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









Convenient, but real molecules fill up space

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

1B6Q







Another view of the same

1B6Q

JMOL cartoon





Space-filling model of the same

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

1B6Q



JMOL

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/





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

http://swissmodel.expasy.org/course/



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 turn5.4 Å repeat along the helix axisEach 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: 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

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



LRR domain



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/

Home · Download · Help/Info

🔅 STRING 9.0

STRING - Known and Predicted Protein-Protein Interactions

\frown	$\neg \frown$			
by name protein sequence r	nultiple multiple names sequences	What it does		
protein name: (examples: #1 #2 #3) fak1		STRING is a database of known and predicted protein interactions. The interactions include direct (physical) and indirect (functional) associations; they are derived from four sources:		
(STRING understands a variety of p and accessions; you can also try a	rotein names <u>random entry</u>)	Genomic High-throughput (Conserved) Previous Context Experiments Coexpression Knowledge		
organism: hum	•	STRING quantitatively integrates interaction data from these		
human Homo sapiens human body lice	Reset GO!	sources for a large number of organisms, and transfers information between these organisms where applicable. The database current covers 5'214'234 proteins from 1133 organisms.		
Pediculus humanus corporis human body louse Pediculus humanus corporis	st	Use Scenarios		
Anaplasma phagocytophilum HZ human lice	val of Interacting Genes, t al. 2011 / 2009 / 20 Robot Access Guide, ST	/ <i>Proteins</i>) is being developed at <u>CPR, EMBL, SIB, KU, TUD</u> and <u>UZH</u> . 107 / <u>2005</u> / <u>2003</u> / <u>Snel et al. 2000</u> . TRING/STITCH Blog, Supported Browsers.		
Pediculus humanus corporis human louse Pediculus humanus corporis agent of human granulocytic	of STRING - now covering more than 1100 organisms (and counting) ! <u>(H</u> and <u>eggNOG</u> - two sister projects built on STRING data! produce an earlier finding? Confused? Refer to our <u>old releases</u> .			
Anaplasma phagocytophilum				

Home · Download · Help/Info

😤 STRING 9.0



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



(requires Flash player 10 or better)

Your Input:

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)

leighborhood sene Fusion ooccurrence Textmining [Homology] atabases (Homo sapiens) Score 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 ٠ 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 • L. 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



Neiahborhood

Fusion

Views:

Occurence Coexpression Experiments

Database Textmining

Summary Network

Publed

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 <u>intrinsically</u> <u>disordered</u>=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 Gonoimeter (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: ¹H, ¹³C, ¹⁵N, ¹⁹F, ³¹P
- 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 nonidentical 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



SIB 38 Swiss Institute of Bioinformatics

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

X-ray Crystallography

High resolution (up to 0.5Å) No protein mass limit

NMR

No crystals needed Conformation of protein in solution **Dynamic aspects** (conformation ensemble view)

Disadvantages

Crystals needed

Artefacts due to crystallization (Enzyme in open vs closed Conformation) Structure is a static average

Highly concentrated solution

(1mM at least) Isotope substitution (¹³C, ¹⁵N) Limited maximum weight (about 60 kD)

Electron Microscopy No 3D-crystals needed **Direct image**

Large radiation damage Need 2D crystals or large complexes Artefacts

SIB 41 Swiss Institute of Bioinformatics

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

PROTEIN DATA BANK	An Information Portal to Biologic As of Tuesday Nov 16, 2010 at 4 PM PST there are 69351 St	A MEMBER OF THE PDB al Macromolecular Structures ructures () ? PDB Statistics ?
Contact Us Print	PDB ID or Text E.coli alkb dna repair	Search 😨 Advanced Search
MyPDB Hide Login to your Account	A Resource for Studying Biological Macromolecules	Customize This Page
Register a New Account Home Hide	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.	Source Organism Browser with Common Names
News & Publications Usage/Reference Policies	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,	Latest reatures released: Website Release Archive: 💌



Advanced search: query with UniProt AC

Contact Us Print	PDB ID or Text Search 3 Advanced Search
Image: MyPDB Hide Login to your Account Register a New Account	Advanced Search Interface
† Home Hide	UniProtKB Accession Number(s)
News & Publications Usage/Reference Policies Deposition Policies Website FAQ Deposition FAQ Contect Us	Search for structures by entering one or more UniProtKB Accession Numbers (UniProtKB AC), e.g. P69905 Accession IDs



