binding Protein**
Compound	**Molecule:** DNA-binding protein HU-alpha **Polymer:** 1 **Type:** polypeptide(L) **Length:** 90 **Chains:** A **Molecule:** DNA-binding protein HU-beta **Polymer:** 2 **Type:** polypeptide(L) **Length:** 90 **Chains:** B
Authors	**Guo, F.**, **Adhya, S.**

1MUL

Crystal structure of the E. coli HU alpha2 protein

Chains	A
Characteristics	Release Date: 05-Aug-2003 Exp. Method: X-RAY DIFFRACTION Resolution: 2.30 Å
Classification	**DNA Binding Protein**
Compound	**Molecule:** DNA binding protein HU-alpha **Polymer:** 1 **Type:** polypeptide(L) **Length:** 90 **Chains:** A
Authors	**Ramstein, J.**, **Hervouet, N.**, **Coste, F.**, **Zelwer, C.**, **Oberto, J.**, **Castaing, B.**

PDB *E. coli* Hu Entry

http://www.rcsb.org/pdb/explore/explore.do?structureId=2O97

Summary | Derived Data | Sequence | Seq. Similarity | Literature | Biol. & Chem. | Methods | Geometry | Links

Crystal Structure of E. coli HU heterodimer

2097

Display Files ▾
Download Files ▾
Print this Page
Share this Page

DOI:10.2210/pdb2o97/pdb

Primary Citation

Spiral structure of Escherichia coli HU{alpha}beta provides foundation for DNA supercoiling.
Guo, F., Adhya, S.
(2007) Proc.Natl.Acad.Sci.Usa **104:** 4309-4314
PubMed: 17360520 **PubMedCentral:** PMC1838598 **DOI:** 10.1073/pnas.0611686104
Search Related Articles in PubMed

PubMed Abstract:

We determined the crystal structure of the Escherichia coli nucleoid-associated HUalphabeta protein by x-ray diffraction and observed that the heterodimers form multimers with octameric units in three potential arrangements, which may serve specialized roles in different DNA transaction reactions. It ...
[Read More & Search PubMed Abstracts]

Molecular Description Hide

Classification: DNA Binding Protein
Structure Weight: 18883.88

Molecule: DNA-binding protein HU-alpha
Polymer: 1 **Type:** polypeptide(L) **Length:** 90
Chains: A

Molecule: DNA-binding protein HU-beta
Polymer: 2 **Type:** polypeptide(L) **Length:** 90
Chains: B

Source Hide

Polymer: 1
Scientific Name: Escherichia coli **Expression System:** Escherichia coli

Polymer: 2
Scientific Name: Escherichia coli **Expression System:** Escherichia coli

Ligand Chemical Component Hide

Identifier	Name	Formula	Interaction View	Links
CL	CHLORIDE ION	Cl	Ligand Explorer	LE D H
NI	NICKEL (II) ION	Ni	Ligand Explorer	LE D H

Derived Data Hide

- SCOP Classification v1.75 - (2 Domains)
- CATH Classification v3.2.0 - (2 Domains)
- PFAM Classification - (2 Domains)
- GO Terms - (13 Terms)

Biological Assembly

More Images...

View in Jmol | SimpleViewer
Other Viewers ▾ | Protein Workshop

Biological assembly assigned by authors

Deposition Summary Hide

Authors: Guo, F., Adhya, S.

Deposition:	2006-12-13
Release:	2007-03-06
Last Modified (REVDAT):	2009-02-24

Experimental Details Hide

Method: X-RAY DIFFRACTION
Experimental Data: [EDS]
Resolution[Å]: 2.45
R-Value: 0.228 (obs.)
R-Free: 0.264
Space Group: $I\,4_1$

Unit Cell:

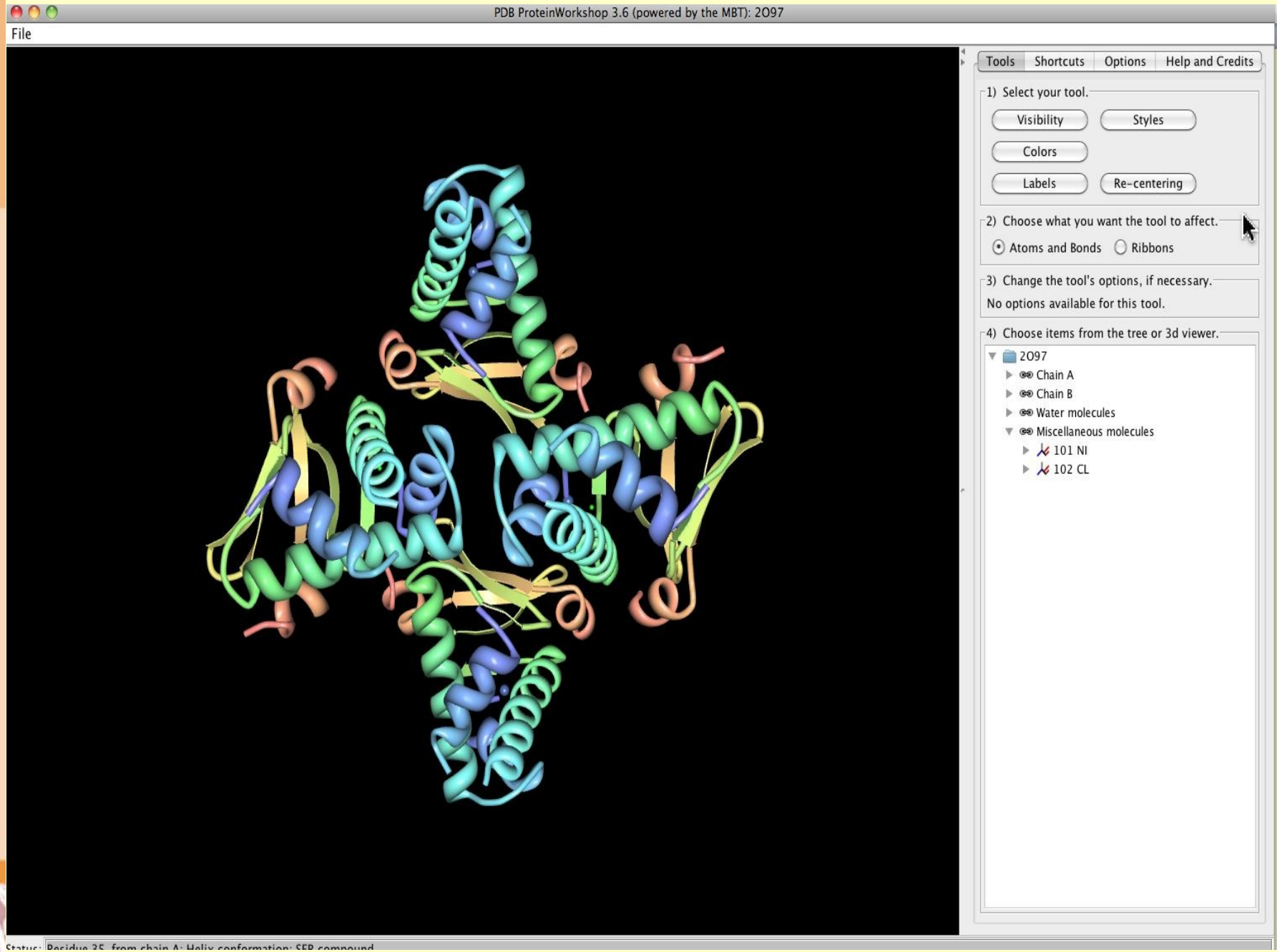
Length [Å]	Angles [°]
a = 82.92	α = 90.00
b = 82.92	β = 90.00
c = 61.05	γ = 90.00

PDB SimpleViewer

http://www.rcsb.org/pdb/

PDB Protein Workshop View

http://www.rcsb.org/pdb/



PDB Derived Data

http://www.rcsb.org/pdb/

Summary | **Derived Data** | Sequence | Seq. Similarity | Literature | Biol. & Chem. | Methods | Geometry | Links

Derived Data

2097

Display Files ▾
Download Files ▾
Print this Page
Share this Page

Derived Data: SCOP Classification (version 1.75) Hide

Domain Info	Class	Fold	Superfamily	Family	Domain	Species
d2o97a1	All alpha proteins	IHF-like DNA-binding proteins	IHF-like DNA-binding proteins	Prokaryotic DNA-bending protein	HU protein	Escherichia coli [TaxId: 562]
d2o97b1	All alpha proteins	IHF-like DNA-binding proteins	IHF-like DNA-binding proteins	Prokaryotic DNA-bending protein	HU protein	Escherichia coli, beta-isoform [TaxId: 562]

Derived Data: CATH Classification (version v3.2.0) Hide

Domain	Class	Architecture	Topology	Homology
2o97A00	Few Secondary Structures	Irregular	HU Protein; Chain A	HU Protein, subunit A
2o97B00	Few Secondary Structures	Irregular	HU Protein; Chain A	HU Protein, subunit A

Derived Data: PFAM Classification Hide

Chain	PFAM Accession	PFAM ID	Description	Type	Clan ID
B	PF00216	Bac_DNA_binding	Bacterial DNA-binding protein	Domain	
A	PF00216	Bac_DNA_binding	Bacterial DNA-binding protein	Domain	

Derived Data: GO Terms Hide

Polymer	Molecular Function	Biological Process	Cellular Component
DNA-binding protein HU-alpha (2O97:A)	∘ DNA binding ∘ protein binding ∘ transcription activator activity ∘ transcription repressor activity	∘ transcription ∘ chromosome condensation	∘ membrane
DNA-binding protein HU-beta (2O97:B)	∘ DNA binding ∘ protein binding ∘ transcription activator activity ∘ transcription repressor activity	∘ transcription ∘ chromosome condensation	∘ none

Molecule of the Month: Enhanceosome

http://www.rcsb.org/pdb/static.do?p=education_discussion/molecule_of_the_month/current_month.html

February 2010 **Molecule of the Month** by David Goodsell
Previous Features
doi: **10.2210/rcsb_pdb/mom_2010_2**

Enhanceosome

keywords: transcription factor, gene expression, CBP, CREB-binding protein, transcriptional enhancers, enhanceosome

Take a moment to ponder the form of your body: the shape of your face, the color of your eyes, the length of your fingers, the perfect articulation of your bones and muscles, the way your hair grows curly or straight. Now let your imagination travel inward, and think of the complex shapes and functions of your different cells, and the teeming molecular world inside each one. Remarkably, this amazing structure and form and function is specified by information in the genome, which encodes a mere 20,000-25,000 protein-coding genes. One of the great puzzles being pieced together by scientists is the mechanism by which these genes, and the methods used to control their expression, specify all of these different aspects of life.

Combinatorial Control

In order to specify which gene will be expressed in a given situation, your cells use a diverse collection of DNA-binding proteins to control access to the DNA. Surprisingly, there are relatively few of these proteins: by some estimates, the human genome encodes about 2,600 of them. But then, the capabilities of this limited set are greatly expanded by using them in combination, by requiring two or more to bind simultaneously to activate a gene. In this way, each protein may be used in many ways and the spectrum of responses is far more varied.

Enhancing Transcription

The assembly of DNA and proteins pictured here is a transcriptional enhanceosome (PDB entries **1t2k**, **2pi0**, **2o6g** and **2o61**) that controls expression of interferon-beta, an important protein for fighting viral infection. When the cell is infected by viruses, several different DNA-binding proteins are produced, including ATF-2/c-Jun (in green at the top), interferon response factors (IRF, shown in turquoise at the center), and nuclear factor kB (NF-kB, shown in blue and magenta at the bottom). Individually, each one is not sufficient to activate the gene, and each one also plays other roles in the activation of other genes (for instance, Nf-kB is also important in immune responses, inflammation, apoptosis, and many other processes). But when they all bind together, they activate the gene and interferon is made.

Next: **Integrating the Signal**

NCBI Structure Database

http://www.ncbi.nlm.nih.gov/Structure/

- Macromolecular Structures
- Related Structures
- View Aligned Structures & Sequences
- Cn3D: Downloadable Structure & Sequence Viewer
- CDD: Conserved Domain Database
  - CD-Search: Protein Sequence Queries
  - CD-TREE: Protein Classification Downloadable Application
  - CDART: Conserved Domain Architecture Tool
- PubChem: Small Molecules and Biological Activity
- Biological Systems: BioCyc, KEGG and Reactome Pathways
- MMDB: Molecular Modeling Database
- CBLAST: BLAST sequence against PDB and Related Structure Database
- IBIS: Inferred Biomolecular Interaction Server
- VAST Search: Structure Alignment Tool

NCBI Structure Database

http://www.ncbi.nlm.nih.gov/Structure/

NCBI Structure

My NCBI
Welcome brutlag. [Sign Out]

All Databases | PubMed | Nucleotide | Protein | Genome | Structure | OMIM | PMC | Journals | Books

Search Structure for [ ] Go Clear

Limits | Preview/Index | History | Clipboard | Details

Molecular Modeling Database (MMDB)

RESOURCES | SEARCH METHODS | HOW TO | HELP | NEWS | FTP | PUBLICATIONS | DISCOVER

Hints on Finding 3D Macromolecular Structures

- This page is used for searching by text term (other search methods allow queries by protein sequence)
- Enter one or more search terms (e.g., chloride channel)
- Use search fields and other Advanced Search options (Limits, Preview/Index, and History) to refine a search
- Boolean operators AND, OR, NOT must be in upper case
- Use quotes to force a phrase search (e.g., "voltage gated")
- Use a wildcard (e.g., glycol*[title]) to search for a word stem
- Search results and structure record displays are described in the help document.

About the Database

Three dimensional structures provide a wealth of information on the biological function, on mechanisms linked to the function, and on the evolutionary history of and relationships between macromolecules. Most 3D-structure data are obtained from X-ray crystallography and NMR-spectroscopy.

The Molecular Modeling DataBase (MMDB), also known as "Entrez Structure," is a database of experimentally determined structures obtained from the RCSB Protein Data Bank (PDB). MMDB is developed by the Structure Group of the NCBI Computational Biology Branch. The data processing procedure at NCBI results in the addition of a number of useful features that facilitate computation on the data and link them to many other data types in the Entrez system. The help document and how to pages provides examples of how the database can be used.

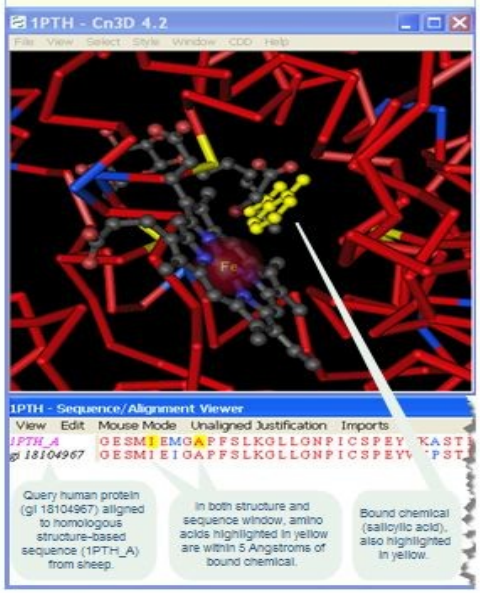
The structure database is considerably smaller than Entrez's Protein or Nucleotide databases, but a large fraction of all known protein sequences have homologs in this set, and one may often learn more about a protein by examining 3-D structures of its homologs. These are accessible as "Related Structures" in the "Links" menu of Entrez Protein sequence records (illustrated example). It is then possible to align the query protein to the structure-based sequence, as shown in the illustration on this page.

Additional resources can be used along with MMDB to interactively view the structures, find similar 3D structures, learn about the types of interactions and bound chemicals that have been found to exist among the similar 3D structures, and more.

RETRIEVE STRUCTURES THAT HAVE:		
Protein Only	DNA Only	RNA Only
Protein + Chemical	DNA + Chemical	RNA + Chemical
Protein + DNA	Protein + RNA	DNA + RNA

The "How To" page provides tips for searching by gene/protein product, molecule type, and more.

SEQUENCE-STRUCTURE-FUNCTION RELATIONSHIPS
Example: structural basis of aspirin activity (1PTH)

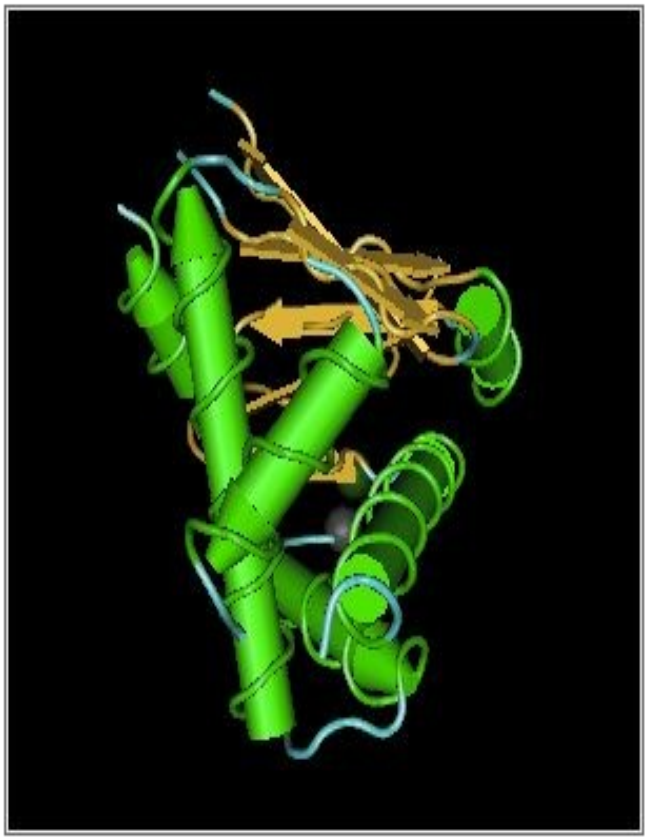


NCBI Structure Database

NCBI

Structure Summary
MMDB

HOME | SEARCH | SITE MAP | Entrez | Structure | Protein | CDD | PubMed | Taxonomy | PubChem | Help | Cn3D



MMDB ID: 44665 **PDB ID:** 2O97

Search PDB or MMDB ID

Reference: Guo F, Adhya S *Spiral structure of Escherichia coli HUalphabeta provides foundation for DNA supercoiling* Proc. Natl. Acad. Sci. U. S. A. v104, p.4309-4314

We determined the crystal structure of the Escherichia coli nucleoid-associated HUalphabeta protein by x-ray diffraction and observed that the heterodimers form multimers with octameric units in three potential arrangements, which may serve specialized roles in different DNA transaction reactions. It is of special importance that one of the structures forms spiral filaments with left-handed rotations....

» View full abstract

Description: Crystal Structure Of E. Coli Hu Heterodimer.
Deposition: 2006/12/13
Taxonomy: Escherichia coli
Related Structure: VAST

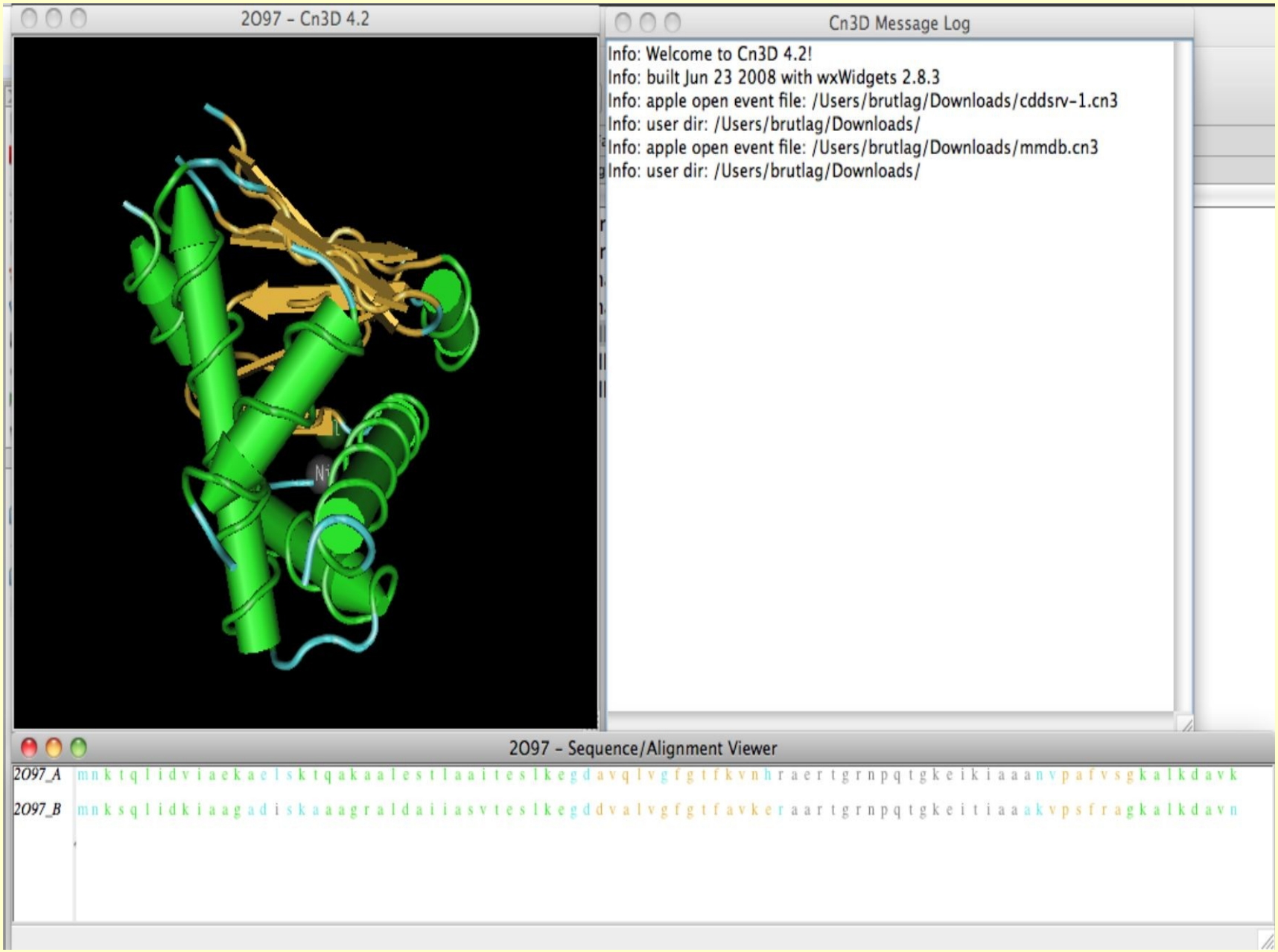
Structure View in Cn3D | Structure View in RasMol

Tasks: Display **Drawing:** Backbone

Download Cn3D | View Cn3D Tutorial

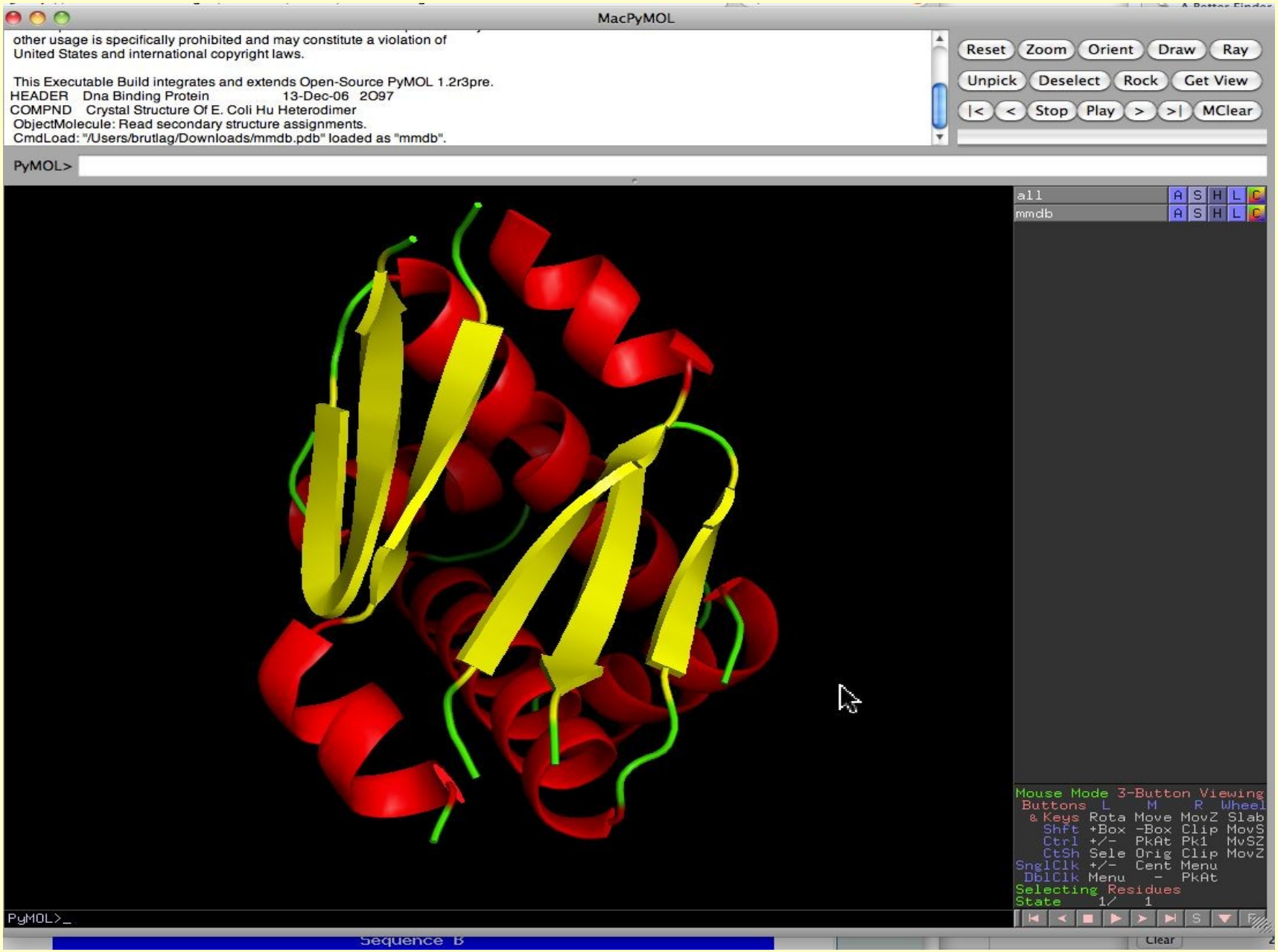
NCBI Cn3D Viewer

http://www.ncbi.nlm.nih.gov/Structure/CN3D/cn3d.shtml



PyMol PDB Structure Viewer

http://www.pymol.org/



Databases of Protein Folds

- SCOP (http://scop.berkeley.edu/)
  - Structural Classification of Proteins
  - Class-Fold-Superfamily-Family
  - Manual assembly by inspection
- Superfamily (http://supfam.mrc-lmb.cam.ac.uk/SUPERFAMILY/)
  - HMM models for each SCOP fold
  - Fold assignments to all genome ORFs
  - Assessment of specificity/sensitivity of structure prediction
  - Search by sequence, genome and keywords
- CATH (http://www.biochem.ucl.ac.uk/bsm/cath/)
  - Class - Architecture - Topology - Homologous Superfamily
  - Manual classification at Architecture level
  - Automated topology classification using SSAP (Orengo & Taylor)
- FSSP (http://www2.embl-ebi.ac.uk/dali/fssp/)
  - Fully automated using the DALI algorithm (Holm & Sander)
  - No internal node annotations
  - Structural similarity search using DALI

SCOP Database of Protein Folds

http://scop.berkeley.edu/

Structural Classification of Proteins

Welcome to **SCOP**: Structural Classification of Proteins.
1.75 release (June 2009)

38221 PDB Entries. 1 Literature Reference. 110800 Domains. (excluding nucleic acids and theoretical models).
Folds, superfamilies, and families statistics here.
New folds superfamilies families.
List of obsolete entries and their replacements.

