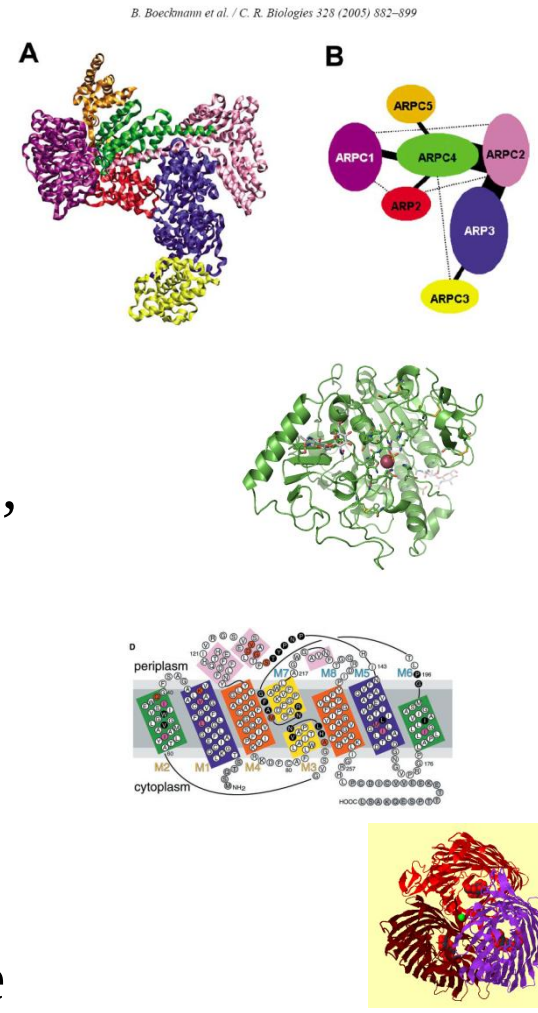
Protein 3D-structure analysis

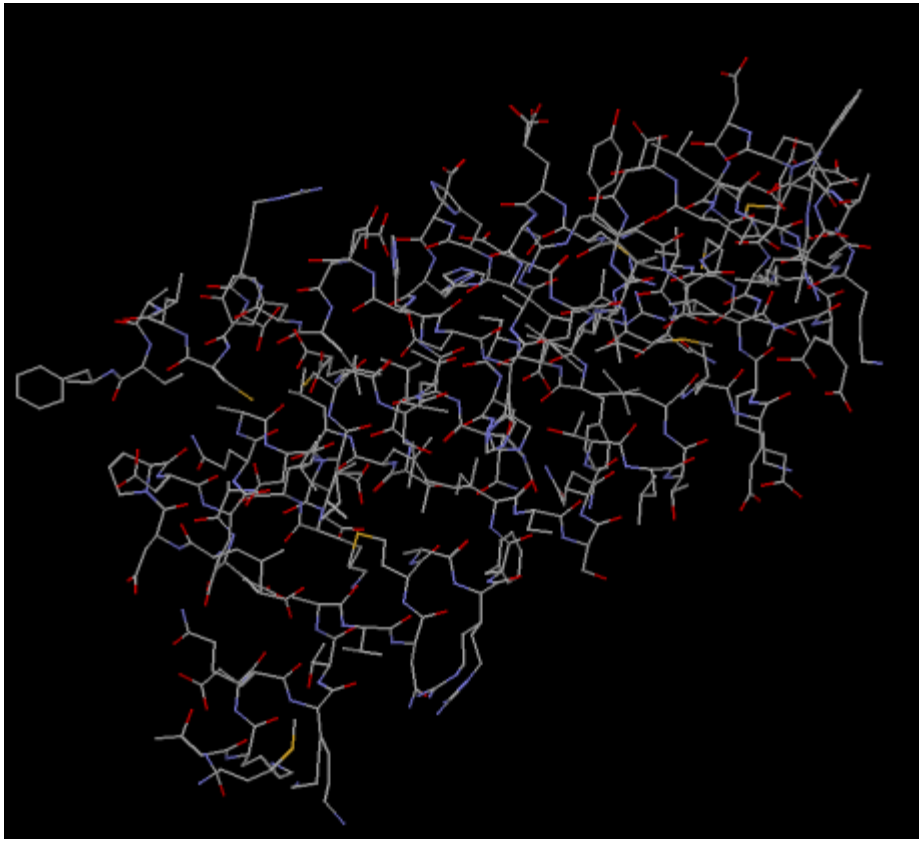
why and how

3D-structures are precious sources of information

- Shape and domain structure
- Protein classification
- Prediction of function for uncharacterized proteins
- Interaction with other macromolecules
- Interactions with small ligands: metal ions, nucleotides, substrates, cofactors and inhibitors
- Evidence for enzyme mechanism
- Structure-based drug development
- Posttranslational modifications: disulfide bonds, N-glycosylation,…
- Experimental evidence for transmembrane domains



Structure of the polypeptide chain



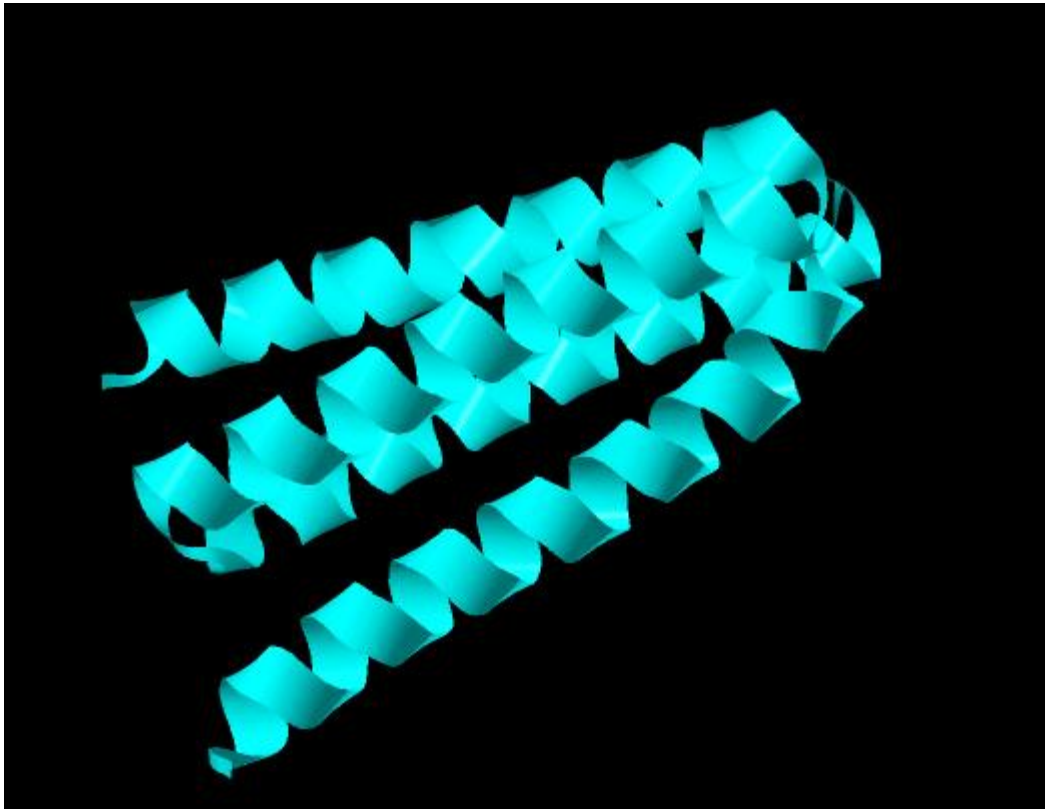
Visualization with DeepView

Convenient, but real molecules fill up space

Colors:
Carbon=light grey
Nitrogen=blue
Oxygen=red
Sulfur=yellow

1B6Q

Structure of the polypeptide chain

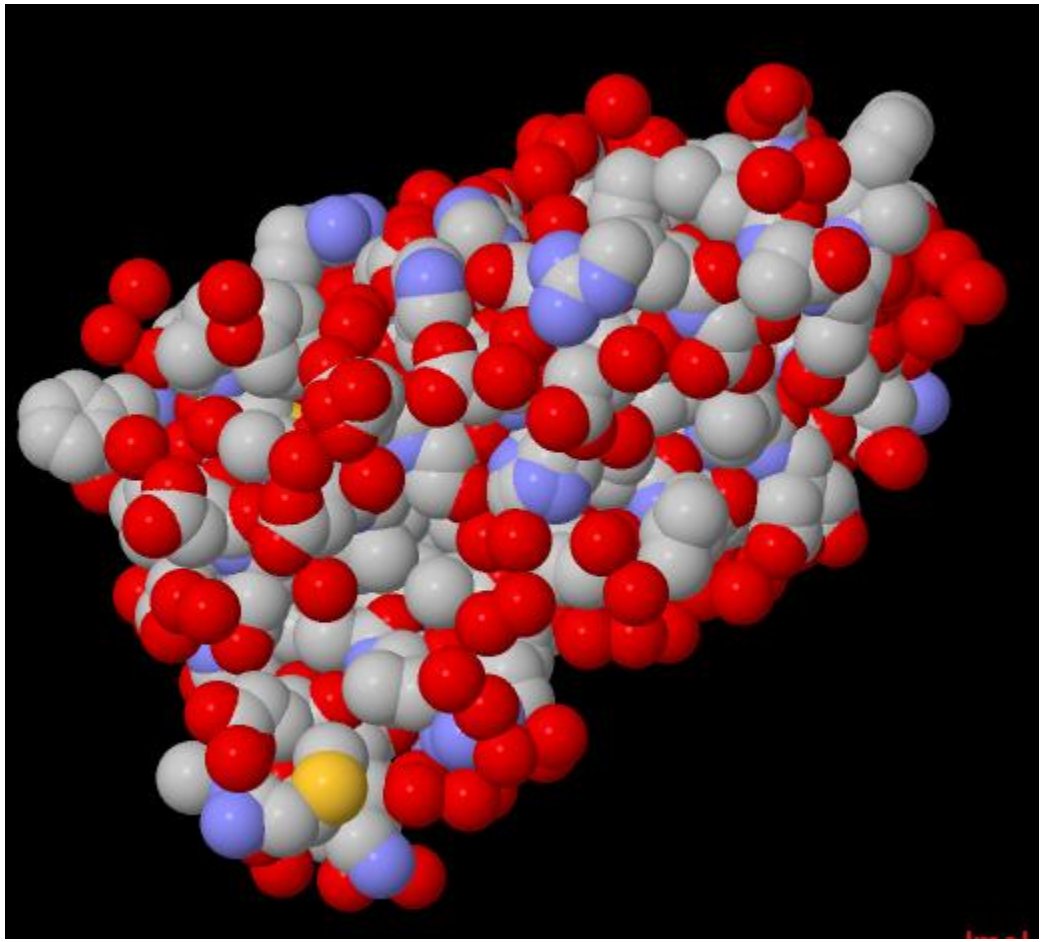


JMOL cartoon

Another view of the same

1B6Q

Structure of the polypeptide chain



JMOL

Space-filling model of the same

Colors:
Carbon=light grey
Nitrogen=blue
Oxygen=red
Sulfur=yellow

1B6Q

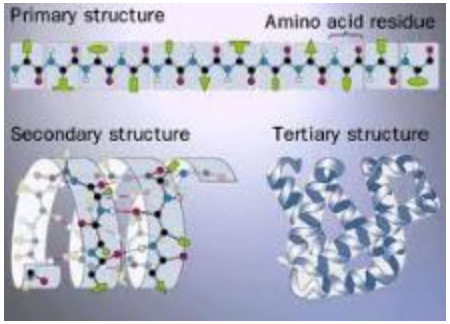
Basics of protein structure

Primary structure

Secondary structure

Tertiary structure

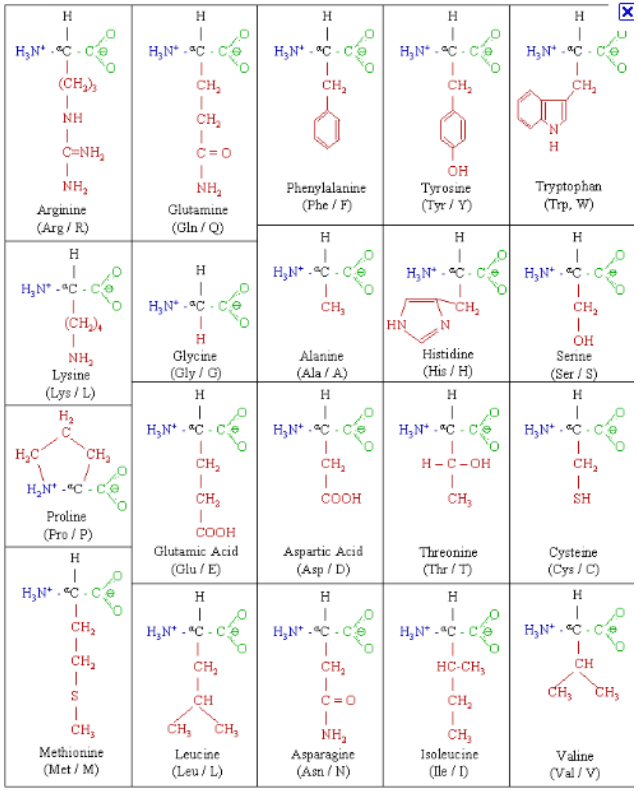
Quaternary structure



Nota bene: some proteins are inherently disordered

Primary structure: the protein sequence

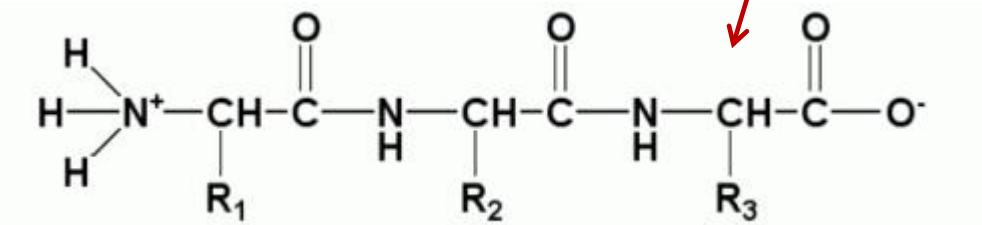
20 amino acids with different characteristics:
small, large, polar, lipophilic, charged,…



- http://schoolworkhelper.net/amino-acids-categories-function/

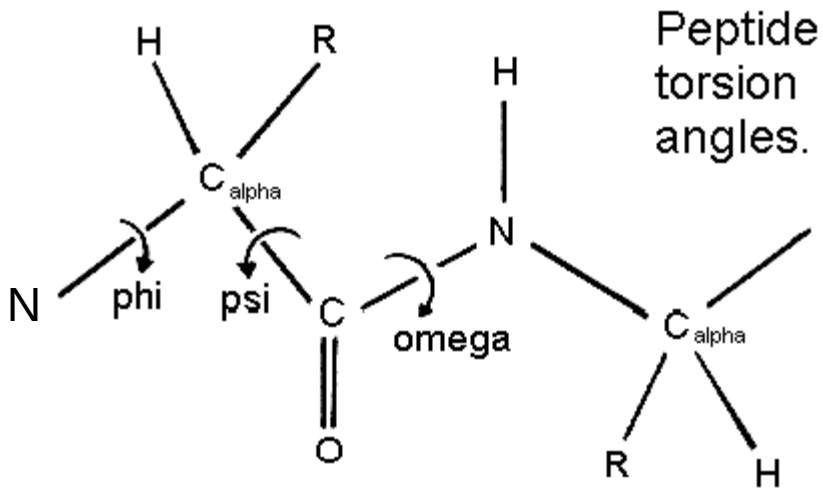
Structure of the polypeptide chain

Alpha carbon atom



http://www.ncbi.nlm.nih.gov/bookshelf/br.fcgi?book=bioinfo&part=A135#A146

The folding pattern of a polypeptide chain can be described in terms of the angles of rotation around the main chain bonds

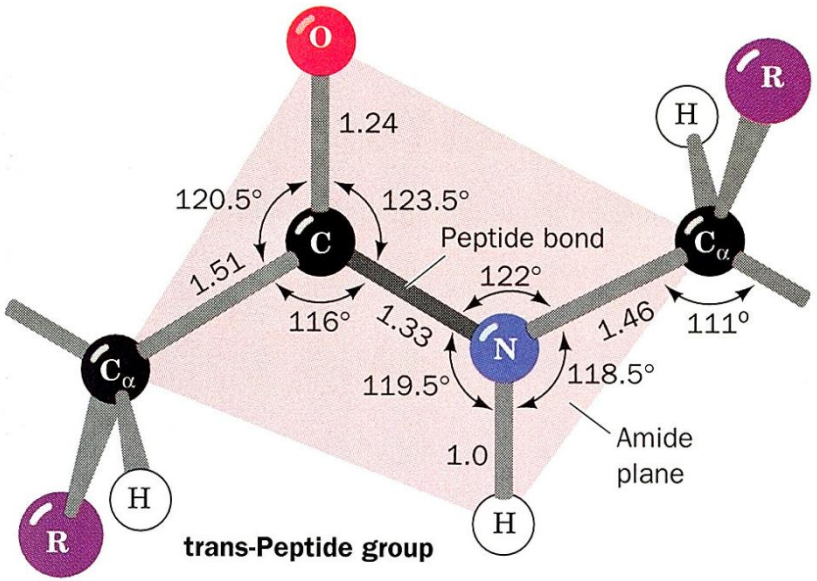


Phi and psi describe the main chain conformation.

Omega corresponds to the trans (omega=180) or cis (omega=0) conformation. Except for Pro, trans is the more stable conformation

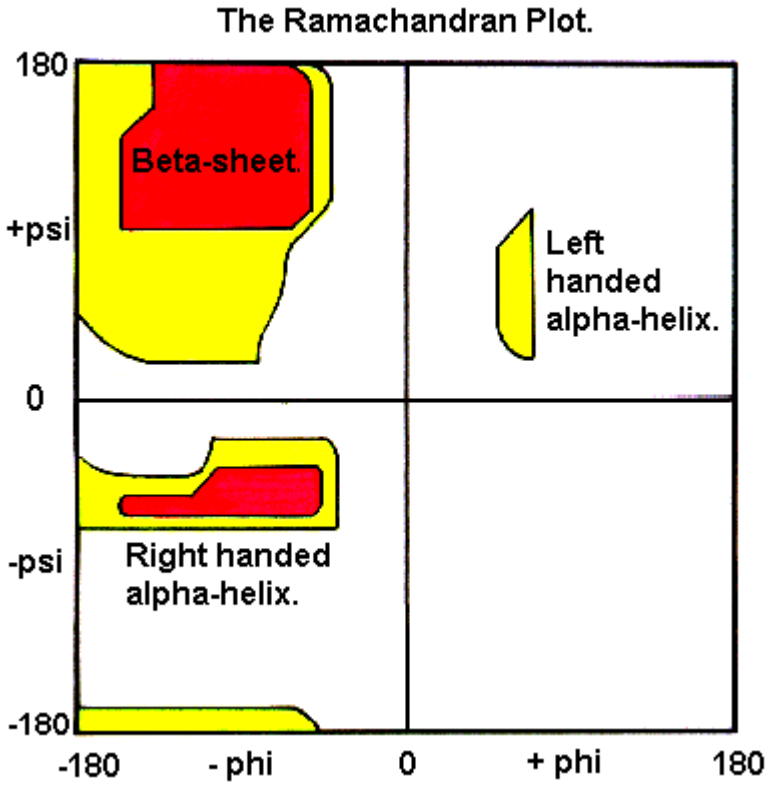
Key facts about a polypeptide chain:

- Chemical bonds have characteristic lengths.
- The peptide bond has partial double-bond character, meaning it is shorter, and rigid
- Other bonds are single bonds (but: restriction of rotation due to steric hindrance)



Ramachandran plot (1) :

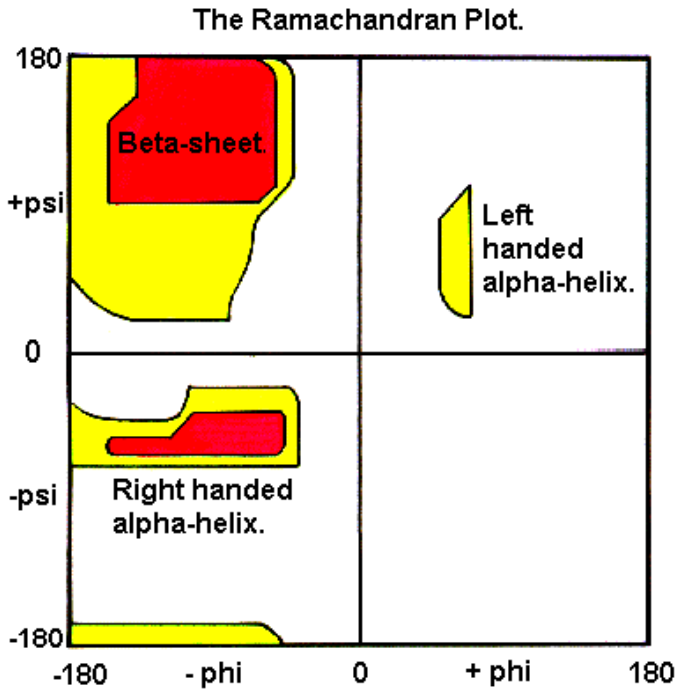
Each type of secondary structure has a characteristic combination of phi and psi angles



http://swissmodel.expasy.org/course/text/chapter1.htm

Ramachandran plot (2) :

For each possible conformation, the structure is examined for close contacts between atoms. Atoms are treated as hard spheres with dimensions corresponding to their van der Waals radii. Angles, which cause spheres to collide correspond to sterically disallowed conformations of the polypeptide backbone (white zone).



