EXPERIMENT 4:
PROTEIN DATA BASES, FINDING AND VIEWING

Dr. Zhiyi Wei
SUSTC

Web based databases of structural information

- **PDB** Protein Data Bank
- SCOP The Structure Classification of Proteins
- CATH Class, Architecture, Topology and Homologous superfamily

PDB (Protein Data Bank)

- http://www.pdb.org
- The single worldwide repository for the processing and distribution of 3-D structure data of large molecules of proteins and nucleic acids
- Goal: to provide a portal to general information of interest to users of macromolecular structure data as well as specific details on a single structure
- Technology used in deposted structures
  - X-ray crystallography
  - Nuclear Magnetic Resonance (NMR)
  - Cryo-electron microscopy

RCSB PDB Deposit Search Visualize Analyze Download Learn More MyPDB Login

RCSB PDB PROTEIN DATA BANK

An Information Portal to 115306 Biological Macromolecular Structures

Search by PDB ID, author, macromolecule, sequence, or ligands Go

Advanced Search | Browse by Annotations | Search History (2) | Previous Results (896)

PDB-101 wwPDB EMDataBank NDB StructuralBiology Knowledgebase

Welcome
Deposit
Search
Visualize
Analyze
Download
Learn

A Structural View of Biology

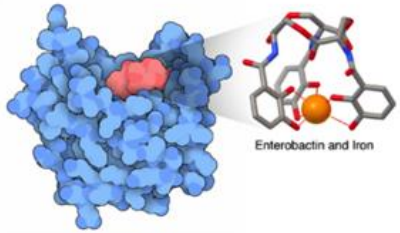
This resource is powered by the Protein Data Bank archive-information about the 3D shapes of proteins, nucleic acids, and complex assemblies that helps students and researchers understand all aspects of biomedicine and agriculture, from protein synthesis to health and disease.

As a member of the wwPDB, the RCSB PDB curates and annotates PDB data.

The RCSB PDB builds upon the data by creating tools and resources for research and education in molecular biology, structural biology, computational biology, and beyond.

Take an Interactive Tour of the PDB

January Molecule of the Month



Siderocalin

Latest Entries

As of Tuesday Jan 19

4V1H PDB Entry

Crystal structure of a mycobacterial ATP synthase rotor ring in complex with Iodo-Bedaquiline

View in 3D

New Features

October 2015 Release

Redesigned Structure Summary Page
New Organization. Improved Layout. Clean. Usable. Simple.

Improved Literature Tab

Better Support for Mobile Browsing

Redesigned Ligand Summary Page

September 2015 Release

Validation Track on Protein Feature View
Mapping validation annotations to sequence

News

Publications

Winter Newsletter Published
Read about 2015 statistics and highlights, the new wwPDB Deposition Tool, RCSB PDB's 2016 protein-drug calendar, and more. The Education Corner describes *Using 815,527 Crystal Structures to Teach Chemistry: The Cambridge Structural Database.* » 01/19/16

wwPDB Deposition & Annotation System Now Available for NMR and 3DEM Structures

» 01/11/16

Time-stamped Copies of the PDB Archive » 01/06/16

PDB ID

- Each structure in the PDB is represented by a 4 character identifier:

  [0-9][a-z,0-9] [a-z,0-9][a-z,0-9]

- Publication papers

PDB statistics

Exp.Method	Proteins	Nucleic Acids	Protein/NA Complexes	Other	Total
X-RAY	97776	1712	4938	4	104430
NMR	9941	1140	231	8	11320
ELECTRON MICROSCOPY	710	29	248	0	987
HYBRID	85	3	2	1	91
other	171	4	6	13	194
Total	108683	2888	5425	26	117022

- Growth of Released Structures Per Year (http://www.rcsb.org/pdb/statistics/contentGrowthChart.do?content=total&seqid=100)