Advanced search: query via BLAST

PROTEIN DATA BANK	An II As of Tuesday Nov 09, 201	A MEMBER OF THE PDB nformation Portal to Biological Macromolecular Structures 0 at 4 PM PST there are 69162 Structures 2 ? PDB Statistics ?
Contact Us Print	PDB ID or Text	Search 🕐 Advanced Search
MyPDB Hide Login to your Account Register a New Account Home Hide News & Publications Usage/Reference Policies Deposition Policies Website FAQ Deposition FAQ Contact Us	Advanced Search Interface Choose a Query Type: Number of Disulfide Bonds Molecular Weight (Structure) Secondary Structure Content Secondary Structure Length SCOP Classification Browser (opens popup) CATH Classification Browser (opens popup)	Sequence search (BLAST or FASTA) Result Count Add Search Criteria C
About Us Careers External Links Sitemap New Website Features t Deposition Hide All Deposit Services Electron Microscopy X-ray NMR Validation Server BioSync Beamline Related Tools t Search Hide	Taxonomy Browser (opens popup) Sequence Features Sequence (BLAST/FASTA/PSI-BLAST) Translated Nucleotide Sequence (BLASTX) Sequence Motif Chain Length Genome Location Browser (opens popup) Chemical Components Chemical Name Chemical ID InChI Descriptor SMILES / SMARTS Molecular Weight (Chemical component) Chemical Formula	Clear All Parameters Submit Query

Swiss Institute of Bioinformatics

Finding protein structures at RCSB PDB

⊠ 3I49 ≛ 🖹 ᡖ 🚱

3B

☑ 3B.

Download PDB F

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

0	Authors:	Yu, B. <i>P</i> ,	Hunt, J.F.🔎							
ile	Release Date:	2009-08	-25	Classification:	Oxidoreductase/dna 🔎					
5	Experiment:	X-RAY DI	X-RAY DIFFRACTION with resolution of 1.60 Å							
Ÿ	Compound:	2 Polyme 2 Ligand	BrS [Display Full Polymer Detail IS [Display Full Ligand Details	ls Display for All Results] Display for All Results]						
	Citation:	Enzymol oxidativ (2009) P	logical and structural studie re DNA repair enzyme AlkB. Proc.Natl.Acad.Sci.USA 106: 1	4315-14320 <i>[Display Full Abstract</i> D	substrate recognition by the isplay for All Results]					
I3 0	X-ray structur Authors:	re of Alk ⊻i, c.,⊃,	(B protein bound to dsD Yang, CG.,P	NA containing 1meA/A with c	ofactors					
	Release Date:	2008-04	22	Classification:	Oxidoreductase/dna 🔎					
R	Experiment:	X-RAY DI	(FFRACTION with resolution o	f 1.90 Å						
R ^{\$}	Compound:	3 Polyme 3 Ligand	BrS [Display Full Polymer Detail IS [Display Full Ligand Details	ls Display for All Results] Display for All Results]						
	Citation:	Crystal 9 (2008) N	structures of DNA/RNA repairs Nature 452: 961-965 (Display	air enzymes AlkB and ABH2 bound t y Full Abstract Display for All Results]	o dsDNA.					
IE Ø	X-ray structur Authors:	reofEc ⊻i,c.,⊃,	oli AlkB bound to dsDN. Yang, CG.P	A containing 1meA/T with Mn	and 2KG					
	Release Date:	2008-04	22	Classification:	Oxidoreductase/dna 🔎					
No.	Experiment:	X-RAY DI	IFFRACTION with resolution o	f 1.68 Å						
ED .	Compound:	3 Polyme 3 Ligand	BrS [Display Full Polymer Detail IS [Display Full Ligand Details	ls Display for All Results] Display for All Results]						
	Citation:	Crystal s	structures of DNA/RNA repa	air enzymes AlkB and ABH2 bound t	o dsDNA					

(2008) Nature 452: 961-965 [Display Full Abstract | Display for All Results]

SIB Swiss Institute o Bioinformatics

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

Biol. & Chem. Methods Derived Data Seq. Similarity 3D Similarity Literature Summary Display Files * 3149 Crystal Structure of AlkB in complex with Fe(II), 2-oxoglutarate and methylated trinucleotide T-meC-T Download Files 🔻 Print this Page Share this Page * DOI:10.2210/pdb3i49/pdb NDB ID: NA0060 **Primary Citation** Biological Assembly $(\mathbf{?})$ Enzymological and structural studies of the mechanism of promiscuous substrate recognition (+)by the oxidative DNA repair enzyme AlkB. Yu, B.P. Hunt, J.F.P 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 More Images... damage to different nucleotide bases to be repaired in a metabolically parsimonious manner. We present enzymological and crystallographic studies of the mechanisms underlying promiscuous 📥 View in Jmol SimpleViewer substrate recognition by Escherichia coli AlkB, a DNA repair enzyme that removes methyl adducts Protein Workshop Other Viewers 🔻 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 Biological assembly assigned by authors N1-methyladenine (m1A) and N3-methylcytosine (m3C), comparatively low activity in repairing and generated by PISA (software)

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.