Red: no steric hindrance

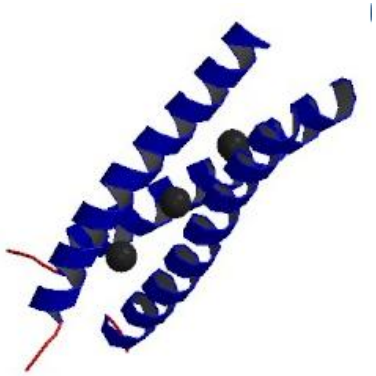
Yellow: some steric constraints

White: “forbidden zone”

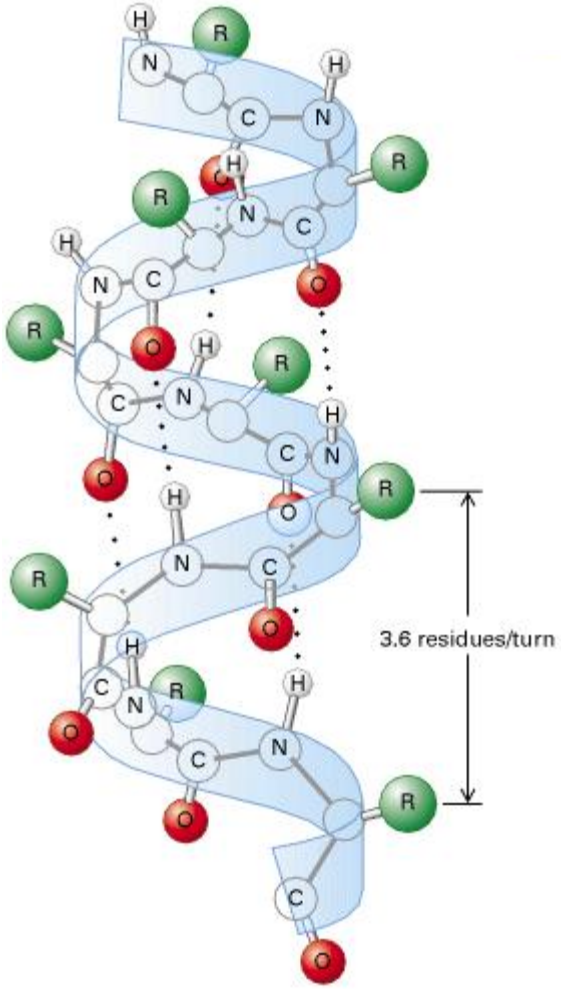
Exception: Gly has no side chain and can be found in the white region

When determining a protein structure, nearly all residues should be in the permitted zone (excepting a few Gly)

Secondary structure:
helices, strands, turns and loops



Secondary structure:
Alpha-helix

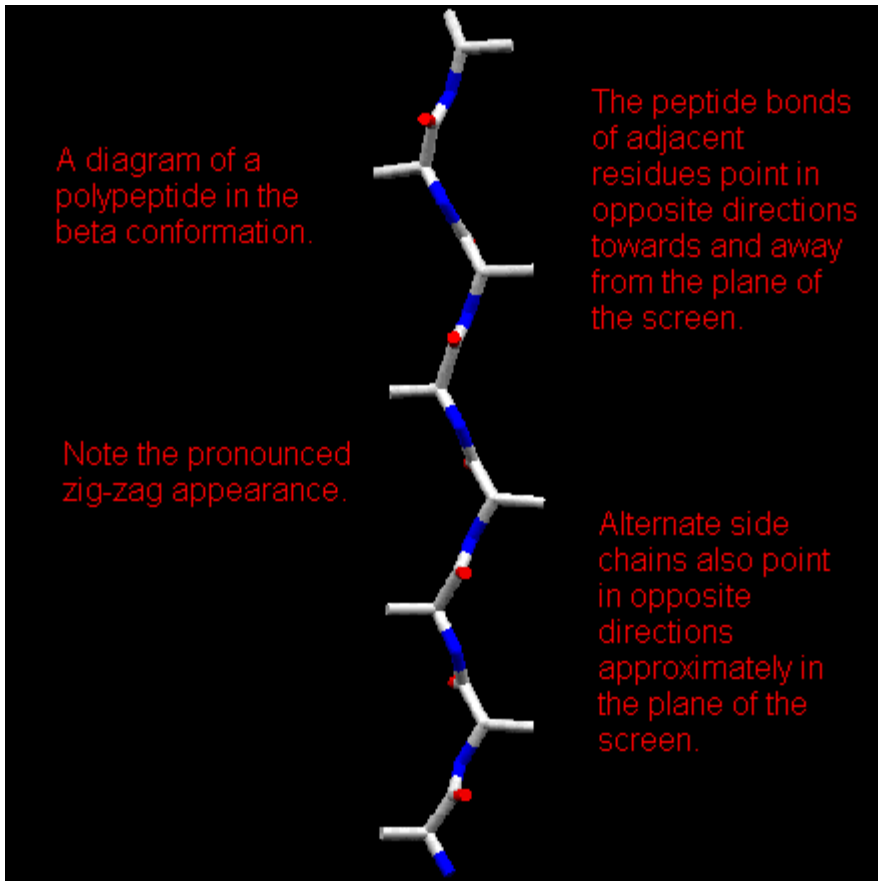


Characteristics:
Helical residues have negative phi and psi angles, typical values being -60 degrees and -50 degrees

Every main chain C=O and N-H group is hydrogen-bonded to a peptide bond 4 residues away (i.e. O_i to N_{i+4}). This gives a very regular, stable arrangement.

3.6 residues per turn
5.4 Å repeat along the helix axis
Each residue corresponds to a rise of ca. 1.5 Å

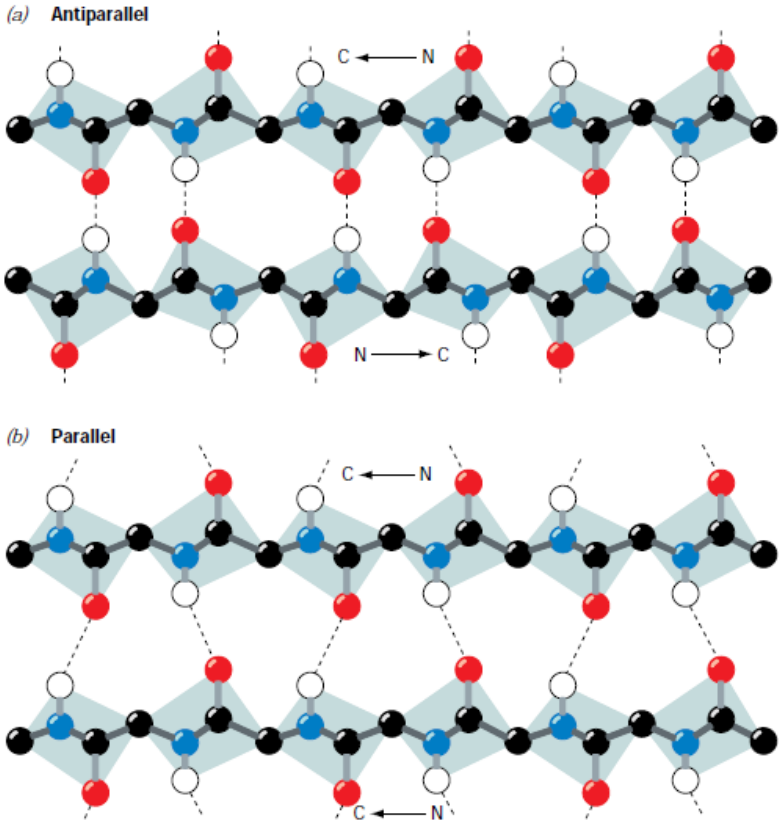
Secondary structure: beta strands



Characteristics:
Positive psi angles, typically ca. 130 degrees, and negative phi values, typically ca. -140 degrees
No hydrogen bonds amongst backbone atoms from the same strand!

http://swissmodel.expasy.org/course/text/chapter1.htm

Beta strands can form parallel or antiparallel beta-sheets



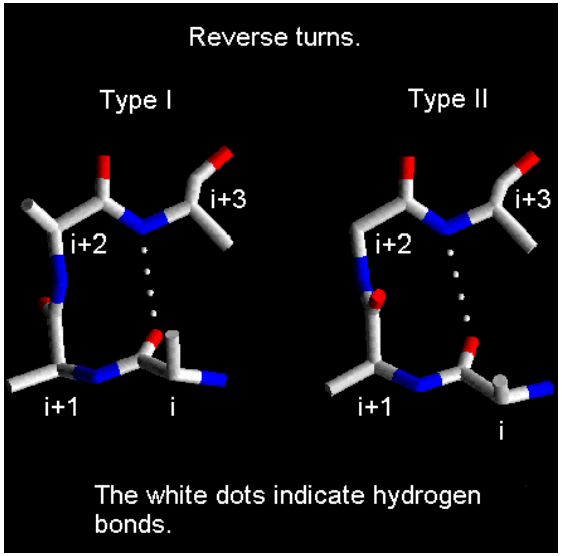
Characteristics:
Stabilized by hydrogen bonds between backbone atoms from adjacent chains
The axial distance between adjacent residues is 3.5 Å.
There are two residues per repeat unit which gives the beta-strand a 7 Å pitch

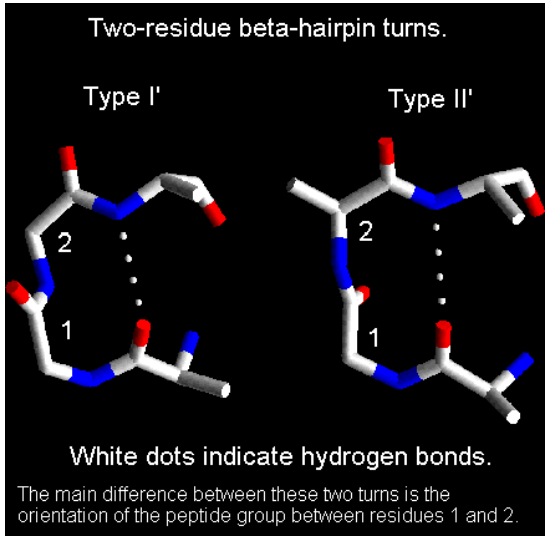
Turns and loops

Loop between beta strands

Loop: general name for a mobile part of the polypeptide with no fixed secondary structure

Turn: several types, defined structure, requirement for specific aa at key positions, meaning they can be predicted. The polypeptide chain "makes a U-turn" over 2-5 residues.

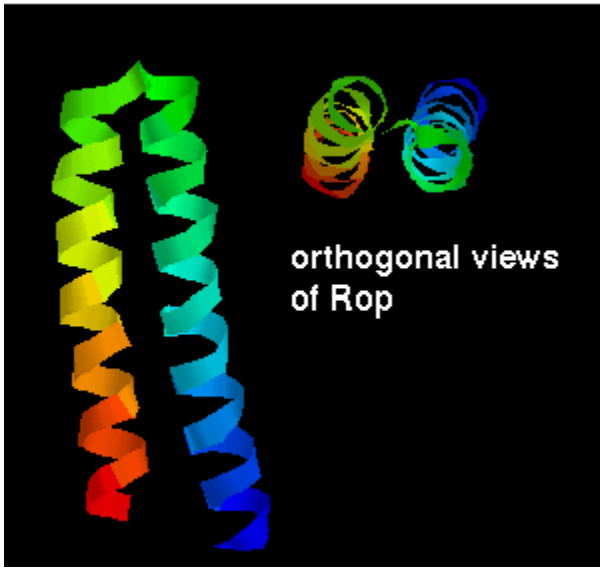




Supersecondary structures:
Composed of 2-3 secondary structure elements

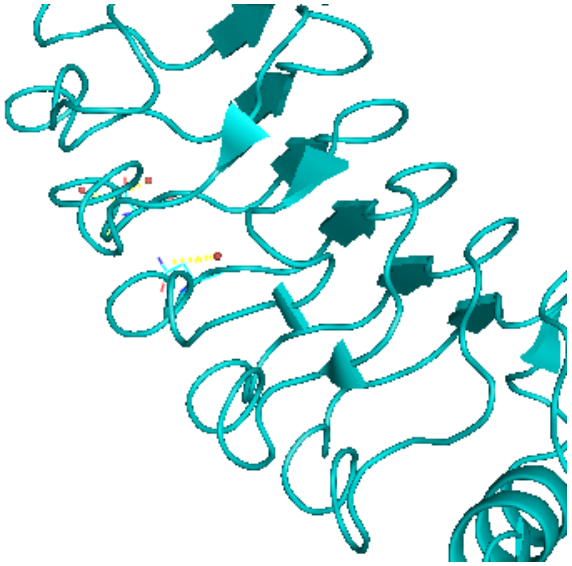
Examples:

Helix-turn-helix motifs, frequent in DNA-binding proteins



Coiled coils, e.g from myosin

Tertiary structure

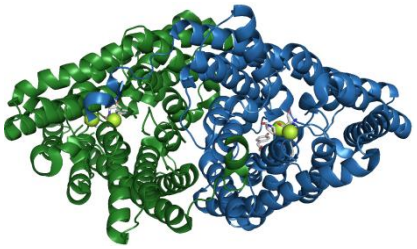


LRR domain

Domains, repeats, zinc fingers…

Domain: independently folded part of a protein. Average size, about 150 aa residues, lower limit ca 50 residues
Repeats: several types: LRR, ANK, HEAT…. Composed of few secondary structure elements. Stabilized by interactions between repeats; can form large structures.
Zinc fingers: several types; structure is stabilized by bound zinc ion
EF Hands: structure is stabilized by bound calcium

Quaternary structure: subunit structure



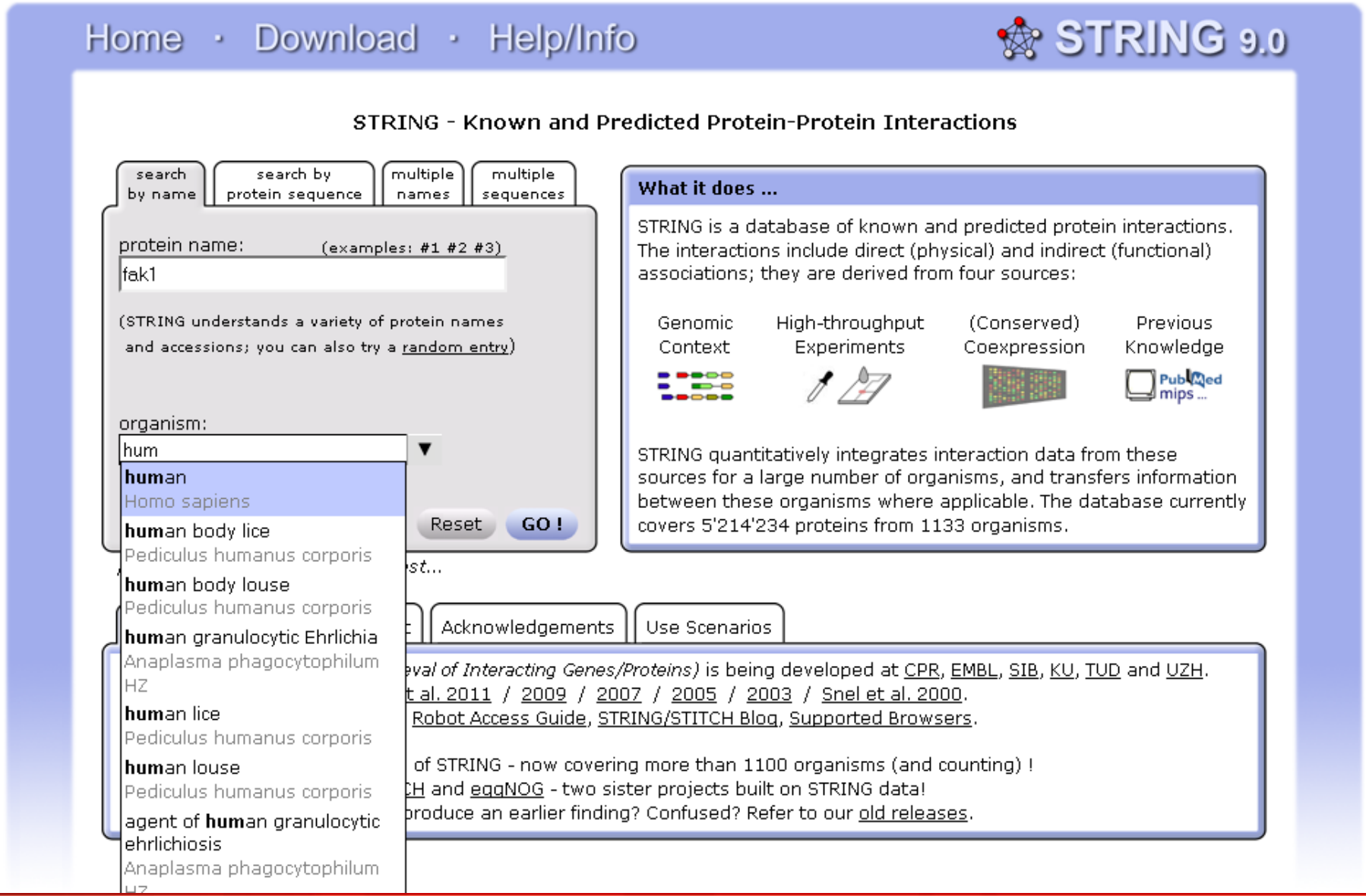
Examples:

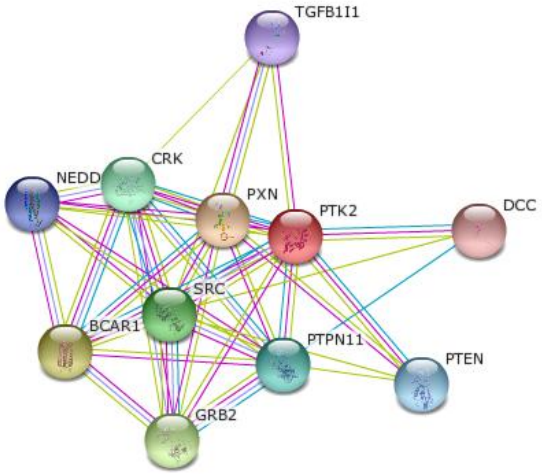
- Homodimer
- Complex between ligand and receptor, enzyme and substrate
- Multisubunit complex

- **STRING database**
- **IntAct, DIP, MINT**

STRING

http://string-db.org/





This is the **evidence view**. Different line colors represent the types of evidence for the association.