Authors. Alexey G. Murzin, John-Marc Chandonia, Antonina Andreeva, Dave Howorth, Loredana Lo Bartlett G. Ailey, Steven E. Brenner, Tim J. P. Hubbard, and Cyrus Chothia. scop@mrc-lmb.cam.ac.uk
Reference: Murzin A. G., Brenner S. E., Hubbard T., Chothia C. (1995). SCOP: a structural classification of p database for the investigation of sequences and structures. *J. Mol. Biol.* 247, 536-540. [PDF]
Recent changes are described in: Lo Conte L., Brenner S. E., Hubbard T.J.P., Chothia C., Murzin A. (2002) database in 2002: refinements accommodate structural genomics. *Nucl. Acid Res.* 30(1), 264-267. [PDF], Andreeva A., Howorth D., Brenner S.E., Hubbard T.J.P., Chothia C., Murzin A.G. (2004). SCOP database i refinements integrate structure and sequence family data. *Nucl. Acid Res.* 32:D226-D229. [PDF], and Andreeva A., Howorth D., Chandonia J.-M., Brenner S.E., Hubbard T.J.P., Chothia C., Murzin A.G. (2007) growth and its impact on the SCOP database: new developments. *Nucl. Acid Res.* advance doi:10.1093/nar/gkm993. [PDF].

Access methods

- Enter SCOP at the **top of the hierarchy**
- Keyword search of SCOP entries
- SCOP parseable files (MRC site)
- All SCOP releases and reclassified entry history (MRC site)
- **pre-SCOP - preview of the next release**
- SCOP domain sequences and pdb-style coordinate files (ASTRAL)
- Hidden Markov Model library for SCOP superfamilies (SUPERFAMILY)
- Structural alignments for proteins with non-trivial relationships (SISYPHUS)

- Online resources of potential interest to SCOP users

SCOP **mirrors** around the world may speed your access.

SCOP Hierarchy

http://scop.berkeley.edu/data/scop.b.html

Structural Classification of Proteins

Root: scop

Classes:

1. All alpha proteins [46456] (284)
2. All beta proteins [48724] (174)
3. Alpha and beta proteins (a/b) [51349] (147)
   Mainly parallel beta sheets (beta-alpha-beta units)
4. Alpha and beta proteins (a+b) [53931] (376)
   Mainly antiparallel beta sheets (segregated alpha and beta regions)
5. Multi-domain proteins (alpha and beta) [56572] (66)
   Folds consisting of two or more domains belonging to different classes
6. Membrane and cell surface proteins and peptides [56835] (58)
   Does not include proteins in the immune system
7. Small proteins [56992] (90)
   Usually dominated by metal ligand, heme, and/or disulfide bridges
8. Coiled coil proteins [57942] (7)
   Not a true class
9. Low resolution protein structures [58117] (26)
   Not a true class
10. Peptides [58231] (121)
    Peptides and fragments. Not a true class
11. Designed proteins [58788] (44)
    Experimental structures of proteins with essentially non-natural sequences. Not a true class

Enter search key: Search

Generated from scop database 1.75 with scopm 1.101 on Wed Jun 3 10:42:06 2009

SCOP Alpha and Beta Proteins

http://scop.berkeley.edu/data/scop.b.d.html

Structural Classification of Proteins

Class: Alpha and beta proteins (a/b)

Mainly parallel beta sheets (beta-alpha-beta units)

Lineage:

1. Root: scop
2. Class: Alpha and beta proteins (a/b) [51349]
   Mainly parallel beta sheets (beta-alpha-beta units)

Folds:

1. TIM beta/alpha-barrel [51350] (33)
   contains parallel beta-sheet barrel, closed; n=8, S=8; strand order 12345678
   the first seven superfamilies have similar phosphate-binding sites
2. NAD(P)-binding Rossmann-fold domains [51734] (1)
   core: 3 layers, a/b/a; parallel beta-sheet of 6 strands, order 321456
   The nucleotide-binding modes of this and the next two folds/superfamilies are similar
3. FAD/NAD(P)-binding domain [51904] (1)
   core: 3 layers, b/b/a; central parallel beta-sheet of 5 strands, order 32145; top antiparallel beta-sheet of 3 strands, meander
4. Nucleotide-binding domain [51970] (1)
   3 layers: a/b/a; parallel beta-sheet of 5 strands, order 32145; Rossmann-like
5. MurCD N-terminal domain [51983] (1)
   3 layers: a/b/a; parallel beta-sheet of 5 strands, order 32145; incomplete Rossmann-like fold; binds UDP group
6. 7-stranded beta/alpha barrel [51988] (3)
   variant of beta/alpha barrel; parallel beta-sheet barrel, closed, n=7, S=8; strand order 1234567; some members may have fewer strands

SCOP TIM Barrels

http://scop.berkeley.edu/data/scop.b.d.b.html

Structural Classification of Proteins

Fold: TIM beta/alpha-barrel

contains parallel beta-sheet barrel, closed; n=8, S=8; strand order 12345678
the first seven superfamilies have similar phosphate-binding sites

Lineage:

1. Root: scop
2. Class: Alpha and beta proteins (a/b) [51349]
   Mainly parallel beta sheets (beta-alpha-beta units)
3. Fold: TIM beta/alpha-barrel [51350]
   contains parallel beta-sheet barrel, closed; n=8, S=8; strand order 12345678
   the first seven superfamilies have similar phosphate-binding sites

Superfamilies:

1. Triosephosphate isomerase (TIM) [51351] (1)
   Superfamily
2. Ribulose-phoshate binding barrel [51366] (6)
   Superfamily
3. Thiamin phosphate synthase [51391] (1)
   Superfamily
4. Pyridoxine 5'-phosphate synthase [63892] (1)
   Superfamily
5. FMN-linked oxidoreductases [51395] (1)
   Superfamily
6. Inosine monophosphate dehydrogenase (IMPDH) [51412] (1)
   The phosphate moiety of substrate binds in the 'common' phosphate-binding site

SCOP Thiamin Phosphate Synthase

http://scop.berkeley.edu/data/scop.b.d.b.d.A.html

Structural Classification of Proteins

Superfamily: Thiamin phosphate synthase

Superfamily

Lineage:

1. Root: scop
2. Class: Alpha and beta proteins (a/b) [51349]
   Mainly parallel beta sheets (beta-alpha-beta units)
3. Fold: TIM beta/alpha-barrel [51350]
   contains parallel beta-sheet barrel, closed; n=8, S=8; strand order 12345678
   the first seven superfamilies have similar phosphate-binding sites
4. Superfamily: Thiamin phosphate synthase [51391]
   Superfamily

Families:

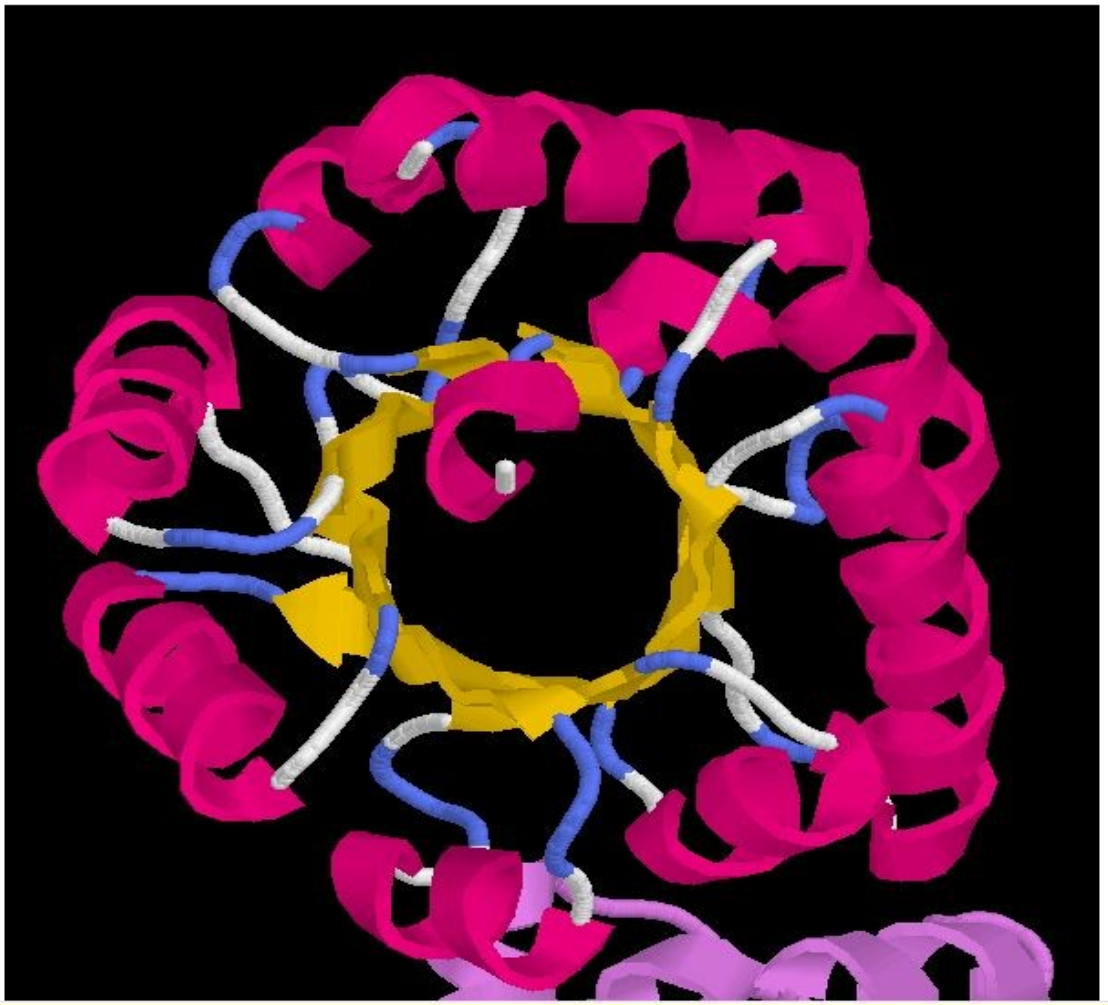
1. Thiamin phosphate synthase [51392] (2)
   1. Thiamin phosphate synthase [51393]
      1. Bacillus subtilis [TaxId: 1423] [51394] (8)
      2. Archaeon (Pyrococcus furiosus) [TaxId: 2261] [110344] (1)
         SQ Q8U192

SCOP Thiamin Phosphate Synthase Entry

http://scop.berkeley.edu/

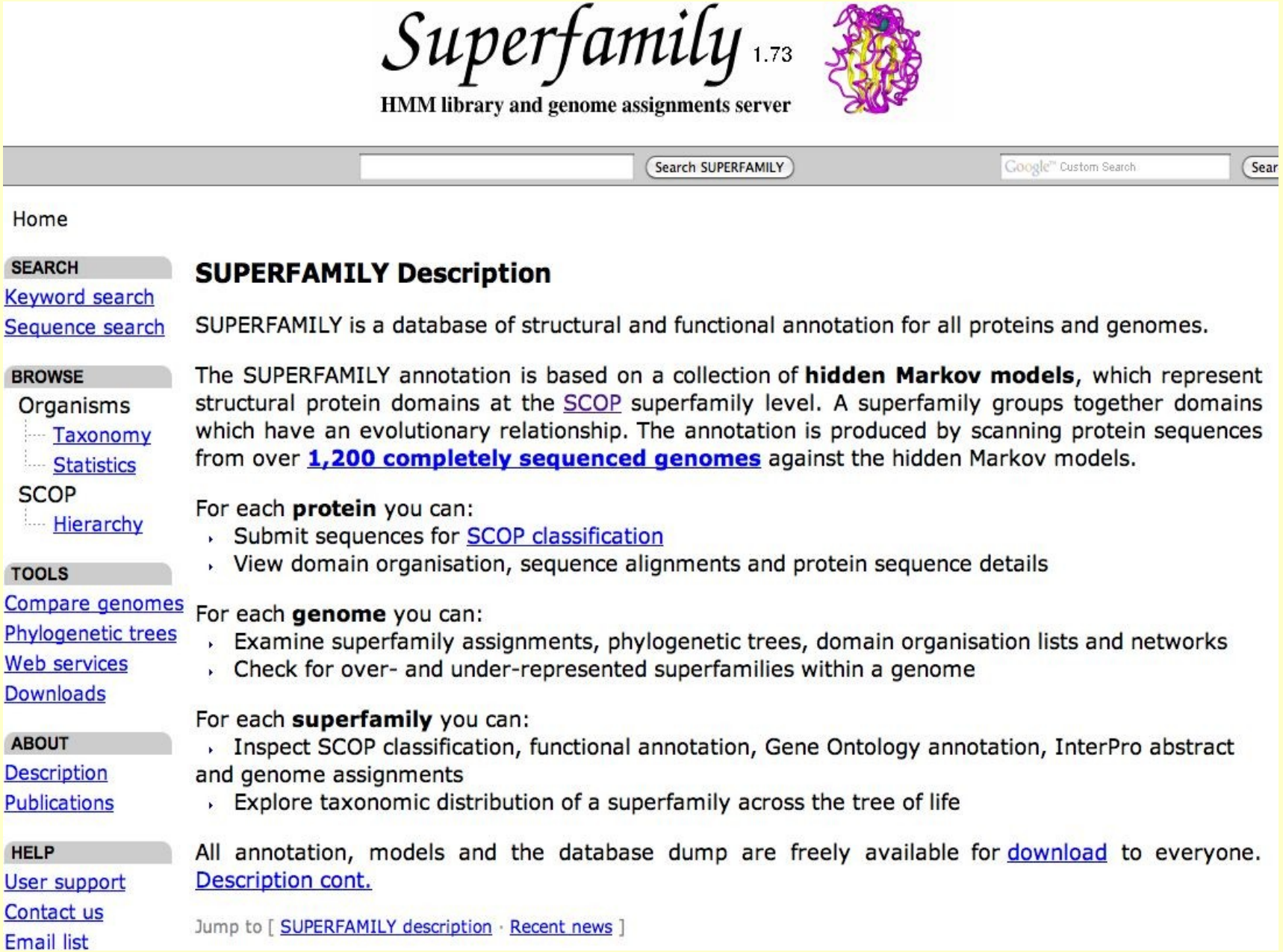
Structural Classification of Proteins

Chime display of PDB entry 1xi3 , chain a :
Click to display: ☒ chain only, ☒ whole structure.



SuperFamily HMM Fold Library

http://supfam.mrc-lmb.cam.ac.uk/SUPERFAMILY/



Superf
HMM library and gen

Major Features

Sequence search	Submit your protein, or DNA, sequence for superfamily and family level classifi
Keyword search	Search for superfamily names, family names, species names, sequence ID IDs, PDB IDs and hidden Markov model IDs.
Domain assignments	Domain assignments, alignments and architectures for completely se eukaryotic, and prokaryotic organisms. Additional sequence collections a included.
Comparative genomics tools	Browse unusual (over- and under-represented) superfamilies and families, domain pair lists and graphs, unique domain pairs, domain combinations, architecture co-occurrence networks and domain distribution across ta kingdoms for each organism.
Genome statistics	For each genome: number of sequences, number of sequences with assi percentage of sequences with assignment, percentage total sequence co number of domains assigned, number of superfamilies assigned, number of assigned, average superfamily size, percentage produced by duplication, sequence length, average length matched, number of domain pairs and nu unique domain architectures.
Superfamily annotation	InterPro have added abstracts for 1,052 superfamilies, and 763 superfamilies h some Gene Ontology (GO) annotation. Mapping file between SUPERFAMILY and SUPERFAMILY2go.
Functional annotation	Functional annotation of SCOP 1.73 superfamilies. By Christine Vogel.
Phylogenetic trees	Genome combinations, or specific clades, can be displayed as individual trees. trees are based on protein domain architecture data for all genomes in SUPERF and are generated using heuristic parsimony methods.
Similar domain architectures	Finds the 10 domain architectures which are most similar to a domain architect interest.
Profile comparison	For finding remote domain matches. Available when the sequence search fails significant match. Profile comparison (PRC) for aligning and scoring two profil Markov models by Martin Madera.
Hidden Markov models	Produce SCOP domain assignments for your sequences using the SUPE models. HMM visualisation by Martin Madera, e.g. model 0045110.
Web services	Distributed Annotation Server and linking to SUPERFAMILY.
Downloads	Sequences, assignments, models, MySQL database and scripts - updated week

UPERFAMILY Assignments for Genomes and Sequence Collections

he assignments are organised into four tables: Model organisms, Strains/versions, Longest transcript per gene and Other
lick on a table heading to sort on that column.
/e put considerable effort into classifying prokaryotes as either model species or strains. Please let us know if yo
assification. The "longest transcript per gene" versions of the eukaryotic genomes may prove useful in eliminating bias f
number of duplicate genomes are included. Originally these duplicate genomes used different sequence identifiers.
lease contact superfamily@mrc-lmb.cam.ac.uk to have a genome or some other data set analysed and/or added.
rganisms can be browsed by taxonomy on the taxonomy page. All data is available for download.

Model organisms

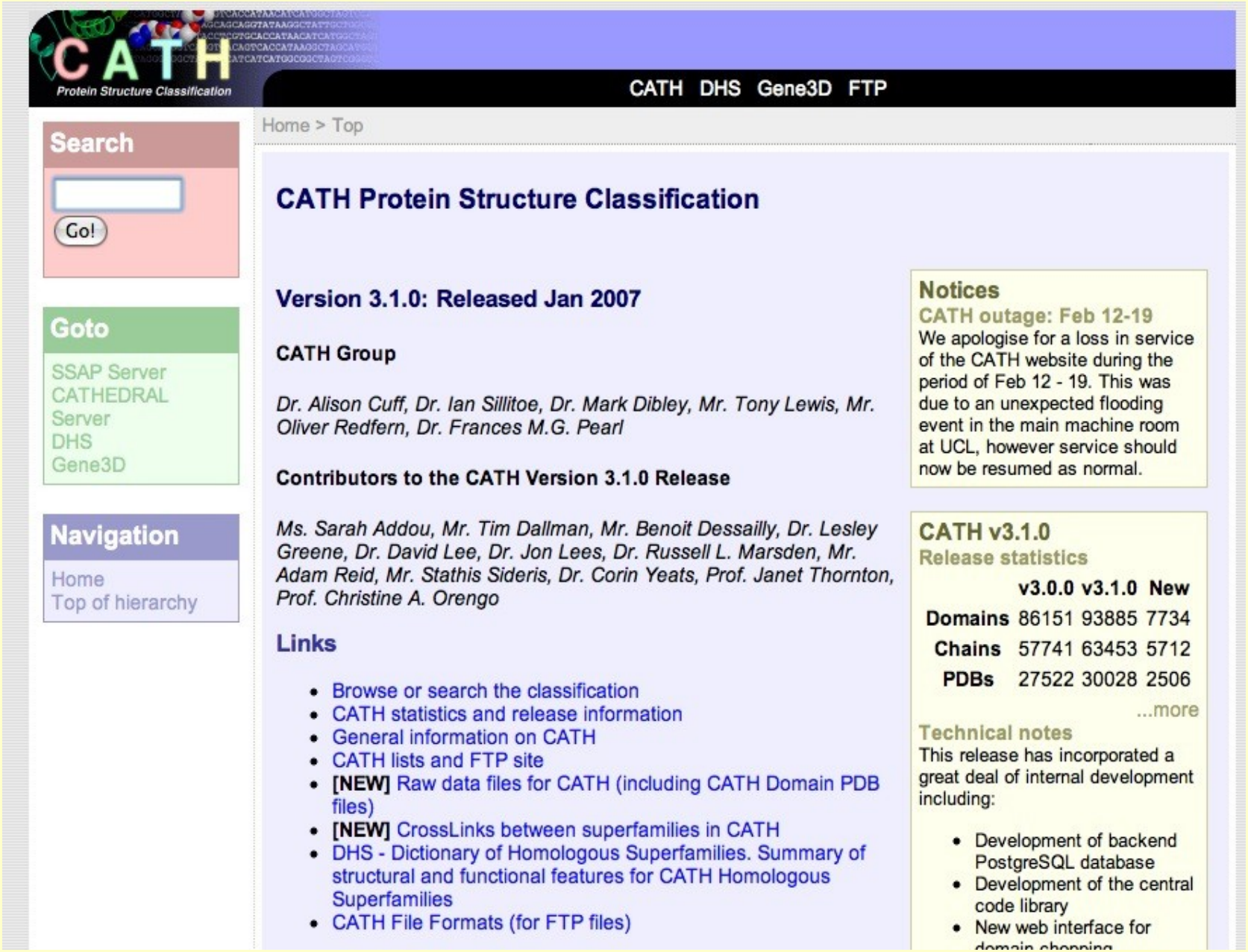
Click on a genome name	Dom	No. of sequences	No. with assignment	% with assign.	% total sequence coverage	No. of domains assigned	No. of supfam.s	No. of fam.s	Average Supfam. size	% produce by duplicati
omo sapiens 49_36k (all transcripts)	E	46591	30712	66	41	64225	1056	1299	60.8	98
an troglodytes 49_21h (all transcripts)	E	33137	21709	66	41	45312	1045	1276	43.4	98
orilla gorilla 52_1 (all transcripts)	E	16782	11231	67	37	21208	987	1110	21.5	95
ongo pygmaeus 49_1 (all transcripts)	E	23409	15489	66	41	31089	1042	1258	29.8	97
acaca mulatta 49_10h (all transcripts)	E	36384	24355	67	41	48417	1052	1278	46	98
allithrix jacchus 56_3 (all transcripts)	E	41941	28547	68	40	58578	1045	1275	56.1	98
tolemur garnettii 49_1c (all transcripts)	E	15448	10483	68	38	19713	965	1070	20.4	95
icrocebus murinus 49_1 (all transcripts)	E	16319	11105	68	39	21224	989	1146	21.5	95
arsius syrichta 51_1 (all transcripts)	E	13561	9238	68	38	17929	934	1023	19.2	95
attus norvegicus 49_34s (all transcripts)	E	32948	23088	70	43	43189	1039	1265	41.6	98
us musculus 49_37b (all transcripts)	E	39665	26987	68	42	53539	1054	1295	50.8	98
permophilus tridecemlineatus 49_1e (all transcripts)	E	14830	9948	67	37	18280	969	1039	18.9	95
ipodomys ordii 51_1 (all transcripts)	E	15750	10967	70	39	20801	1003	1148	20.7	95
avia porcellus 51_3 (all transcripts)	E	19774	14139	72	43	26512	1045	1266	25.4	96
ryctolagus cuniculus 49_1f (all transcripts)	E	15438	10414	67	38	18779	948	1073	19.8	95
chotona princeps 49_1 (all transcripts)	E	15843	10899	69	39	20729	997	1139	20.8	95
upaia belangeri 49_1d (all transcripts)	E	15462	10477	68	38	19113	972	1092	19.7	95
us scrofa 56_9 (all transcripts)	E	19082	13058	68	43	23298	986	1144	23.6	96
os taurus 49_3f (all transcripts)	E	27694	18751	68	42	36189	1042	1272	34.7	97
icugna pacos 51_1 (all transcripts)	E	11704	7963	68	38	14748	916	993	16.1	94
ursiops truncatus 51_1 (all transcripts)	E	16492	11494	70	39	23236	1026	1223	22.6	96
anis familiaris 49_2q (all transcripts)	E	25558	18341	72	43	33636	1042	1273	32.3	97
elis catus 49_1c (all transcripts)	E	14843	9993	67	42	18561	967	1064	19.2	95
quus caballus 49_2 (all transcripts)	E	22748	15979	70	42	34094	1027	1243	33.2	97
yotis lucifugus 49_1e (all transcripts)	E	16232	11114	68	38	20935	1005	1127	20.8	95
eropus vampyrus 51_1 (all transcripts)	E	16930	11839	70	39	23263	1024	1224	22.7	96
orex araneus 49_1c (all transcripts)	E	13192	9130	69	39	16252	941	1008	17.3	94