PDB file

1	HEADER	OXIDORED	JCTASE/DNA	01-JUL-09	3149
	TITLE	CRYSTAL :	STRUCTURE OF ALKB IN	COMPLEX WITH FE(II), 2	-
	TITLE	2 OXOGLUT	ARATE AND METHYLATED	TRINUCLEOTIDE T-MEC-T	
	COMPND	MOL_ID: :	1;		
(COMPND	2 MOLECULE	E: ALPHA-KETOGLUTARA	FE-DEPENDENT DIOXYGENAS	SE ALKB;
0	COMPND	3 CHAIN: J	A;		
(COMPND	4 FRAGMENT	r: UNP RESIDUES 12-23	16;	
	COMPND	5 SYNONYM	: ALKYLATED DNA REPA	IR PROTEIN ALKB;	
0	COMPND	6 EC: 1.14	4.11;		
0	COMPND	7 ENGINEED	RED: YES;		
	COMPND	8 MOL_ID:	2;		
	COMPND	9 MOLECULI	E: DNA (5'-D(P*TP*(M)	E6)P*T)−3');	
0	COMPND 1	LO CHAIN: 1	3;		
0	COMPND 1	1 ENGINEE	RED: YES		
	SOURCE	MOL_ID: :	1;		
	SOURCE	2 ORGANIS	M_SCIENTIFIC: ESCHER:	ICHIA COLI;	
2	SOURCE	3 ORGANISI	M_TAXID: 83333;		
:	SOURCE	4 STRAIN:	K-12;		
;	SOURCE	5 GENE: A	IDD, ALKB, B2212, JW2	2200;	
	SOURCE	6 EXPRESS	ION_SYSTEM: ESCHERIC	HIA COLI;	
:	SOURCE	7 EXPRESS:	ION_SYSTEM_TAXID: 562	2;	
2	SOURCE	8 EXPRESS	ION_SYSTEM_STRAIN: BI	L21(DE3);	
2	SOURCE	9 EXPRESS	ION_SYSTEM_VECTOR_TY	PE: PLASMID;	
2	SOURCE 1	LO EXPRESS:	ION_SYSTEM_PLASMID: 1	PET26B;	
:	SOURCE 1	1 MOL_ID:	2;		
2	SOURCE 1	2 SYNTHET:	IC: YES		
1	KEYWDS	BETA JELI	LYROLL, PROTEIN-DNA (COMPLEX, DIOXYGENASE, 1	DNA
]	KEYWDS	2 DAMAGE,	DNA REPAIR, IRON, MI	ETAL-BINDING, OXIDORED	UCTASE,
1	KEYWDS	3 OXIDOREI	DUCTASE/DNA COMPLEX		
	EXPDTA	X-RAY DI	FRACTION		
L L	AUTHOR	B.YU,J.F	HUNT		
	REVDAT	2 08-SEI	2-09 3149 1	JRNL	
1	REVDAT	1 25-AU	3-09 3I49 0		
i.	JRNL	AUTH	B.YU,J.F.HUNT		
i.	JRNL	TITL	ENZYMOLOGICAL AND S'	FRUCTURAL STUDIES OF T	HE
,	JRNL	TITL 2	MECHANISM OF PROMIS	CUOUS SUBSTRATE RECOGN	ITION BY
,	JRNL	TITL 3	THE OXIDATIVE DNA R	EPAIR ENZYME ALKB.	
	JRNL	REF	PROC.NATL.ACAD.SCI.	JSA V. 106 1431	5 2009
	JRNL	REFN	ISSI	1 0027-8424	
	JRNL	PMID	19706517		
	JRNL	DOI	10.1073/PNAS.0812938	3106	