(requires Flash player 10 or better)

Your Input:

		Neighborhood	Gene Fusion	Cooccurrence	Coexpression	Experiments	Databases	Textmining	[Homology]	Score
PTK2	PTK2 protein tyrosine kinase 2; Non-receptor protein-tyrosine kinase implicated in signaling pathways involved in cell motility, proliferation and apoptosis. Activated by tyrosine-phosphorylation in response to either integrin clustering induced by cell adhesion or antibody cross-linking, or via G-protein coupled receptor (GPCR) occupancy by ligands such as bombesin or lysophosphatidic acid, or via LDL receptor occupancy. Plays a potential role in oncogenic transformations resulting in increased kinase activity (1052 aa) *(Homo sapiens)*									

Predicted Functional Partners:

PXN	paxillin; Cytoskeletal protein involved in actin-membrane attachment at sites of cell adhesion [...] (591 aa)					•	•	•		0.999
BCAR1	breast cancer anti-estrogen resistance 1; Docking protein which plays a central coordinating ro [...] (870 aa)					•	•	•		0.999
GRB2	growth factor receptor-bound protein 2; Adapter protein that provides a critical link between c [...] (217 aa)					•		•		0.999
SRC	v-src sarcoma (Schmidt-Ruppin A-2) viral oncogene homolog (avian) (536 aa)					•	•	•	•	0.999
CRK	v-crk sarcoma virus CT10 oncogene homolog (avian); The Crk-I and Crk-II forms differ in their b [...] (304 aa)					•	•	•		0.999
PTPN11	protein tyrosine phosphatase, non-receptor type 11; Acts downstream of various receptor and cyt [...] (593 aa)					•	•	•		0.998
PTEN	phosphatase and tensin homolog; Tumor suppressor. Acts as a dual-specificity protein phosphatas [...] (403 aa)					•	•	•		0.998
NEDD9	neural precursor cell expressed, developmentally down-regulated 9; Docking protein which plays [...] (834 aa)					•		•		0.998
TGFB1I1	transforming growth factor beta 1 induced transcript 1; Functions as a molecular adapter coordi [...] (461 aa)					•		•		0.998
DCC	deleted in colorectal carcinoma; Receptor for netrin required for axon guidance. Mediates axon [...] (1447 aa)					•	•	•		0.996

Views:

Protein folding

Many proteins can fold rapidly and spontaneously (msec range)

The physicochemical properties of the polypeptide chain (=the protein sequence) determines protein structure

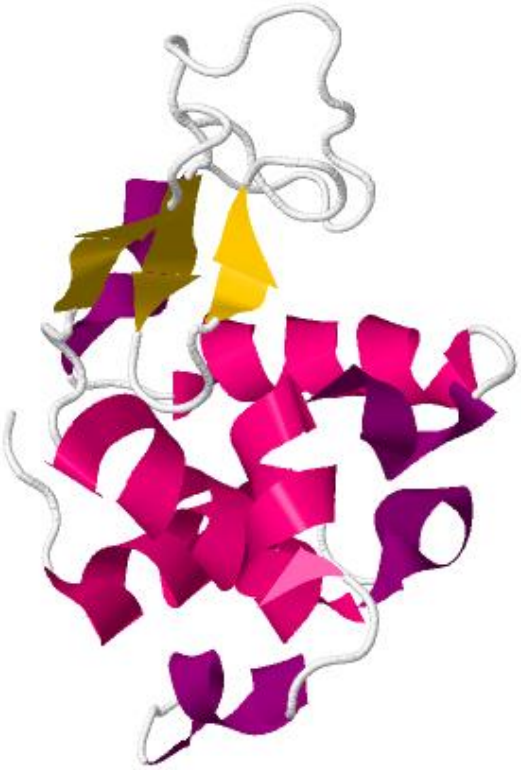
One sequence -> one stable fold

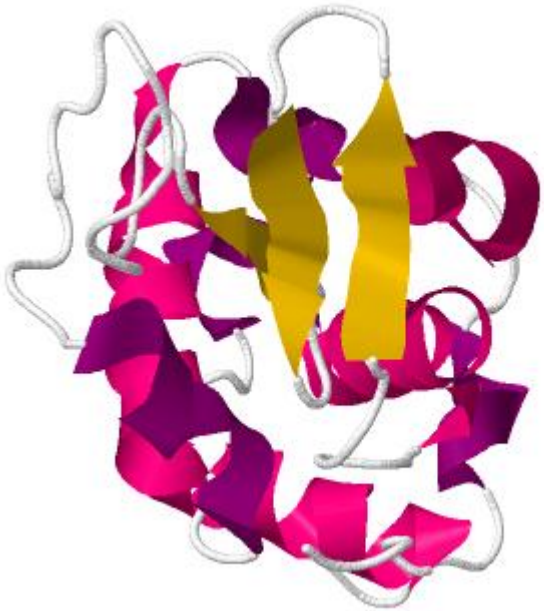
NB: some proteins or parts of proteins are intrinsically disordered=unstructured in the absence of a specific ligand (e.g. Mineralocorticoid receptor ligand-binding domain)

See also: http://en.wikipedia.org/wiki/Protein_folding

Protein structures: the need for classification

How similar/dissimilar are these proteins?

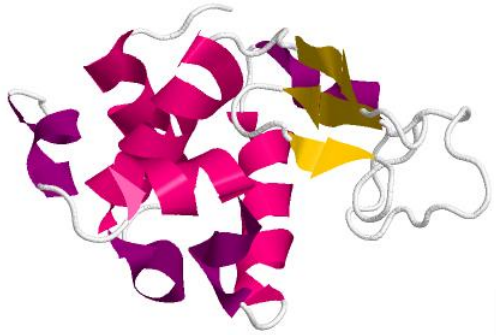




Protein structure classification: quantitative criteria

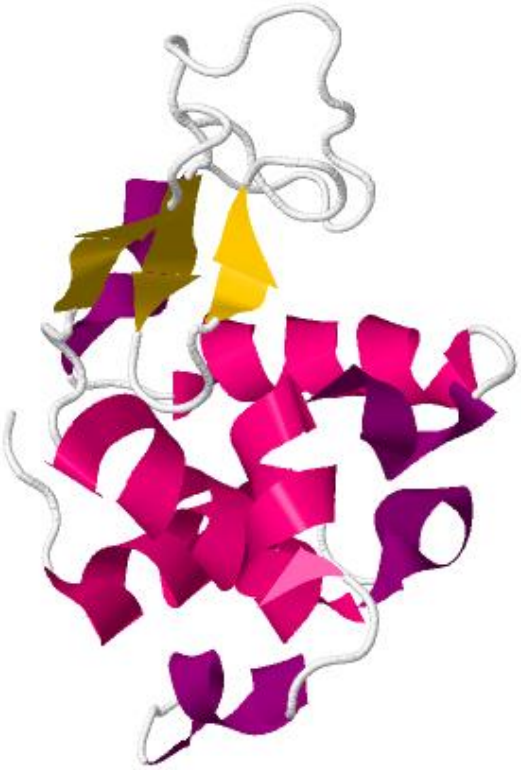
Purely alpha-helical structure
Purely beta-strand structure,
Mixed
Topology (= orientation & connectivity of structural elements)

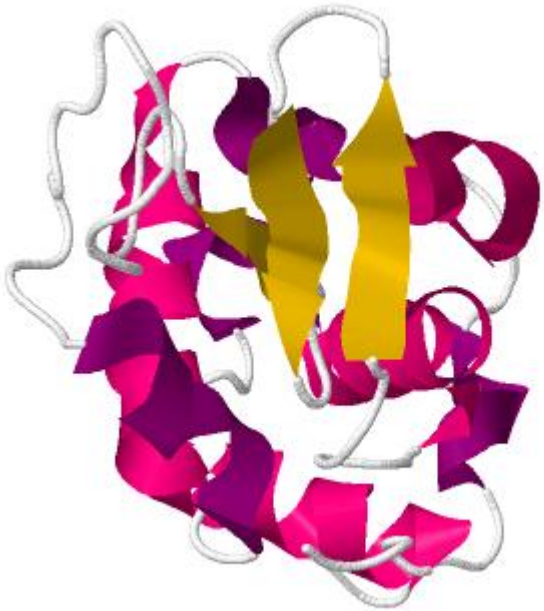
Single domain vs multidomain proteins



Implications of structural similarity:

Evolution
Function prediction





Reasons for structural similarity

- Similarity arises due to divergent evolution (homologues) from a common ancestor - structure much more highly conserved than sequence

- Similarity due to convergent evolution (analogues)
  - Similarity due to there being a limited number of ways of packing helices and strands in 3D space
  - no significant sequence similarity, but proteins may use similar structural locations as active sites
  - NB: at low sequence identity, it is difficult to know whether 2 sequences share a common ancestor, or not

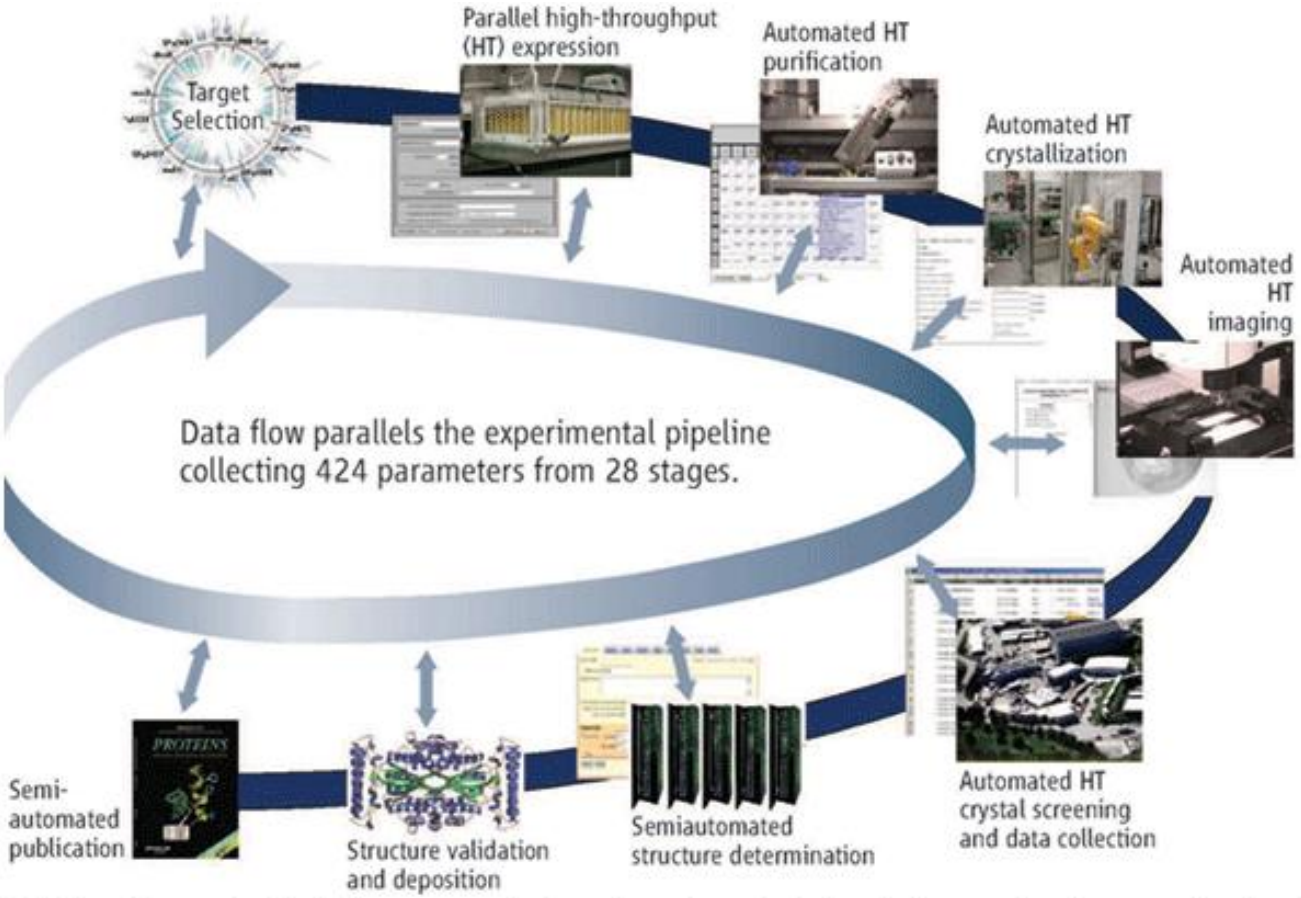
Current dogma

- Proteins with similar sequences have similar 3D-structures
- Proteins with similar 3D-structure are likely to have similar function (generally true, but exceptions exist)
- Proteins with similar function can have entirely different sequences (subtilisin vs chymotrypsin: same active site geometry, no detectable sequence similarity)

Protein 3D-structure analysis

methodology

Protein structure initiatives: technical progress and automatisation



Building a bigger pipeline. PSI groups created a series of new technologies to speed up the many steps involved in determining a protein's structure, such as robots to purify and crystallize proteins.
CREDIT: JOINT CENTER FOR STRUCTURAL GENOMICS

Most protein structures are determined

using X-ray crystallography

Parameters affecting crystallisation:

Physico-chemical: find the right conditions

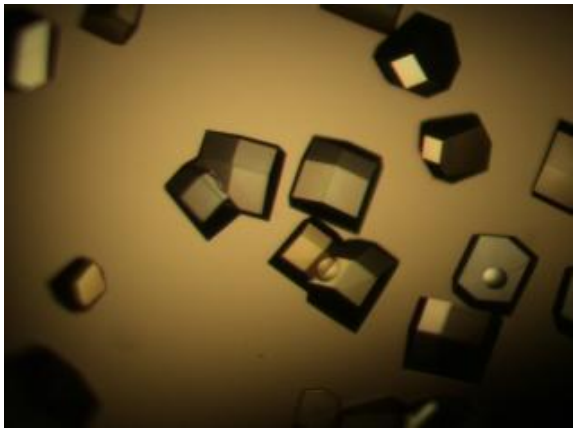
Precipitants: type and concentration

pH

Solvant, buffer composition

Temperature, Pressure

Time



Tetragonal lysozyme crystals

Biological

Protein purity, presence of ligands

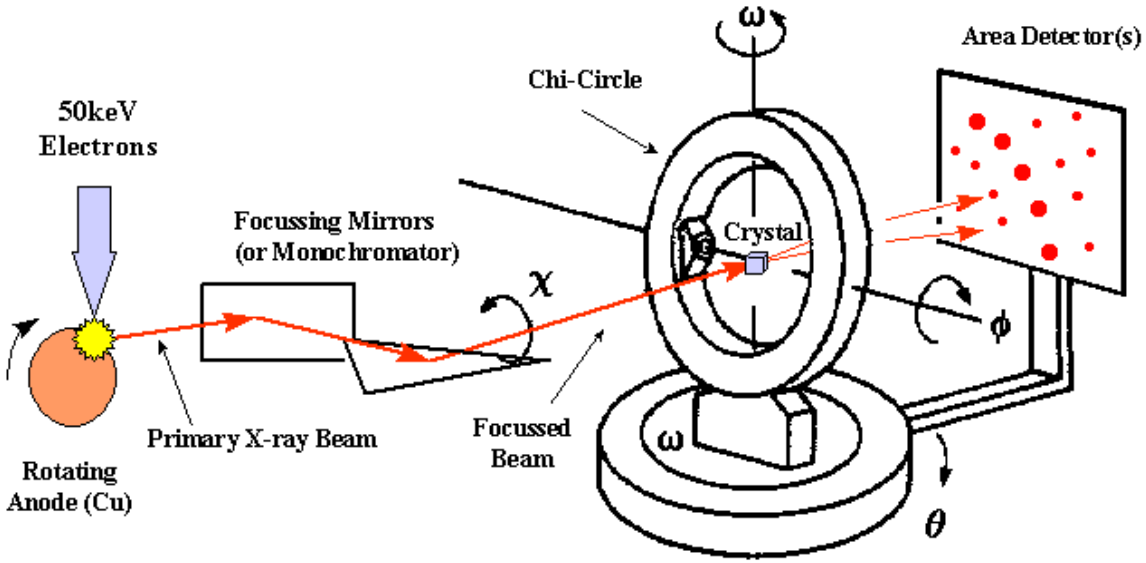
Biological source (native vs heterologous expression, PTMs)

Sequence (micro) heterogeneities

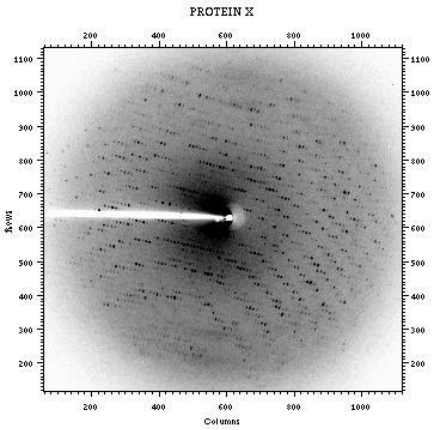
Conformational (micro) heterogeneities

Some proteins are intrinsically unstructured

Data acquisition



4-Circle Goniometer (Eulerian or Kappa Geometry)



Monochrome X-ray beam focused on a crystal

Atoms within the crystal diffract the beam; each type of crystal gives a characteristic diffraction pattern that can be recorded and used to calculate the structure

The more complex the sample, the more complex the diffraction pattern (practical Problems due to weak and/or diffuse spots)

Resolution and structural knowledge

The resolution affects the amount of information that can be obtained

The resolution depends on the quality of the crystals, how similar protein molecules in the crystal are to each other, and how well ordered they are throughout the entire crystal.

Res (A)	Structural information	X-ray
4.0	Global fold, some indication of secondary structure	Useful
3.5	Secondary structure	
3.0	Most side chains are positioned	
2.5	All side-chains, phi-psi angles constrained, waters located	Typical
1.5	phi-psi angles well defined, hydrogen atoms begin to appear	Very good
1.0	Hydrogens are visible	Possible

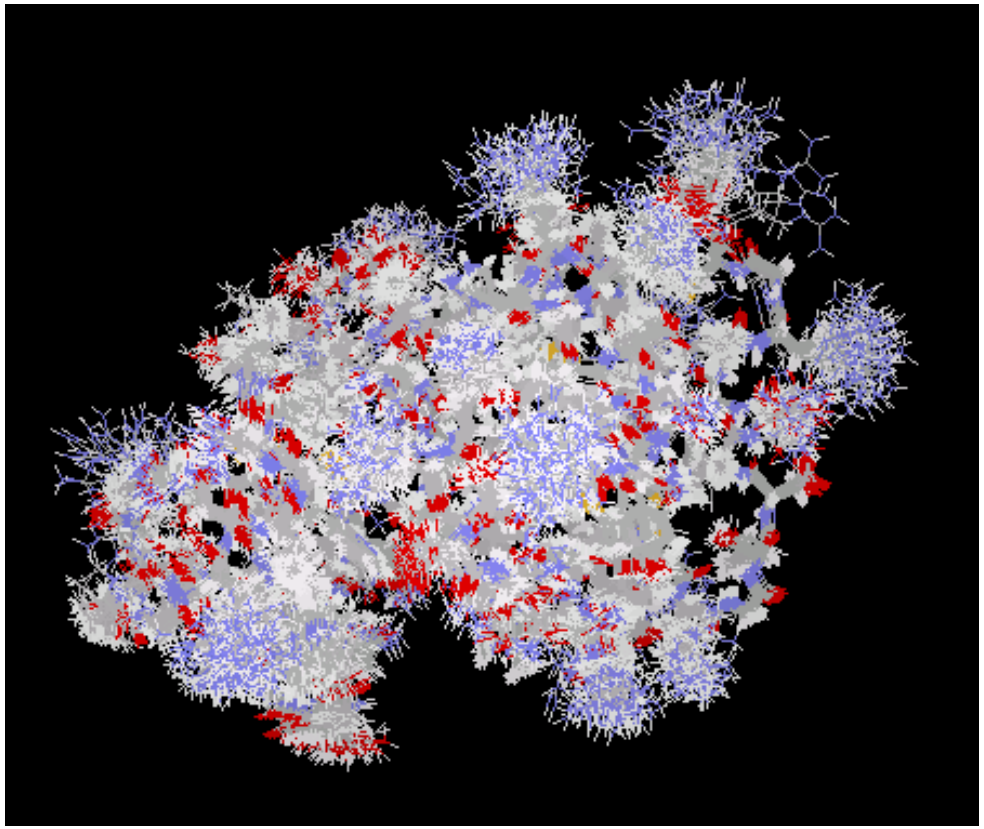
Nuclear magnetic resonance (NMR)

- Measures the energy levels of magnetic atoms, i.e. atoms with odd electron numbers: ^{1}H, ^{13}C, ^{15}N, ^{19}F, ^{31}P
- Energy levels of an atom are influenced by the local environment (chemical shifts)
  - Via covalent bonds
  - Through space, max. 5A apart: Nuclear Overhauser Effect (NOE)
- NMR can identify atoms that are close together, also those that are close in space but not linked by direct covalent bonds
- Chemical shifts can define secondary structures
- NMR spectra yield a set of peaks that correspond to the interactions between pairs of atoms
- From these, one can calculate the protein structure

Nuclear magnetic resonance (NMR)

- Advantage: done with proteins in solution
- But: still requires high protein concentrations
- Advantage/Disadvantage: conformational heterogeneity – proteins move
- NMR results usually yield ca 20 closely similar but non-identical structures
- Disadvantage: cannot determine the structures of large proteins (ok up to 30 kDa, feasible up to 60 kDa)
- NMR spectra are used to study small proteins or isolated domains