Databases of Protein Folds

- SCOP (http://scop.berkeley.edu/)
  - Structural Classification of Proteins
  - Class-Fold-Superfamily-Family
  - Manual assembly by inspection
- Superfamily (http://supfam.mrc-lmb.cam.ac.uk/SUPERFAMILY/)
  - HMM models for each SCOP fold
  - Fold assignments to all genome ORFs
  - Assessment of specificity/sensitivity of structure prediction
  - Search by sequence, genome and keywords
- CATH (http://www.biochem.ucl.ac.uk/bsm/cath/)
  - Class - Architecture - Topology - Homologous Superfamily
  - Manual classification at Architecture level
  - Automated topology classification using SSAP (Orengo & Taylor)
- FSSP (http://www2.embl-ebi.ac.uk/dali/fssp/)
  - Fully automated using the DALI algorithm (Holm & Sander)
  - No internal node annotations
  - Structural similarity search using DALI

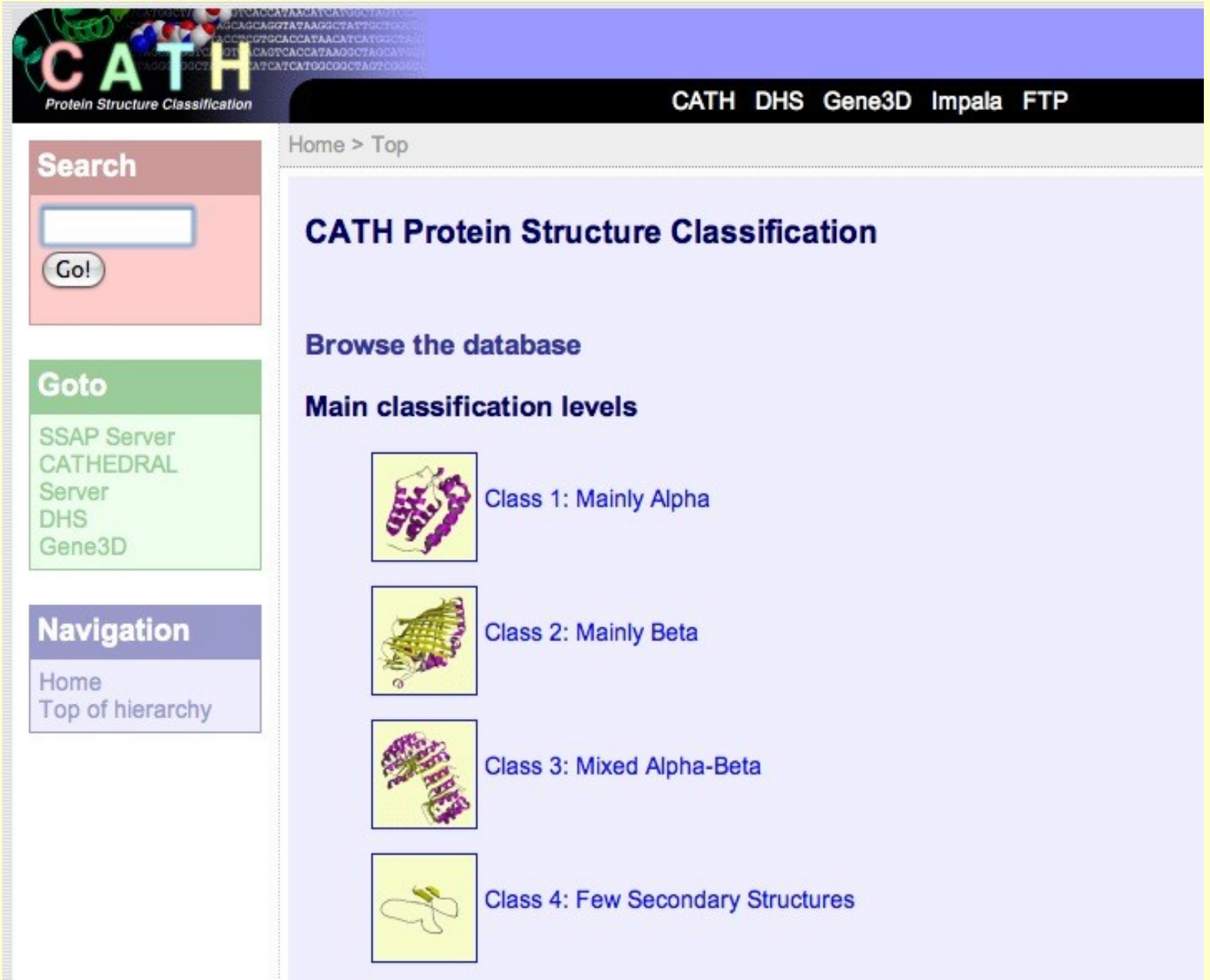
CATH Protein Structure Classification

http://www.biochem.ucl.ac.uk/bsm/cath/



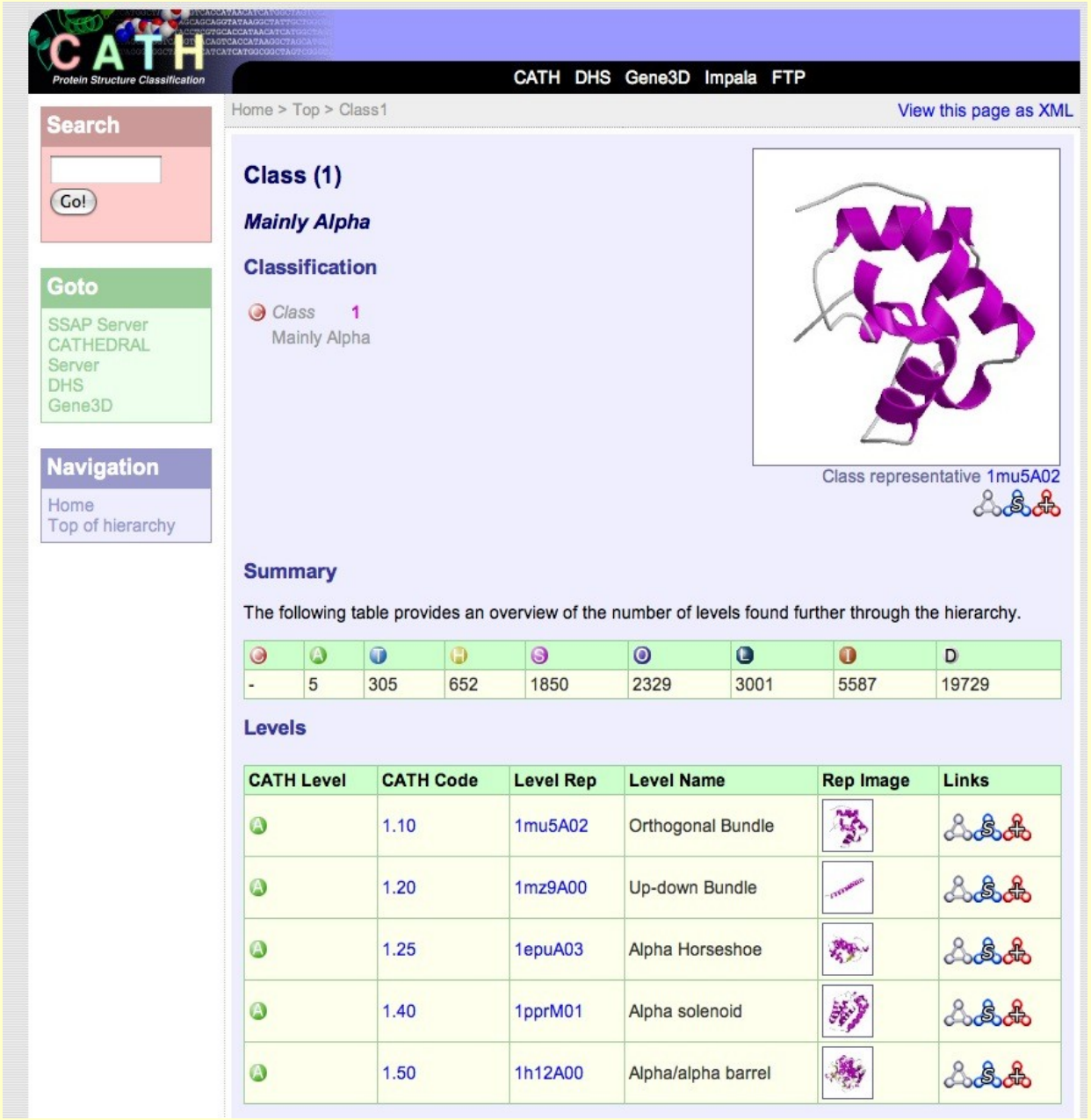
CATH Protein Structure Hierarchy

http://www.biochem.ucl.ac.uk/bsm/cath/



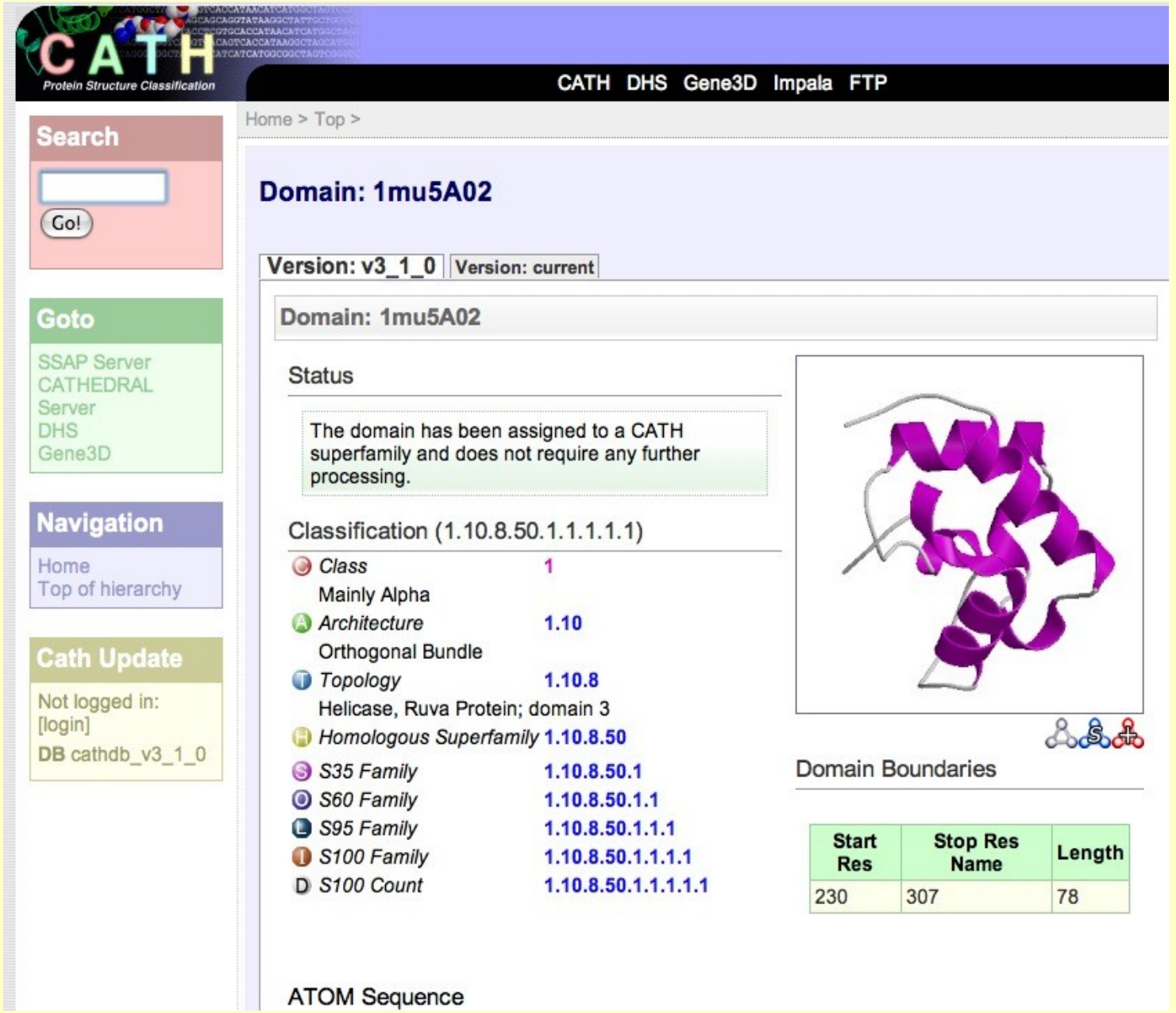
CATH Protein Class Level

http://www.biochem.ucl.ac.uk/bsm/cath/

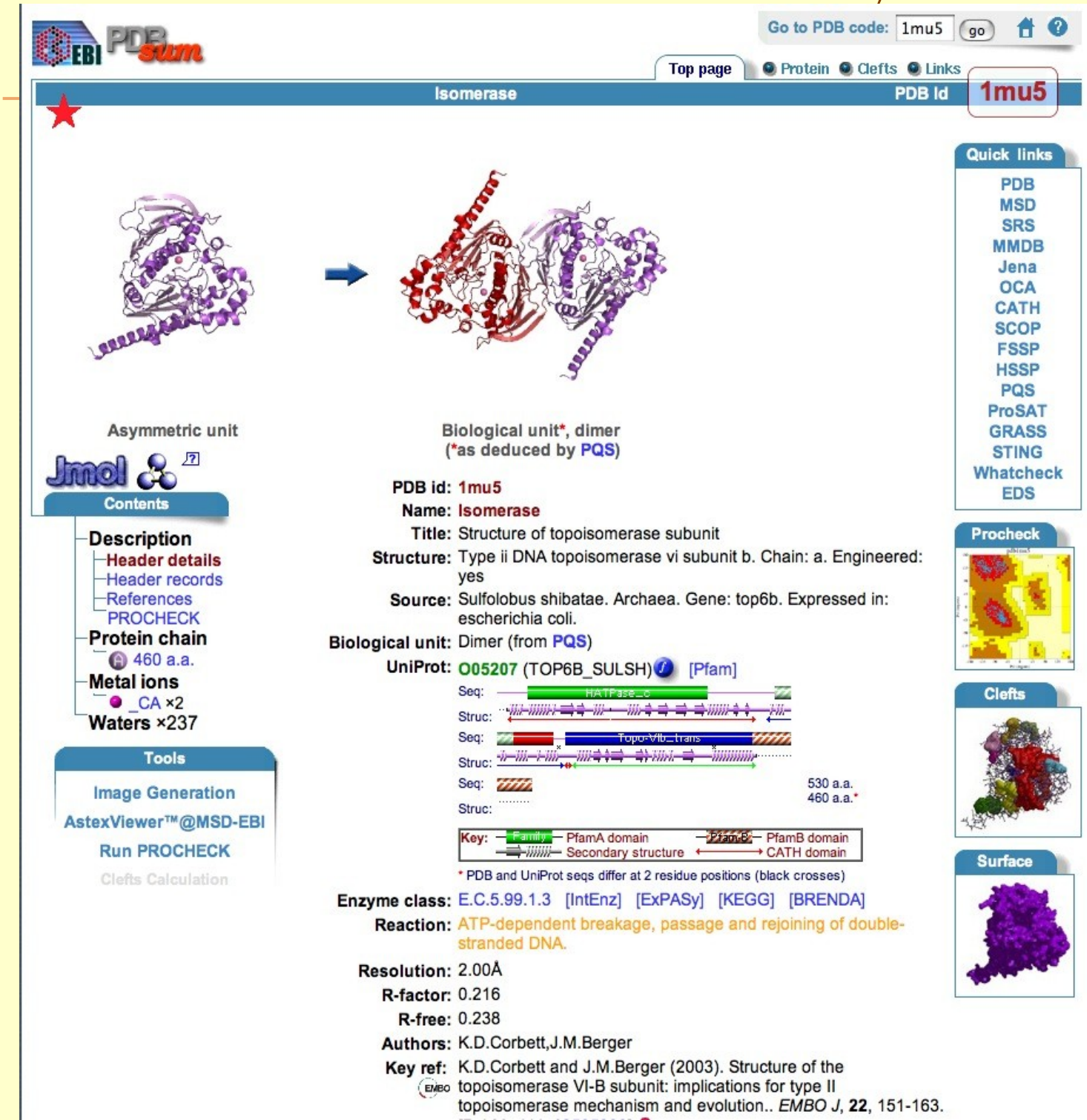


CATH Orthogonal Bundle

http://www.biochem.ucl.ac.uk/bsm/cath/

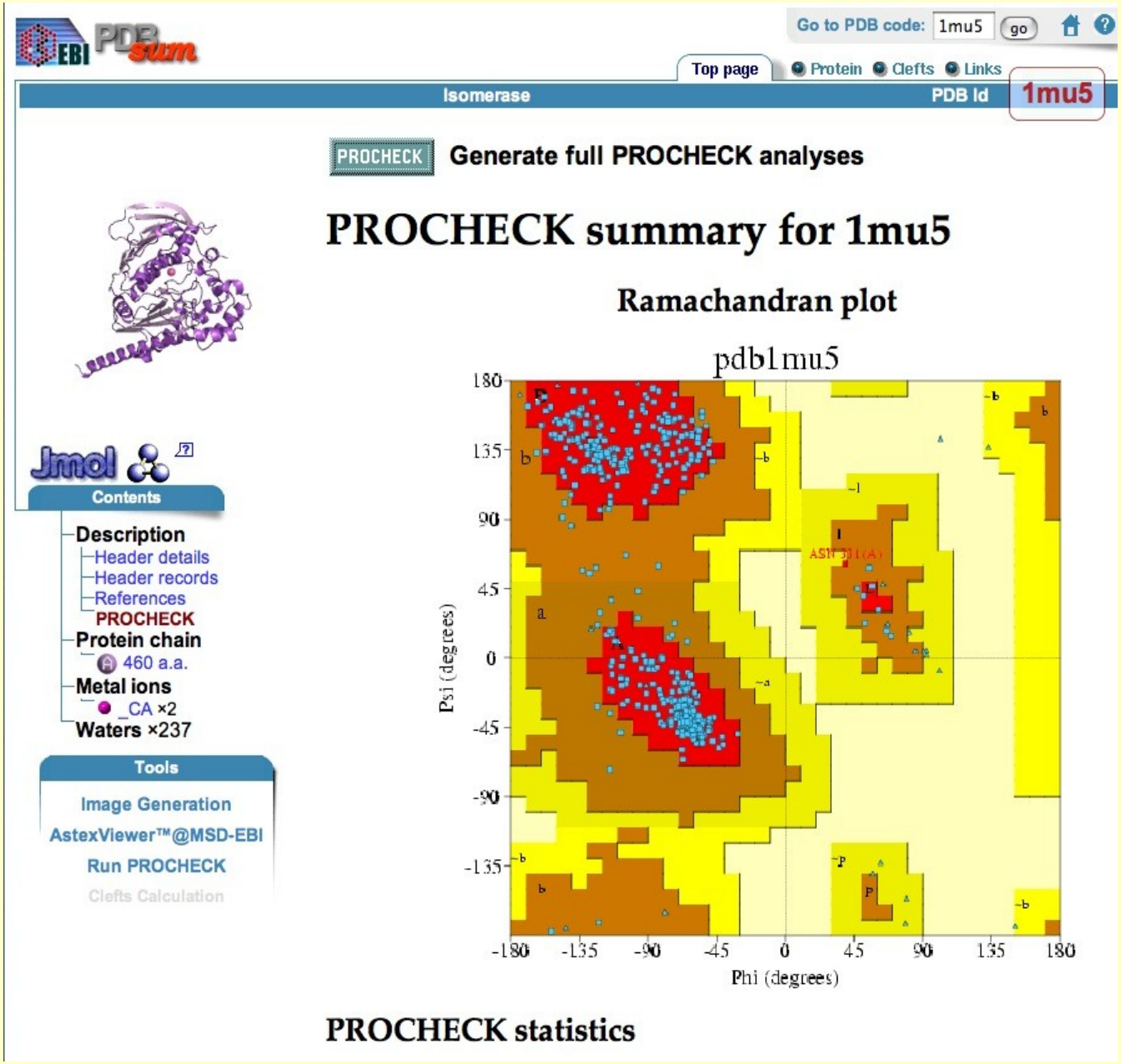


CATH Protein Summary



CATH Protein Summary

http://www.biochem.ucl.ac.uk/bsm/cath/

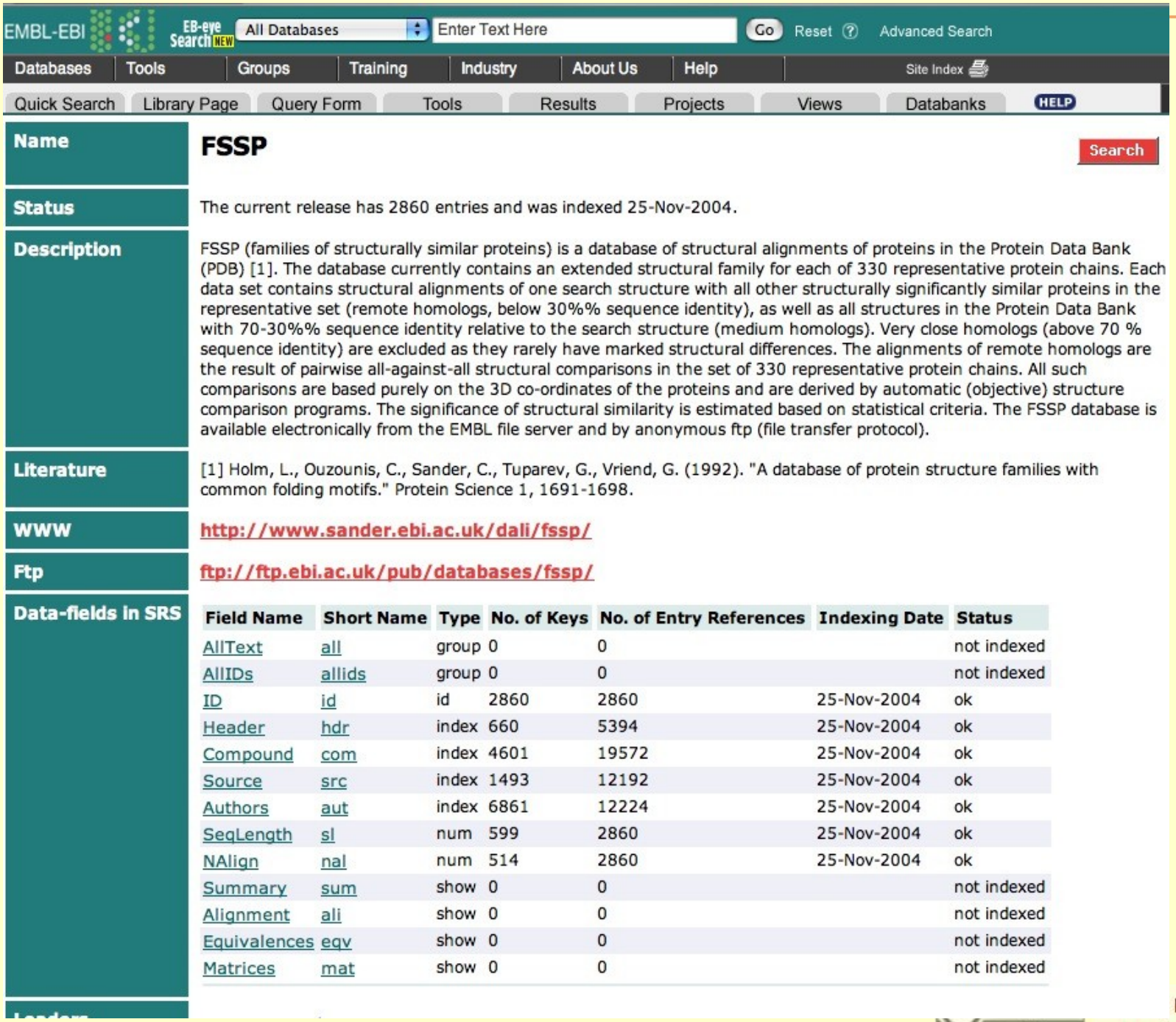


Databases of Protein Folds

- SCOP (http://scop.berkeley.edu/)
  - Structural Classification of Proteins
  - Class-Fold-Superfamily-Family
  - Manual assembly by inspection
- Superfamily (http://supfam.mrc-lmb.cam.ac.uk/SUPERFAMILY/)
  - HMM models for each SCOP fold
  - Fold assignments to all genome ORFs
  - Assessment of specificity/sensitivity of structure prediction
  - Search by sequence, genome and keywords
- CATH (http://www.biochem.ucl.ac.uk/bsm/cath/)
  - Class - Architecture - Topology - Homologous Superfamily
  - Manual classification at Architecture level
  - Automated topology classification using SSAP (Orengo & Taylor)
- FSSP (http://www2.embl-ebi.ac.uk/dali/fssp/)
  - Fully automated using the DALI algorithm (Holm & Sander)
  - No internal node annotations
  - Structural similarity search using DALI