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	Ν	LEU	A	15	22.640	-10.051	3.565	1.00 30.09
ATOM	2	CA	LEU	A	15	22.265	-9.579	2.201	1.00 29.42
ATOM	3	С	LEU	A	15	23.495	-9.407	1.320	1.00 28.67
ATOM	4	0	LEU	A	15	23.473	-9.733	0.133	1.00 29.97
ATOM	5	СВ	LEU	A	15	21.514	-8.249	2.291	1.00 29.55
ATOM	6	CG	LEU	A	15	21.294	-7.521	0.963	1.00 30.49
ATOM	7	CD1	LEU	A	15	20.638	-8.459	-0.032	1.00 31.68
ATOM	8	CD2	LEU	A	15	20.442	-6.283	1.190	1.00 30.60
ATOM	9	Ν	ALA	A	16	24.563	-8.887	1.913	1.00 26.31
ATOM	10	CA	ALA	A	16	25.818	-8.660	1.209	1.00 24.06
ATOM	11	С	ALA	A	16	26.815	-8.107	2.214	1.00 22.57
ATOM	12	0	ALA	A	16	26.452	-7.797	3.348	1.00 21.88
ATOM	13	СВ	ALA	A	16	25.616	-7.664	0.078	1.00 23.31
ATOM	14	Ν	ALA	A	17	28.072	-7.989	1.802	1.00 21.94
ATOM	15	CA	ALA	A	17	29.097	-7.453	2.685	1.00 22.11
ATOM	16	С	ALA	A	17	28.716	-6.033	3.091	1.00 20.73
ATOM	17	0	ALA	A	17	28.613	-5.141	2.244	1.00 21.03
ATOM	18	СВ	ALA	A	17	30.450	-7.451	1.979	1.00 23.22
ATOM	19	Ν	GLY	A	18	28.499	-5.833	4.388	1.00 19.10
ATOM	20	CA	GLY	A	18	28.137	-4.521	4.892	1.00 18.73
ATOM	21	С	GLY	A	18	26.661	-4.197	4.763	1.00 19.12
ATOM	22	0	GLY	A	18	26.236	-3.087	5.089	1.00 17.79
ATOM	23	Ν	ALA	A	19	25.875	-5.162	4.295	1.00 18.49
ATOM	24	CA	ALA	A	19	24.440	-4.952	4.128	1.00 18.93
ATOM	25	С	ALA	A	19	23.622	-5.987	4.885	1.00 19.68
ATOM	26	0	ALA	A	19	24.049	-7.129	5.075	1.00 21.19
ATOM	27	СВ	ALA	A	19	24.074	-4.982	2.647	1.00 17.60
ATOM	28	Ν	VAL	A	20	22.432	-5.585	5.311	1.00 18.94
ATOM	29	CA	VAL	A	20	21.561	-6.483	6.046	1.00 18.89
ATOM	30	С	VAL	A	20	20.122	-6.015	5.964	1.00 17.95
ATOM	31	0	VAL	A	20	19.854	-4.824	5.798	1.00 16.19
ATOM	32	СВ	VAL	A	20	21.968	-6.564	7.534	1.00 20.99
ATOM	33	CG1	VAL	A	20	21.871	-5.189	8.178	1.00 22.02
ATOM	34	CG2	VAL	A	20	21.079	-7.562	8.264	1.00 24.32

SIB Swiss Institute of Bioinformatics

С

Ν С С 0 С С С С Ν С С 0 С Ν С С 0 С Ν С С 0 Ν С С 0 С Ν С С 0 С С

Header: protein sequence and ligand information

AT NO ROOT T

DDDEE

DDREF	2143	, A	14	4.	10 (лчP	PU	JSUSC	· .	APVD	_ECOI	- L	-	14	210	2
DBREF	3149	ЭВ	501	50)3 F	DB	30	I49	;	3149	_		50	01	503	3
SEQADV	3149	M	ET A	11	UNF	P P(05050)			II	VITI.	ATING	G MET	THION	VINE
SEQADV	3149	A:	SN A	217	UNF	P P(05050)			E	XPRE:	SSION	J TAC	3	
SEQADV	3149) LI	EU A	218	UNF	P P(05050)			E	KPRE:	SSION	J TAC	3	
SEQADV	3149) T	YR A	219	UNF	P P(05050)			E	KPRE:	SSION	J TAC	3	
SEQADV	3149) PI	HE A	220	UNF	P P(05050)			E	KPRE:	SSION	J TAC	3	
SEQADV	3149) G1	LN A	221	UNF	P P(05050)			E	KPRE:	SSION	J TAC	3	
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	$\operatorname{GL} Y$	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	$\operatorname{GL} \mathtt{Y}$	ALA	LYS
SEQRES	10	A	211	LEU	SER	LEU	HIS	GLN	ASP	LYS	ASP	GLU	PRO	ASP	LEU	ARG
		_														
HET	ME 6	В	502		20											
HET	AKG	A	400		10											

DOFORO

TELET

1

PHOSPHATE

FE2 FE (II) ION

AKG 2-OXYGLUTARIC ACID

C5 H6 O5

*281(H2 O)

FE 2+

	~	

HETNAM HETNAM HETNAM HETNAM HETNAM FORMUL

FORMUL

FORMUL FORMUL

HET

FE2 A 300

2 ME6

3 ME6

2 ME6

3 AKG

4 FE2

5 HOH

ME6 [(2R,3S,5R)-5-(4-AZANYL-3-METHYL-2-OXO-PYRIMIDIN-3-IUM-1-YL)-3-HYDROXY-OXOLAN-2-YL]METHYL DIHYDROGEN **HET: Heteroatoms=non-protein** atoms: small molecules and ions, each C10 H17 N3 O7 P 1+ with a unique abbreviation

040



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

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.P. Hunt, J.F.P

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.