NMR output

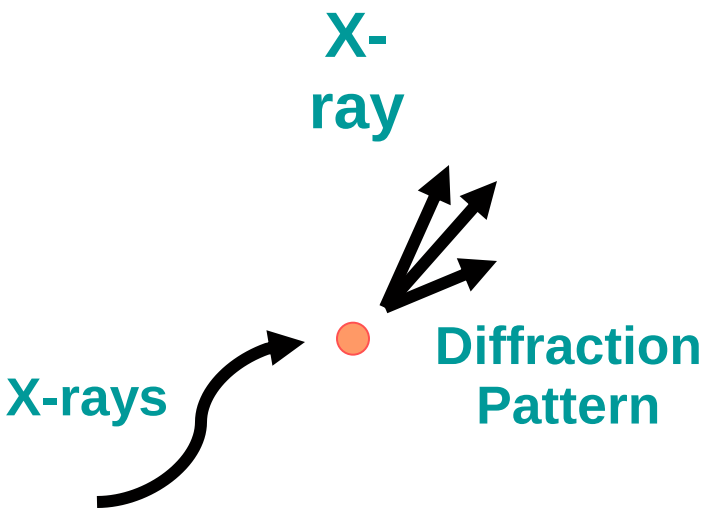


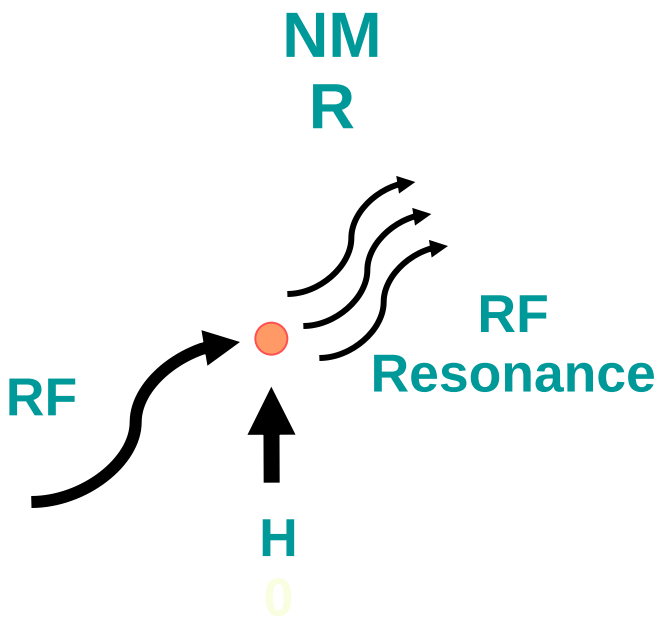
An ensemble of 15-20 closely similar structures

Dynamic aspects of the structure

Less “precise” than rigid X-ray structures

X-ray vs NMR: principles

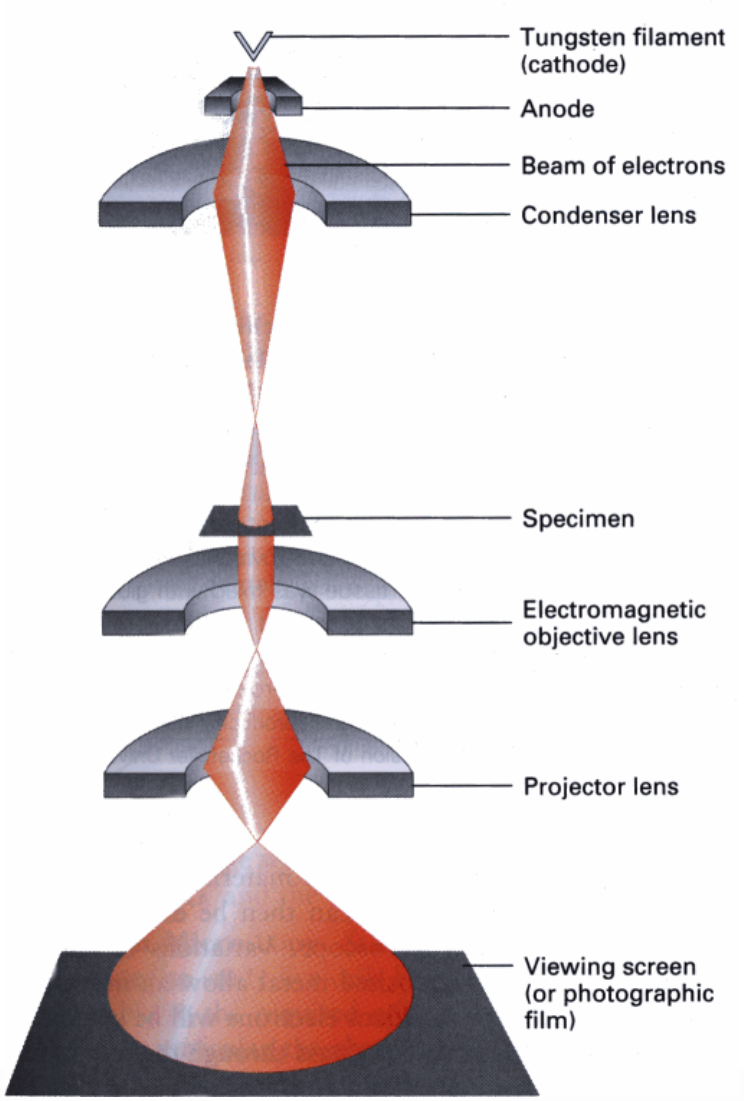




Direct detection of atom positions
Crystals

Indirect detection of H-H distances
In solution

Electron Microscopy



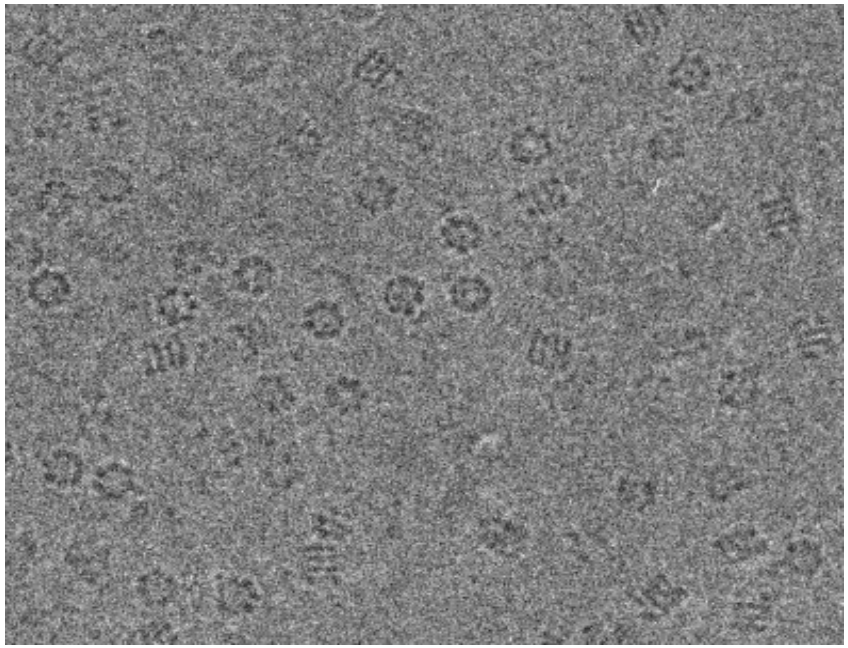
Developed in 1930 to overcome limitations of Light Microscopes

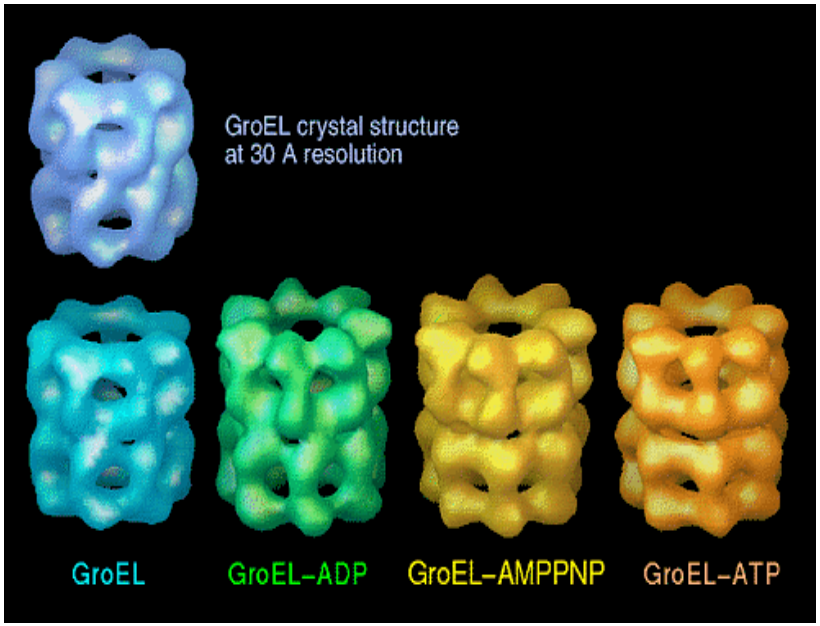
Based on Light Transmission Microscope principle

Potentially resolutions of 1Å are possible

Allows to reconstruct 3D structures from 2D projections

Transmission Electron Microscope





Cryo-EM image of GroEL chaperonin complexes, showing end views (rings) and side views (stripes).

Methods for determining 3D structures

	Advantages	Disadvantages
X-ray Crystallo-graphy	High resolution (up to 0.5Å) No protein mass limit	**Crystals needed** Artefacts due to crystallization (Enzyme in open vs closed Conformation) Structure is a static average
NMR	No crystals needed Conformation of protein in solution **Dynamic aspects (conformation ensemble view)**	**Highly concentrated solution** (1mM at least) Isotope substitution (^{13}C, ^{15}N) Limited maximum weight (about 60 kD)
Electron Microscopy	No 3D-crystals needed **Direct image**	Large radiation damage Need 2D crystals or large complexes Artefacts

Working with protein structures
Databases and tools

One central archive for 3D-structure data: wwPDB (**www.wwpdb.org**)

- What can you find there?
- How to find a structure for protein x?
- How to find a structure with bound z?

What if ?

- If the structure has not been determined, is there a structure for a similar protein?
- Can we predict the structure of a protein? How?

World-wide PDB (wwPDB)

Four member sites:
same data, but different presentation and tools

- **RCSB PDB** (www.rcsb.org/pdb/home/home.do)
- **PDBe** (http://www.ebi.ac.uk/pdbe/)
- **PDBj** (www.pdbj.org/)
- **BMRB** (Biological Magnetic Resonance Data Bank, www.bmrb.wisc.edu)

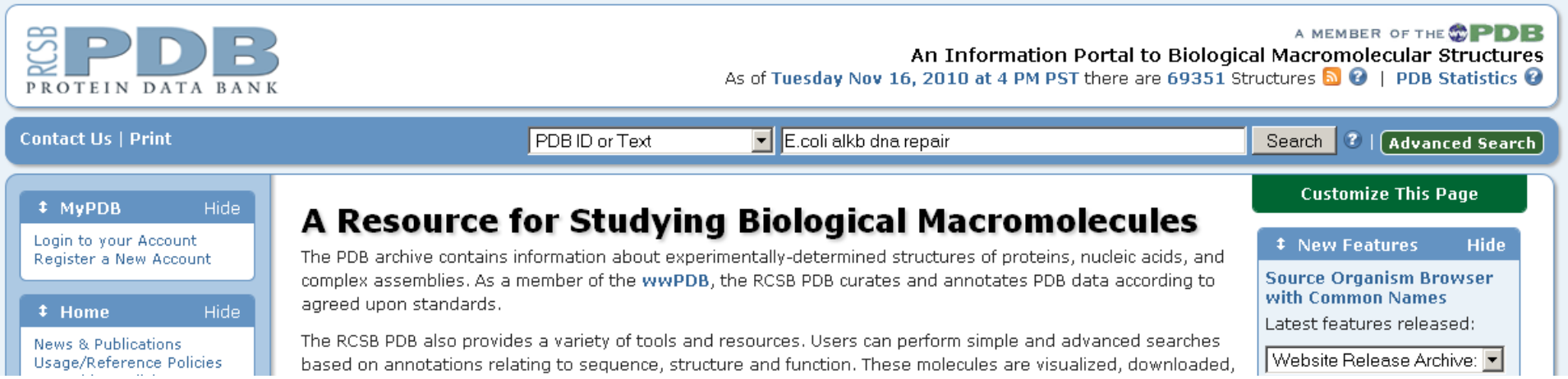
Information and tools you can find at RCSB PDB

PDB file and Header document
Structure viewer
Links

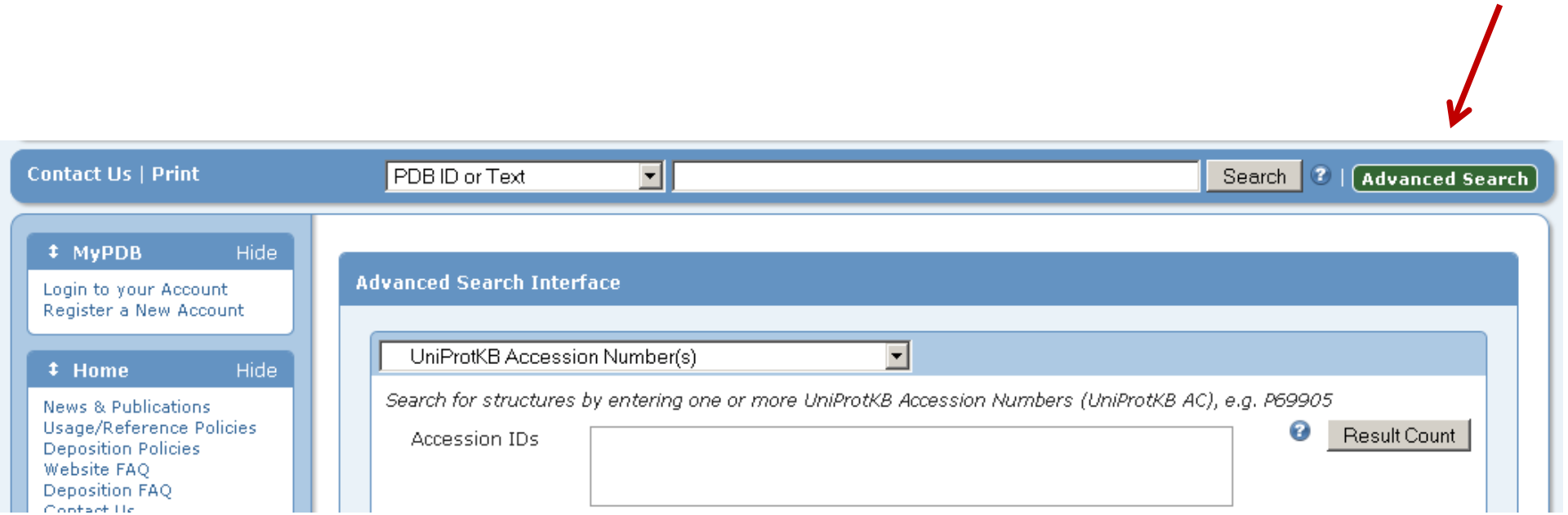
Good documentation and help:
http://www.rcsb.org/robohelp_f/#search_database/how_to_search.htm
http://www.ebi.ac.uk/pdbe/#m=2&h=0&e=1&r=0&l=0&a=0&w=1-3-2
http://www.ebi.ac.uk/pdbe/#m=2&h=0&e=0&r=0&l=0&a=0&w=0

Finding protein structures at RCSB PDB

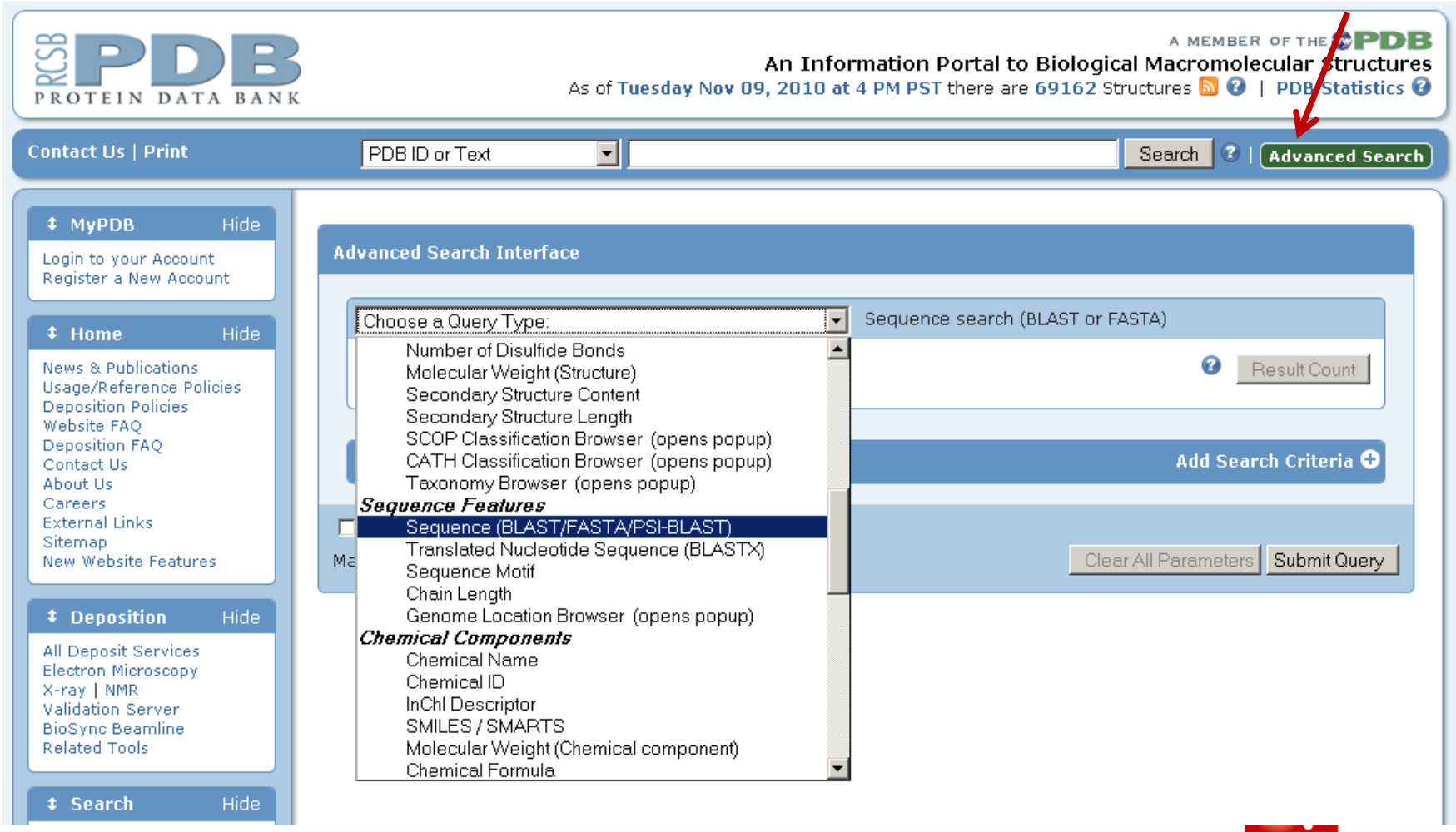
Via main query window: PDB ID, or text



Advanced search: query with UniProt AC



Advanced search: query via BLAST



Finding protein structures at RCSB PDB

3I49 — Crystal Structure of AlkB in complex with Fe(II), 2-oxoglutarate and methylated trinucleotide T-meC-T

Download PDB File

Authors: Yu, B., Hunt, J.F.

Release Date: 2009-08-25 **Classification:** Oxidoreductase/dna

Experiment: X-RAY DIFFRACTION with resolution of 1.60 Å

Compound: 2 Polymers [*Display Full Polymer Details* | *Display for All Results*]
2 Ligands [*Display Full Ligand Details* | *Display for All Results*]

Citation: **Enzymological and structural studies of the mechanism of promiscuous substrate recognition by the oxidative DNA repair enzyme AlkB.**
(2009) Proc.Natl.Acad.Sci.USA **106:** 14315-14320 [*Display Full Abstract* | *Display for All Results*]

3BI3 — X-ray structure of AlkB protein bound to dsDNA containing 1meA/A with cofactors

Authors: Yi, C., Yang, C.-G.

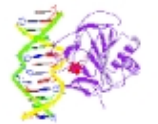
Release Date: 2008-04-22 **Classification:** Oxidoreductase/dna

Experiment: X-RAY DIFFRACTION with resolution of 1.90 Å

Compound: 3 Polymers [*Display Full Polymer Details* | *Display for All Results*]
3 Ligands [*Display Full Ligand Details* | *Display for All Results*]

Citation: **Crystal structures of DNA/RNA repair enzymes AlkB and ABH2 bound to dsDNA.**
(2008) Nature **452:** 961-965 [*Display Full Abstract* | *Display for All Results*]

3BIE — X-ray structure of E coli AlkB bound to dsDNA containing 1meA/T with Mn and 2KG



Authors: Yi, C., Yang, C.-G.