FSSP Database

http://srs.ebi.ac.uk/srsbin/cgi-bin/wgetz?-page+LibInfo+-lib+FS



EMBL-EBI | EB-eye Search NEW | All Databases | Enter Text Here | Go | Reset | Advanced Search

Databases | Tools | Groups | Training | Industry | About Us | Help | Site Index

Quick Search | Library Page | Query Form | Tools | Results | Projects | Views | Databanks | HELP

Name FSSP — Search

Status The current release has 2860 entries and was indexed 25-Nov-2004.

Description FSSP (families of structurally similar proteins) is a database of structural alignments of proteins in the Protein Data Bank (PDB) [1]. The database currently contains an extended structural family for each of 330 representative protein chains. Each data set contains structural alignments of one search structure with all other structurally significantly similar proteins in the representative set (remote homologs, below 30%% sequence identity), as well as all structures in the Protein Data Bank with 70-30%% sequence identity relative to the search structure (medium homologs). Very close homologs (above 70 % sequence identity) are excluded as they rarely have marked structural differences. The alignments of remote homologs are the result of pairwise all-against-all structural comparisons in the set of 330 representative protein chains. All such comparisons are based purely on the 3D co-ordinates of the proteins and are derived by automatic (objective) structure comparison programs. The significance of structural similarity is estimated based on statistical criteria. The FSSP database is available electronically from the EMBL file server and by anonymous ftp (file transfer protocol).

Literature [1] Holm, L., Ouzounis, C., Sander, C., Tuparev, G., Vriend, G. (1992). "A database of protein structure families with common folding motifs." Protein Science 1, 1691-1698.

WWW http://www.sander.ebi.ac.uk/dali/fssp/

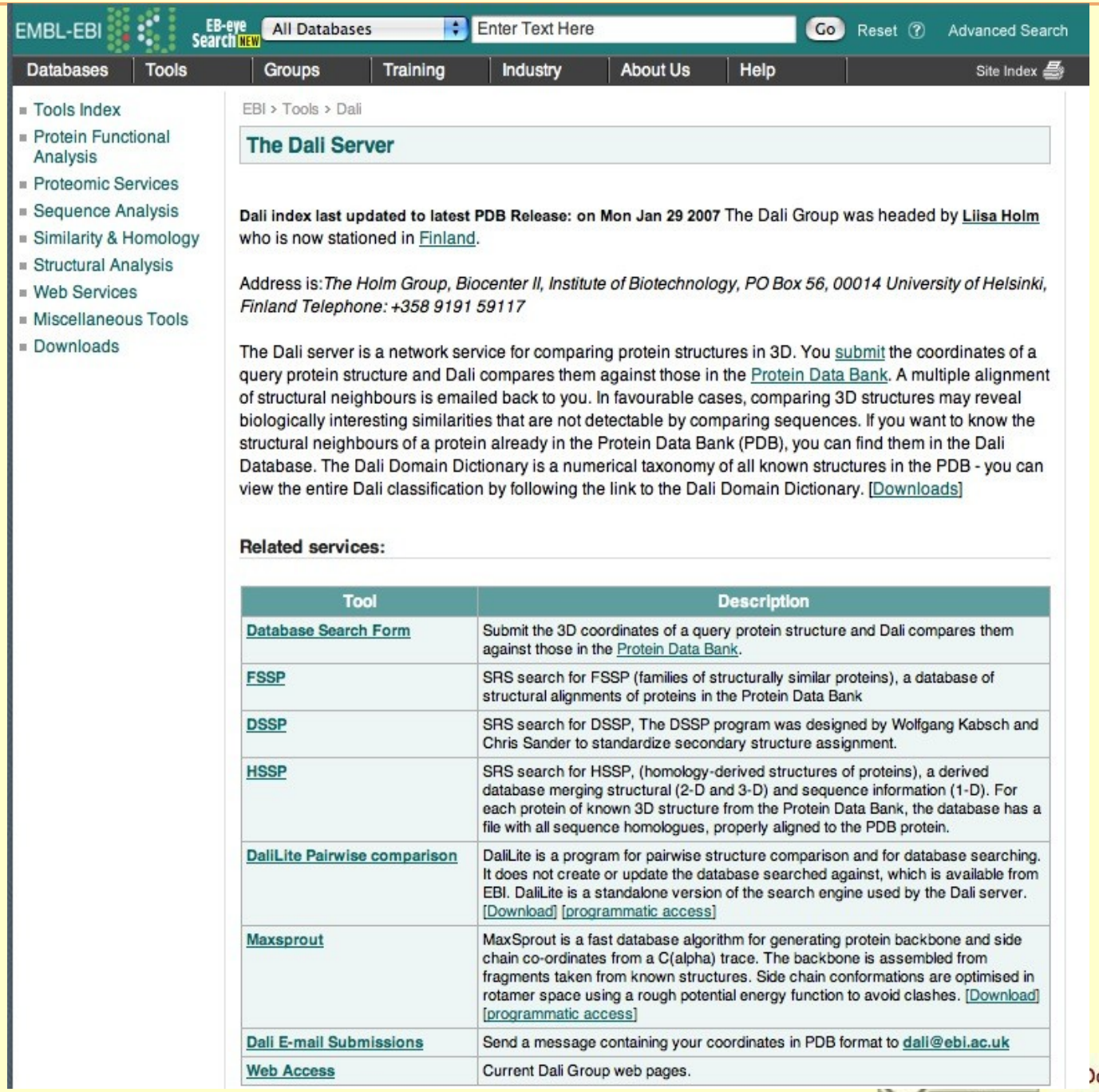
Ftp ftp://ftp.ebi.ac.uk/pub/databases/fssp/

Data-fields in SRS

Field Name	Short Name	Type	No. of Keys	No. of Entry References	Indexing Date	Status
AllText	all	group	0	0		not indexed
AllIDs	allids	group	0	0		not indexed
ID	id	id	2860	2860	25-Nov-2004	ok
Header	hdr	index	660	5394	25-Nov-2004	ok
Compound	com	index	4601	19572	25-Nov-2004	ok
Source	src	index	1493	12192	25-Nov-2004	ok
Authors	aut	index	6861	12224	25-Nov-2004	ok
SeqLength	sl	num	599	2860	25-Nov-2004	ok
NAlign	nal	num	514	2860	25-Nov-2004	ok
Summary	sum	show	0	0		not indexed
Alignment	ali	show	0	0		not indexed
Equivalences	eqv	show	0	0		not indexed
Matrices	mat	show	0	0		not indexed

Dali Server

http://www.ebi.ac.uk/dali/



EMBL-EBI | EB-eye Search NEW | All Databases | Enter Text Here | Go | Reset | Advanced Search

Databases | Tools | Groups | Training | Industry | About Us | Help | Site Index

- Tools Index
- Protein Functional Analysis
- Proteomic Services
- Sequence Analysis
- Similarity & Homology
- Structural Analysis
- Web Services
- Miscellaneous Tools
- Downloads

EBI > Tools > Dali

The Dali Server

Dali index last updated to latest PDB Release: on Mon Jan 29 2007 The Dali Group was headed by **Liisa Holm** who is now stationed in Finland.

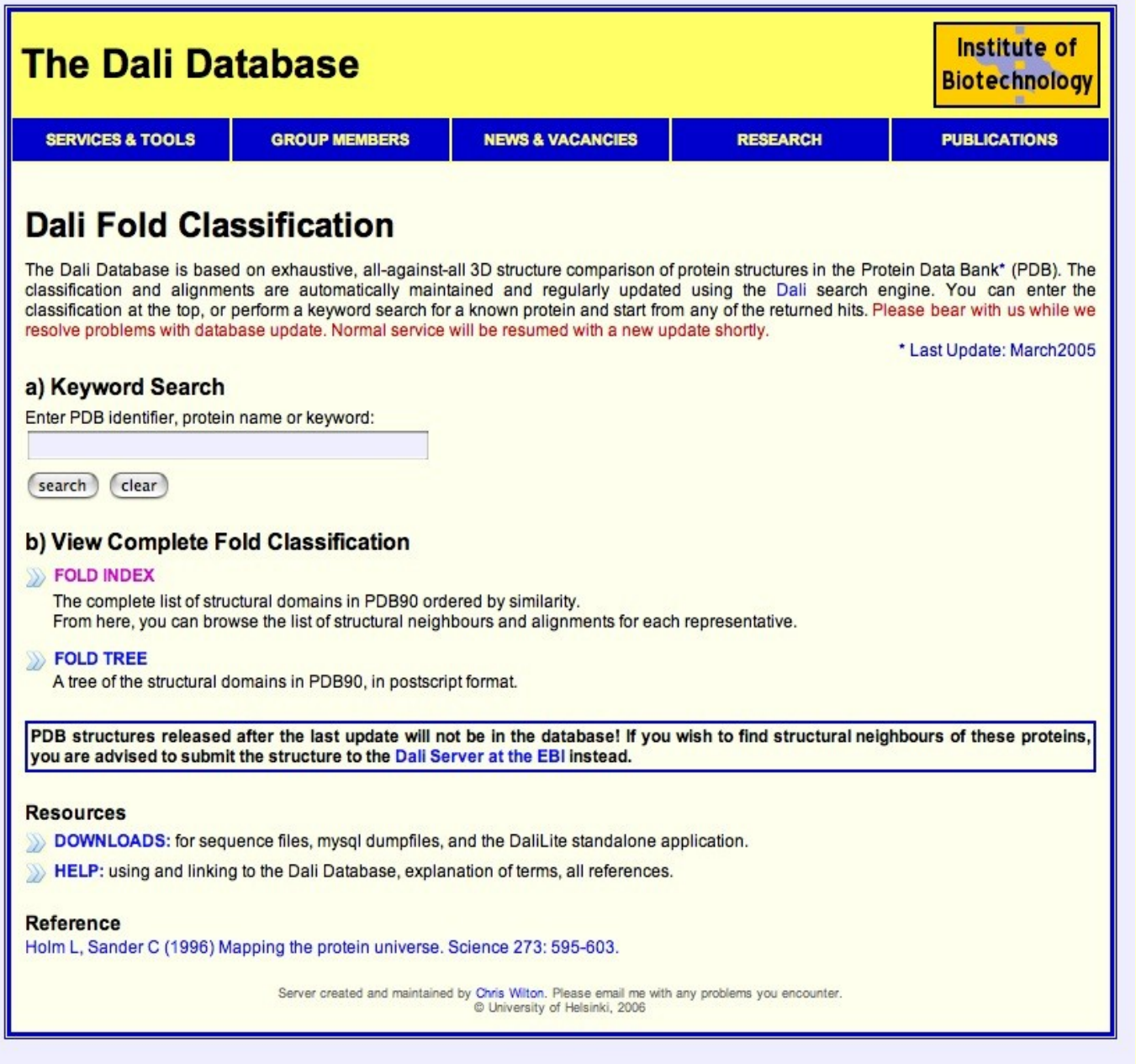
Address is: *The Holm Group, Biocenter II, Institute of Biotechnology, PO Box 56, 00014 University of Helsinki, Finland Telephone: +358 9191 59117*

The Dali server is a network service for comparing protein structures in 3D. You submit the coordinates of a query protein structure and Dali compares them against those in the Protein Data Bank. A multiple alignment of structural neighbours is emailed back to you. In favourable cases, comparing 3D structures may reveal biologically interesting similarities that are not detectable by comparing sequences. If you want to know the structural neighbours of a protein already in the Protein Data Bank (PDB), you can find them in the Dali Database. The Dali Domain Dictionary is a numerical taxonomy of all known structures in the PDB - you can view the entire Dali classification by following the link to the Dali Domain Dictionary. [Downloads]

Related services:

Tool	Description
Database Search Form	Submit the 3D coordinates of a query protein structure and Dali compares them against those in the Protein Data Bank.
FSSP	SRS search for FSSP (families of structurally similar proteins), a database of structural alignments of proteins in the Protein Data Bank
DSSP	SRS search for DSSP, The DSSP program was designed by Wolfgang Kabsch and Chris Sander to standardize secondary structure assignment.
HSSP	SRS search for HSSP, (homology-derived structures of proteins), a derived database merging structural (2-D and 3-D) and sequence information (1-D). For each protein of known 3D structure from the Protein Data Bank, the database has a file with all sequence homologues, properly aligned to the PDB protein.
DaliLite Pairwise comparison	DaliLite is a program for pairwise structure comparison and for database searching. It does not create or update the database searched against, which is available from EBI. DaliLite is a standalone version of the search engine used by the Dali server. [Download] [programmatic access]
Maxsprout	MaxSprout is a fast database algorithm for generating protein backbone and side chain co-ordinates from a C(alpha) trace. The backbone is assembled from fragments taken from known structures. Side chain conformations are optimised in rotamer space using a rough potential energy function to avoid clashes. [Download] [programmatic access]
Dali E-mail Submissions	Send a message containing your coordinates in PDB format to dali@ebi.ac.uk
Web Access	Current Dali Group web pages.

DALI Database (Liisa Holm)
http://ekhidna.biocenter.helsinki.fi/dali/start



The Dali Database

Institute of Biotechnology

SERVICES & TOOLS | GROUP MEMBERS | NEWS & VACANCIES | RESEARCH | PUBLICATIONS

Dali Fold Classification

The Dali Database is based on exhaustive, all-against-all 3D structure comparison of protein structures in the Protein Data Bank* (PDB). The classification and alignments are automatically maintained and regularly updated using the Dali search engine. You can enter the classification at the top, or perform a keyword search for a known protein and start from any of the returned hits. Please bear with us while we resolve problems with database update. Normal service will be resumed with a new update shortly.

* Last Update: March2005

a) Keyword Search

Enter PDB identifier, protein name or keyword:

search clear

b) View Complete Fold Classification

FOLD INDEX
The complete list of structural domains in PDB90 ordered by similarity.
From here, you can browse the list of structural neighbours and alignments for each representative.

FOLD TREE
A tree of the structural domains in PDB90, in postscript format.

PDB structures released after the last update will not be in the database! If you wish to find structural neighbours of these proteins, you are advised to submit the structure to the Dali Server at the EBI instead.

Resources

DOWNLOADS: for sequence files, mysql dumpfiles, and the DaliLite standalone application.

HELP: using and linking to the Dali Database, explanation of terms, all references.

Reference

Holm L, Sander C (1996) Mapping the protein universe. Science 273: 595-603.

Server created and maintained by Chris Wilton. Please email me with any problems you encounter.
© University of Helsinki, 2006

Protein Fold Prediction: Swiss Model

http://swissmodel.expasy.org/

- Amos Bairoch, Swiss Bioinformatics Institute, SBI
- Threading and Template Discovery
- Workspace for saving Template Results
- Domain Annotation
- Structure Assessment
- Template Library
- Structures & Models
- Documentation and Tutorials

Protein Fold Prediction: Swiss Model

http://swissmodel.expasy.org/

SIB Swiss Institute of Bioinformatics | BIOZENTRUM Universität Basel | SWISS-MODEL

Modelling
- myWorkspace
- Automated Mode
- Alignment Mode
- Project Mode

Tools
- Template Identification
- Domain Annotation
- Structure Assessment
- Template Library

Repository
- Search by Sequence
- Search by AC

Documentation
- SWISS-MODEL Workspace
- SWISS-MODEL Repository
- Structures & Models
- Helpdesk

SWISS-MODEL is a fully automated protein structure homology-modeling server, accessible via the ExPASy web server, or from the program DeepView (Swiss Pdb-Viewer). The purpose of this server is to make Protein Modelling accessible to all biochemists and molecular biologists WorldWide.

What's new?

- New automated modeling pipeline with improved hierarchical approach for template selection.
- Increased sensitivity of template detection (sequence to profile search using an adapted HHSearch protocol)
- New tools for model and structure quality assessment: Dfire and Qmean global scores; ProQres residue based assessment scores

SWISS-MODEL Team

Torsten Schwede:	Project Leader
Florian Kiefer:	SWISS-MODEL Repository
Lorenza Bordoli:	Method Development and user support
Konstantin Arnold:	SWISS-MODEL Workspace

References:

When you publish or report results using SWISS-MODEL, please cite the relevant publications:

- Arnold K., Bordoli L., Kopp J., and Schwede T. (2006). The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling. Bioinformatics, 22,195-201.
- Kiefer F, Arnold K, Künzli M, Bordoli L, Schwede T (2009). The SWISS-MODEL Repository and associated resources. Nucleic Acids Research. 37, D387-D392.
- Peitsch, M. C. (1995) Protein modeling by E-mail Bio/Technology 13: 658-660.

Automatic Protein Fold Prediction

http://swissmodel.expasy.org/

SIB BIOZENTRUM

SWISS-MODEL Workspace

Modelling Tools Repository

[myWorkspace]

SwissModel Automatic Modelling Mode

Email: brutlag@stanford.edu

Project Title: HU E. coli

Provide a protein sequence or a UniProt AC Code:

```
>HU-NS1 P0ACF0
MNKSQLIDKIAAGADISKAAAGRALDAIIASVTESLKEGDDVALVGFGTFAVKERAARTGRNPQTGKEIT
IAAAKVPSFRAGKALKDAVN
```

Submit Modelling Request

Options:

Use a specific template:

SWISS-MODEL

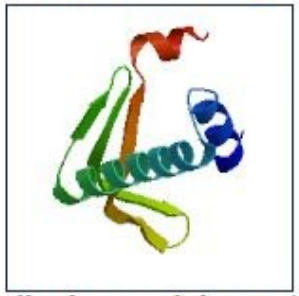
Automatic Protein Fold Prediction Results

http://swissmodel.expasy.org/

Workunit: P000002 Title:HU E. coli

Go to: [Template Selection] [Alignment] [Modelling Log] [Evaluation]

Model Details: **Segment 1**



Model info:
modelled residue range: 1 to 90
based on template **1mulA** (2.30 Å)
Sequence Identity [%]: 84.444
Evalue: 3.25e-22

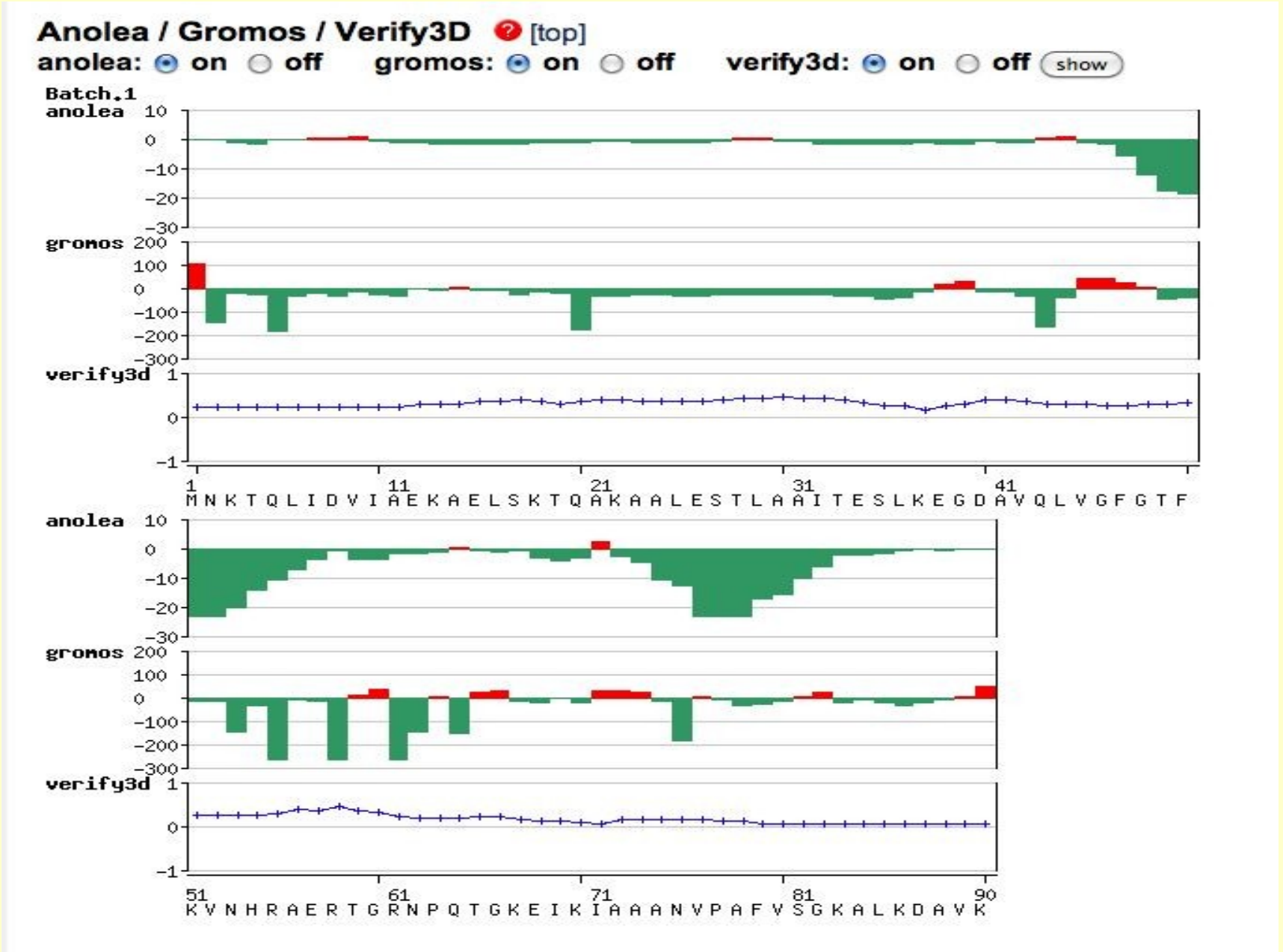
display model: as pdb - as DeepView project
download model: as pdb - as Deepview project - as text

Alignment [top]

```
TARGET    1       MNKTQLID VIAEKAELSK TQAKAALEST LAAITESLKE GDAVQLVGFG
1mulA     1       mnktqlid viaekaelsk tqakaalest laaiteslke gdavqlvgfg

TARGET              hhhhhh hhhhhh   h hhhhhhhhhh hhhhhhhhhh    sss  ss
1mulA               hhhhhh hhhhhh   h hhhhhhhhhh hhhhhhhhhh    sss  ss


TARGET    49    TFKVNHRAER TGRNPQTGKE IKIAAANVPA FVSGKALKDA VK
1mulA     49    tfkvnhrae- ---------- ---aaanvpa fvsgkalkda vk-

TARGET          sssssss    sssss  sss ss   sssss sssshhhhhh h
1mulA           sssssss                     sssss sssshhhhhh
```

Automatic Protein Fold Prediction Results

http://swissmodel.expasy.org/



Automatic Protein Fold Prediction Results

http://swissmodel.expasy.org/

SIB BIOZENTRUM

SWISS-MODEL Workspace

Modelling Tools Repository Documentation

[myWorkspace] [Settings] [logout]

Computation of this workunit has stopped.