3149

Display Files
 Download Files
 Print this Page
 Share this Page



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

Viewing a structure (JMOL): Right-click in window to change parameters

Display Cartoon

👚 Previous 🔊 Highlight all 🗌 Match

Surface C Off Toggle 📃 Selection

Color

					2575	modèle 1/1 Configurations (2) Sélectionner (1'932) Vue Style Couleur	 ✓ Profondeur de la perspective □ Boîte englobante 	
						Surfaces	Cellule unitaire	
		Tion with		1		Symetrie 🕨	Stéréographique	•
		iip: rigr You m	nt-mouse click on . Iay also drag the I	right-bottom corner (o additional Jmol fund of the Jmol area to re	Off	Combinaison	•
Da	cot Dicplay	Export Ima		-		Squelette	Atomes	•
Cartoon	O Backhone	Exportination (O Ball and Stick	Otigands	Roquettes Cartoon	Étiquettes	•
Secondary Structure	Chain	(Rainhow	© By Element	C By Amino Acid	Rubans	Liens	•
• Off	C Solvent Arr	cessible (Solvent Exclude	d Cavities	by Finnie Fisia	Roquettes	Liens Hydrogène	•
Selection	H-Bonds	Г	SS Bonds		🗸 Antialias Display	Rives	Liens disulfure	•
			00 001100	- Notation		Trace	Structures	
; 🖌 Highlight all 🔲 Match c	ase					Fichier	Boîte englobante	
							Cellule unitaire	•

Access to Ligand explorer

Summary Sequence

Derived Data | Seq. Similarity |

3D Similarity | Literature | Biol. & Chem.

Methods Geometry Link

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



DOI:10.2210/pdb3i49/pdb NDB ID: NA0060

Further down on same page..

🗘 Ligand	Chemical Component					Hide
Identifier	Name	Formula	Binding Affinity (BindingDB 🗗)	Interaction V	/iew	
AKG 🔎	2-OXOGLUTARIC ACID	C5 H6 O5		ᡖ Ligand Ex	plorer	
FE2 🔎	FE (II) ION	Fe		ᡖ Ligand Ex	plorer	
‡ Modifie	ed Residues					Hide
Identifier	Name			Formula	Parent	Туре
ME6 🔎	[(2R,3S,5R)-5-(4-azanyl-3-methyl-2-o 2-yl]methyl dihydrogen phosphate	xo-pyrimidin-3-ium	-1-yl)-3-hydroxy-oxolan-	C10 H17 N3 O7 P	DC	dnaLinking



Ligand interactions



Very easy to use, excellent tools and links:



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



Ligand interactions





EMBL-EBI				Enter Text H	lere		Find Help Feedback		
Databases	Tools	Research	Training	Industry	About Us	Help	Site Index <u>ର</u> 🚑		
PD	-				Go t	o PDB code: 🕄	3i49 go 🔒 😗		
		🔍 Top paç	ge 🥥 Protein 🛛	DNA/RNA	🖲 Ligands 🛛 🛚 🖉	detals 🧕 Clefts			
			Oxidoreduct	ase/DNA			PDB Id 3149		
			Links to o	other datab	ases				
		$\overline{\mathcal{A}}$	Structure d	atabases					
~	A CARA	$\nabla \lambda$	PDBe	Protein D	ata Bank, Europ	e at the EBI			
51	Sol -	EN O	RCSB	Protein D	ata Bank at the F	RCSB			
20		N N	SRS	SRS at the	e EBI				
~		21	MMDB	MMDB en	try at the NCBI				
1				Jena Libr	ary of Biological	Macromolecule	es at the Fritz-Lipman		
1 2		nN		Institute					
2			<u>0CA</u>	OCA at the Weizmann Institute					
1	$S \prec \mathbf{N}$	9							
			Wikis						
	Ø.,		PDBWiki	dge base of biological					
	- 🦿		DUTIN	molecula	r structures				
			PROTEOPE	DIA Proteope	dia - collaborativ	e 3D encycloped	lia of proteins and other		
	Co muai 🦓	and the second se		molecule	5				
Conte	nts		Fold databa	200					
Protein	chain		CATH	CATH stri	ictural classificat	tion			
	200 a.a.		SCOP	SCOP str	ictural classifica	tion			
-DNA/RM	A		ESSP	ESSP stru	ctural alignment	ls			
-B			1.001	1001 040	otarar anginnon				
Ligand	s								
			Protein seq	uence					
Metals			PDBSWS	PDBSWS	mapping of PD	B protein chains	to SwissProt entries		
	=2								
Waters	 ×281		Secondary	structure					
weater a			HSSD	HSSD: Ho	mology derived	Secondary Struc	ture of Proteins		
			<u>1155F</u>	HOOF. HU	mology derived	occondary out			