Release Date: 2008-04-22 **Classification:** Oxidoreductase/dna

Experiment: X-RAY DIFFRACTION with resolution of 1.68 Å

Compound: 3 Polymers [*Display Full Polymer Details* | *Display for All Results*]
3 Ligands [*Display Full Ligand Details* | *Display for All Results*]

Citation: **Crystal structures of DNA/RNA repair enzymes AlkB and ABH2 bound to dsDNA**
(2008) Nature **452:** 961-965 [*Display Full Abstract* | *Display for All Results*]

Information from protein 3D-structures about E.coli alkB DNA repair dioxygenase

Summary | Sequence | Derived Data | Seq. Similarity | 3D Similarity | Literature | Biol. & Chem. | Methods | Geometry | Links

Crystal Structure of AlkB in complex with Fe(II), 2-oxoglutarate and methylated trinucleotide T-meC-T

3I49

Display Files
Download Files
Print this Page
Share this Page

DOI: 10.2210/pdb3i49/pdb NDB ID: NA0060

Primary Citation

Enzymological and structural studies of the mechanism of promiscuous substrate recognition by the oxidative DNA repair enzyme AlkB.

Yu, B., Hunt, J.F.

Journal: (2009) Proc.Natl.Acad.Sci.USA **106:** 14315-14320

PubMed: 19706517
PubMedCentral: PMC2725012
DOI: 10.1073/pnas.0812938106
Search Related Articles in PubMed

PubMed Abstract:

Promiscuous substrate recognition, the ability to catalyze transformations of chemically diverse compounds, is an evolutionarily advantageous, but poorly understood phenomenon. The promiscuity of DNA repair enzymes is particularly important, because it enables diverse kinds of damage to different nucleotide bases to be repaired in a metabolically parsimonious manner. We present enzymological and crystallographic studies of the mechanisms underlying promiscuous substrate recognition by Escherichia coli AlkB, a DNA repair enzyme that removes methyl adducts and some larger alkylation lesions from endocyclic positions on purine and pyrimidine bases. In vitro Michaelis-Menten analyses on a series of alkylated bases show high activity in repairing N1-methyladenine (m1A) and N3-methylcytosine (m3C), comparatively low activity in repairing 1,N(6)-ethenoadenine, and no detectable activity in repairing N1-methylguanine or N3-methylthymine. AlkB has a substantially higher k(cat) and K(m) for m3C compared with m1A.

Biological Assembly

More Images...

View in Jmol | SimpleViewer
Other Viewers | Protein Workshop

Biological assembly assigned by authors and generated by PISA (software)

PDB file

```
HEADER    OXIDOREDUCTASE/DNA                      01-JUL-09   3I49
TITLE     CRYSTAL STRUCTURE OF ALKB IN COMPLEX WITH FE(II), 2-
TITLE    2 OXOGLUTARATE AND METHYLATED TRINUCLEOTIDE T-MEC-T
COMPND    MOL_ID: 1;
COMPND   2 MOLECULE: ALPHA-KETOGLUTARATE-DEPENDENT DIOXYGENASE ALKB;
COMPND   3 CHAIN: A;
COMPND   4 FRAGMENT: UNP RESIDUES 12-216;
COMPND   5 SYNONYM: ALKYLATED DNA REPAIR PROTEIN ALKB;
COMPND   6 EC: 1.14.11.-;
COMPND   7 ENGINEERED: YES;
COMPND   8 MOL_ID: 2;
COMPND   9 MOLECULE: DNA (5'-D(P*TP*(ME6)P*T)-3');
COMPND  10 CHAIN: B;
COMPND  11 ENGINEERED: YES
SOURCE    MOL_ID: 1;
SOURCE   2 ORGANISM_SCIENTIFIC: ESCHERICHIA COLI;
SOURCE   3 ORGANISM_TAXID: 83333;
SOURCE   4 STRAIN: K-12;
SOURCE   5 GENE: AIDD, ALKB, B2212, JW2200;
SOURCE   6 EXPRESSION_SYSTEM: ESCHERICHIA COLI;
SOURCE   7 EXPRESSION_SYSTEM_TAXID: 562;
SOURCE   8 EXPRESSION_SYSTEM_STRAIN: BL21(DE3);
SOURCE   9 EXPRESSION_SYSTEM_VECTOR_TYPE: PLASMID;
SOURCE  10 EXPRESSION_SYSTEM_PLASMID: PET26B;
SOURCE  11 MOL_ID: 2;
SOURCE  12 SYNTHETIC: YES
KEYWDS    BETA JELLYROLL, PROTEIN-DNA COMPLEX, DIOXYGENASE, DNA
KEYWDS   2 DAMAGE, DNA REPAIR, IRON, METAL-BINDING, OXIDOREDUCTASE,
KEYWDS   3 OXIDOREDUCTASE/DNA COMPLEX
EXPDTA    X-RAY DIFFRACTION
AUTHOR    B.YU,J.F.HUNT
REVDAT   2   08-SEP-09 3I49    1       JRNL
REVDAT   1   25-AUG-09 3I49    0
JRNL        AUTH   B.YU,J.F.HUNT
JRNL        TITL   ENZYMOLOGICAL AND STRUCTURAL STUDIES OF THE
JRNL        TITL 2 MECHANISM OF PROMISCUOUS SUBSTRATE RECOGNITION BY
JRNL        TITL 3 THE OXIDATIVE DNA REPAIR ENZYME ALKB.
JRNL        REF    PROC.NATL.ACAD.SCI.USA        V. 106 14315 2009
JRNL        REFN                   ISSN 0027-8424
JRNL        PMID   19706517
JRNL        DOI    10.1073/PNAS.0812938106
```

Header: Atom coordinates

http://www.rcsb.org/pdb/static.do?p=education_discussion/Looking-at-Structures/coordinates.html

X, Y, Z, occupancy and temperature factor

```
ATOM      1  N   LEU A  15      22.640 -10.051   3.565  1.00 30.09           N
ATOM      2  CA  LEU A  15      22.265  -9.579   2.201  1.00 29.42           C
ATOM      3  C   LEU A  15      23.495  -9.407   1.320  1.00 28.67           C
ATOM      4  O   LEU A  15      23.473  -9.733   0.133  1.00 29.97           O
ATOM      5  CB  LEU A  15      21.514  -8.249   2.291  1.00 29.55           C
ATOM      6  CG  LEU A  15      21.294  -7.521   0.963  1.00 30.49           C
ATOM      7  CD1 LEU A  15      20.638  -8.459  -0.032  1.00 31.68           C
ATOM      8  CD2 LEU A  15      20.442  -6.283   1.190  1.00 30.60           C
ATOM      9  N   ALA A  16      24.563  -8.887   1.913  1.00 26.31           N
ATOM     10  CA  ALA A  16      25.818  -8.660   1.209  1.00 24.06           C
ATOM     11  C   ALA A  16      26.815  -8.107   2.214  1.00 22.57           C
ATOM     12  O   ALA A  16      26.452  -7.797   3.348  1.00 21.88           O
ATOM     13  CB  ALA A  16      25.616  -7.664   0.078  1.00 23.31           C
ATOM     14  N   ALA A  17      28.072  -7.989   1.802  1.00 21.94           N
ATOM     15  CA  ALA A  17      29.097  -7.453   2.685  1.00 22.11           C
ATOM     16  C   ALA A  17      28.716  -6.033   3.091  1.00 20.73           C
ATOM     17  O   ALA A  17      28.613  -5.141   2.244  1.00 21.03           O
ATOM     18  CB  ALA A  17      30.450  -7.451   1.979  1.00 23.22           C
ATOM     19  N   GLY A  18      28.499  -5.833   4.388  1.00 19.10           N
ATOM     20  CA  GLY A  18      28.137  -4.521   4.892  1.00 18.73           C
ATOM     21  C   GLY A  18      26.661  -4.197   4.763  1.00 19.12           C
ATOM     22  O   GLY A  18      26.236  -3.087   5.089  1.00 17.79           O
ATOM     23  N   ALA A  19      25.875  -5.162   4.295  1.00 18.49           N
ATOM     24  CA  ALA A  19      24.440  -4.952   4.128  1.00 18.93           C
ATOM     25  C   ALA A  19      23.622  -5.987   4.885  1.00 19.68           C
ATOM     26  O   ALA A  19      24.049  -7.129   5.075  1.00 21.19           O
ATOM     27  CB  ALA A  19      24.074  -4.982   2.647  1.00 17.60           C
ATOM     28  N   VAL A  20      22.432  -5.585   5.311  1.00 18.94           N
ATOM     29  CA  VAL A  20      21.561  -6.483   6.046  1.00 18.89           C
ATOM     30  C   VAL A  20      20.122  -6.015   5.964  1.00 17.95           C
ATOM     31  O   VAL A  20      19.854  -4.824   5.798  1.00 16.19           O
ATOM     32  CB  VAL A  20      21.968  -6.564   7.534  1.00 20.99           C
ATOM     33  CG1 VAL A  20      21.871  -5.189   8.178  1.00 22.02           C
ATOM     34  CG2 VAL A  20      21.079  -7.562   8.264  1.00 24.32           C
```

Header: protein sequence and ligand information

```
DBREF  3I49 A    12   216  UNP    P05050   ALKB_ECOLI      12    216
DBREF  3I49 B   501   503  PDB    3I49     3I49           501    503
SEQADV 3I49 MET A   11  UNP  P05050              INITIATING METHIONINE
SEQADV 3I49 ASN A  217  UNP  P05050              EXPRESSION TAG
SEQADV 3I49 LEU A  218  UNP  P05050              EXPRESSION TAG
SEQADV 3I49 TYR A  219  UNP  P05050              EXPRESSION TAG
SEQADV 3I49 PHE A  220  UNP  P05050              EXPRESSION TAG
SEQADV 3I49 GLN A  221  UNP  P05050              EXPRESSION TAG
SEQRES   1 A  211  MET GLN GLU PRO LEU ALA ALA GLY ALA VAL ILE LEU ARG
SEQRES   2 A  211  ARG PHE ALA PHE ASN ALA ALA GLU GLN LEU ILE ARG ASP
SEQRES   3 A  211  ILE ASN ASP VAL ALA SER GLN SER PRO PHE ARG GLN MET
SEQRES   4 A  211  VAL THR PRO GLY GLY TYR THR MET SER VAL ALA MET THR
SEQRES   5 A  211  ASN CYS GLY HIS LEU GLY TRP THR THR HIS ARG GLN GLY
SEQRES   6 A  211  TYR LEU TYR SER PRO ILE ASP PRO GLN THR ASN LYS PRO
SEQRES   7 A  211  TRP PRO ALA MET PRO GLN SER PHE HIS ASN LEU CYS GLN
SEQRES   8 A  211  ARG ALA ALA THR ALA ALA GLY TYR PRO ASP PHE GLN PRO
SEQRES   9 A  211  ASP ALA CYS LEU ILE ASN ARG TYR ALA PRO GLY ALA LYS
SEQRES  10 A  211  LEU SER LEU HIS GLN ASP LYS ASP GLU PRO ASP LEU ARG

HET    ME6  B 502      20
HET    AKG  A 400      10
HET    FE2  A 300       1
HETNAM     ME6 [(2R,3S,5R)-5-(4-AZANYL-3-METHYL-2-OXO-PYRIMIDIN-3-IUM-
HETNAM   2 ME6  1-YL)-3-HYDROXY-OXOLAN-2-YL]METHYL DIHYDROGEN
HETNAM   3 ME6  PHOSPHATE
HETNAM     AKG 2-OXYGLUTARIC ACID
HETNAM     FE2 FE (II) ION
FORMUL   2  ME6    C10 H17 N3 O7 P 1+
FORMUL   3  AKG    C5 H6 O5
FORMUL   4  FE2    FE 2+
FORMUL   5  HOH   *281(H2 O)
```

HET: Heteroatoms=non-protein atoms: small molecules and ions, each with a unique abbreviation

Information from protein 3D-structures about E.coli alkB DNA repair dioxygenase

Summary | Sequence | Derived Data | Seq. Similarity | 3D Similarity | Literature | Biol. & Chem. | Methods | Geometry | Links

Crystal Structure of AlkB in complex with Fe(II), 2-oxoglutarate and methylated trinucleotide T-meC-T

3I49

Display Files
Download Files
Print this Page
Share this Page

DOI: 10.2210/pdb3i49/pdb NDB ID: NA0060

Primary Citation

Enzymological and structural studies of the mechanism of promiscuous substrate recognition by the oxidative DNA repair enzyme AlkB.

Yu, B., Hunt, J.F.

Journal: (2009) Proc.Natl.Acad.Sci.USA **106:** 14315-14320

PubMed: 19706517
PubMedCentral: PMC2725012
DOI: 10.1073/pnas.0812938106
Search Related Articles in PubMed

PubMed Abstract:

Promiscuous substrate recognition, the ability to catalyze transformations of chemically diverse compounds, is an evolutionarily advantageous, but poorly understood phenomenon. The promiscuity of DNA repair enzymes is particularly important, because it enables diverse kinds of damage to different nucleotide bases to be repaired in a metabolically parsimonious manner. We present enzymological and crystallographic studies of the mechanisms underlying promiscuous substrate recognition by Escherichia coli AlkB, a DNA repair enzyme that removes methyl adducts and some larger alkylation lesions from endocyclic positions on purine and pyrimidine bases. In vitro Michaelis-Menten analyses on a series of alkylated bases show high activity in repairing N1-methyladenine (m1A) and N3-methylcytosine (m3C), comparatively low activity in repairing 1,N(6)-ethenoadenine, and no detectable activity in repairing N1-methylguanine or N3-methylthymine. AlkB has a substantially higher k(cat) and K(m) for m3C compared with m1A.

Biological Assembly

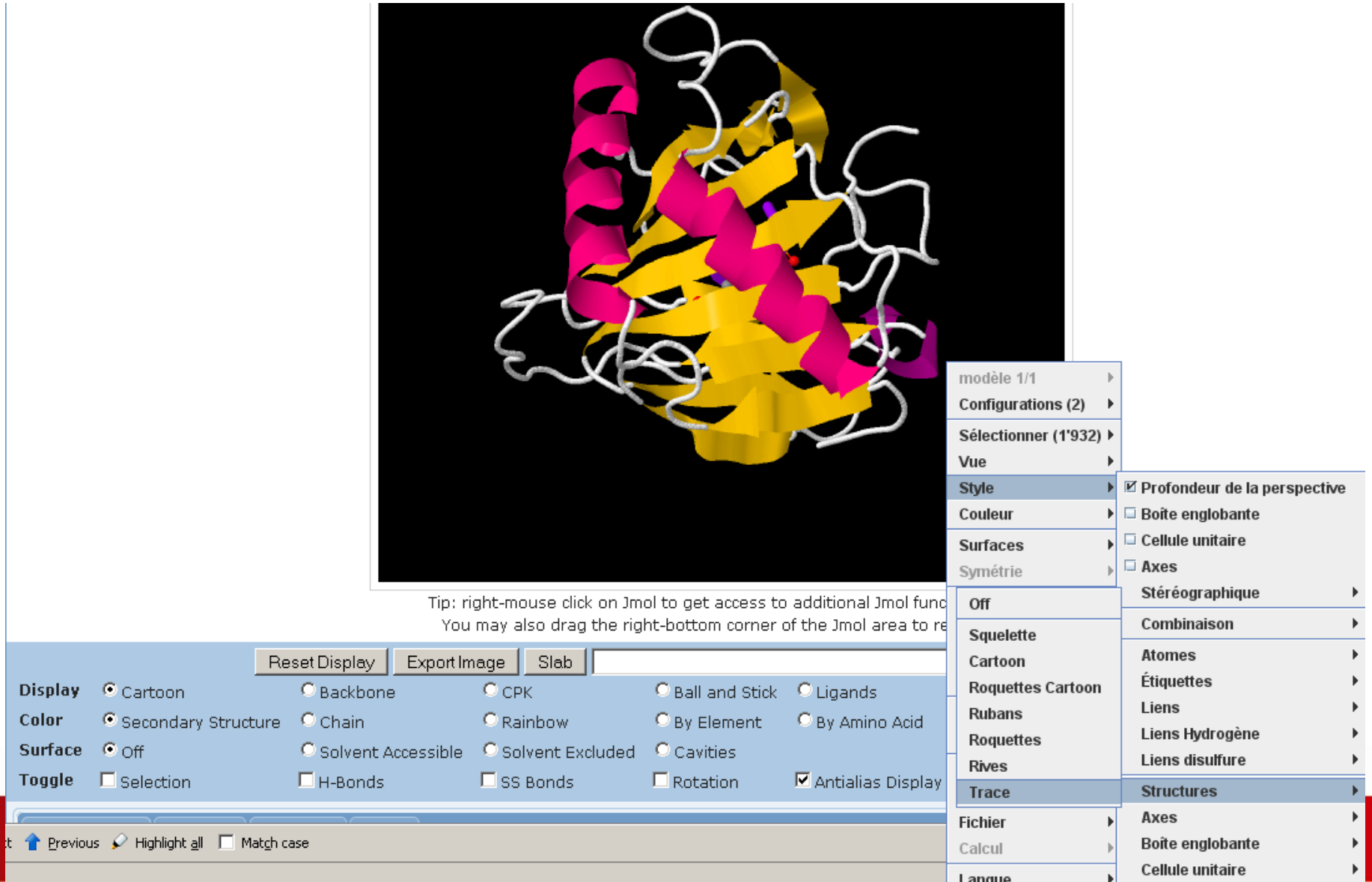
More Images...

View in Jmol SimpleViewer
Other Viewers Protein Workshop

Biological assembly assigned by authors and generated by PISA (software)

Viewing a structure (JMOL):

Right-click in window to change parameters



Access to Ligand explorer

Summary | Sequence | Derived Data | Seq. Similarity | 3D Similarity | Literature | Biol. & Chem. | Methods | Geometry | Links

Crystal Structure of AlkB in complex with Fe(II), 2-oxoglutarate and methylated trinucleotide T-meC-T

3I49

DOI: 10.2210/pdb3i49/pdb NDB ID: NA0060

Further down on same page..

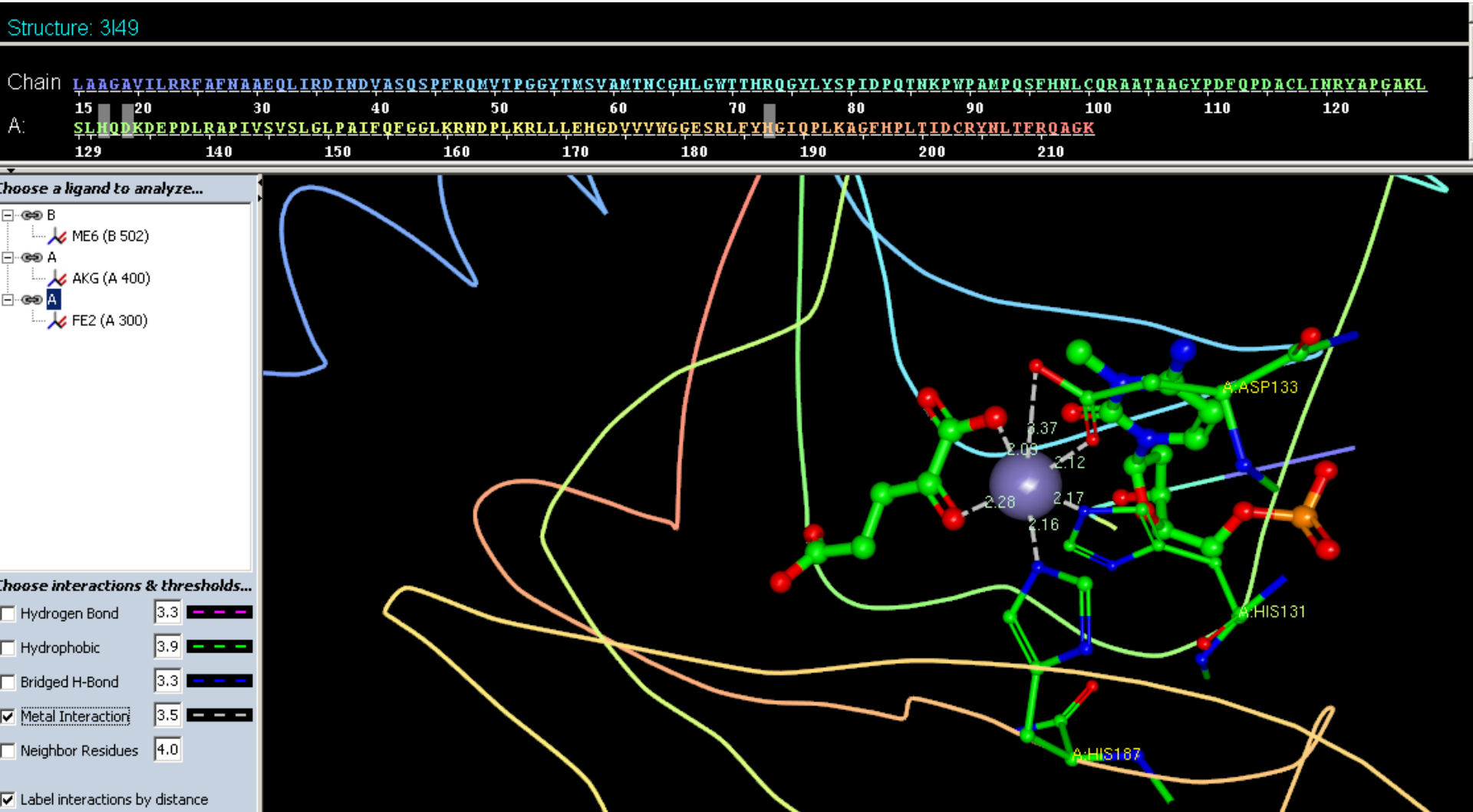
Ligand Chemical Component Hide

Identifier	Name	Formula	Binding Affinity (BindingDB)	Interaction View
AKG	2-OXOGLUTARIC ACID	C5 H6 O5		Ligand Explorer
FE2	FE (II) ION	Fe		Ligand Explorer