Content of the Data in the PDB

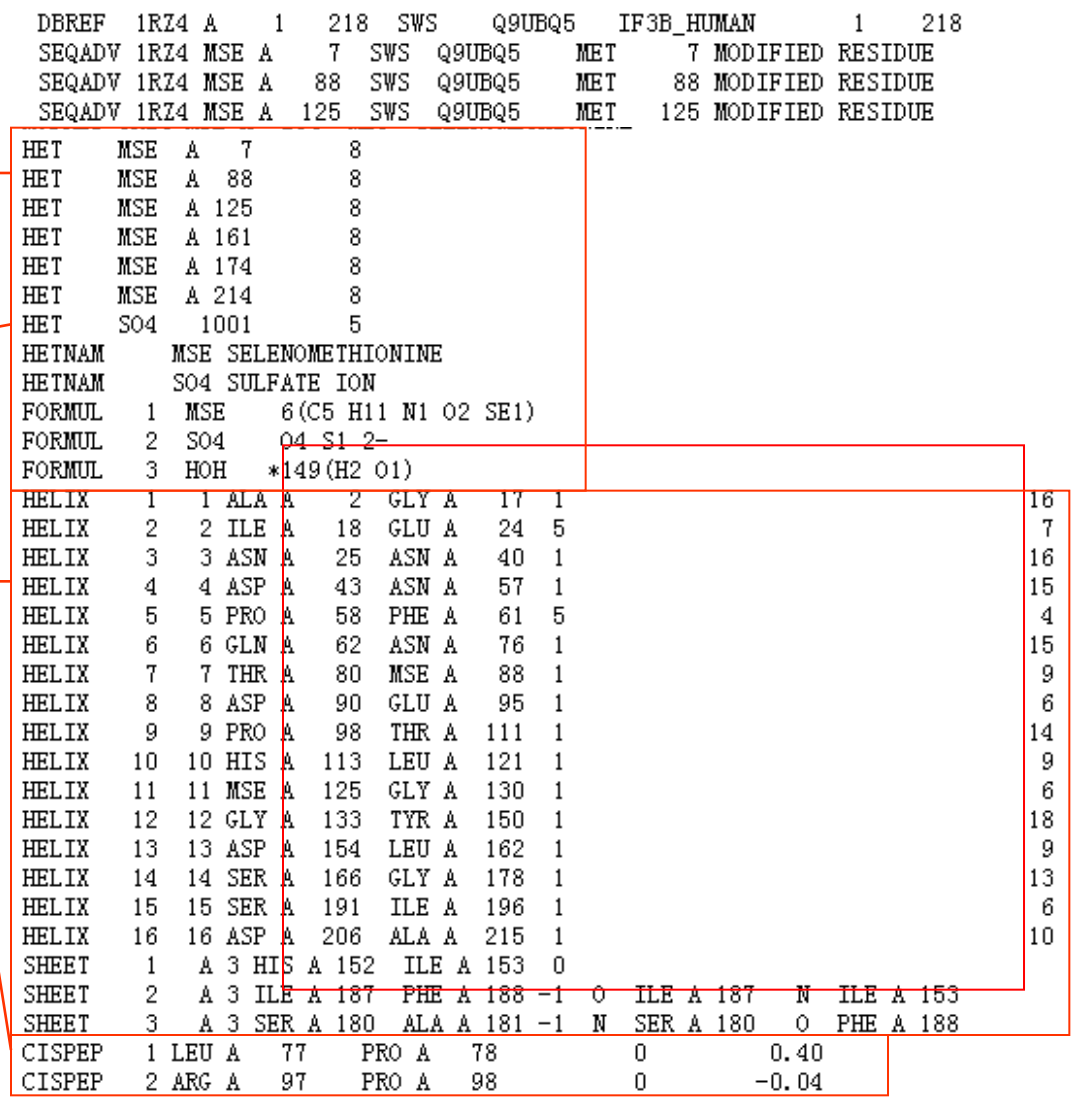
- Content of All Depositions (X-ray and NMR)
  - Source: Specifications such as genus, species, strain, or variant of gene (cloned or synthetic); expression vector and host, or description of method of chemical synthesis
  - Squence: Full sequence of all macromolecular components
  - Chemical structure of cofactors and prosthetic groups
  - Names of all components of the structure
  - Qualitative description of the characteristics of the structure
  - Literature citations for the structure submitted
  - Three-dimensional coordinates
- Additional Items for X-ray Structure Determinations
- Additional Items for NMR Structure Determinations