Please see the following log report for details:

```
Started: Tue Feb  2 06:07:29 2010 (sms_automode)
Reading user input sequence


No Templates found.
====================

Simple automated template selection could not identify suitable templates.

Please use advanced Template Selection under [Tools] to select a template
and prepare a workunit using the project mode.
```

phyre
Protein Homology/analogY Recognition Engine

Protein Fold Prediction: phyre

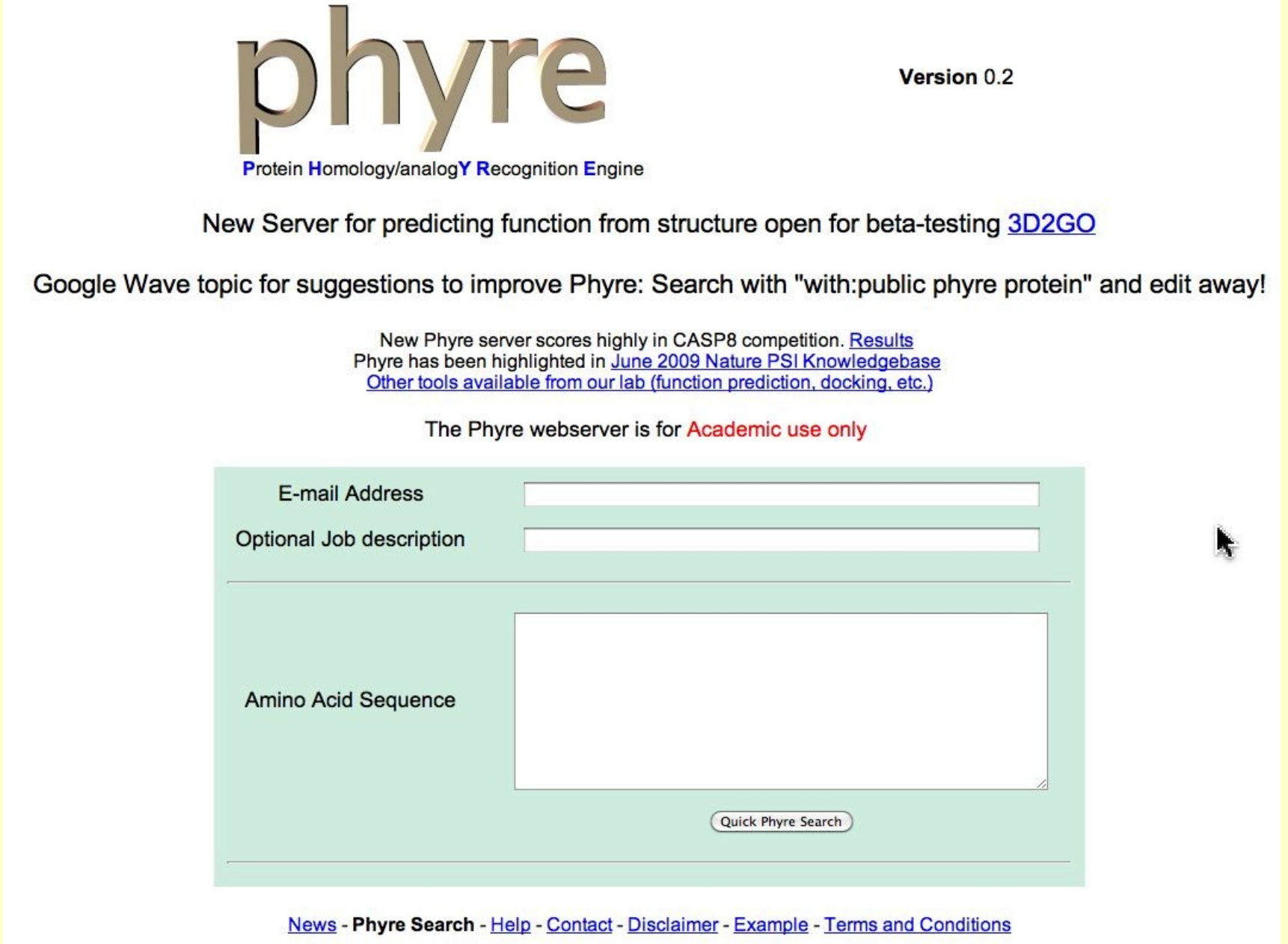
http://www.sbg.bio.ic.ac.uk/~phyre/

- Michael Sternberg, Structural Bioinformatics Group, Imperial College London
- Protein structure prediction on the web: a case study using the Phyre server Kelley LA and Sternberg MJE. *Nature Protocols 4, 363 - 371 (2009)*
- **P**rotein **H**omology/analog**Y** **R**ecognition **E**ngine

phyre
Protein Homology/analogY Recognition Engine

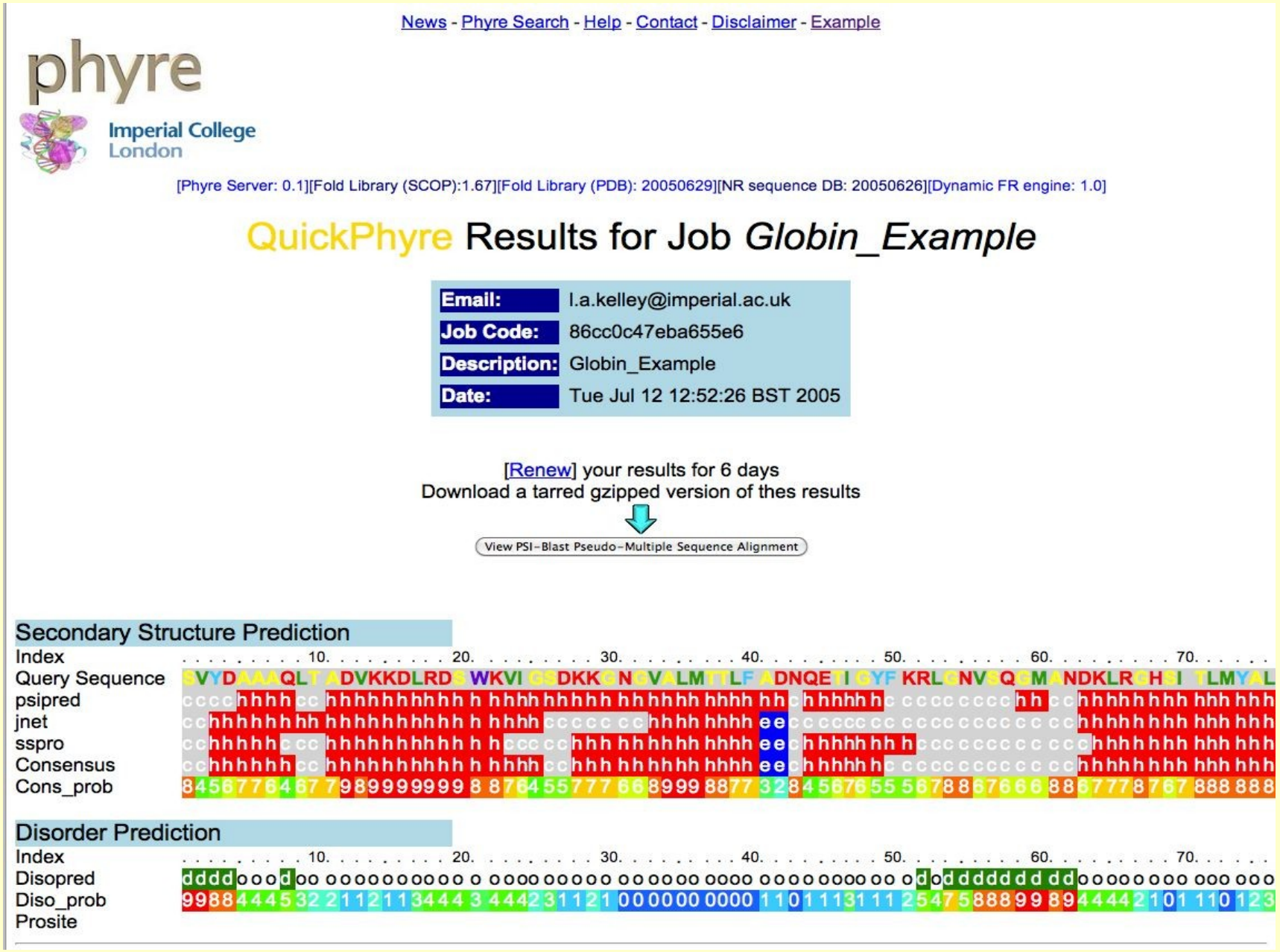
Protein Fold Prediction: phyre

http://www.sbg.bio.ic.ac.uk/~phyre/



Protein Fold Prediction: phyre

http://www.sbg.bio.ic.ac.uk/~phyre/



phyre
Protein Homology/analogY Recognition Engine

Protein Fold Prediction: phyre

http://www.sbg.bio.ic.ac.uk/~phyre/

Fold Recognition

View Alignments	SCOP Code	View Model	E-value	Estimated Precision	BioText	Fold/PDB descriptor	Superfamily	Family
	d3sdha_ (length:145) **100% i.d.**		9.3e-20	100 %	0.90 Biotext	Globin-like	Globin-like	Globins
	c2bk9A_ (length:153) **23% i.d.**		7.7e-17	100 %	0.89 Biotext	**PDB header:**oxygen transport	**Chain:** A: **PDB Molecule:**cg9734-pa;	**PDBTitle:** drosophila melanogaster globin
	d1itha_ (length:141) **16% i.d.**		2.1e-16	100 %	0.88 Biotext	Globin-like	Globin-like	Globins
	c1x3kA_ (length:152) **17% i.d.**		4.1e-16	100 %	0.89 Biotext	**PDB header:**oxygen storage/transport	**Chain:** A: **PDB Molecule:**hemoglobin component v;	**PDBTitle:** crystal structure of a hemoglobin component (ta-v) from2 tokunagayusurika akamusi
	d1irda_ (length:141) **17% i.d.**		7.8e-16	100 %	0.94 Biotext	Globin-like	Globin-like	Globins
	d1fhja_ (length:141) **17% i.d.**		1.4e-15	100 %	0.94 Biotext	Globin-like	Globin-like	Globins

Bloomsbury Centre for Bioinformatics

Protein Fold Prediction: PsiPred

http://bioinf4.cs.ucl.ac.uk:3000/psipred/

- Kevin Bryson and David Jones, University College London
- Predicts Secondary Structure of single molecules
- Predicts Transmembrane Topology
- Three Fold Recognition methods

Protein Fold Prediction: PsiPred

http://bioinf4.cs.ucl.ac.uk:3000/psipred/

UCL Department Of Computer Science

Bioinformatics Group

Search Group

UCL Home ›› Departments of Computer Science ›› Bioinformatics Group ›› psipred

Site Navigation

Server Navigation

- PSIPRED Server
- PSIPRED help
- Server Overview
- Server Citation
- News
- History
- Software Download
- Login

The PSIPRED Protein Structure Prediction Server

The PSIPRED protein structure prediction server allows you to submit a protein sequence, perform a prediction of your choice and receive the results of the prediction via e-mail. You may select one of three prediction methods to apply to your sequence: PSIPRED - a highly accurate method for protein secondary structure prediction, MEMSAT - our widely used transmembrane topology prediction method and GenTHREADER - a sequence profile based fold recognition method. More...

For queries regarding PSIPRED: psipred@cs.ucl.ac.uk

Choose Prediction Method

- (•) Predict Secondary Structure (PSIPRED v2.6)
- ( ) Predict Transmembrane Topology (MEMSAT3 & MEMSAT-SVM)
- ( ) Fold Recognition (GenTHREADER - quick)
- ( ) Fold Recognition (pGenTHREADER - with profiles and predicted secondary structure)
- ( ) Fold Recognition (pDomTHREADER - annotates multiple domain on chains)

Help...

Input Sequence (single letter amino acid code)

Help...

If you wish to test these services follow this link to retrieve a test fasta sequence.

Filtering Options

- [x] Mask low complexity regions
- [ ] Mask transmembrane helices
- [ ] Mask coiled-coil regions

Help...

Warning: No sequence filters are applied when running MEMSAT

Submission Details

Email Address for job completion alert (optional)

Help...

Password (only required for licenced commercial e-mail addresses)

Help...

Protein Fold Prediction: PsiPred

http://bioinf4.cs.ucl.ac.uk:3000/psipred/

Bloomsbury Centre for Bioinformatics

Web Servers

The Bioinformatics Group places a great deal of emphasis on developing Web services which are widely used by many groups and institutions. These include

For commercial enquiries about our software services, please visit **Ebisu**

Protein Fold Recognition

These methods will identify putative folds and structural domains for a sequence without known structure.

GenTHREADER	Rapid fold recognition, matching your sequence against a library of whole PDB chains.
pGenTHREADER	Highly sensitive fold recognition using profile-profile comparison (whole chain library).
pDomTHREADER	Highly sensitive homologous domain recognition using profile-profile comparison (domain library).

Protein Structure Feature Recognition

These methods annotate sequences with the location of important structural features.

PSIPRED	Accurate protein secondary structure prediction.
METSITE	A method for identifying metal binding sites in protein 3D structures

Transmembrane Topology Prediction

These methods identify the topology of putative transmembrane proteins.

MEMSAT3	Transmembrane protein topology predcition.
MEMSATSVM	Improved transmembrane protein topology prediction using SVMs. This method is capable of differentiating signal peptides from transmembrane helices.

3D Protein Modelling

Our latest experimental method will generate a homology model for a given sequence.

BioSerf	Automated homology and de-novo modelling server, utilising Modeller and FragFold.

Annotation of Protein Sequence Features

These methods identify protein sequence features which are important for protein structure and function.

DomPred	Identifies putative domain boundaries in sequences.
DISOPRED	Identifies residues which are likely to be natively unfolded (disordered).

Function Prediction

Feature-based protein function prediction from amino acid sequence.

ffpred	Integrated function prediction

Protein Fold Prediction: Predict Protein

http://www.predictprotein.org/

- Burkhard Rost, Columbia
- Methods
  - *MaxHom : multiple alignment*
  - *PSI-BLAST : iterated profile*
  - *searchProSite : functional motifs*
  - *SEG : composition-bias*
  - *ProDom : domain assignment*
  - *PredictNLS : nuclear localisation signal*
  - *PHDsec : secondary structure*
  - *PHDacc : solvent accessibility*
  - *Globe : globularity of proteins*
  - *PHDhtm : transmembrane helices*
  - *PROFsec : secondary structure*
  - *PROFacc : solvent accessibility Coils : coiled-coil regions*
  - *CYSPRED : cysteine bridges*
  - *Topits : fold recognition by threading*

Protein Fold Prediction: Predict Protein

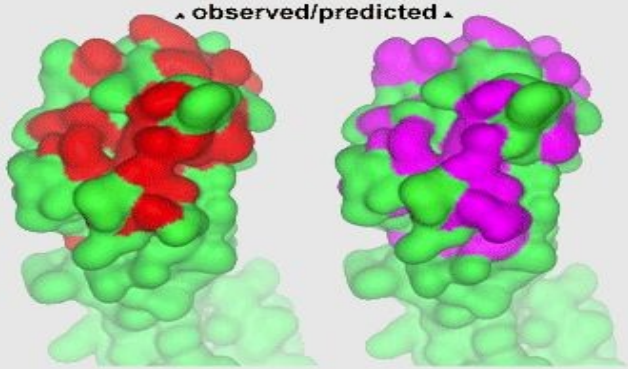
http://www.predictprotein.org/

PredictProtein

About | Submission | Help | Downloads | Register | MetaPP

sign in

Google™ Custom Search | Search



About PredictProtein

PredictProtein is a service for sequence analysis, structure and function prediction. When you **submit** any protein sequence PredictProtein retrieves similar sequences in the database and predicts aspects of protein structure and function (**more**)

News

10/04/2007

PredictProtein upgrade PredictProtein has been upgraded! We have integrated many new methods into the system; you can now get predictions of disordered/natively unstructured regions, of inter-residue contacts, of domain assignments, and protein-protein interaction and protein-DNA binding residues unsing our newer and **faster** server. The new system requires **registration**. Note that registration is free, and the use of PredictProtein remains free for academia.

Citing

In citing PredictProtein please refer to: PredictProtein: B Rost, G Yachdav and J Liu (2004) **The PredictProtein Server. Nucleic Acids Research 32(Web Server issue):W321-W326**.

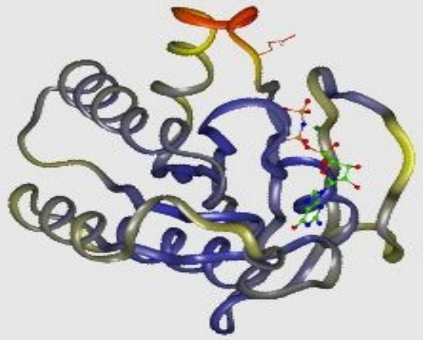
Discussion Board

If you have a publication which makes use of PredictProtein, please let us know by posting a message to the **PredictProtein discussion board**. If that is your first visit to the discussion board you will need to **register** in order to post messages.

XML RSS

Government Support

The development of the methods and the datbases in PredictProtein is supported by R01 LM07329-01 from the **National Library of Medicine.**

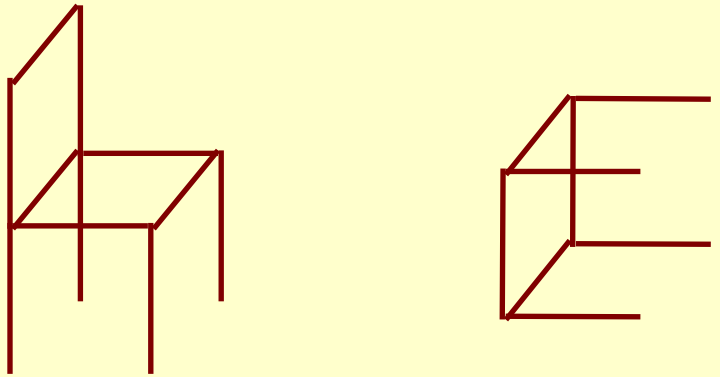


Explicitly predicting normalized B-values enables the implicit identification of flexible and rigid regions that relate to protein function. The crucial residues in the switch II region of ras need to be very flexible for this protein to function properly (more red=more flexible) ***more***

Automating Structure Classification, Fold & Function Detection

- Growth of PDB demands automated techniques for classification and fold detection
- Protein Structure Comparison
  - computing structure similarity based on metrics (distances)
  - identifying protein function
  - understanding functional mechanism
  - identifying structurally conserved regions in the protein
  - finding binding sites or other functionally important regions of the protein

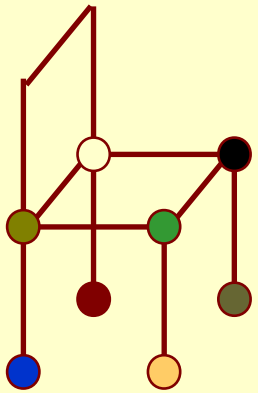
Structure Superposition



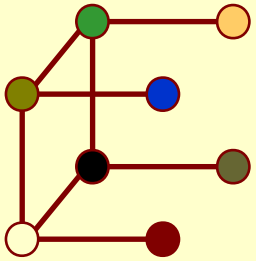
- Find the transformation matrix that *best* overlaps the table and the chair
- i.e. Find the transformation matrix that minimizes the root mean square deviation between **corresponding points of the table and the chair**
- Correspondences:
  - Top of chair to top of table
  - Front of chair to front of table, etc.