Modified Residues Hide

Identifier	Name	Formula	Parent	Type
ME6	[(2R,3S,5R)-5-(4-azanyl-3-methyl-2-oxo-pyrimidin-3-ium-1-yl)-3-hydroxy-oxolan-2-yl]methyl dihydrogen phosphate	C10 H17 N3 O7 P	DC	dnaLinking

Ligand interactions

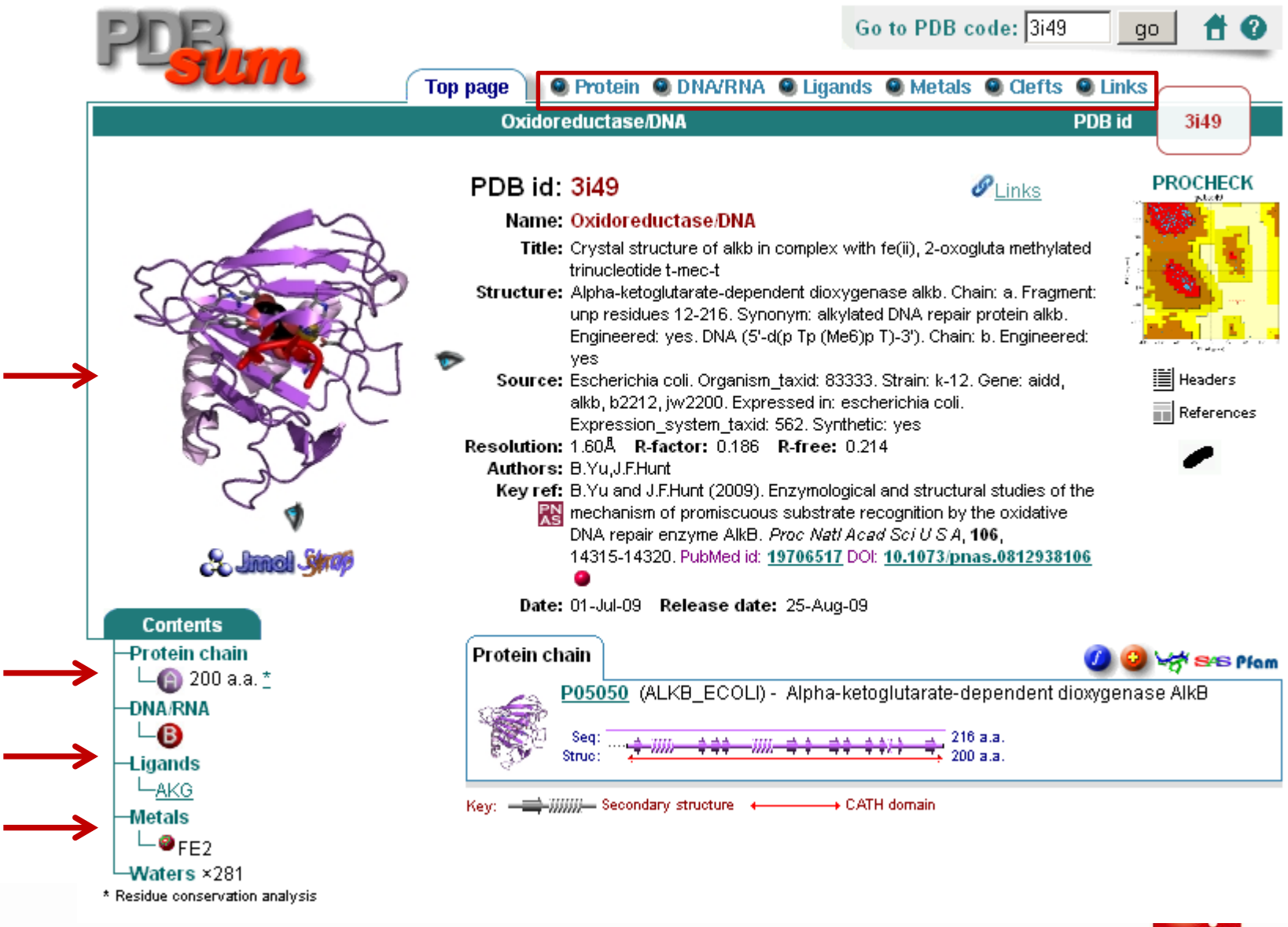


Very easy to use, excellent tools and links:

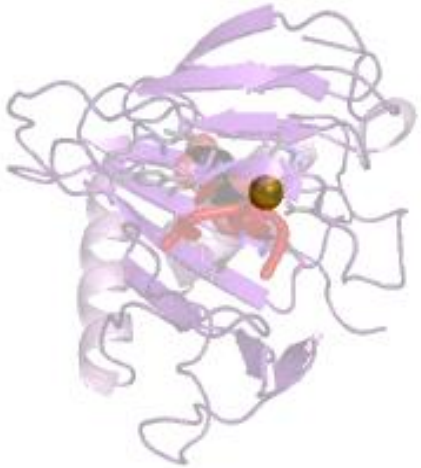
PDB*sum*

www.ebi.ac.uk/thornton-srv/databases/pdbsum/

Ligand interactions

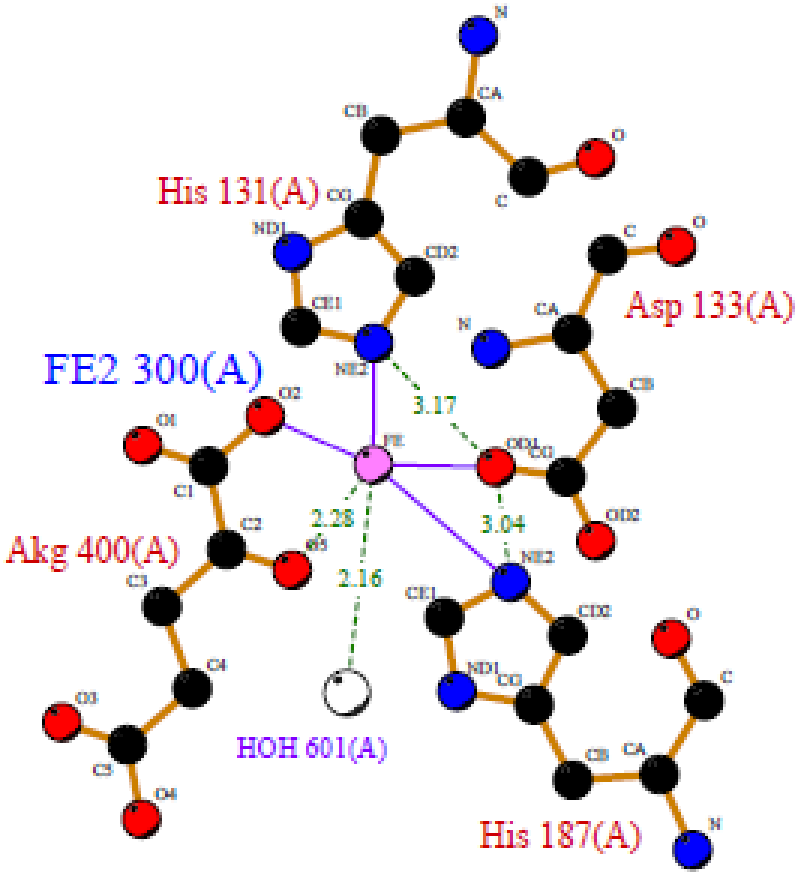


Metal ion FE2 - *Fe (ii) ion*



Metal highlighted

FE2



Ligands
- AKG
  - AKG 400(A)

Metals
- FE2
  - FE2 300(A)

Jmol

Postscript version

PDF

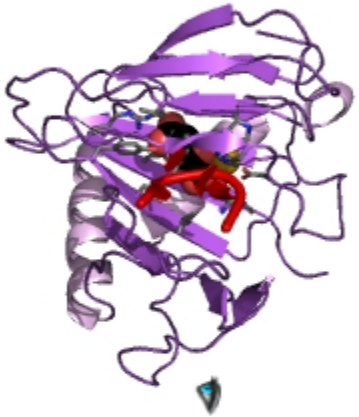
EMBL-EBI | Enter Text Here | Find | Help | Feedback

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

PDBsum | Go to PDB code: 3i49 go

Top page | Protein | DNA/RNA | Ligands | Metals | Clefts | Links

Oxidoreductase/DNA | PDB id 3i49



Contents
- Protein chain
  - A 200 a.a.
- DNA/RNA
  - B
- Ligands
  - AKG
- Metals
  - FE2
- Waters ×281

Links to other databases

Structure databases

PDBe	**Protein Data Bank, Europe** at the **EBI**
RCSB	**Protein Data Bank** at the **RCSB**
SRS	**SRS** at the **EBI**
MMDB	**MMDB entry** at the NCBI
JenaLib	**Jena Library of Biological Macromolecules** at the Fritz-Lipman Institute
OCA	**OCA** at the Weizmann Institute

Wikis

PDBWiki	**PDBWiki** - a community annotated knowledge base of biological molecular structures
PROTEOPEDIA	**Proteopedia** - collaborative 3D encyclopedia of proteins and other molecules

Fold databases

CATH	**CATH** structural classification
SCOP	**SCOP** structural classification
FSSP	**FSSP** structural alignments

Protein sequence

PDBSWS	**PDBSWS**: mapping of PDB protein chains to SwissProt entries

Secondary structure

HSSP	**HSSP**: Homology derived Secondary Structure of Proteins

Finding structures with ligands: substrates, inhibitors, drugs…

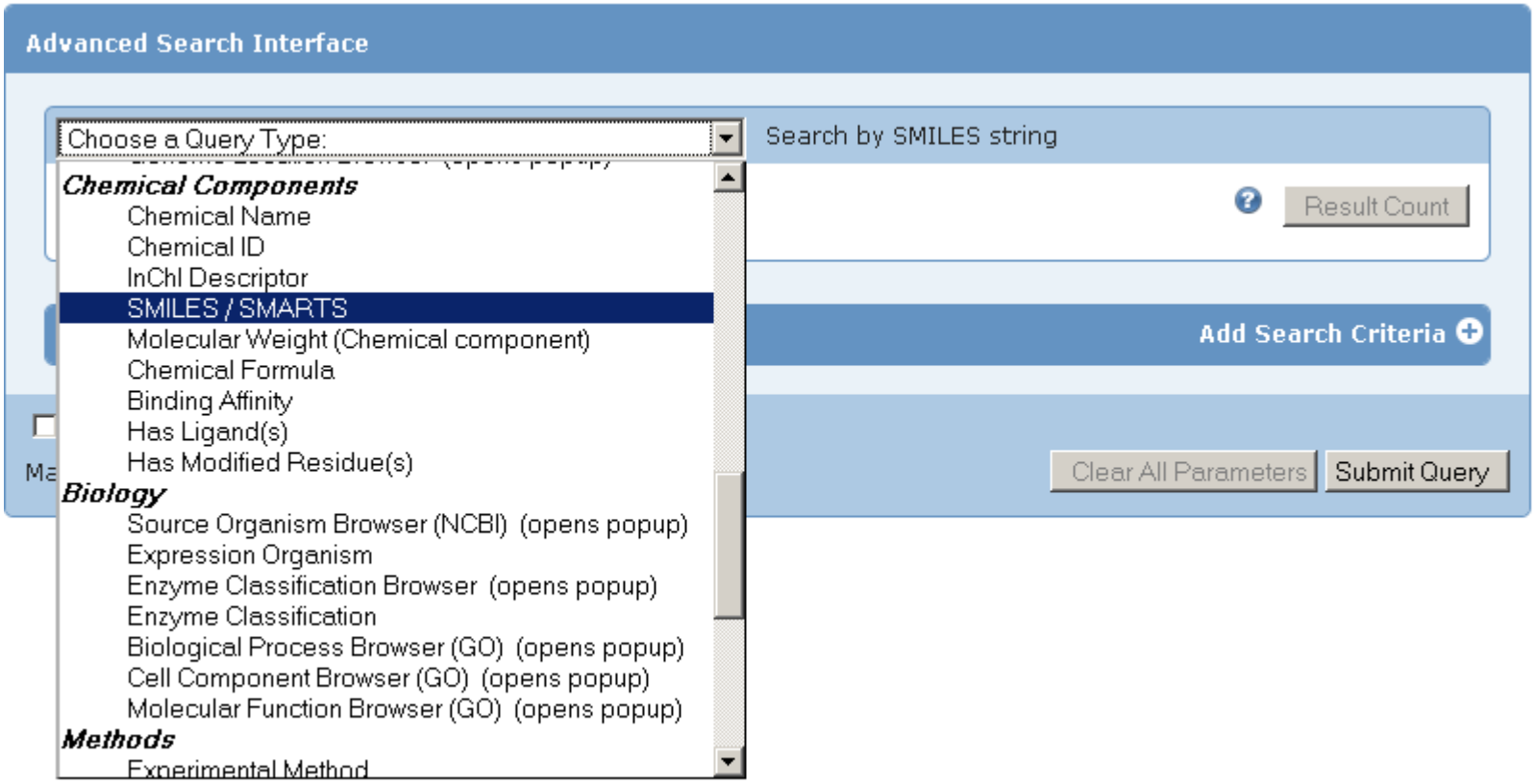
Often, people use non-hydrolyzable substrate analogs in their structure, e.g. ATP analogs. There are many different ATP analogs!

In PDB, every chemical compound has its own abbreviation

If you want to study proteins with bound ATP or ATP analogs, you have to use the adequate tools

Finding structures with ligands

http://www.rcsb.org/pdb/static.do?p=help/advancedSearch/index.html



Naming & finding chemical entities

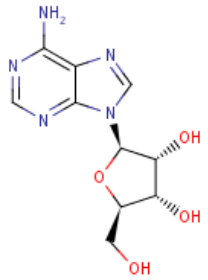
EBI > Databases > Small Molecules > ChEBI > Main

Search ChEBI here! Search ChEBI

Search for ★★★ only

adenosine (CHEBI:16335)

Main | Automatic Xrefs



ChEBI Name	**adenosine**
ChEBI ID	**CHEBI:16335**
Definition	A ribonucleoside composed of a molecule of adenine attached to a ribofuranose moiety via a β–N⁹-glycosidic bond.
Stars	★★★ This entity has been manually annotated by the ChEBI Team.
Secondary ChEBI IDs	CHEBI:2472, CHEBI:40906, CHEBI:40911, CHEBI:40558, CHEBI:40825, CHEBI:13734, CHEBI:105797, CHEBI:22237

See structure as: ☑ Image ☐ Applet

Download Molfile

- Find compounds which contain this structure
- Find compounds which resemble this structure

more structures >>

Formula	Source
C10H13N5O4	KEGG COMPOUND

Net Charge	0
Average Mass	267.24152
InChI	InChI=1S/C10H13N5O4/c11-8-5-9(13-2-12-8)15(3-14-5)10-7(18)6(17)4(1-16)19-10/h2-4,6-7,10,16-18H,1H2,(H2,11,12,13)/t4-,6-,7-,10-/m1/s1
InChIKey	InChIKey=OIRDTQYFTABQOQ-KQYNXXCUSA-N
SMILES	Nc1ncnc2n(cnc12)[C@@H]1O[C@H](CO)[C@@H](O)[C@H]1O

Swiss Institute of Bioinformatics

E.coli alkB DNA repair dioxygenase: finding similar proteins/structures

Summary | Sequence | Derived Data | Seq. Similarity | 3D Similarity | Literature | Biol. & Chem. | Methods | Geometry | Links

Crystal Structure of AlkB in complex with Fe(II), 2-oxoglutarate and methylated trinucleotide T-meC-T

3I49

Display Files
Download Files
Print this Page
Share this Page

DOI: 10.2210/pdb3i49/pdb NDB ID: NA0060

Primary Citation

Enzymological and structural studies of the mechanism of promiscuous substrate recognition by the oxidative DNA repair enzyme AlkB.

Yu, B., Hunt, J.F.

Journal: (2009) Proc.Natl.Acad.Sci.USA **106:** 14315-14320

PubMed: 19706517
PubMedCentral: PMC2725012
DOI: 10.1073/pnas.0812938106
Search Related Articles in PubMed

PubMed Abstract:

Promiscuous substrate recognition, the ability to catalyze transformations of chemically diverse compounds, is an evolutionarily advantageous, but poorly understood phenomenon. The promiscuity of DNA repair enzymes is particularly important, because it enables diverse kinds of damage to different nucleotide bases to be repaired in a metabolically parsimonious manner. We present enzymological and crystallographic studies of the mechanisms underlying promiscuous substrate recognition by Escherichia coli AlkB, a DNA repair enzyme that removes methyl adducts and some larger alkylation lesions from endocyclic positions on purine and pyrimidine bases. In vitro Michaelis-Menten analyses on a series of alkylated bases show high activity in repairing N1-methyladenine (m1A) and N3-methylcytosine (m3C), comparatively low activity in repairing 1,N(6)-ethenoadenine, and no detectable activity in repairing N1-methylguanine or N3-methylthymine. AlkB has a substantially higher k(cat) and K(m) for m3C compared with m1A.

Biological Assembly

More Images...

View in Jmol | SimpleViewer
Other Viewers | Protein Workshop

Biological assembly assigned by authors and generated by PISA (software)

Redundancy in the Protein Data Bank

As the single worldwide repository for macromolecular structures, the Protein Data Bank holds a body of data that contains considerable redundancy in regard to both sequence and structure. We have incorporated into the query interface the ability to select a subset of structures from which similar sequences have been largely removed. In most cases, the selected subset will contain far fewer structures than the complete result set. However, the following caveats should be kept in mind:

- Sequence similarity is defined on a chain basis, but results are returned on a structure basis.
- Many structures in the PDB contain multiple protein chains, or even hybrids of DNA and protein chains.
- Sequence similarity is only assessed for protein chains.
- The relationship between sequence similarity and structure similarity is complex. Users seeking structure similarity should refer to the options available on the Structure Summary page under "External Links" (in the left hand navigation menu) in the "Structure Classification" section.

The primary purpose of this feature is to filter a list of likely highly similar structures to provide one or more representatives. Results may differ from other so-called non-redundant sets (e.g. PDB_SELECT [Hobohm U., and Sander C.,*Protein Science*, 3: 522-524, 1994]).

Sequence clustering in the PDB is performed via blastclust. Detailed information for the clusters containing a given structure is available on the "Seq. Similarity" tab of the Structure Summary page for the respective structure, e.g.

Statistics

The following table shows the number of non-redundant sequences as determined by blastclust at several levels of sequence identity.

Method	Description	# of Clusters
blast	100% identity	39724
blast	95% identity	28436
blast	90% identity	27339
blast	70% identity	24656
blast	50% identity	21647
blast	40% identity	19353
blast	30% identity	16678

Notes on Blast Clustering

Blast clustering is performed with the following parameters (example 95%):

-p T -b T -S 95

Structural Similarities for PDB 3I3Q

The following structural similarities have been found using the jFATCAT-rigid algorithm. In order to reduce the number of hits, **a 40% sequence identity clustering has been applied** and a representative chain is taken from each cluster.