PDB format files

- **1. Title Section**
  - HEADER
  - OBSLTE
  - TITLE
  - CAVEAT
  - COMPND
  - SOURCE
  - KEYWDS
  - EXPDTA
  - AUTHOR
  - REVDAT
  - SPRSDE
  - JRNL
  - REMARK
  - REMARK 1
  - REMARK 2
  - REMARK 3
  - REMARK 4 - 999

```
HEADER    BIOSYNTHETIC PROTEIN                    23-DEC-03   1RZ4
TITLE     CRYSTAL STRUCTURE OF HUMAN EIF3K
COMPND    MOL_ID: 1;
COMPND   2 MOLECULE: EUKARYOTIC TRANSLATION INITIATION FACTOR 3
COMPND   3 SUBUNIT 11;
COMPND   4 CHAIN: A;
COMPND   5 SYNONYM: EUKARYOTIC INITIATION FACTOR 3 SUBUNIT K, EIF3K;
COMPND   6 ENGINEERED: YES
SOURCE    MOL_ID: 1;
SOURCE   2 ORGANISM_SCIENTIFIC: HOMO SAPIENS;
SOURCE   3 ORGANISM_COMMON: HUMAN;
SOURCE   4 EXPRESSION_SYSTEM: ESCHERICHIA COLI;
SOURCE   5 EXPRESSION_SYSTEM_COMMON: BACTERIA
KEYWDS    HEAT ANALOGOUS MOTIF, WINGED-HELIX
EXPDTA    X-RAY DIFFRACTION
AUTHOR    Z.WEI,P.ZHANG,Z.ZHOU,W.GONG
REVDAT   1   21-SEP-04 1RZ4    0
JRNL        AUTH   Z.WEI,P.ZHANG,Z.ZHOU,Z.CHENG,M.WAN,W.GONG
JRNL        TITL   CRYSTAL STRUCTURE OF HUMAN EIF3K, THE FIRST
JRNL        TITL 2 STRUCTURE OF EIF3 SUBUNITS
JRNL        REF    J.BIOL.CHEM.                  V. 279 34983 2004
JRNL        REFN   ASTM JBCHA3  US ISSN 0021-9258
REMARK   1
REMARK   2
REMARK   2 RESOLUTION. 2.10 ANGSTROMS.
REMARK   3
REMARK   3 REFINEMENT.
REMARK   3   PROGRAM     : CNS 1.0
REMARK   3   AUTHORS     : BRUNGER,ADAMS,CLORE,DELANO,GROS,GROSSE-
REMARK   3               : KUNSTLEVE,JIANG,KUSZEWSKI,NILGES,  PANNU,
REMARK   3               : READ,RICE,SIMONSON,WARREN
REMARK   3
REMARK   3  REFINEMENT TARGET : NULL
REMARK   3
REMARK   3  DATA USED IN REFINEMENT.
REMARK   3   RESOLUTION RANGE HIGH (ANGSTROMS) : 2.10
REMARK   3   RESOLUTION RANGE LOW  (ANGSTROMS) : 30.00
```

- **2. Primary Structure Section**
  - DBREF
  - SEQADV
  - SEQRES
  - MODRES
- **3. Heterogen Section**
  - HET
  - HETNAM
  - HETSYN
  - FORMUL
- **4. Secondary Structure Section**
  - HELIX
  - SHEET
  - TURN
- **6. Connectivity Annotation Section**
  - SSBOND
  - LINK
  - HYDBND
  - SLTBRG
  - CISPEP
- **7. Miscellaneous Features Section**
  - SITE