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

Ac	Ivanced Search Interface	
ſ		
	Choose a Query Type:	Search by SMILES string
	Chemical Components	
	Chemical Name	Result Count
l	Chemical ID	
	InChI Descriptor	
1	SMILES / SMARTS	Add Casueh Cuitauis 🗛
	Molecular Weight (Chemical component)	Auu Search Chteria 🗸
	Chemical Formula Directions Affinites	
Г	Binding Attinity	
	Has Liganu(s) Has Modified Residue(s)	
Ma	Rinlaav	Clear All Parameters Submit Query
	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	
	Molecular Function Browser (GO) (opens popup) Methods Experimental Method	



Naming & finding chemical entities

EBI > Databases > Small I	Molecules > ChEB	I > Main						
adenosine (CHEBI:16335)			Search ChEBI here! Search ChEBI					
				Search for 🗰 only 🗹				
Main Automatic Xrefs								
NH ₂		ChEBI Name 🕜	adenosine					
N		ChEBI ID	CHEBI:16335					
	>	Definition 🕜	A ribonucleoside composed of a molecule of adenine attached to a ribofuranose moiety via a eta -	-N ⁹ -glycosidic bond.				
N .	ОН	<u>Stars</u> Ø	*** 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					
	∽он	See structure as: 🗹 Ima	ge 🗆 Applet					
	он	🔡 Download Molfile						
 Find compounds Find compounds 	which contain the which resemble whi	<u>his structure</u> e this structure						
more structures >>								
Formula 🕜			So	Irce				
C10H13N5O4			KE	GG COMPOUND				
Net Charge 🞯	0							
Average Mass 🚱	267.241	52						
InChl 🕜 💷	n Chl=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								
<u>SMILES</u> 🔮 🗎	Nc1ncnc	:2n(cnc12)[C@@H]10[C@	I(CO)[C@@H](O)[C@H]1O					



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



N3-methylthymine. AlkB has a substantially higher k(cat) and K(m) for m3C compared with m1A.

Redundancy in the Protein Data Bank

Cryst	
méth	

		As the single worldwide repository for macromolecular structures, the Protein Data Bank holds a body of data that contains considerable		Statistics		
	Summary Cryst meth	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	T r t j	The follow Ton-redur Dastclust dentity.		
1	Sequence Sequenc	mind:		Method		
	Entity #1	 Sequence similarity is defined on a chain basis, but results are returned on a structure basis. 		blast		
	Cluster Se	 Many structures in the PDB contain multiple protein chains, or even 		blast		
	100%	hybrids of DNA and protein chains.		blast		
	95%	 Sequence similarity is only assessed for protein chains. The relationship between sequence similarity and structure similarity is 		blast		
l	90%	complex. Users seeking structure similarity should refer to the options		blast		
	70%	available on the Structure Summary page under "External Links" (in the left hand navigation menu) in the "Structure Classification" section		blast		
l	50%	The primary purpose of this feature is to filter a list of likely highly similar		blast		
	40%	structures to provide one or more representatives. Results may differ from	l			
	30%	other so-called non-redundant sets (e.g. PDB_SELECT [Hobohm U., and				
	Entity #2	Sander C., <i>Protein Science</i> , <u>3</u> : 522-524, 1994]).		Notes on		
l	Cluster Se	Sequence ductoring in the DDB is performed via blastdust. Detailed				
	Docum	information for the clusters containing a given structure is available on the	E f	3last clust ollowing p		
0	Click here	structure, e.g.				

ing table shows the number of ndant sequences as determined by at several levels of sequence

)

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

Blast Clustering

tering is performed with the parameters (example 95%):

-b T -S 95 ·μ



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.

313Q.A (ch	13Q.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	30N7.D	Oxidoreductase, iron/ascorbate	4.22E-4	204.64	3,11	203	273	7	68
11	view	20PW.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



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





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/







Finding similar protein structures via the DALI database

Dali Datal	oase			Institute of Biotechnology
SERVICES & TOOLS	GROUP MEMBERS	NEWS & VACANCIES	RESEARCH	PUBLICATIONS
Dali struct	ural neighbo	ours		
The Dali Database is (PDB). The structural r search engine.	based on all-against-all neighbourhoods and alig	3D structure comparison nments are automatically	n of protein structures i / maintained and regular	n the Protein Data Bank ly updated using the Dali
 Please note that structural neight If you want to su 	at PDB structures releas bours of these proteins, iperimpose two particulai	ed after the last update you are advised to submi r structures, you can do it	e will not be in the data t the structure to the Dal : in the pairwise DaliLite s	basel If you wish to find i Server instead. erver.
			* Updat	Last Update: 11 July 2010 :e frequency: twice a year
	Quer	y structure not found	in PDB	
Enter PDB identifie (Keyword search for PDB ide Dali Database entries to an existing results p	er: 3i49 chain: ntifiers) are retrieved on demand age are much faster.	(optional) submit	: lear Ilts page may take up to	one minute. Return visits
Example Structural neighbours	of 1tu9, a globin-like prot	tein in bacteria. Tutorial		Swice Instit

Swiss Institute of Bioinformatics

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.