Root mean square deviation

3I3Q.A (chain 1) vs. representatives of other sequence clusters (chain 2)										
Rank	Results	Chain 2	Title	P-value	Score	Rmsd	Len1	Len2	%ID	%Cov1
1	view	2IUW.A	ALKYLATED REPAIR PROTEIN ALKB	2.12E-7	268.58	3.08	203	205	15	90
2	view	3H8R.A	Alpha-ketoglutarate-dependent d	2.56E-7	261.15	3.02	203	204	17	90
3	view	2RG4.B	Uncharacterized protein	2.68E-6	244.65	3.02	203	206	7	77
4	view	2HBT.A	Egl nine homolog 1	8.43E-6	242.55	3.04	203	224	9	74
5	view	3LFM.A	Protein fto	3.63E-5	291.43	3.03	203	420	9	91
6	view	3ITQ.A	Prolyl 4-hydroxylase, alpha subuni	5.73E-5	193.94	3.13	203	198	10	77
7	view	3DKQ.A	PKHD-type hydroxylase Sbal_3634	1.15E-4	193.30	3.02	203	230	9	78
8	view	2JIG.B	PROLYL-4 HYDROXYLASE	1.85E-4	176.60	3.09	203	195	9	73
9	view	2A1X.A	Phytanoyl-CoA dioxygenase	4.09E-4	184.59	3.09	203	253	9	77
10	view	3ON7.D	Oxidoreductase, iron/ascorbate	4.22E-4	204.64	3.11	203	273	7	68
11	view	2OPW.A	PHYHD1 protein	4.32E-4	195.11	3.04	203	286	8	77
12	view	1DGW.A	CANAVALIN	6.44E-4	171.72	3.04	203	178	8	54
13	view	2RDQ.A	1-deoxypentalenic acid 11-beta h	7.47E-4	186.12	3.27	203	265	6	74
14	view	1H2K.A	FACTOR INHIBITING HIF1	7.56E-4	209.20	3.11	203	332	6	74
15	view	1LR5.A	Auxin binding protein 1	9.06E-4	150.59	3.00	203	160	7	58

Root Mean Square Deviation (rmsd):

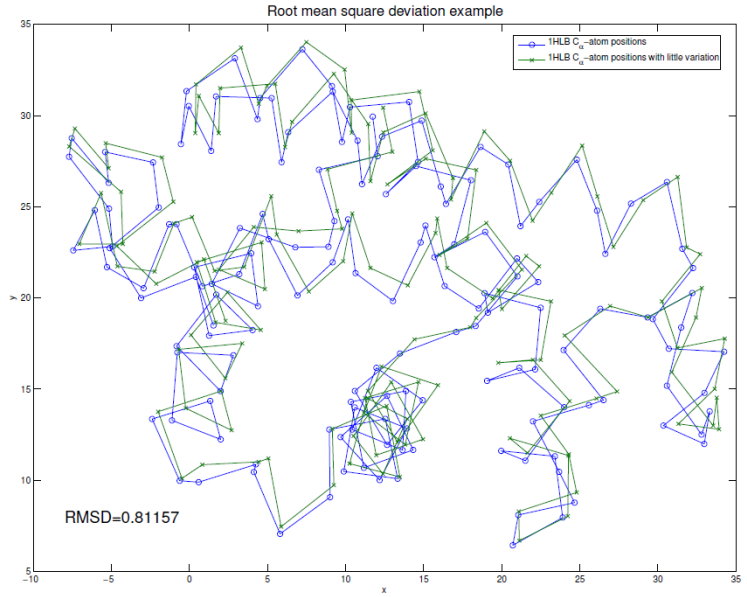
describes how well the alpha-carbon atoms of 2 proteins superimpose

.

The formula for root mean square deviation is defined as

$$RMSD = \sqrt{\frac{1}{N}\sum_{i=1}^{N} |r_i^{model} - r_i^{real}|^2}$$

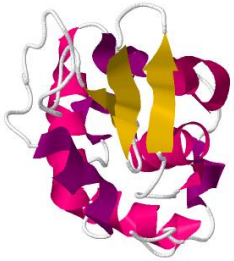
where r_i^{model}, and r_i^{real} are the positions of i:th C_α atoms in model and real protein.



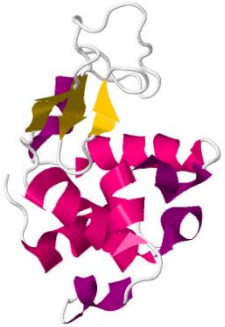
www.lce.hut.fi/teaching/S-114.500/Protein_Structure1.pdf

3D-structure classification and alignment

Structure description1



Structure description 2



Comparison algorithm

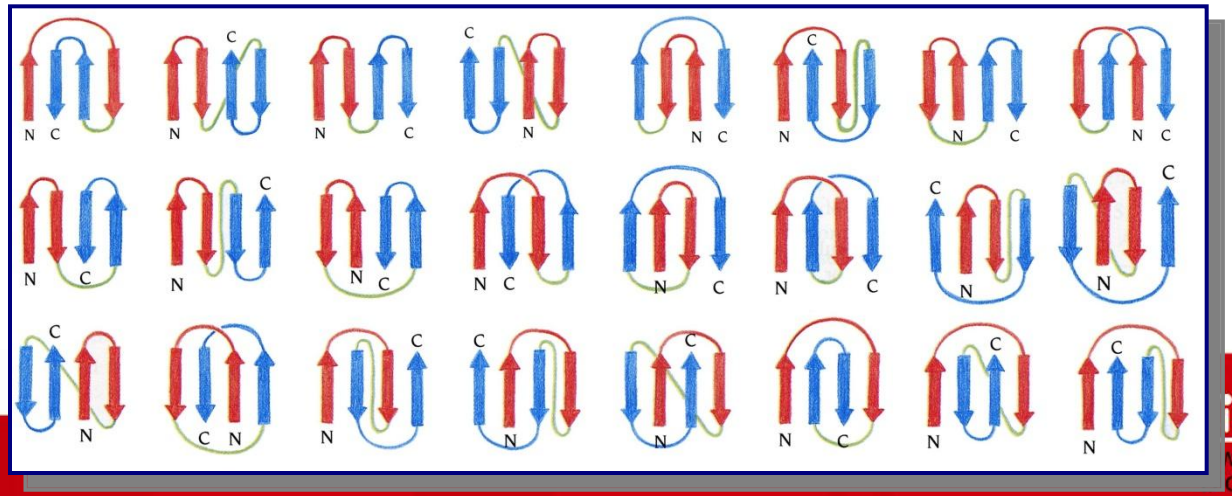
Scores

Similarity

Protein structure classification: quantitative criteria

Purely alpha-helical structure
Purely beta-strand structure,
Mixed
Topology

Single domain vs multidomain proteins

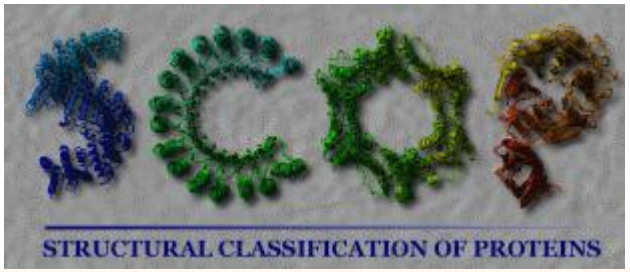


Structure Classification Databases

- SCOP (MRC Cambridge)
  - Structural Classification of Proteins
  - Murzin et al. 1995
  - Largely manual (visual inspection), last update June 2009
- CATH (University College, London)
  - Class, Architecture, Topology and Homologous superfamily
  - Orengo et al. 1993, 1997
  - Manual and automatic method, last update Sept 2011
- DALI/FSSP (EBI, Cambridge)
  - Fold classification based on Structure-Structure alignment of Proteins
  - Holm and Sander, 1993
  - Completely automatic, updated every 6 months (last in 2011)

SCOP database

- **Classes**
- All alpha proteins
- All beta proteins
- Alpha and beta proteins (a/b) - *Mainly parallel beta sheets*
- Alpha and beta proteins (a+b) - *Mainly antiparallel beta sheets (**segregated alpha and beta regions**)*
- Multi-domain proteins (alpha and beta) - *Folds consisting of two or more domains belonging to different classes*
- Membrane and cell surface proteins
- Small proteins



http://scop.berkeley.edu/

CATH – Protein Structure Classification

- Hierarchical classification of protein domain structures in PDB.
- Mostly automated classification
- Domains are clustered at four major levels:

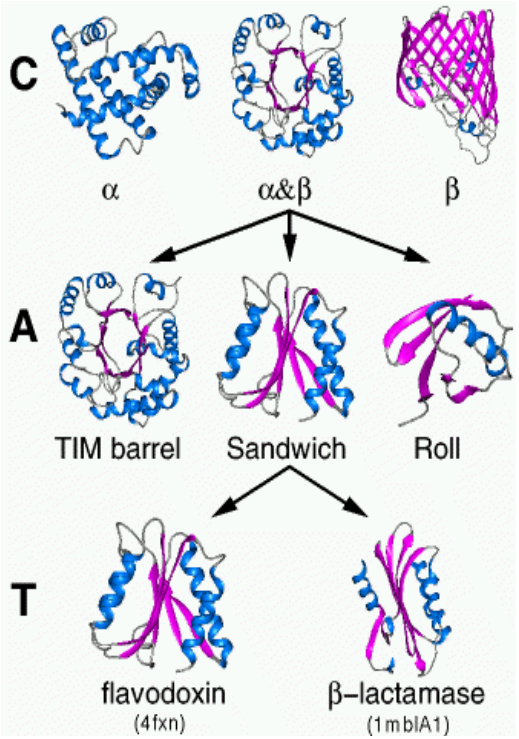
Class

Architecture

Topology

Homologous superfamily

Sequence family

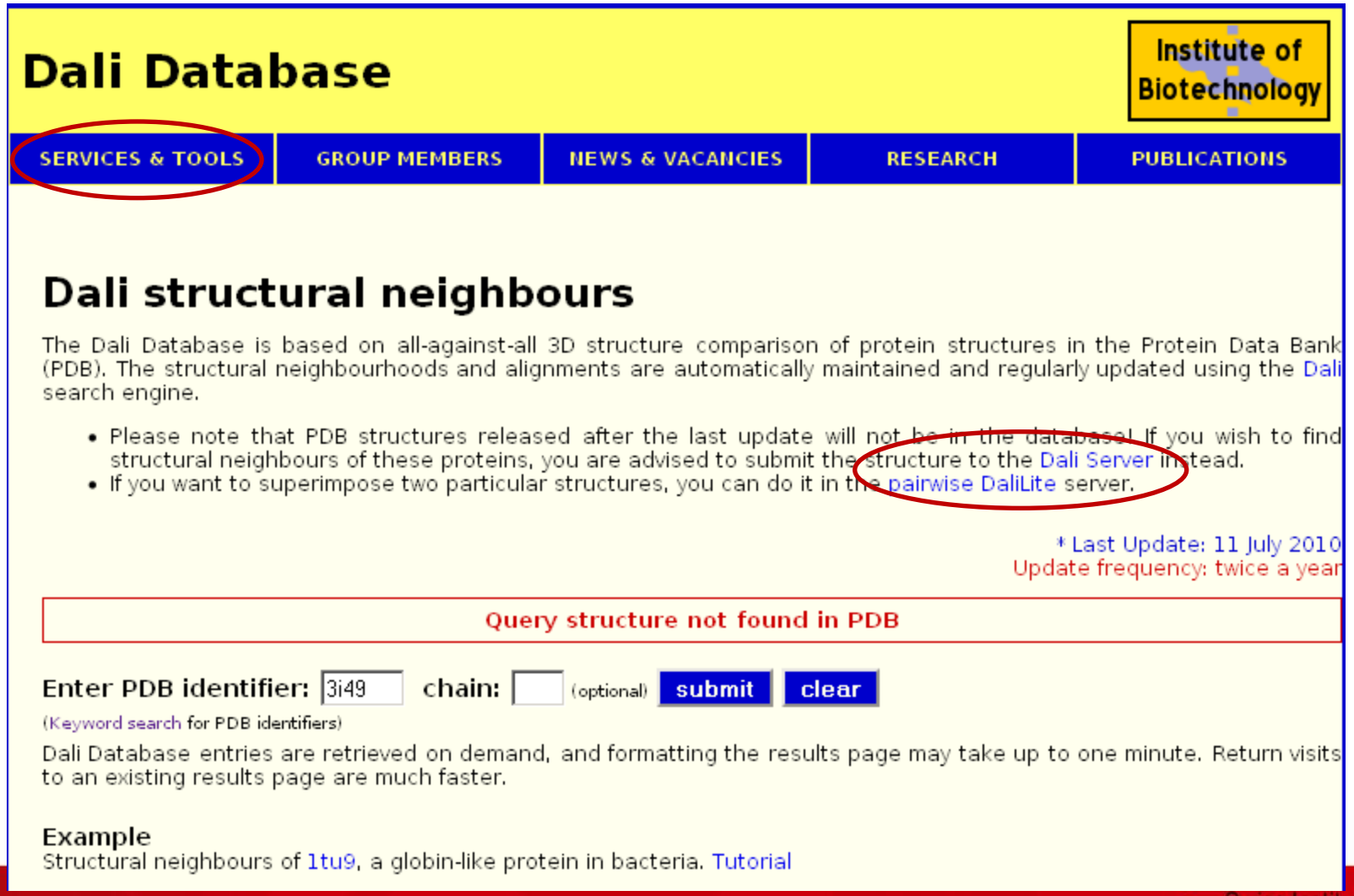


www.cathdb.info/

http://nar.oxfordjournals.org/content/27/1/275.full

Swiss Institute of Bioinformatics

Finding similar protein structures via the DALI database



Finding similar protein structures via the DALI server

Select neighbours (check boxes) for viewing as multiple structural alignment or 3D superimposition. The list of neighbours is sorted by Z-score (A measure of the statistical significance of the result relative to an alignment of random structures). Similarities with a Z-score lower than 2 are spurious.

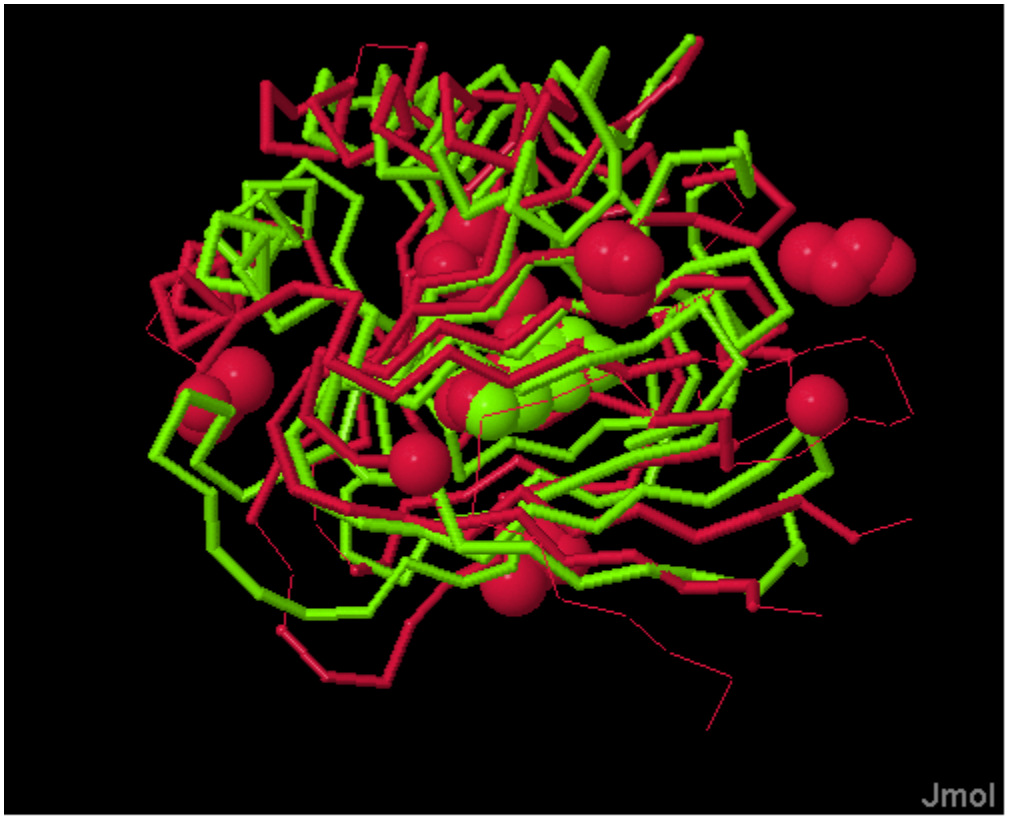
Structural Alignment | ☑ Expand gaps | 3D Superimposition (Jmol Applet) | Reset Selection

Summary

No:	Chain	Z	rmsd	lali	nres	%id	PDB	Description
1:	3i49-A	41.6	0.0	200	200	100	PDB	MOLECULE: ALPHA-KETOGLUTARATE-DEPENDENT DIOXYGENASE ALKB;
2:	3i3m-A	40.2	0.2	200	200	100	PDB	MOLECULE: ALPHA-KETOGLUTARATE-DEPENDENT DIOXYGENASE ALKB;
3:	2fdf-A	39.4	0.3	199	199	100	PDB	MOLECULE: ALKYLATED DNA REPAIR PROTEIN ALKB;
4:	3bkz-A	39.1	0.3	200	201	100	PDB	MOLECULE: ALPHA-KETOGLUTARATE-DEPENDENT DIOXYGENASE ALKB;
5:	3bie-A	39.0	0.4	200	202	100	PDB	MOLECULE: ALPHA-KETOGLUTARATE-DEPENDENT DIOXYGENASE ALKB;
6:	3bi3-A	38.8	0.4	199	201	99	PDB	MOLECULE: ALPHA-KETOGLUTARATE-DEPENDENT DIOXYGENASE ALKB;
7:	3i3q-B	38.6	0.5	200	203	100	PDB	MOLECULE: ALPHA-KETOGLUTARATE-DEPENDENT DIOXYGENASE ALKB;
8:	3khc-B	38.3	0.4	200	207	100	PDB	MOLECULE: ALPHA-KETOGLUTARATE-DEPENDENT DIOXYGENASE ALKB;
9:	3i3q-A	38.2	0.7	200	203	100	PDB	MOLECULE: ALPHA-KETOGLUTARATE-DEPENDENT DIOXYGENASE ALKB;
10:	3khc-A	37.5	0.6	200	207	100	PDB	MOLECULE: ALPHA-KETOGLUTARATE-DEPENDENT DIOXYGENASE ALKB;
11:	2fdi-A	36.9	0.7	199	199	100	PDB	MOLECULE: ALKYLATED DNA REPAIR PROTEIN ALKB;
12:	3khb-B	36.9	0.5	200	206	100	PDB	MOLECULE: ALPHA-KETOGLUTARATE-DEPENDENT DIOXYGENASE ALKB;
13:	2fdk-A	36.7	0.7	198	198	100	PDB	MOLECULE: ALKYLATED DNA REPAIR PROTEIN ALKB;
14:	3khb-A	36.7	0.7	200	203	100	PDB	MOLECULE: ALPHA-KETOGLUTARATE-DEPENDENT DIOXYGENASE ALKB;
15:	2fdh-A	36.6	0.8	199	199	100	PDB	MOLECULE: ALKYLATED DNA REPAIR PROTEIN ALKB;
16:	3i2o-A	36.5	0.8	199	199	100	PDB	MOLECULE: ALPHA-KETOGLUTARATE-DEPENDENT DIOXYGENASE ALKB;
17:	2fd8-A	36.5	0.8	199	199	100	PDB	MOLECULE: ALKYLATED DNA REPAIR PROTEIN ALKB;
18:	2fdg-A	36.1	0.8	199	199	100	PDB	MOLECULE: ALKYLATED DNA REPAIR PROTEIN ALKB;

Alignment and 3D-superposition of 3i49A and 2iuwA

(another family member: 18% seq identity, z-score 17.7 and 2.8 A rmsd)



Toggle: ☐ spinning ☑ superimpose all ligands Clear labels

First structure's backbone: ◉ CA trace ○ Cartoon Rockets. Matched structures' backbone

Colour backbone of first structure: ◉ monochrome ○ rainbow ○ sequence conservation

Alignment and 3D-superposition of 3i49A and 2iuwA

(another family member: 18% seq identity, z-score 17.7 and 2.8 A rmsd)

Conservation of secondary structure

Sequence alignment: residues essential for catalysis are conserved

DaliLite Results: Multiple structural alignment

Each neighbour is shown in the pairwise Dali-alignment to 3i49A. Inserted segments relative to the top structure are hidden. You can check the 'Expand gaps' option in the su
proteins. Uppercase means structurally equivalent positions with 3i49A. Lowercase means insertions relative to 3i49A. The first part shows the amino acid sequences of the s
structure assignments by DSSP (H/h: helix, E/e: strand, L/l: coil). The most frequent amino acid type is coloured in each column.

```
                 :         :         :         :         |         :         :         :         :       100       :
0001 3i49A   LAAGAVILRRFAFNAAEQLIRDINDVASQSPFRQMVTPGGYTMSVAMTNCGHLGWTTHRQGYLYSPIDPQTNKPWPAMPQSFHNLCQRAATAAGYPDFQPDACLINRYAPGAKLSLH
0002 2iuwA   YEIRVCLYPGFVDVEADWILEQLCQDV---PWKQRTGIITYQQPRLTAWYG-ELPY------TYSRIT-----EPNPWHPVLRTLKNRIEEN--TGHT-FNSLXCNLYREKDSVDWH

                 :         :         :         :         |         :         :         :         :       100       :
0001 3i49A   LLLLEEEELLLLLLLHHHHHHHHHHHHHHHLLLLLLLLLLLLLLLLEEEEELLEEEEELLEEEEELLLLLLLLLLLLLLHHHHHHHHHHHHHHLLLLLLLLEEEEEEELLLLLEEEE
0002 2iuwA   EEEEEEEELLLLLLHHHHHHHHHHHHHHL---LLLLLEEELEEELLLEEEEEE-LLLL------LLLHHH-----LLLLLLHHHHHHHHHHHHH--HLLL-LLEEEEEEELLLLLEEEE
```