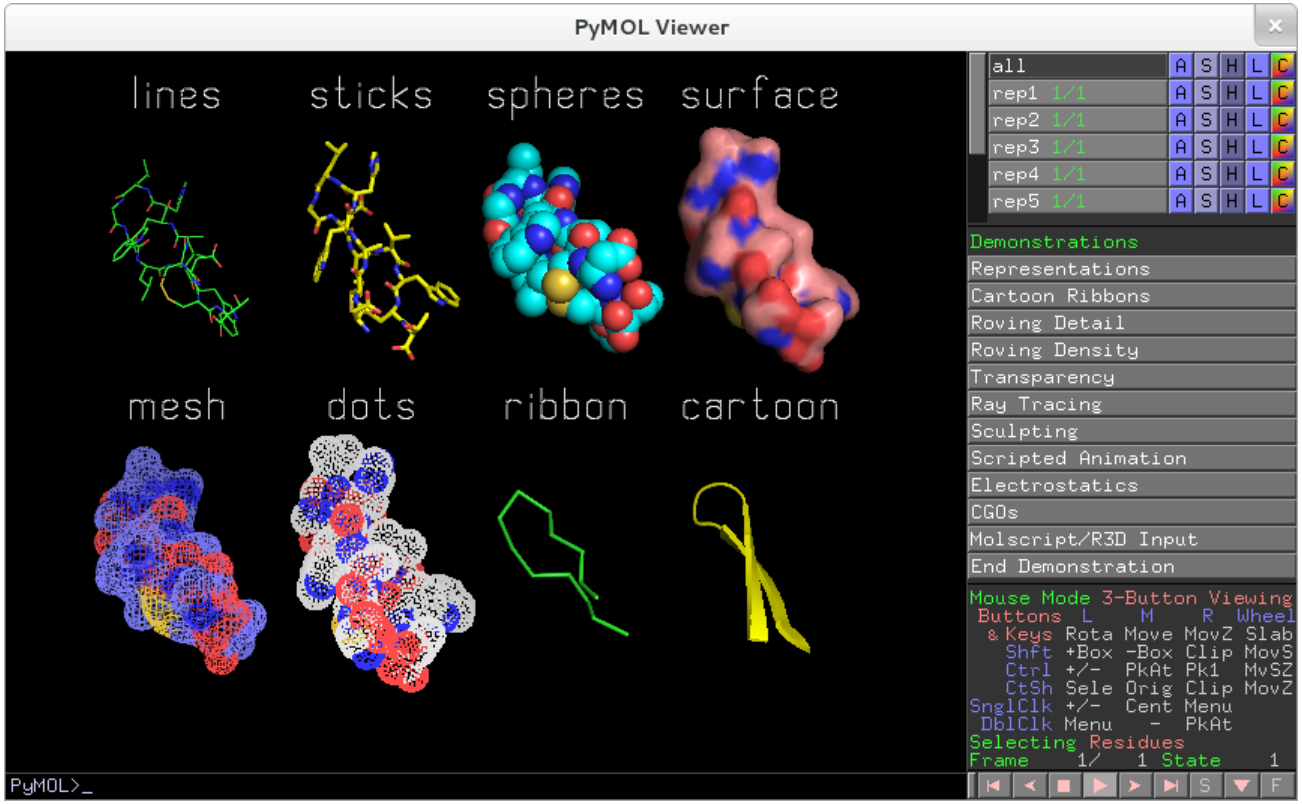
```
DBREF  1RZ4 A    1   218  SWS    Q9UBQ5   IF3B_HUMAN       1    218
SEQADV 1RZ4 MSE A    7  SWS  Q9UBQ5    MET     7 MODIFIED RESIDUE
SEQADV 1RZ4 MSE A   88  SWS  Q9UBQ5    MET    88 MODIFIED RESIDUE
SEQADV 1RZ4 MSE A  125  SWS  Q9UBQ5    MET   125 MODIFIED RESIDUE
HET    MSE  A   7       8
HET    MSE  A  88       8
HET    MSE  A 125       8
HET    MSE  A 161       8
HET    MSE  A 174       8
HET    MSE  A 214       8
HET    SO4    1001      5
HETNAM     MSE SELENOMETHIONINE
HETNAM     SO4 SULFATE ION
FORMUL   1  MSE    6(C5 H11 N1 O2 SE1)
FORMUL   2  SO4    O4 S1 2-
FORMUL   3  HOH   *149(H2 O1)
HELIX    1   1 ALA A    2  GLY A   17  1                                  16
HELIX    2   2 ILE A   18  GLU A   24  5                                   7
HELIX    3   3 ASN A   25  ASN A   40  1                                  16
HELIX    4   4 ASP A   43  ASN A   57  1                                  15
HELIX    5   5 PRO A   58  PHE A   61  5                                   4
HELIX    6   6 GLN A   62  ASN A   76  1                                  15
HELIX    7   7 THR A   80  MSE A   88  1                                   9
HELIX    8   8 ASP A   90  GLU A   95  1                                   6
HELIX    9   9 PRO A   98  THR A  111  1                                  14
HELIX   10  10 HIS A  113  LEU A  121  1                                   9
HELIX   11  11 MSE A  125  GLY A  130  1                                   6
HELIX   12  12 GLY A  133  TYR A  150  1                                  18
HELIX   13  13 ASP A  154  LEU A  162  1                                   9
HELIX   14  14 SER A  166  GLY A  178  1                                  13
HELIX   15  15 SER A  191  ILE A  196  1                                   6
HELIX   16  16 ASP A  206  ALA A  215  1                                  10
SHEET    1   A 3 HIS A 152  ILE A 153  0
SHEET    2   A 3 ILE A 187  PHE A 188 -1  O  ILE A 187   N  ILE A 153
SHEET    3   A 3 SER A 180  ALA A 181 -1  N  SER A 180   O  PHE A 188
CISPEP   1 LEU A   77    PRO A   78          0         0.40
CISPEP   2 ARG A   97    PRO A   98          0        -0.04
```

```
CRYST1   83.074   44.671   55.524  90.00  90.00  90.00 P 21 21 2     4