Reset Selection

Structural Alignment

Expand gaps 3D Superimposition (Jmol Applet)

Summary

	_		
-			
_			
_			
_			

	No:	Chain	Z	rmsd	lali	nres	%id PDB	Description	n								
	<u>1</u> :	<u>3149-A</u>	41.6	0.0	200	200	100 <u>PDB</u>	MOLECULE:	ALPI	HA-KET(OGLUT	ARATE	-DEPE	NDENT	DIO	XYGENASH	ALKB;
	<u>2</u> :	<u>3i3m-A</u>	40.2	0.2	200	200	100 <u>PDB</u>	MOLECULE:	ALPI	HA-KET(OGLUT	ARATE	-DEPE	NDENT	DIO	XYGENASH	ALKB;
	<u>3</u> :	<u>2fdf-A</u>	39.4	0.3	199	199	100 <u>PDB</u>	MOLECULE:	ALK	/LATED	DNA	REPAI	R PRC)TE IN	ALKB	;	
	<u>4</u> :	<u>3bkz-A</u>	39.1	0.3	200	201	100 <u>PDB</u>	MOLECULE:	ALPI	HA-KET(OGLUT	ARATE	-DEPE	NDENT	DIO	XYGENASH	ALKB;
\Box	<u>5</u> :	<u>3bie-A</u>	39.0	0.4	200	202	100 <u>PDB</u>	MOLECULE:	ALPI	HA-KET(OGLUT	ARATE	-DEPE	NDENT	DIO	XYGENASH	ALKB;
\Box	<u>6</u> :	<u>3bi3-A</u>	38.8	0.4	199	201	99 <u>PDB</u>	MOLECULE:	ALPI	HA-KET(OGLUT	ARATE	-DEPE	NDENT	DIO	XYGENASI	ALKB;
	<u>7</u> :	<u>3i3q-B</u>	38.6	0.5	200	203	100 <u>PDB</u>	MOLECULE:	ALPI	HA-KET(OGLUT	ARATE	-DEPE	NDENT	DIO	XYGENASI	ALKB;
	<u>8</u> :	<u>3khc-B</u>	38.3	0.4	200	207	100 <u>PDB</u>	MOLECULE:	ALPI	HA-KET(OGLUT	ARATE	-DEPE	NDENT	DIO	XYGENASI	ALKB;
	<u>9</u> :	<u>3i3q-A</u>	38.2	0.7	200	203	100 <u>PDB</u>	MOLECULE:	ALPI	HA-KET(OGLUT	ARATE	-DEPE	NDENT	DIO	XYGENASI	ALKB;
	<u>10</u> :	<u>3khc-A</u>	37.5	0.6	200	207	100 <u>PDB</u>	MOLECULE:	ALPI	HA-KET(OGLUT	ARATE	-DEPE	NDENT	DIO	XYGENASI	ALKB;
	<u>11</u> :	<u>2fdi-A</u>	36.9	0.7	199	199	100 <u>PDB</u>	MOLECULE:	ALK	/LATED	DNA	REPAI	R PRC	TEIN	ALKB	;	
	<u>12</u> :	<u>3khb-B</u>	36.9	0.5	200	206	100 <u>PDB</u>	MOLECULE:	ALPI	HA-KET(OGLUI	ARATE	-DEPE	NDENT	DIO	XYGENASH	ALKB;
	<u>13</u> :	<u>2fdk-A</u>	36.7	0.7	198	198	100 <u>PDB</u>	MOLECULE:	ALK	/LATED	DNA	REPAI	R PRC)TE IN	ALKB	;	
	<u>14</u> :	<u>3khb-A</u>	36.7	0.7	200	203	100 <u>PDB</u>	MOLECULE:	ALPI	HA-KET(OGLUI	ARATE	-DEPE	NDENT	DIO	XYGENASI	ALKB;
	<u>15</u> :	2fdh-A	36.6	0.8	199	199	100 <u>PDB</u>	MOLECULE:	ALK	/LATED	DNA	REPAI	R PRC)TE IN	ALKB	;	
	<u>16</u> :	<u>3120-A</u>	36.5	0.8	199	199	100 <u>PDB</u>	MOLECULE:	ALPI	HA-KET(OGLUT	ARATE	-DEPE	NDENT	DIO	XYGENASI	ALKB;
	<u>17</u> :	<u>2fd8-A</u>	36.5	0.8	199	199	100 <u>PDB</u>	MOLECULE:	ALK	/LATED	DNA	REPAI	R PRC	TEIN	ALKB	;	
	18:	2fdq-A	36.1	0.8	199	199	100 PDB	MOLECULE:	ALK	/LATED	DNA	REPAI	R PRC	TEIN	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 <u>Clear labels</u>

First structure's backbone: 💿 CA trace 🛛 Cartoon Rockets. Matched structures' backbone

Colour backbone of first structure: I monochrome C rainbow C 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/I: coil). The most frequent amino acid type is coloured in each column.