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

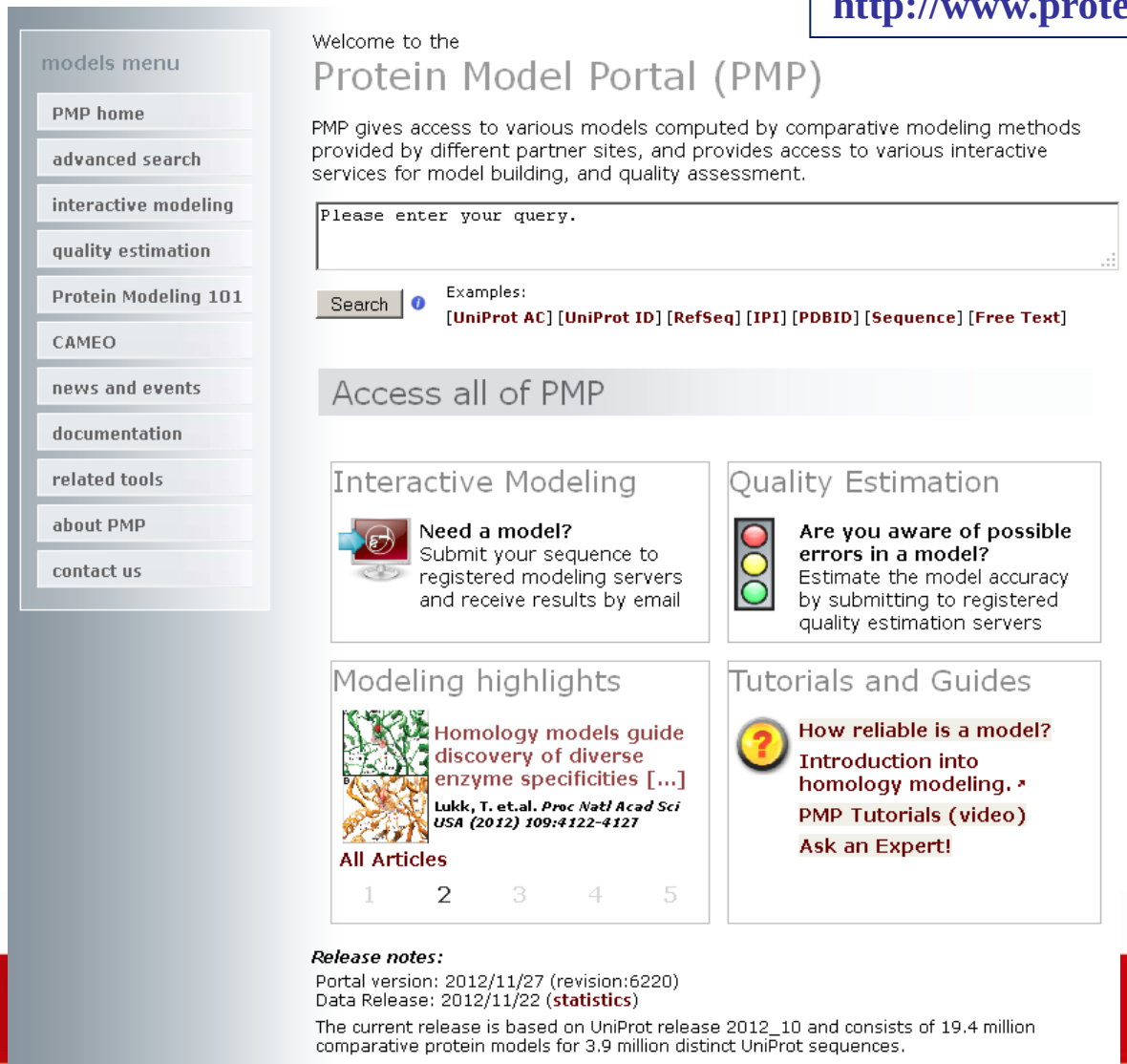
Protein structure prediction

- Ab initio: only meaningful for small proteins (up to ca 120 residues)
- Homology modeling: can give highly valuable results
- The higher the sequence similarity, the higher the chance that proteins have similar structures
- Proteins with similar structures often (but not always!!) have similar functions
- For function prediction, you need a basis: characterized proteins

The Protein Model Portal: access to all publicly accessible protein models and 3D-structures

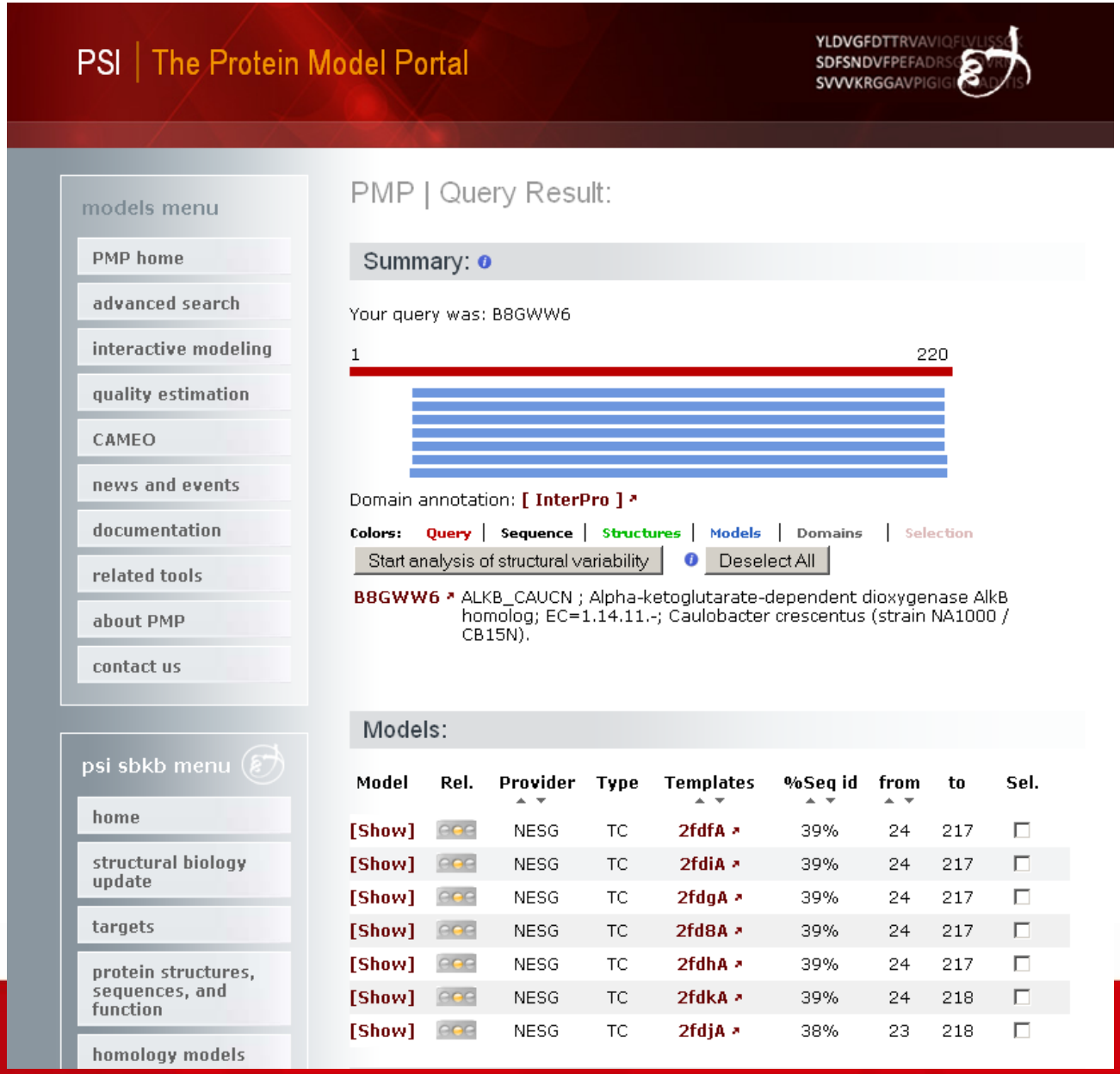
Out of 538'000 UniProtKB/Swiss-Prot entries, 429'000 have a link to Protein Model Portal

http://www.proteinmodelportal.org/?



SIB Swiss Institute of Bioinformatics

Example: alkB fromC.crescentus

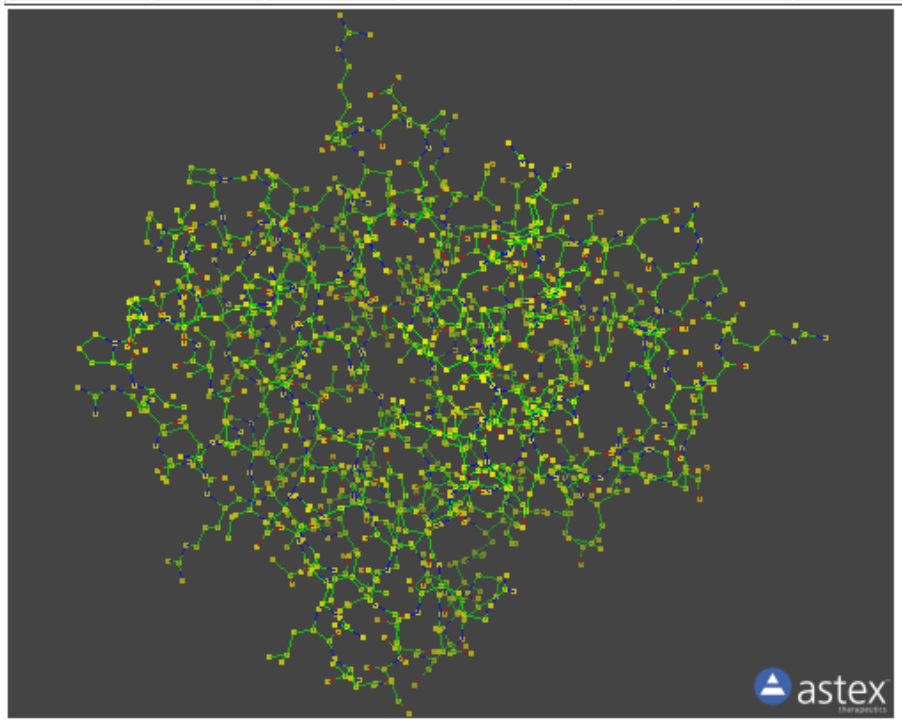


SIB Swiss Institute of Bioinformatics

NESG MODELS DATABASE

The similarity search retrieved the following modeled proteins from the NESG database

Target ID	Template	Model Quality(pG)	Target/Template Sequence Identity	E Value	Target Coverage	Species	Target Description
AAK21997.1	2FDF	1.00	39%	2.1e-49	88%	Caulobacter crescentus	DNA alkylation damage repair protein AlkBRecName: Full=Alpha-ketoglutarate-dependent dioxygenase AlkB homologRecName: Full=Alpha-ketoglutarate-dependent dioxygenase AlkB homologDNA alkylation damage repair protein AlkB



Colour...

Atoms

Surface

Opacity...

Cartoon

template:2FDF

target:AAK21997.1

download:model

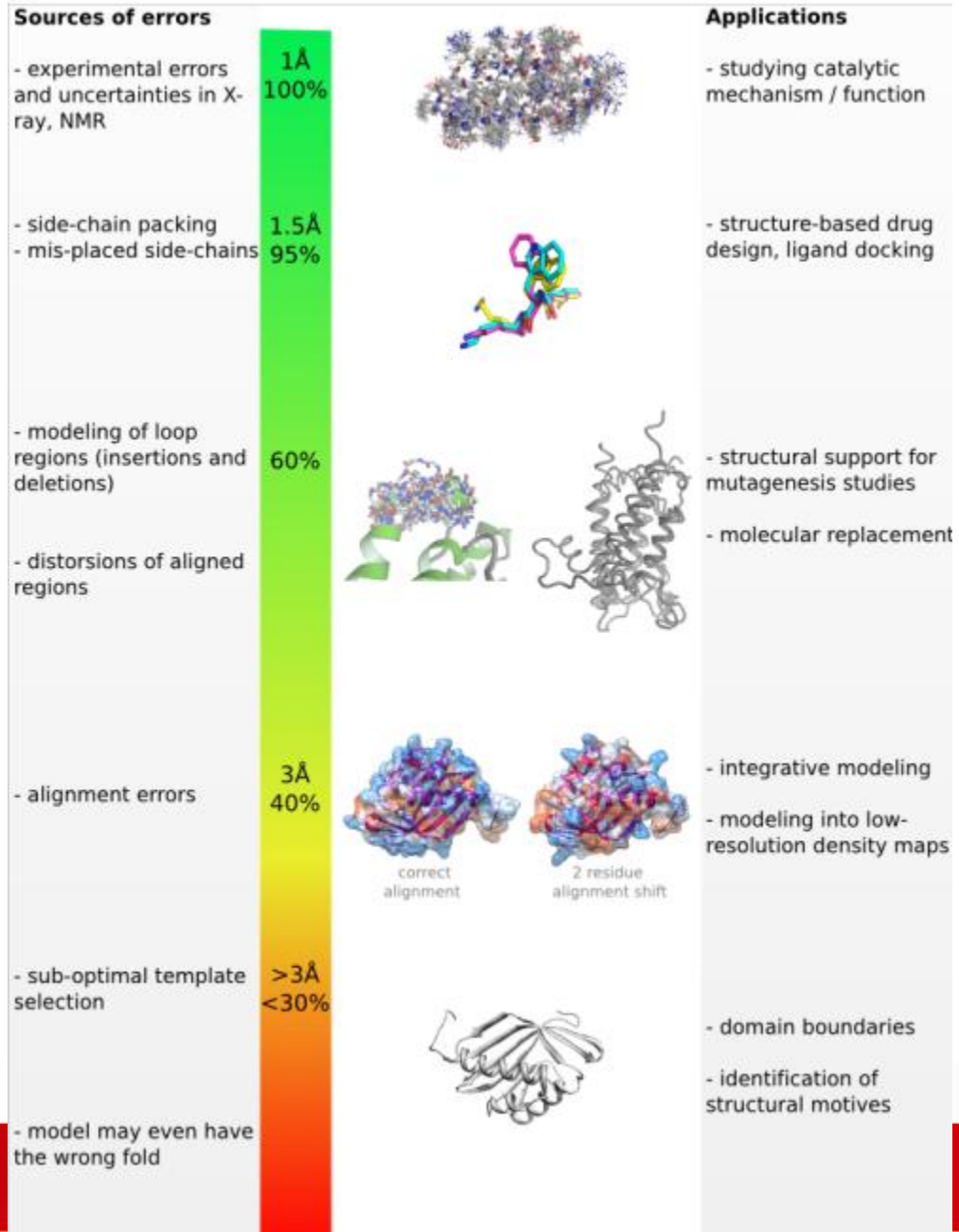
Side chain optimization

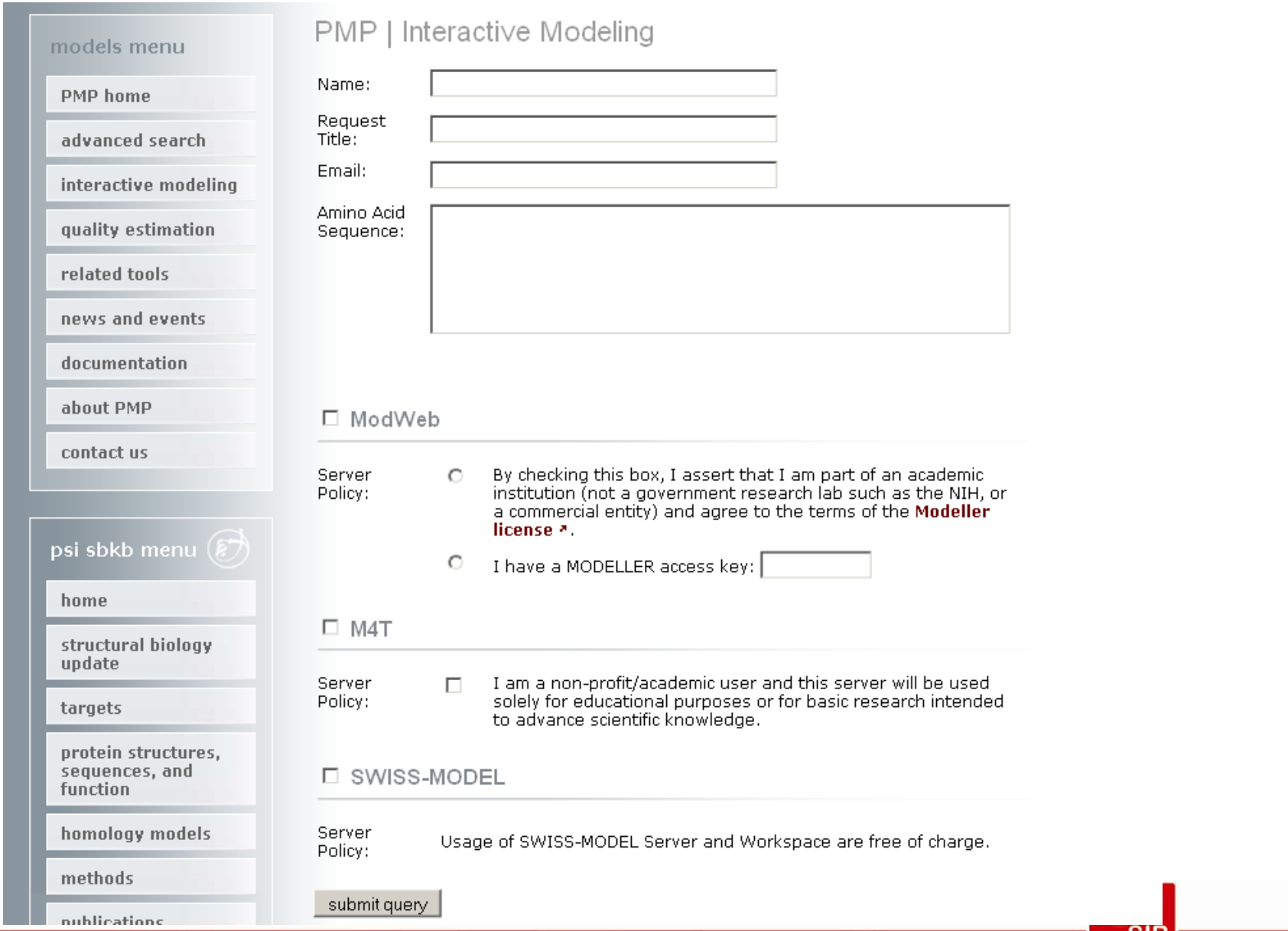
Adv.Refinement

Function Analysis

SIB
Swiss Institute of Bioinformatics

http://www.proteinmodelportal.org/?pid=documentation#modelquality





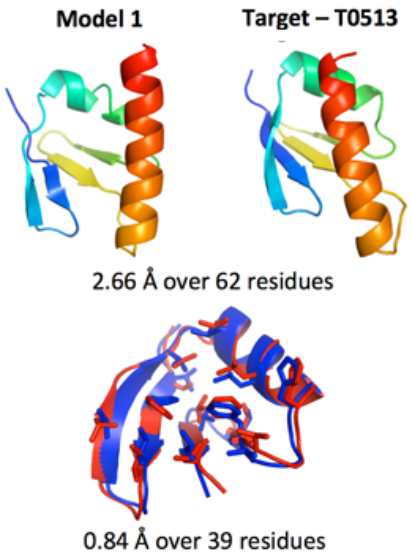
SIB Swiss Institute of Bioinformatics

Ab initio and comparative protein structure prediction

http://robetta.bakerlab.org/

ROBETTA BETA

Full-chain Protein Structure Prediction Server



de novo prediction by Robetta in CASP-8

REGISTRATION

[Register / Update] [Login]

DOCUMENTATION

[Docs / FAQs]

SERVICES

Domain Parsing & 3-D Modeling
[Queue] [Submit]

Interface Alanine Scanning
[Queue] [Submit]

Fragment Libraries
[Queue] [Submit]

DNA Interface Residue Scanning
[Queue] [Submit]

RELATED SITES

RosettaBackrub Server
RosettaAntibody Server
RosettaDesign Server
RosettaDock Server
Rosetta Commons
FoldIt
Rosetta@home
Human Proteome Folding Project

SIB Swiss Institute of Bioinformatics

Protein 3D-structure analysis

and now it is up to you:

Practicals