Absolute Orientation Algorithm

http://www-mtl.mit.edu/researchgroups/itrc/ITRC_publication/horn_publications.html

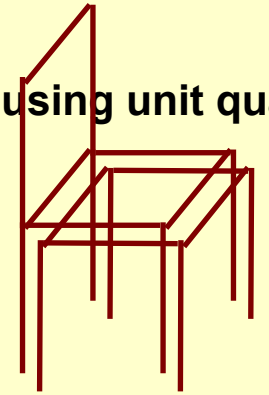


+



Closed-form solution of absolute orientation using unit quat

Berthold K.P. Horn, J.Opt.Soc.Am, April 1987, Vol 4, No. 4

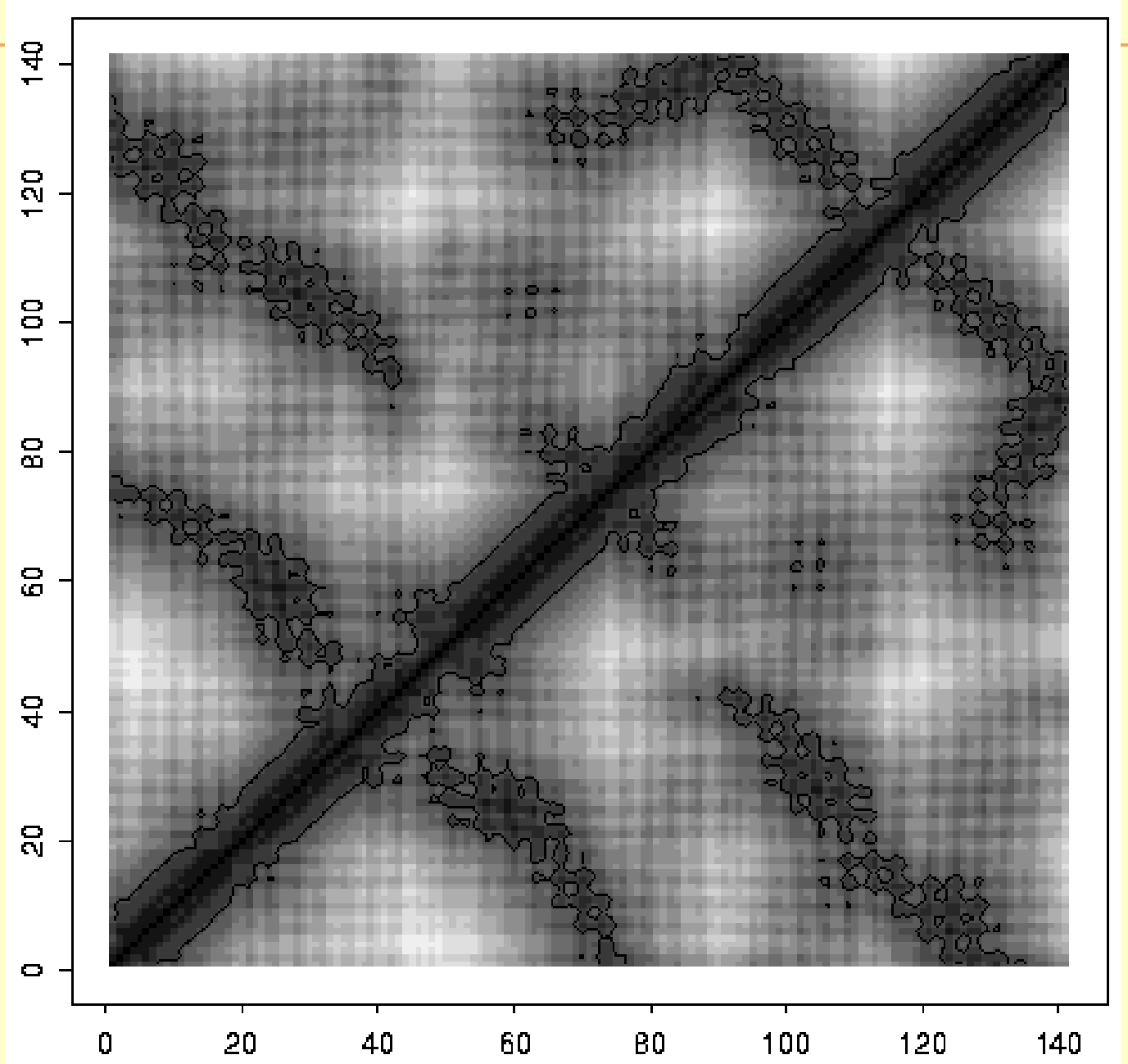


The key is finding corresponding points between the two structures

Algorithms for Structure Superposition

- Distance based methods:
  - DALI (Holm & Sander): Aligning scalar distance plots
  - STRUCTAL (Gerstein & Levitt): Dynamic programming using pair-wise inter-molecular distances
  - SSAP (Orengo & Taylor): Dynamic programming using intra-molecular vector distances
  - MINAREA (Falicov and Cohen): Minimizing soap-bubble surface area
  - CE (Shindyalov & Bourne)
- Vector based methods:
  - VAST (Bryant): Graph theory based secondary structure alignment
  - 3D Search (Singh and Brutlag) & 3D Lookup (Holm and Sander): Fast secondary structure index lookup
- Both
  - LOCK (Singh & Brutlag) LOCK2 (Ebert & Brutlag): Hierarchically uses both secondary structure vectors and atomic distances

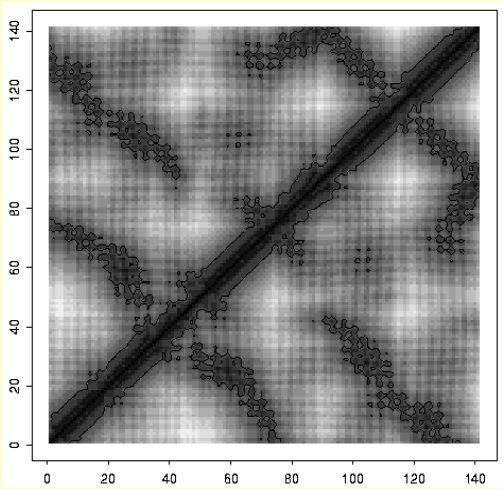
DALI



An intra-molecular distance plot for myoglobin

DALI

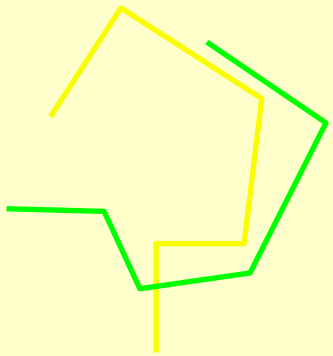
- Based on aligning 2-D intra-molecular distance matrices
- Computes the best subset of corresponding residues from the two proteins such that the similarity between the 2-D distance matrices is maximized
- Searches through all possible alignments of residues using Monte-Carlo and Branch-and-Bound algorithms

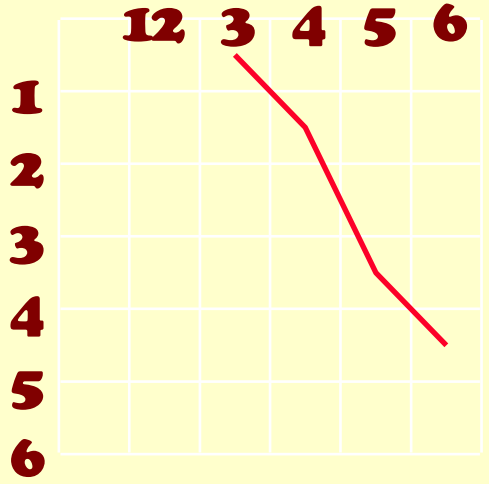


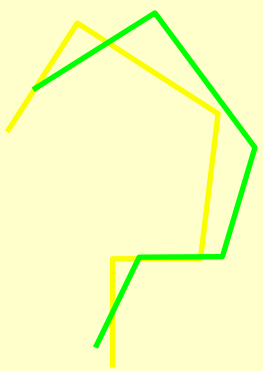
$$Score(i, j) = 1.5 - |distance^{A}(i, j) - distance^{B}(i, j)|$$

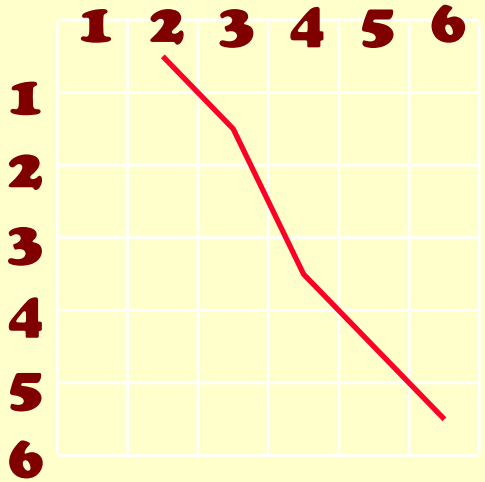
STRUCTAL

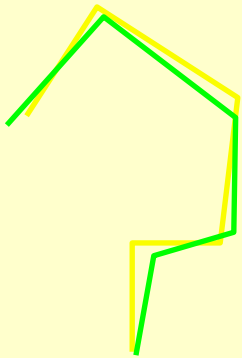
- Based on Iterative Dynamic Programming to align inter-molecular distances
- Pair-wise alignment score in each square of the matrix is inversely proportional to distance between the two atoms





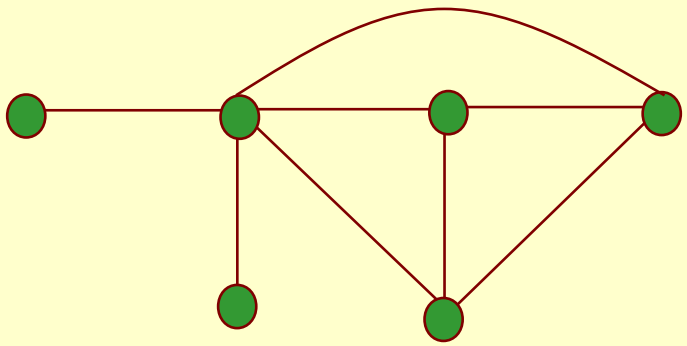






VAST - Vector Alignment Search Tool

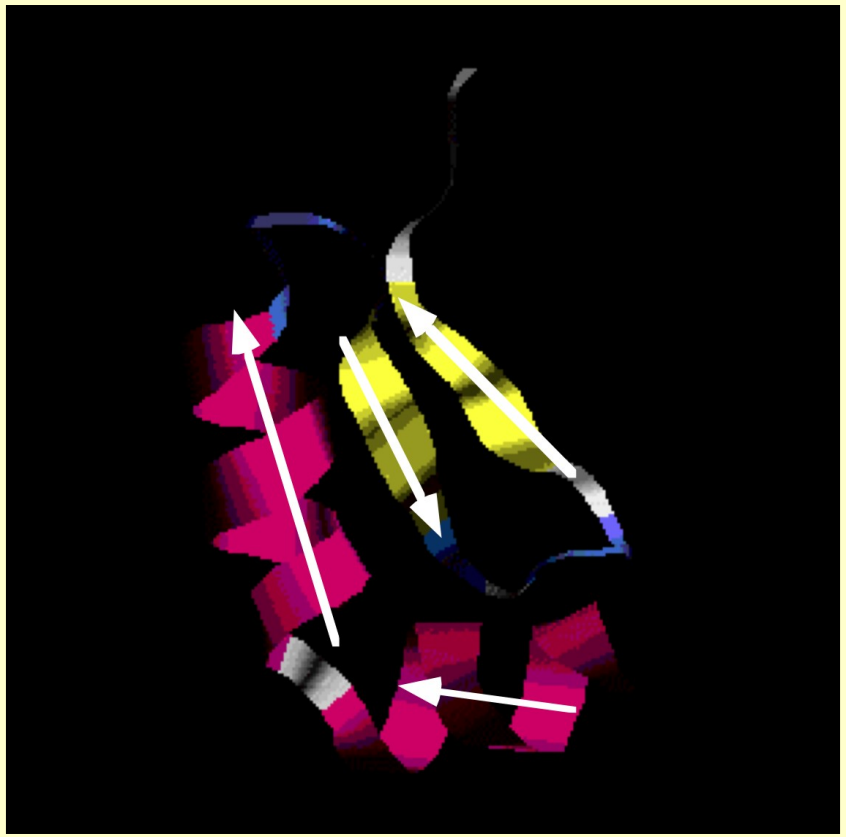
- Aligns only secondary structure elements (SSE)
- Represents each SSE as a vector
- Finds all possible pairs of vectors from the two structures that are similar
- Uses a graph theory algorithm to find maximal subset of similar vector pairs
- Overall alignment score is based on the number of similar pairs of vectors between the two structures

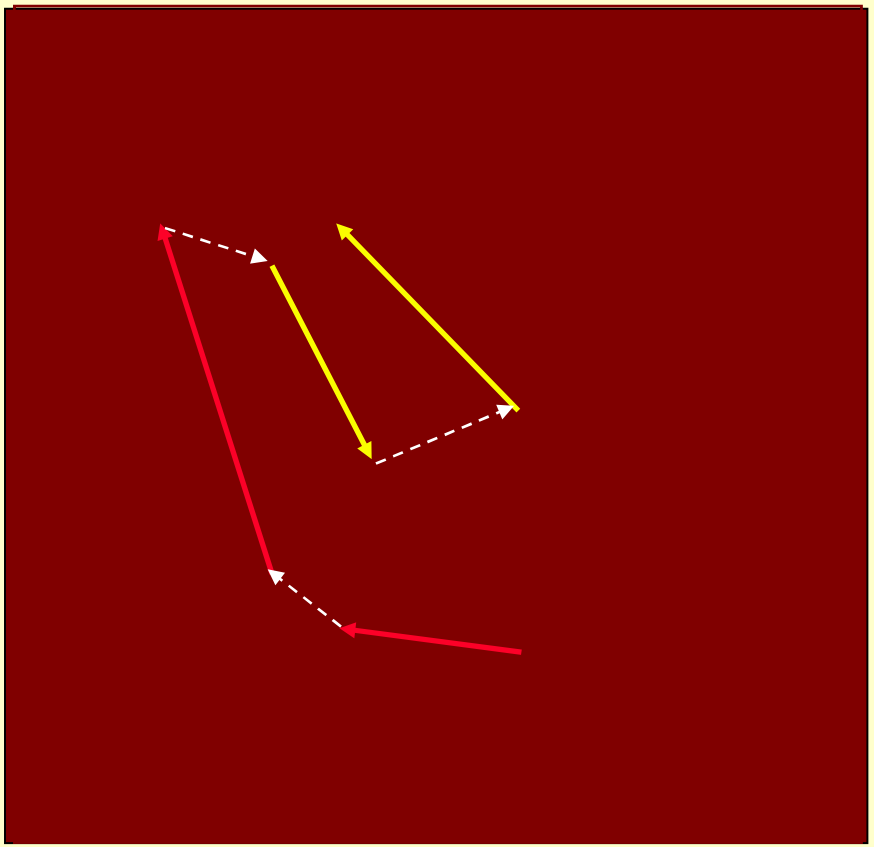


Algorithms for Structure Superposition

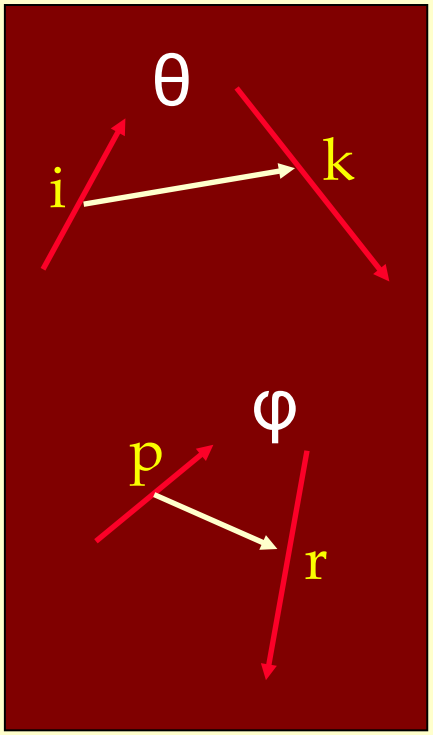
- Atomic distance based methods:
  - DALI (Holm and Sander): Aligning scalar distance plots
  - STRUCTAL (Gerstein and Levitt): Dynamic programming using pair wise inter-molecular distances
  - SSAP (Orengo and Taylor): Dynamic programming using intra-molecular vector distances
  - MINAREA (Falicov and Cohen): Minimizing soap-bubble surface area
- Vector based methods:
  - VAST (Bryant): Graph theory based secondary structure alignment
  - 3dSearch (Singh and Brutlag): Fast secondary structure index lookup
- Use both SSE vectors and atomic distances
  - LOCK (Singh and Brutlag): Hierarchically uses both secondary structure vectors and atomic distances

LOCK - Creating Secondary Structure Vectors





Comparing Secondary Structure Vectors



Orientation Independent Scores:

S = S(|angle θ(i,k) - angle φ(p,r)|)

S = S(|distance(i,k) - distance(p,r)|)

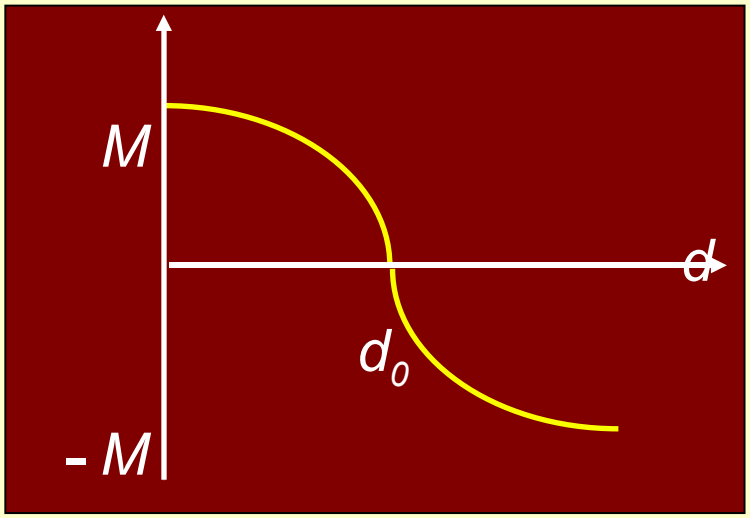
S = S(|length(i) - length(p)|)+
S(|length(p) - length(r)|)

Orientation Dependent Scores:

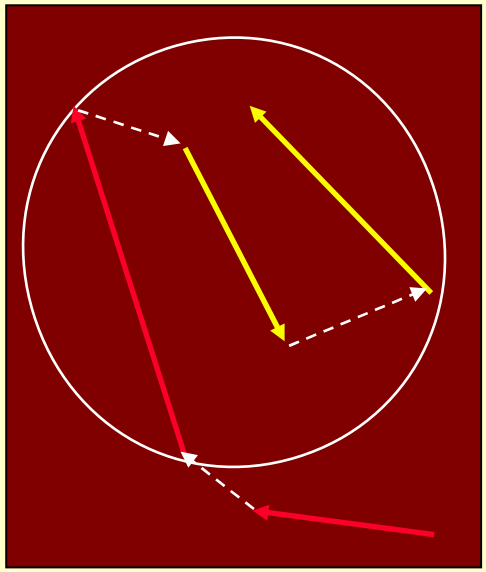
S = S(angle(k,r))

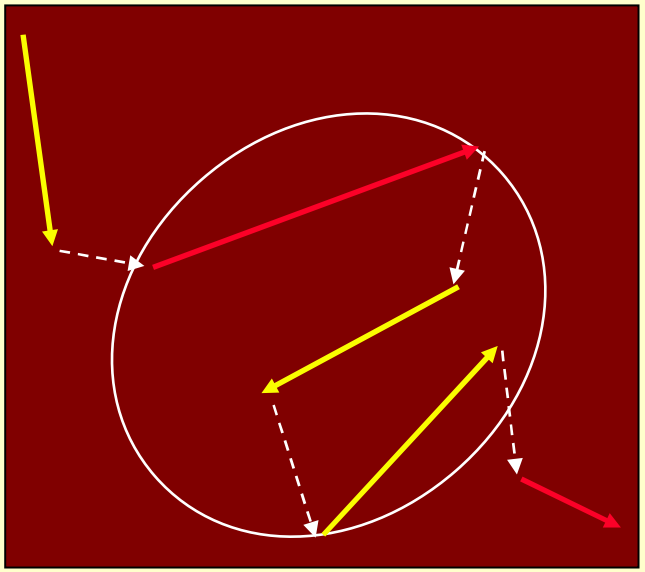
S = S(distance(k,r))

$$S(d) = \left[\frac{2M}{1 + \left[\frac{d}{d_0} \right]^2} - M \right]$$



Aligning Secondary Structure Vectors





		H	H	S	S
	S				
	H				
	S				
	S				
	H				

Best local alignment :

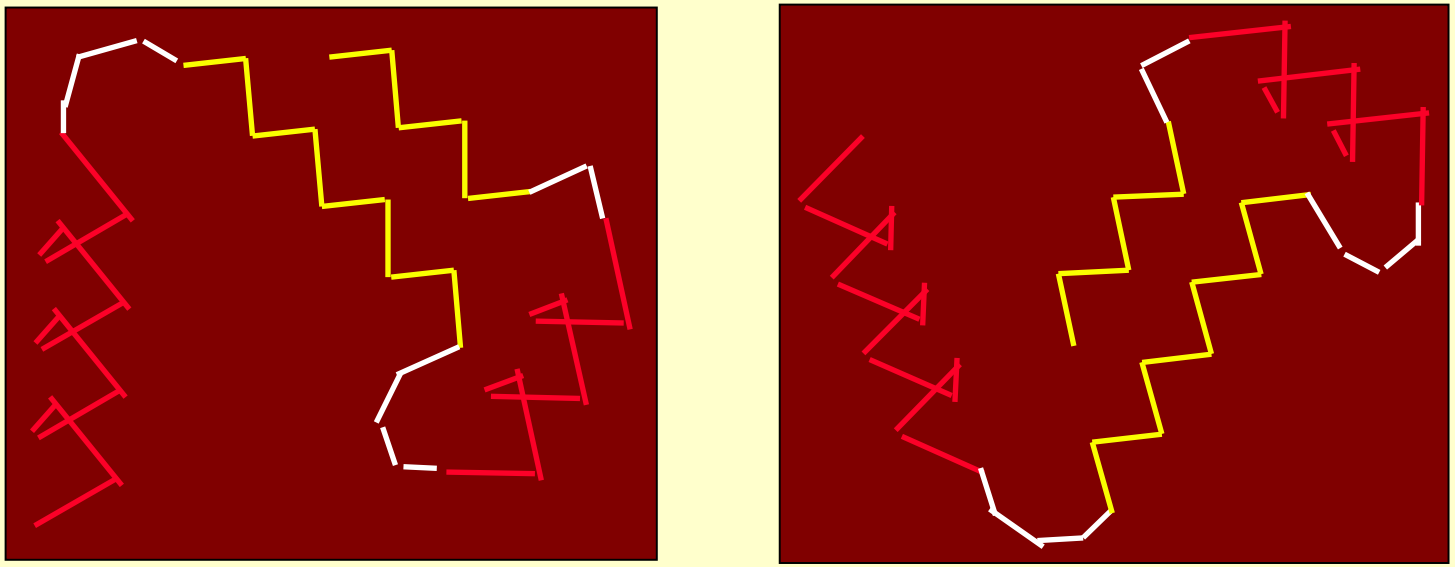
HHSS

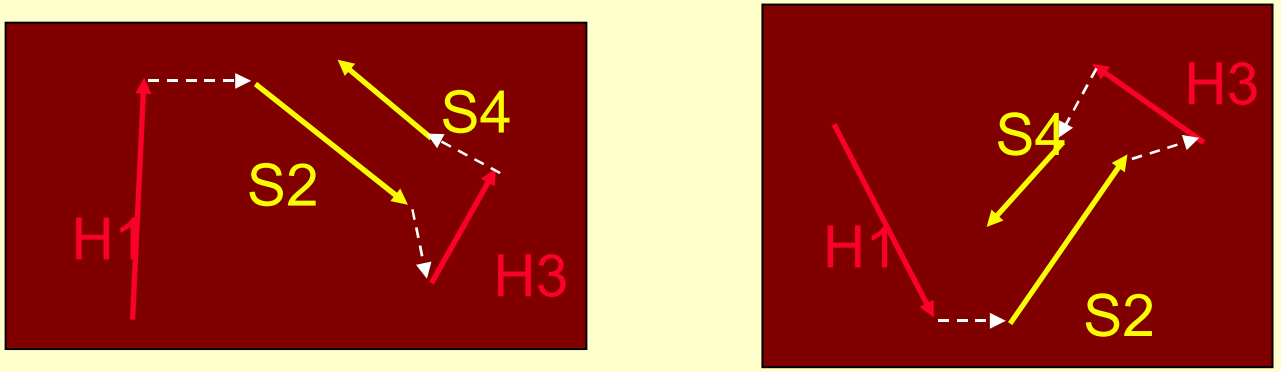
SHSSH

Three Step Algorithm

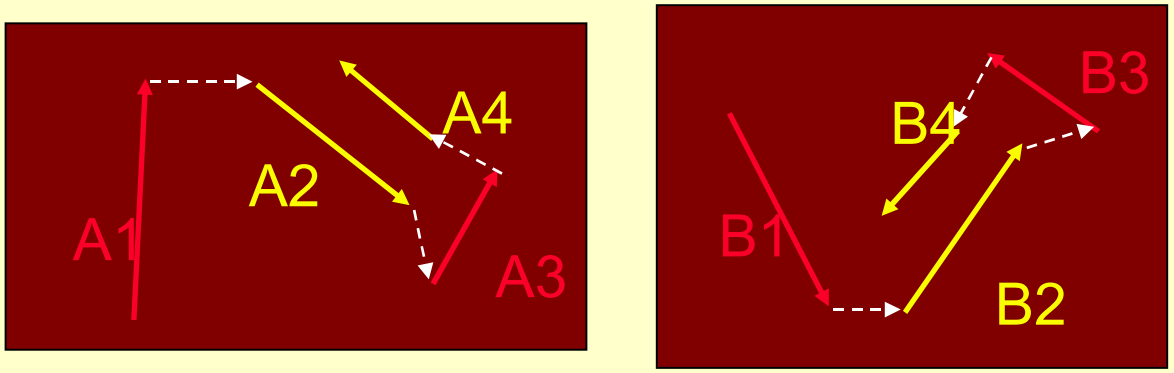
- Local Secondary Structure Superposition
  - Find an initial superposition of the two proteins by using dynamic programming to align the secondary structure vectors
- Atomic Superposition
  - Apply a greedy nearest neighbor method to minimize the RMSD between the C-α atoms from query and the target (i.e. find the nearest local minimum in the alignment space)
- Core Superposition
  - Find the best sequential core of aligned C-α atoms and minimize the RMSD between them