0001 3i49A 0002 2iuwA	: LAAGAVILRRFAFNA. YEIRV <mark>C</mark> LYPGFVDVE.	: AEQLIRDINI ADWILEQLC(: DVASQSPFRQM DVPWKQR	: VTPGGYTMSV TGIITYQOPI	 / amtnc ghlgu Rltawyg- e lf	: TTHRQGYLY: YTY:	: SPIDPQTNKPU SRITEH	: PAMPOSFHNI NPWHPVLRTI	: Coraataagy KnrieenT	100 PDFQPDACLIN GHT-FNSLXCN	: RYAPGAKLSLH LYREKDSVDWH
0001 3i49A 0002 2iuwA	LLLLEEEELLLLLLL EEEEEEELLLLLHH	: нннннннни ннннннни	: IHHHHLLLLLL IHLLLLLL	: LLLLLLLLI EEELEEELLI	 LEEEEELLLEE LEEEEEE-LLL	: EEELLEEEE LLL	ELLLLLLLL LHHHLI	: .LLLLHHHHHH .LLLLHHHHHH	: HHHHHHHLLL HHHHHHHHH	100 LLLLLEEEEE LLL-LLEEEEE	: EELLLLEEEE EELLLLEEEE

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/?

Portal version: 2012/11/27 (revision:6220) Data Release: 2012/11/22 (statistics)

The current release is based on UniProt release 2012_10 and consists of 19.4 million comparative protein models for 3.9 million distinct UniProt sequences.

Example: alkB fromC.crescentus

PSI | The Protein Model Portal



models menu	PMP	Que	ery Resu	ult:					
PMP home	Summ	ary: 🛛							
advanced search	Your que	ry was:	B8GWW6						
interactive modeling	1						22	20	
quality estimation									
CAMEO									
news and events	Domain a	nnotati	ion: [Inter l	Pro 1 ×					
documentation	Colors:	Query	Sequence	Struct	ures Models	Domains	Sele	ection	
related tools	Start ar	alysis o	of structural v	ariability	🕐 🕐 Desele	ect All			
about PMP	B8GWW	6 * ALK hor CB	(B_CAUCN ; molog; EC=: 15N).	Alpha-k 1.14.11.	etoglutarate-c -; Caulobacter	crescentus	dioxygei (strain	nase Al NA100	кв 0 /
contact us			·						
	Mode	s:							
psi sbkb menu 🗭	Model	Rel.	Provider	Type	Templates	%Seq id	from	to	Sel.
home	[Show]	000	NESG	тс	2fdfA 🔺	39%	24	217	
structural biology	[Show]	e e e	NESG	TC	2fdiA 🔺	39%	24	217	
update	[Show]	000	NESG	тс	2fdgA *	39%	24	217	
targets	[Show]	000	NESG	тс	2fd8A *	39%	24	217	
protein structures,	[Show]	000	NESG	тс	2fdhA 🔻	39%	24	217	
sequences, and function	[Show]	000	NESG	тс	2fdkA 🔹	39%	24	218	IkB 0 / Sel.
homology models	[Show]	000	NESG	тс	2fdjA 🔺	38%	23	218	



NESG MODELS DATABASE

	The simila	rity search retr	ieved the following	modeled	proteins fror	n the NESG da	tabase
Target ID	Template	Model Quality(pG)	Target/Template Sequence Identity	E Value	Target Coverage	Species	Target Description
<u>AAK21997.1</u>	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: <u>2FDF</u> target: <u>AAK21997.1</u> download: <u>model</u> Side chain optimization Adv.Refinement Function Analysis
						astex	



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





models menu	PMP Interacti∨e Modeling	
PMP home	Name:	
advanced search	Request Title:	
interactive modeling	Email:	
quality estimation	Amino Acid Sequence:	
related tools		
news and events		
documentation		
about PMP	□ ModWeb	
contact us	Server O By checking this box, I assert that I am part of an academic Policy: institution (not a government research lab such as the NIH, or	
psi sbkb menu 😥	I have a MODELLER access key:	
home	Пиат	
structural biology update		
targets	Server I am a non-profit/academic user and this server will be used Policy: solely for educational purposes or for basic research intended to advance scientific knowledge.	
protein structures, sequences, and function	SWISS-MODEL	
homology models	Server Bolicy: Usage of SWISS-MODEL Server and Workspace are free of charge.	
methods	roncy.	
nublications	submit query	SIB
		Swiss Institu Bioinformatir

Ab initio and comparative protein structure prediction http://robetta.bakerlab.org/



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



Protein 3D-structure analysis

and now it is up to you: Practicals