HETATM 1854  O   HOH   135      27.950   3.588  10.081  1.00 40.08           O
ATOM   1422  N   TYR A 177      35.015   2.061  36.397  1.00 24.50           N
ATOM   1423  CA  TYR A 177      36.442   2.138  36.087  1.00 25.21           C
ATOM   1424  C   TYR A 177      37.344   1.431  37.098  1.00 25.27           C
ATOM   1425  O   TYR A 177      38.568   1.568  37.057  1.00 25.36           O
ATOM   1426  CB  TYR A 177      36.873   3.601  35.927  1.00 25.16           C
ATOM   1427  CG  TYR A 177      36.239   4.280  34.733  1.00 25.73           C
ATOM   1428  CD1 TYR A 177      34.922   4.741  34.782  1.00 24.85           C
ATOM   1429  CD2 TYR A 177      36.937   4.411  33.531  1.00 26.86           C
ATOM   1430  CE1 TYR A 177      34.311   5.310  33.660  1.00 23.22           C
ATOM   1431  CE2 TYR A 177      36.335   4.978  32.403  1.00 26.09           C
ATOM   1432  CZ  TYR A 177      35.023   5.421  32.476  1.00 24.89           C
ATOM   1433  OH  TYR A 177      34.420   5.956  31.358  1.00 24.64           O
ATOM   1434  N   GLY A 178      36.736   0.668  37.998  1.00 24.77           N
ATOM   1435  CA  GLY A 178      37.504  -0.060  38.990  1.00 25.13           C
ATOM   1436  C   GLY A 178      38.193   0.776  40.055  1.00 25.06           C
ATOM   1437  O   GLY A 178      39.251   0.389  40.548  1.00 24.04           O
MASTER      211    0    1   10    3    0    0    6 1861    1   83   16
END
```

Other protein structure based database

- Protein-Ligand Interaction: ReliBase
- Metalloprotein Database & Browser (MDB)
- Enzyme Structures Database
- Membrane Proteins of Known Structure

Viewing PDB with PyMOL



Tasks

1. Find the PDB id for human hemoglobin
2. Download the PDB coordinate file and check the detailed information in the file
   - How many protein chains?
   - Which compounds in structure?
   - Which residues are missed in the coordinate?
3. Edit the coordinate file to make a new coordinate file that contain only one β subunit of hemoglobin (β-globin)
4. Visualize the β-globin structure in different viewing modes
   - Save picture for each mode

Lab report format

- Title
- Your name and student No.
- Introduction
- Methods
- Results
- Conclusions