Step 1: Local Secondary Structure Superposition



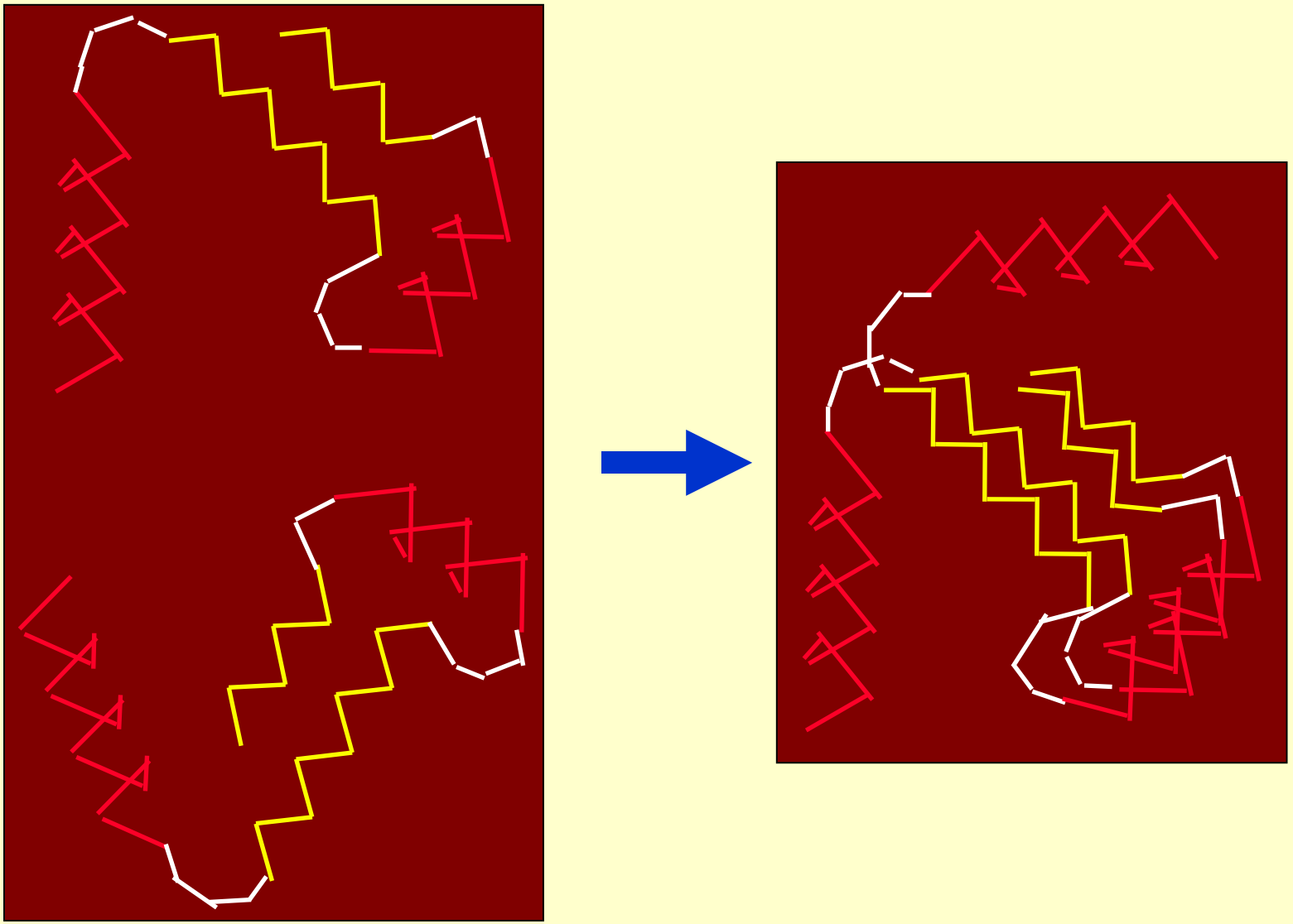


Step 1: Local Secondary Structure Superposition

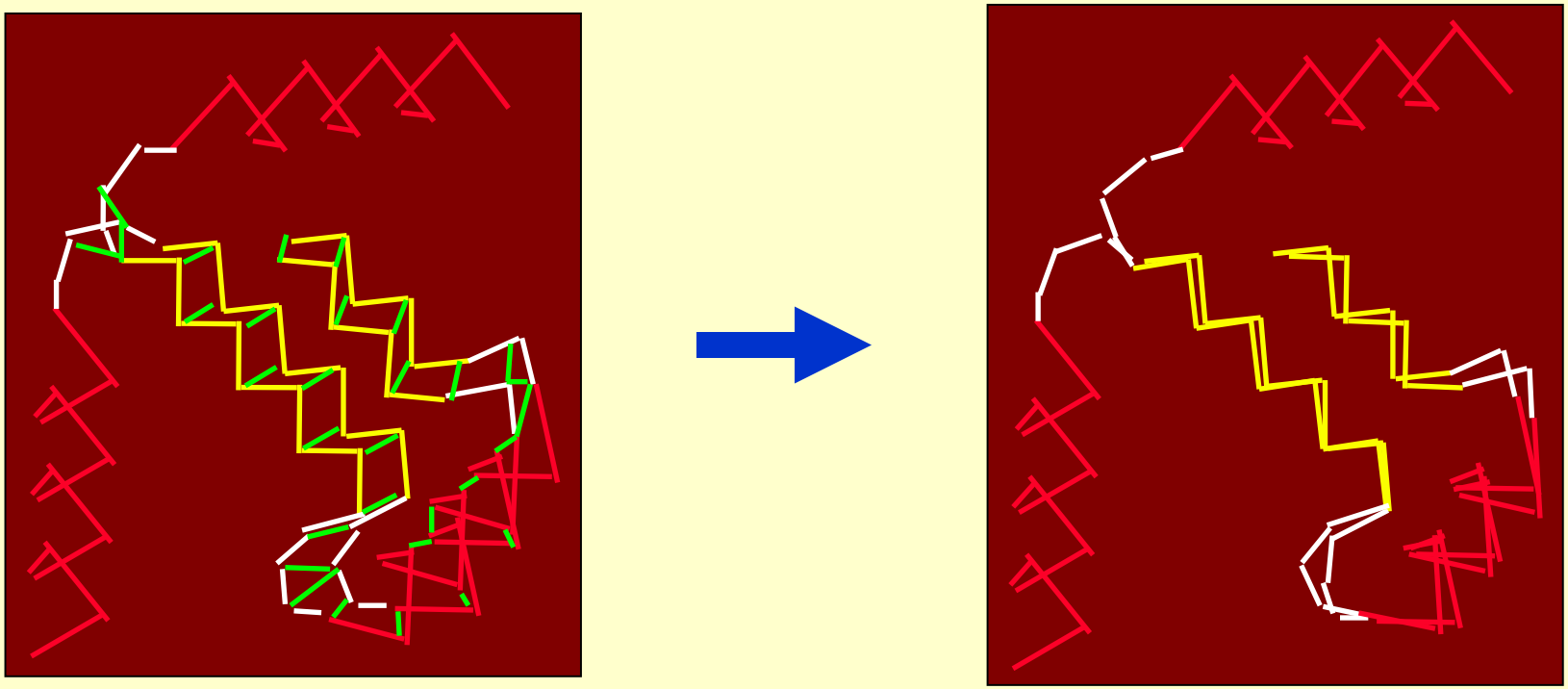


pair	# of aligned vectors	total alignment score
A1,A2 B2,B3	2	32
A3,A4 B3,B4	3	71

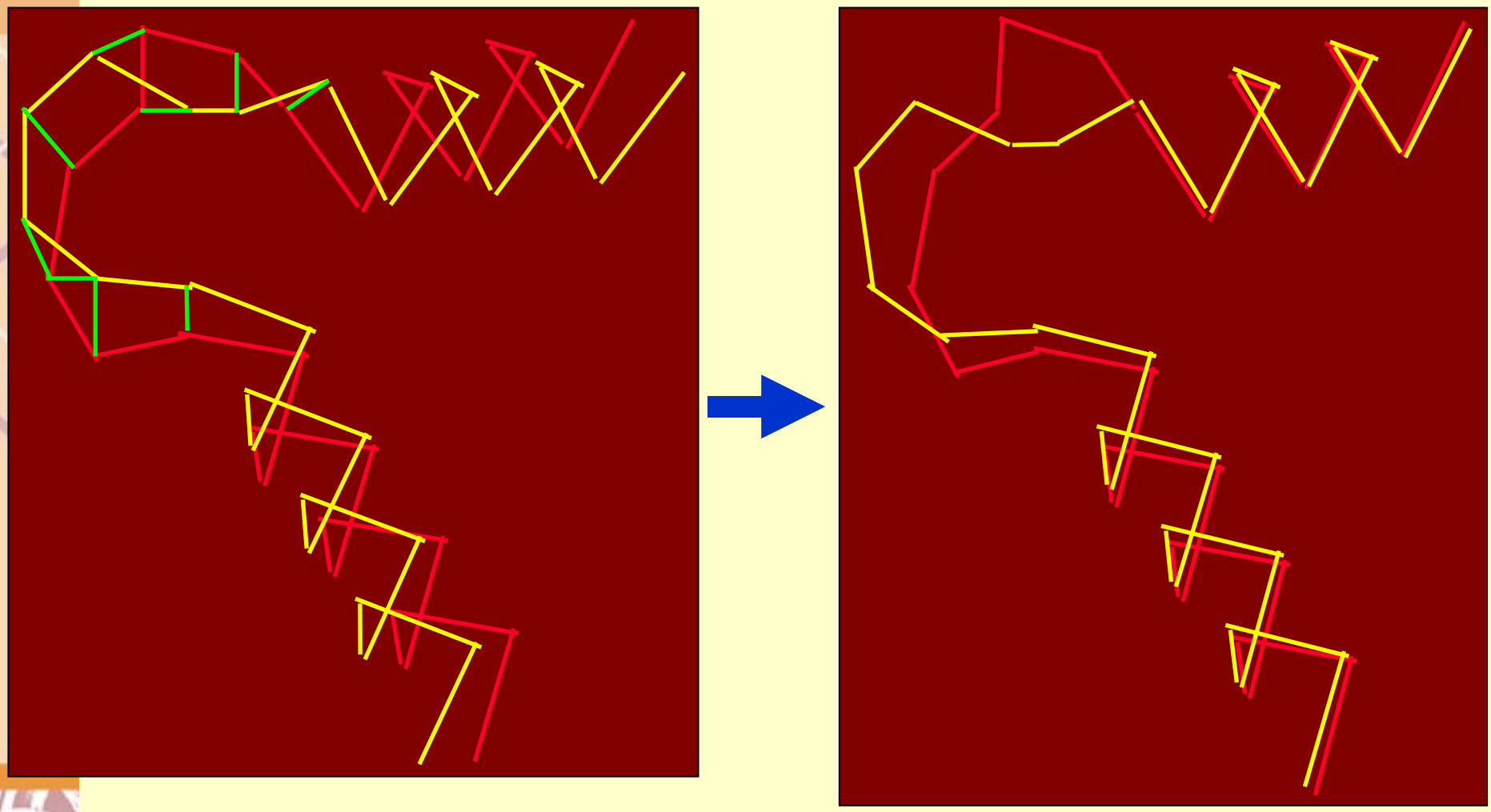
Step 1: Local Secondary Structure Superposition



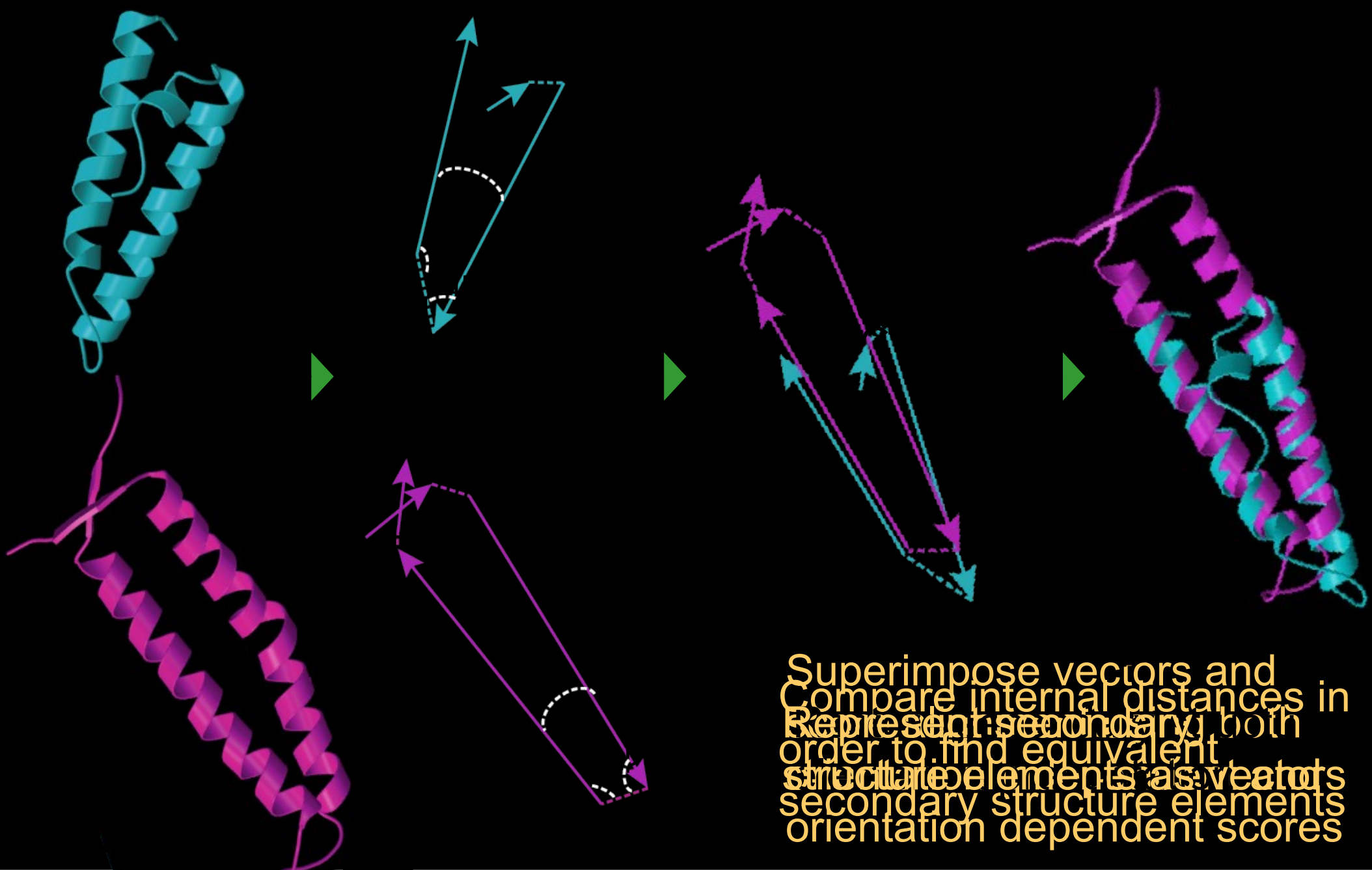
Step 2: Atomic Superposition



Step 3: Core Superposition



LOCK 2: Secondary Structure Element Alignment

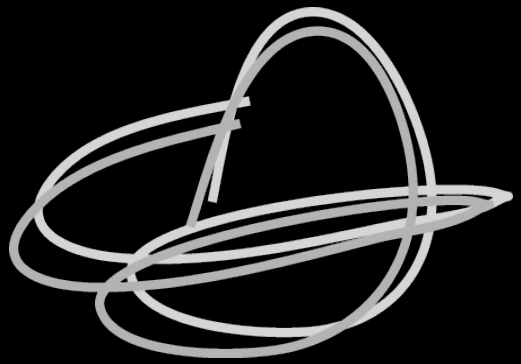


Represent secondary structure elements as vectors

Superimpose vectors and Compare internal distances in order to find equivalent secondary structure elements

orientation dependent scores

Residue Alignment

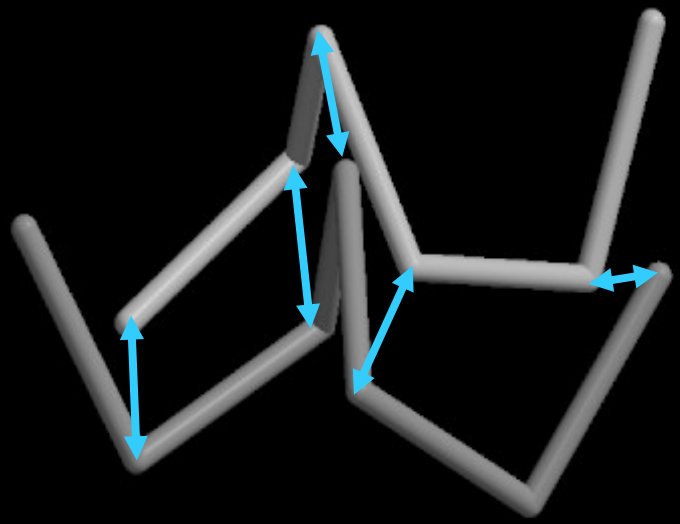


EEKSAVTALWGKV--
GDKKAINKIWPKIYK

superposition

residue registration

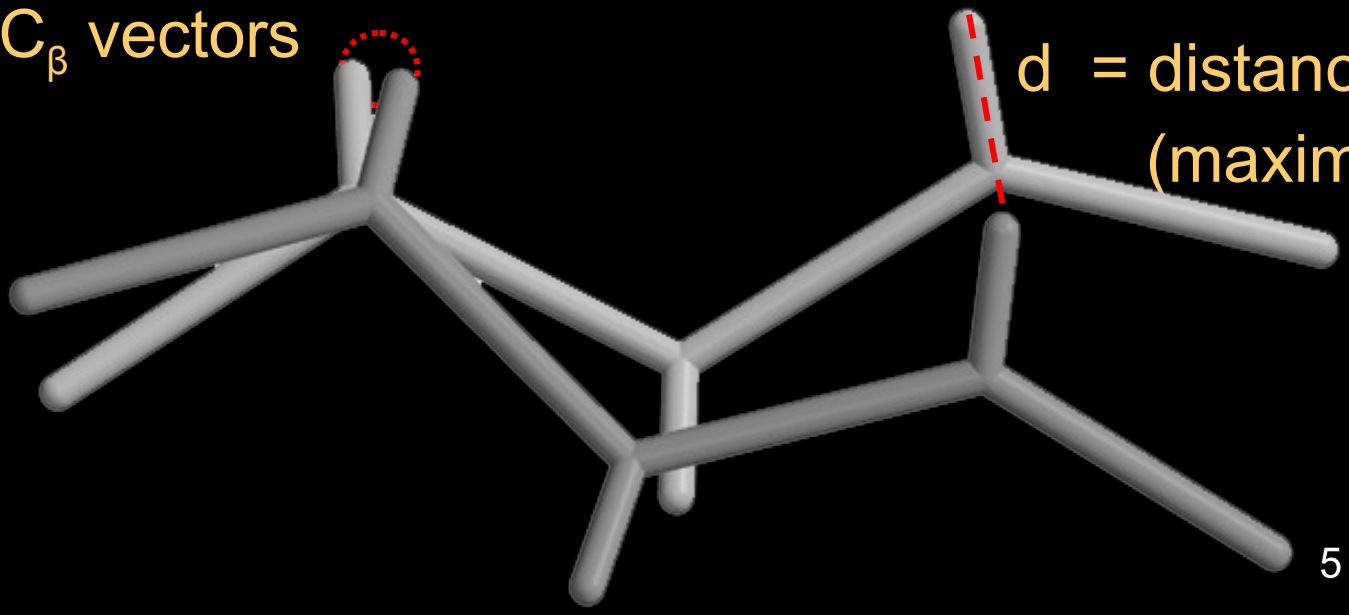
- Naïve approach: Nearest neighbor alpha carbons

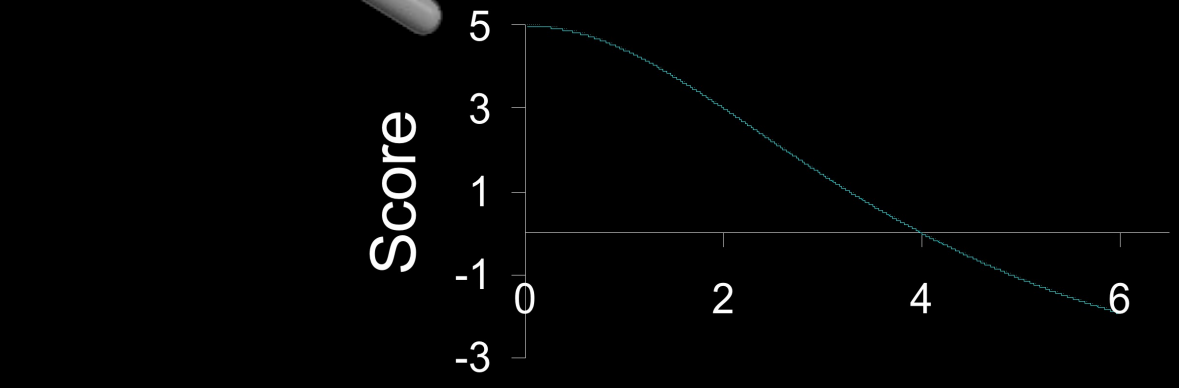


Beta Carbons Encode Directional Information

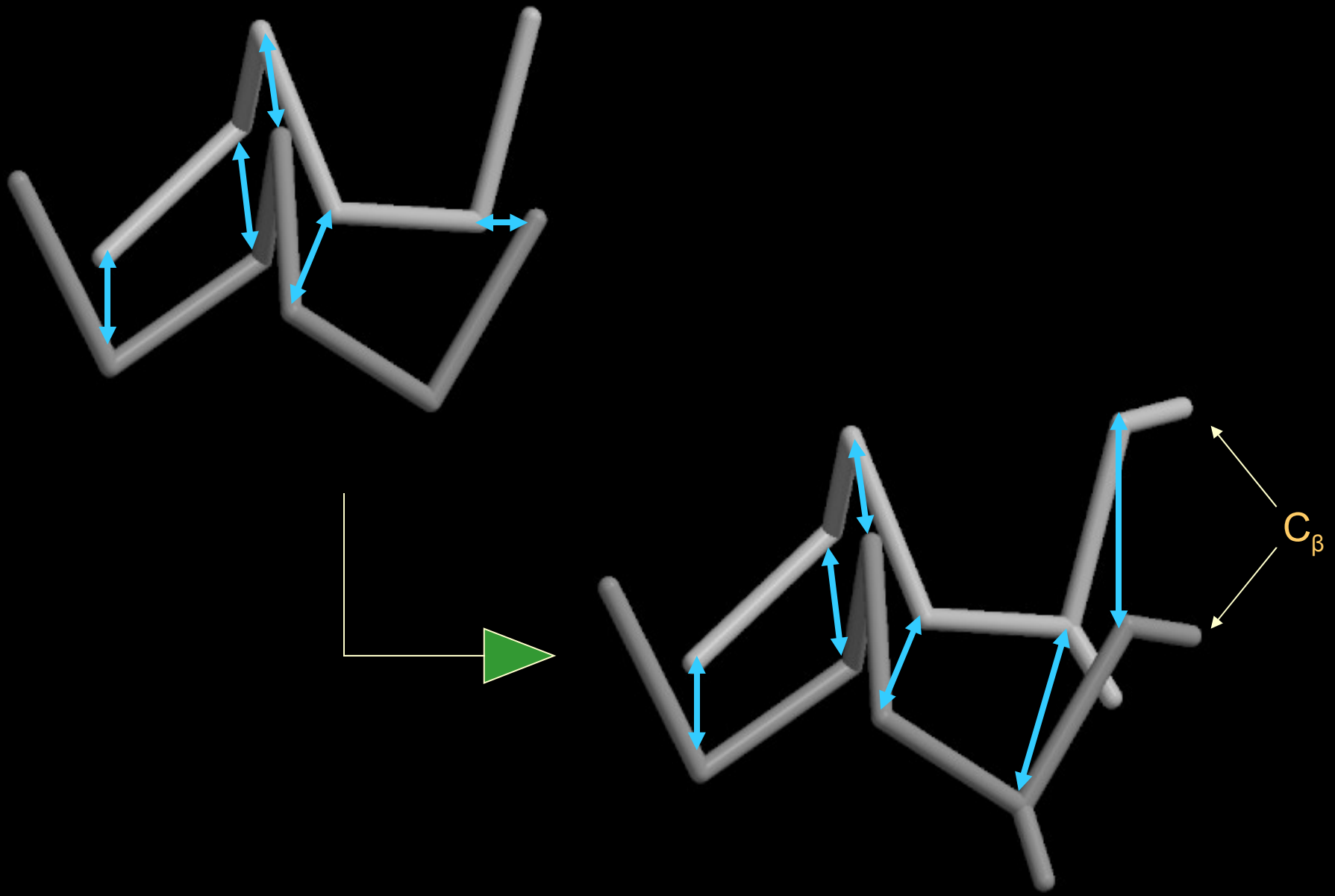
θ = Angle between C_α and C_β vectors

d = distance between C_β atom (maximum 6Å)





New Residue Alignment



Improvements in Consistency

- Consistency: measures the adherence to the transitivity property among all triples of protein structures in a given superfamily

	Globin Superfamily	Immunoglobulin Superfamily
Alpha carbon distances	74.3%	58.6%
Beta carbon positions	80%	59.9%
% increase in aligned residues	37.0%	77.8%

(less than 10% pairwise sequence identity)

New LOCK 2 Properties

- Changes to secondary structure element alignment phase allow for recognition of more distant structural relationships
- Metric scoring function:

  $$1\text{-score}(A,B) + 1\text{-score}(B,C) \leq 1\text{-score}(A,C)$$

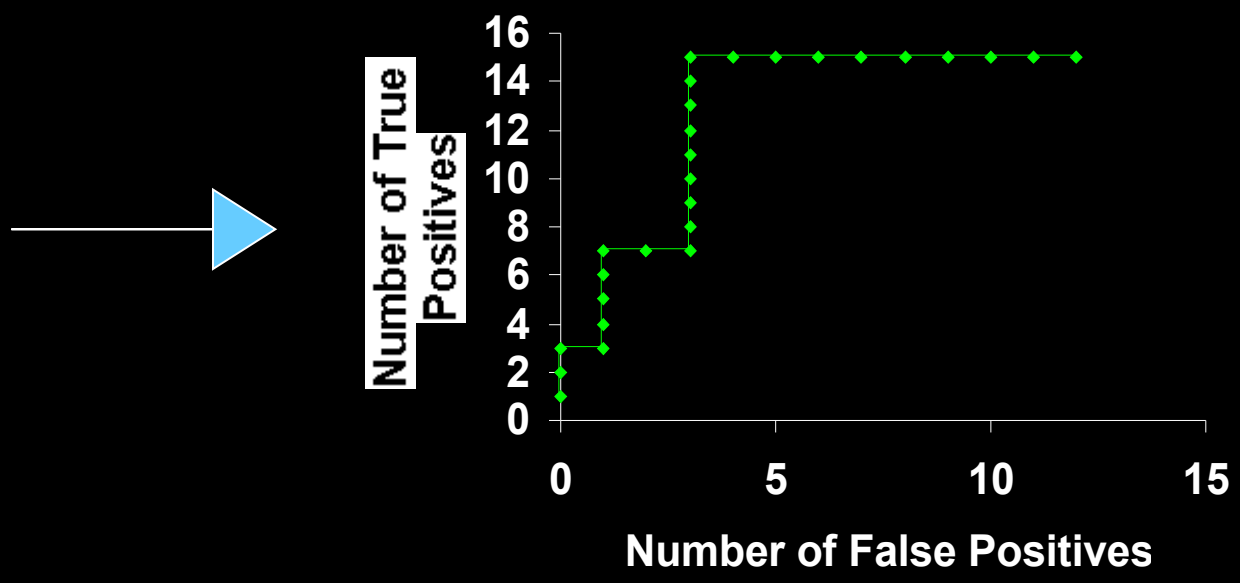
- Biologically relevant residue alignment
- Highly consistent alignments
- Symmetric
- Assessment of statistical significance

FoldMiner: Structure Similarity Search Based on LOCK2 Alignment

- FoldMiner aligns query structure with all database structures using LOCK2
- FoldMiner up weights secondary structure elements in query that are aligned more often
- FoldMiner outperforms CE and VAST is searches for structure similarity

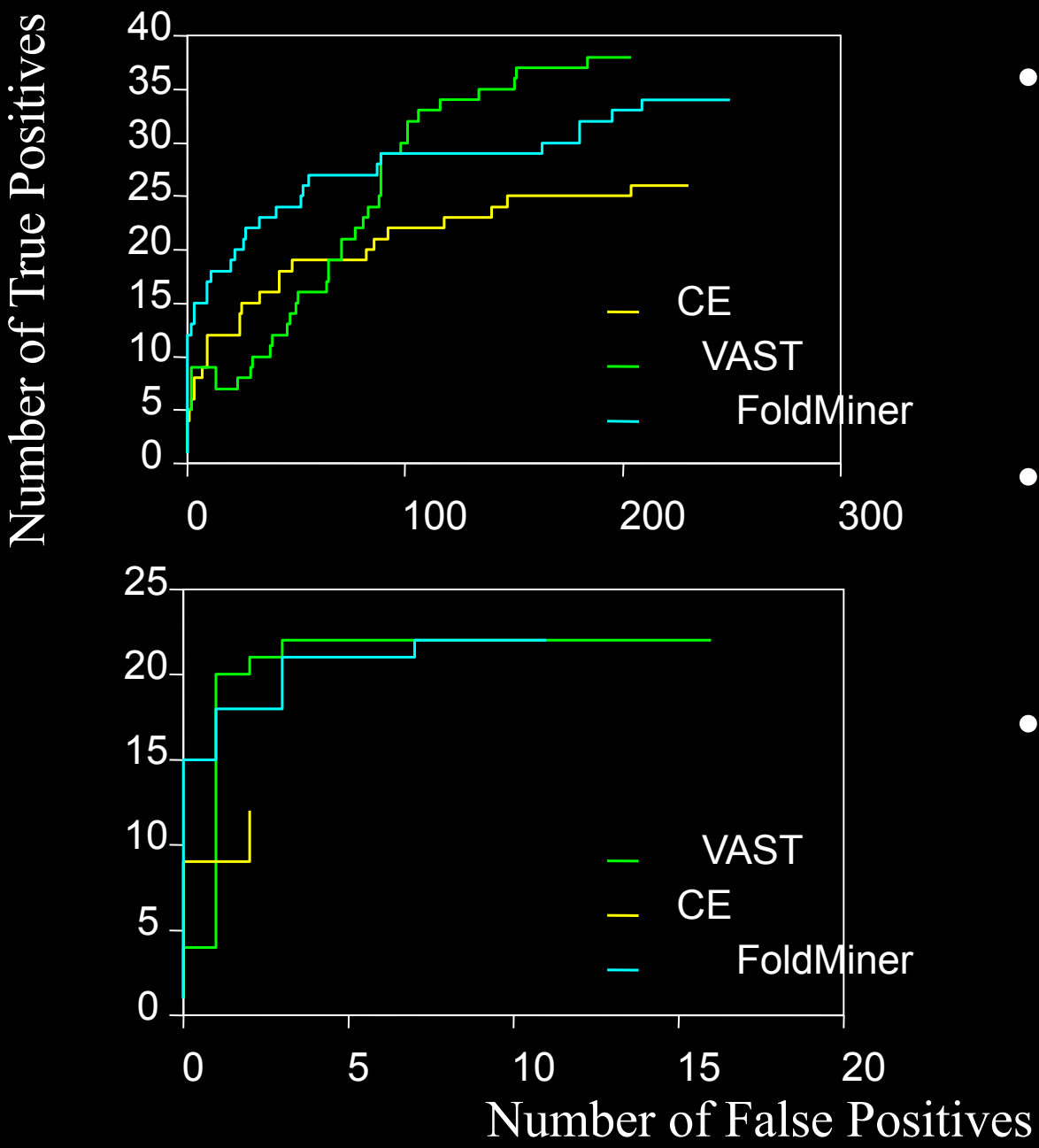
Receiver-Operating Characteristic (ROC) Curves

Rank	Fold
1	Immunoglobulin
2	Immunoglobulin
⋮	⋮
3	p53



- Gold standard: Structural Classification of Proteins (SCOP)
  - SCOP folds: similar arrangement and connectivity of secondary structure elements

Comparing ROC Curves



- Area under the ROC curve correlates with the property of ranking true positives ahead of false positives
- Curves may terminate at different numbers of true and false positives
- Areas can only be directly compared if calculated at points where the two curves cross over one another

Comprehensive Analysis of ROC Curves

	Comparison to VAST			Comparison to CE		
Fold	FoldMiner wins	VAST wins	Ties	FoldMiner wins	CE wins	Ties
Globin-like	6	7	3	7	3	6
Immunoglobulin-like	80	15	21	58	20	38
SH3-like barrels	10	0	7	6	0	11
Flavodoxin-like	34	2	4	18	7	15
Thioredoxins	6	1	19	7	1	18
beta-Grasps	7	8	15	5	4	21
Ferredoxin-like	33	15	25	25	9	39

(less than 25% pairwise sequence identity)

Motif Alignment Results

	Families	Superfamilies
*e*MOTIFs	96.4%	91.6%
Prosite patterns	97.4%	92.6%

LOCK 2
Hierarchical Protein Structure Superposition

LOCK2 Superposition Web Site

http://brutlag.stanford.edu/lock2/

LOCK 2

Hierarchical Protein Structure Superposition

Jessica Ebert
Amit P. Singh
Douglas L. Brutlag
Bioinformatics Group
Stanford University

Query
Enter PDB or SCOP code: 5mbn
OR
Upload file: Choose File no file selected
Specify chain (if present):

Target
Enter PDB or SCOP code: 1gdi
OR
Upload file: Choose File no file selected
Specify Chain (if present):

Alignment Options

Gap Opening Penalty	Gap Extension Penalty	Distance Threshold (for aligned residues)	Geometric Hashing
3.0	2.7	4.0	Fewer Initial Superpositions ○
Reset to Default	Reset to Default	Reset to Default	More Initial Superpositions ◉

Submit Query Clear form

LOCK 2
Hierarchical Protein Structure Superposition

LOCK2 Superposition Web Site

http://brutlag.stanford.edu/lock2/

LOCK 2

Hierarchical Protein Structure Superposition

Jessica Shapiro
Amit P. Singh
Douglas L. Brutlag
Bioinformatics Group
Stanford University

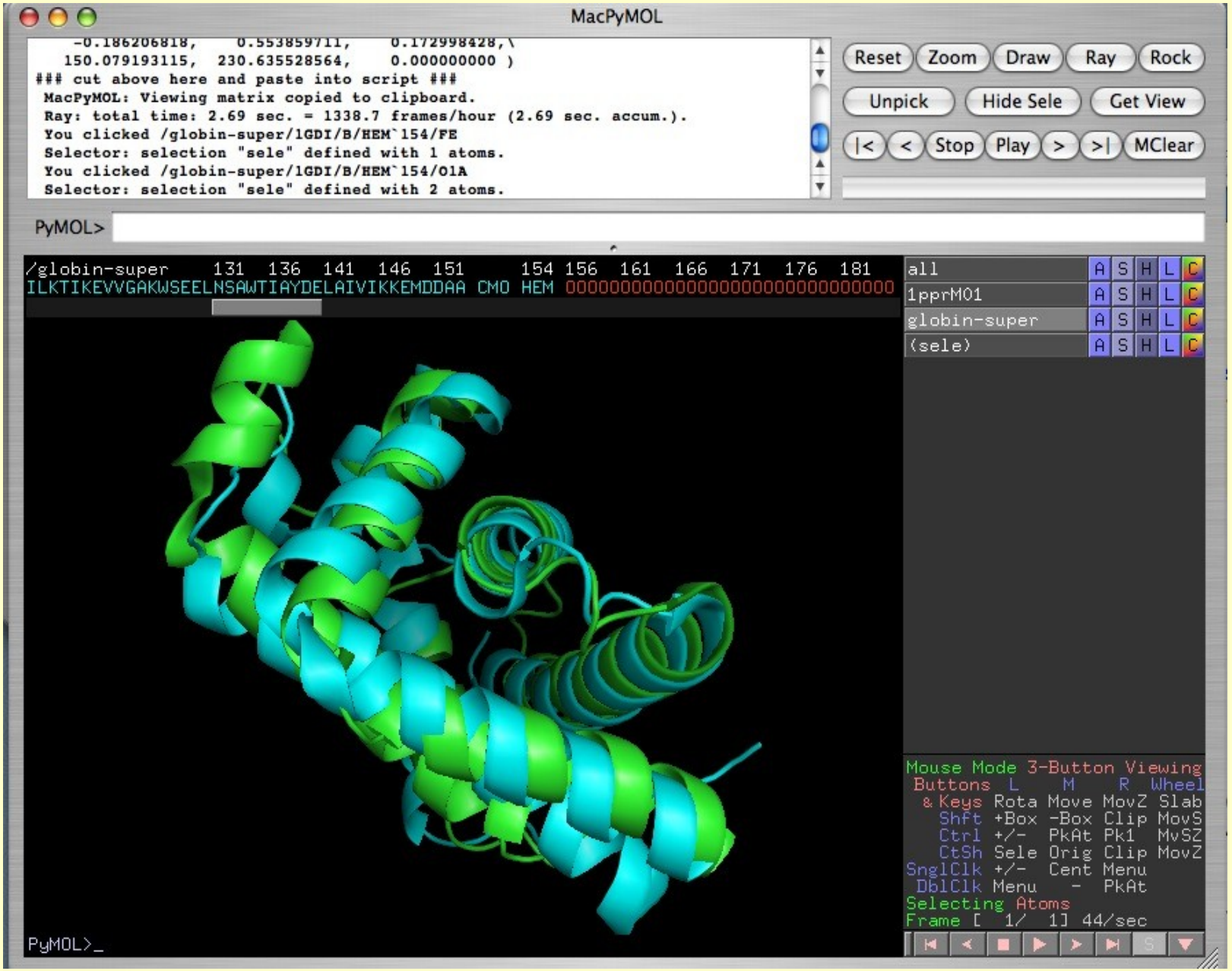
View the aligned structures in RasMol , Chime , or any other molecular viewer of your choice. For best results in RasMol or Chime, select *display:backbone* and *color:chain*. This will color the query molecule in blue and the target in red.

LOCK 2 search results for query: 5mbn

Execution Log
Search results (non-html)

Target	Results	Score	P value	SSEs Aligned	CA Atoms Aligned	RMSD	PDB HEADER
1gdi	Text , PDB , Chime	0.77	9.04e-10	7	142	2.400	OXYGEN TRANSPORT

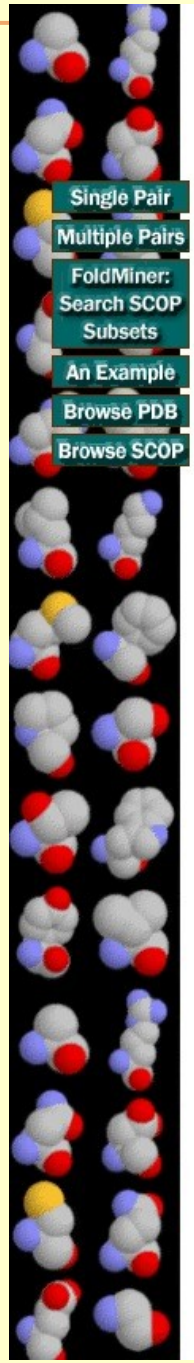
PyMol Display of LOCK2 Superposition



FoldMiner
Structural Similarity Search and Motif Discovery

FoldMiner Structure Search

http://brutlag.stanford.edu/foldminer/



FoldMiner
Structural Similarity Search and Motif Discovery

Jessica Ebert
Amit P. Singh
Douglas L. Brutlag
Bioinformatics Group
Stanford University

Click to run an example structural similarity search: Example

Step 1:

Query
Enter PDB or SCOP code:
OR
Upload file: Choose File no file selected
Specify chain (if present):

Select a Target Database
Pairwise sequence identity less than or equal to 25 (4044 targets) %
Hierarchy representatives at the Superfamily (1227 targets) level

Scroll down to set options, or Submit Query

(If you wish to be notified by email when your search is finished, please enter your email address at the bottom of the page.)

Step 2: Structural Similarity Search Options

Expectation
1.0 e 1
Reset to Default

Conservation Profile's Contribution to Recalculated SSE Scores
0.75 (Value must lie on the interval [0,1])
Reset to Default

Scroll down to set additional options, or Submit Query

(If you wish to be notified by email when your search is finished, please enter your email address at the bottom of the page.)

Step 3: Alignment Options

Gap Opening Penalty
3.0
Reset to Default

Gap Extension Penalty
2.7
Reset to Default

Distance Threshold *(for aligned residues)*
4.0
Reset to Default

Geometric Hashing
Fewer Initial Superpositions
More Initial Superpositions

FoldMiner
Structural Similarity Search and Motif Discovery

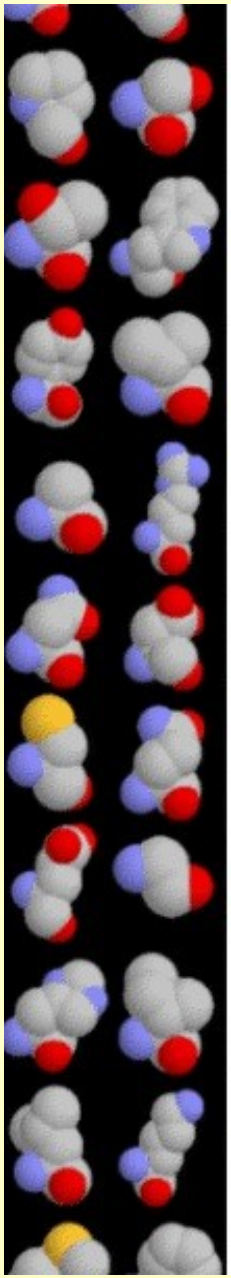
FoldMiner Myoglobin Structure Search

http://brutlag.stanford.edu/foldminer/

FoldMiner
Structural Similarity Search and Motif Discovery

FoldMiner Myoglobin Structure Search

http://brutlag.stanford.edu/foldminer/



Step 3: Alignment Options

Gap Opening Penalty	*Gap Extension Penalty*	*Distance Threshold* (for aligned residues)	*Geometric Hashing*
3.0	2.7	4.0	Fewer Initial Superpositions ○
Reset to Default	Reset to Default	Reset to Default	More Initial Superpositions ◉

To be informed by email when your results are ready, enter your full email address below. If you leave this space blank, the results will appear on the following page as soon as they are available.

Email:

Submit Query Clear form

References:

- Shapiro, J.C. and Brutlag, D.L. (2004). FoldMiner: Structural Motif Discovery Using an Improved Superposition Algorithm *Protein Science* 13(1) 278-294. Abstract Full Text
- Shapiro, J. and Brutlag, D. (2004). FoldMiner and LOCK 2: protein structure comparison and motif discovery on the web. *Nucleic Acids Research* 32(Web server issue) W536-W541. Abstract Full Text
- Singh, A.P. and Brutlag, D.L. (1997). Hierarchical Protein Structure Superposition using both Secondary Structure and Atomic Representations. *Proc. Intelligent Systems for Molecular Biology* 5:284-293. Abstract

Suggestions, comments, bugs to: Jessica Ebert.

FoldMiner
Structural Similarity Search and Motif Discovery

FoldMiner Myoglobin Structure Search

http://brutlag.stanford.edu/foldminer/

FoldMiner
Structural Similarity Search and Motif Discovery

FoldMiner search results for query: 1mbn
(Bring control panel to front)
View: SCOP fold statistics Execution Log (very large file!) Search results summary (help) Alignment results (help)

	Target Sort by fold	View Results	Score Sort	P value Sort	SSEs aligned Sort	CA Atoms Aligned Sort	RMSD Sort	PDB HEADER Sort
1	d1a6m__	Text , PDB , Chime	0.94	8.3e-10	8	151	0.604	SCOP/ASTRAL domain d1a6m__ [15018]
2	d1hlb__	Text , PDB , Chime	0.84	8.7e-09	8	138	2.220	SCOP/ASTRAL domain d1hlb__ [15625]
3	d1mba__	Text , PDB , Chime	0.83	9.7e-09	8	139	2.303	SCOP/ASTRAL domain d1mba__ [15149]
4	d1irda_	Text , PDB , Chime	0.82	1.4e-08	7	140	1.620	SCOP/ASTRAL domain d1irda_ [66286]
5	d2gdm__	Text , PDB , Chime	0.80	2e-08	7	144	2.502	SCOP/ASTRAL domain d2gdm__ [15212]
6	d1ash__	Text , PDB , Chime	0.79	2.7e-08	8	133	1.916	SCOP/ASTRAL domain d1ash__ [15622]
7	d1gcwb_	Text , PDB , Chime	0.78	3.5e-08	7	127	1.925	SCOP/ASTRAL domain d1gcwb_ [15591]
8	d1itha_	Text , PDB , Chime	0.77	4.7e-08	8	138	1.922	SCOP/ASTRAL domain d1itha_ [15623]
9	d3sdha_	Text , PDB , Chime	0.76	6.3e-08	8	133	2.047	SCOP/ASTRAL domain d3sdha_ [14984]
10	d1kr7a_	Text , PDB , Chime	0.74	1e-07	7	105	2.480	SCOP/ASTRAL domain d1kr7a_ [72890]
11	d1it2a_	Text , PDB , Chime	0.73	1.2e-07	7	133	1.862	SCOP/ASTRAL domain d1it2a_ [66365]
12	d1cqxa1	Text , PDB , Chime	0.72	1.7e-07	7	131	2.287	SCOP/ASTRAL domain d1cqxa1 [15635]
13	d1h97a_	Text , PDB , Chime	0.71	1.9e-07	7	138	2.437	SCOP/ASTRAL domain d1h97a_ [60812]
14	d1jl7a_	Text , PDB , Chime	0.68	3.9e-07	7	135	1.901	SCOP/ASTRAL domain d1jl7a_ [71726]
15	d1ew6a_	Text , PDB , Chime	0.68	4.1e-07	8	130	1.960	SCOP/ASTRAL domain d1ew6a_ [15637]
16	d1dlwa_	Text , PDB , Chime	0.59	3.6e-06	6	103	2.636	SCOP/ASTRAL domain d1dlwa_ [14982]
17	d1b0b__	Text , PDB , Chime	0.49	3.9e-05	5	135	1.810	SCOP/ASTRAL domain d1b0b__ [15010]
18	d1phna_	Text , PDB , Chime	0.40	0.00041	6	107	3.174	SCOP/ASTRAL domain d1phna_ [15641]
19	d1cuk_2	Text , PDB , Chime	0.39	0.00052	5	52	3.382	SCOP/ASTRAL domain d1cuk_2 [17946]
20	d1nekb1	Text , PDB , Chime	0.35	0.0012	6	74	3.350	SCOP/ASTRAL domain d1nekb1 [80429]
21	d1qaxa1	Text , PDB , Chime	0.35	0.0014	4	54	2.900	SCOP/ASTRAL domain d1qaxa1 [39376]
22	d1e3oc2	Text , PDB , Chime	0.34	0.0016	4	56	2.754	SCOP/ASTRAL domain d1e3oc2 [59198]

Select	SCOP Fold	Count
☑	Globin-like	19
☐	Ferredoxin-like	1
☐	SAM domain-like	1
☐	lambda repressor-like DNA-binding domains	1

FoldMiner
Structural Similarity Search and Motif Discovery

FoldMiner Myoglobin Structure Search

http://brutlag.stanford.edu/foldminer/

This site requires popup windows to function properly. Click here if your browser blocks them, or disable the popup blocker and reload

Bring control panel to front Bring search results table to front

FoldMiner identified the motif displayed in the left frame by determining which secondary structure elements frequently align well to homologous structures. It may be possible to identify additional structural neighbors or additional motifs by excluding poorly or well conserved secondary structural elements, respectively.

To attempt to identify another motif, please select from the list below seconday structure elements that should be *excluded* from the motif discovery algorithm.

Exclude	Residues	Conservation	Type	Blink
☑	*3-18*	--	*Helix*	
☐	20-35	0.64	Helix	
☐	36-42	0.70	Helix	
☑	*51-57*	--	*Helix*	
☐	58-77	0.71	Helix	
☐	86-94	0.66	Helix	
☐	100-118	0.83	Helix	
☐	125-148	0.63	Helix	

To reanalyze your results using different search parameters: If you do not exclude any secondary structure elements, your results will be reanalyzed using the search parameters you specify. (Leave all checkboxes above unchecked and specify your search parameters below.)

Structural Similarity Search Options

Expectation: 1.0 e 1 — Reset to Default

Alignment Scope: (•) Global ( ) Local

Conservation Profile's Contribution to Recalculated SSE Scores: 0.75 (Value must lie on the interval [0,1]) — Reset to Default

The query structure is either not in SCOP or is a member of a SCOP fold for which a background score distribution does not exist. By default, a background distribution encompassing all of SCOP will be used to analyze your results. You may specifiy a more specific distribution here.

Use background distribution for: All of SCOP

Submit

ModLink+

http://sbi.imim.es/modlink/