GOVT. DEGREE COLLECGE FOR WOMEN GAJWEL, SIDDIPET

Department of Biotechnology
Certificate course- Protein Engineering and Basics of Proteomics

2021-22

Name of the Certificate course Conducted: Protein Engineering and Basics of Proteomics

Name of the Course coordinator : Mrs.K.Geethanjali

No. of Students enrolled : 14

Date of commencement of classes : 15/02/2022

Date of ending of the course : 15/04/2022

No. of classes conducted : 32

No. of students appeared for final exam : 14

No. of students passed final exam : 14

Over all participation of the students : Satisfactory

GOVT. DEGREE COLLECGE FOR WOMEN GAJWEL, SIDDIPET

Department of Biotechnology
Certificate course- Protein Engineering and Basics of Proteomics

2020-21

No. of Hrs. 30

Course Objectives:

- To enable the students to retrieve protein data from data bases
- To impart the knowledge of comparison of sequences and structure prediction
- To analyze the proteomics and interpret using UNIPROT and other secondary structure prediction tools.
- To give the basic structure of protein, levels of organization of protein help the student to search the sequence data form available data bases.
- To give them hands on training on three dimensional structure prediction and analysis of proteins using soft ware is a rapidly growing technology for designing drugs, creates lot of enthusiasm among the students to take up career in Bioinformatics.

Course contents:

UNIT 1

- Introduction to SDS PAGE and 2-D gel Electrophoresis
- Nucleic acid sequence databases: GenBank, EMBL, DDBJ
- Protein sequence databases: UNIPROT , SWISS-PROT, TrEMBL, PIR, PDB
- Genome Databases at NCBI, EBI, TIGR, SANGER

UNIT 2

- Various file formats for bio-molecular sequences: GENBANK, FASTA,PDB
- Basic concepts of sequence similarity, identity and homology, definitions of homologues,
- orthologues, paralogues.
- Scoring matrices: basic concept of a scoring matrix, PAM and BLOSUM series.
- Sequence-based Database Searches: what are sequence-based database searches, BLAST and
- FASTA algorithms, various versions of basic BLAST and FASTA.

UNIT 3

- Use of pair wise alignments and Multiple sequence alignment for analysi of Nucleic acid and protein sequences and interpretation of results.
- Exploring UNIPROT for proteomics
- EXPASY-PROTPARAM, TMHMM, SOPMA, Swiss modelling

Course Outcomes:

- The basic structure of protein, levels of organization of protein helped the student to retrieve the sequence data form available data bases.
- The three dimensional structure prediction and analysis of proteins using soft ware is a rapidly growing technology for designing drugs, creates lot of enthusiasm among the students to take up career in Bioinformatics
- Exploring the UNIPROT and other secondary structure prediction are important in understanding the differential protein expression in several disease conditions

References:

1. Introduction to Bioinformatics by Aurther M lesk
2. Developing Bioinformatics Computer Skills By: Cynthia Gibas, Per Jambeck
3. Tutorials at NCBI https://www.ncbi.nlm.nih.gov/
4. PDB 101 https://pdb101.rcsb.org/
5. Uniprot http://www.uniprot.org/

List of students Enrolled:

S.No	Regd.No	Name Of The Student	No. of classes attended
1	602920489001	CH.SOWMYA	30
2	602920489002	D.SAI DEEPTHI	30
3	602920489003	G.USHA	30
4	602020489004	ANUPAMA	30
5	602920489005	K.SANDHYA	30
6	602920572003	D.KALYANI	30
7	602920572004	G.HARSHITHA	30
8	602920572005	K.POOJA	30
9	602920572006	NAGASRESHTA	30
10	602920572007	SAILAJA	30
11	602920572008	N.VINODHA	30
12	602920572009	RAMYA	30
13	602920572011	S.CHOOSEN GRACE	30
14	602920572012	PAVANI	30

Students Attendance

S.NO	Date	15/2	16/2	17/2	18/2	19/2	4/4	5/4	6/4	7/4	8/4	9/4	10/4	11/4	12/4	15/4
	No. of Hours	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**
1.	CH.SOWMYA	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**
2.	D.SAI DEEPTHI	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**
3.	G.USHA	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**
4.	ANUPAMA	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**
5.	K.SANDHYA	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**
6.	D.KALYANI	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**
7.	G.HARSHITHA	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**
8.	K.POOJA	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**
9.	NAGASRESHTA	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**
10.	SAILAJA	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**
11.	N.VINODHA	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**
12.	RAMYA	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**
13.	S.CHOOSEN GRACE	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**
14.	PAVANI	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**

Students Evaluation

II Year B.SC Biotechnology

S.No	Regd.No	Name Of The Student	Max. Marks (20)
1	602920489001	CH.SOWMYA	18
2	602920489002	D.SAI DEEPTHI	16
3	602920489003	G.USHA	18
4	602020489004	ANUPAMA	19
5	602920489005	K.SANDHYA	20
6	602920572003	D.KALYANI	17
7	602920572004	G.HARSHITHA	20
8	602920572005	K.POOJA	19
9	602920572006	NAGASRESHTA	17
10	602920572007	SAILAJA	19
11	602920572008	N.VINODHA	18
12	602920572009	RAMYA	17
13	602920572011	S.CHOOSEN GRACE	19
14	602920572012	PAVANI	19

GOVERNMENT COLLEGE FOR WOMEN, GAJWEL

Department of Biotechnology
CERTIFICATE COURSE

SUBJECT: BASICS OF BIOINFORMATICS MAX.MARKS: 20
TIME: 30 MIN

CHOOSE CORRECT ANSWERS TO THE FOLLOWING (20 X 1= 20 MARKS)

1. The term used to refer something 'performed on computer or computer simulation (D)
   a. Dry lab
   b. Web lab
   c. Invitro
   d. Insilico
2. Application of bioinformatics include (D)
   a. Data storage management
   b. Drug design
   c. Understanding relationship between organisms
   d. All of the above.
3. Proteomics is the study of (C)
   a. Set of proteins
   b. Set of proteins in a specific region of the cell
   c. Entire set of expressed protein in a cell
   d. None of these
4. The tool for identification of motifs (A)
   a. Pfam
   b. Pattern hunter
   c. PROSPECT
   d. BLAST
5. Deposition of cDNA into inert structure is called (D)
   a. DNA fingerprinting
   b. DNA polymerase
   c. DNA probe
   d. DNA microarray
6. Human genome contains about (B)
   a. 2.5billion basepairs
   b. 3.3 billion basepairs
   c. 4 billion basepairs
   d. 5 billion basepairs
7. Identification of drugs through genomic study (D)
   a. Genomics
   b. Chemiinformatics
   c. Pharmacogenomics
   d. Pharmacogenetics

8. Analysing or comparing entire genome of species (B)
   a. Bioinformatics
   b. Genomics
   c. Proteomics
   d. Pharmacogenomics

9. Which among the following is a protein structure data base (C)
   a. ddbj
   b. TrEMBL
   c. PDB
   d. SCOP

10. One of the following tools compares the DNA sequences with EST database and protein database (A)
   a. AAT
   b. GRAIL
   c. Genscan
   d. None

11. Who has developed first data base (B)

   A. Richard Durbin
   B. Dayhoff
   C. Michael j.Dunn
   D. Pearson

12. one of the following is a primary sequence database. (B)

   a) SCOP
   b) DDBJ
   c) PDB
   d) Carb Bank

13. Clustal W is a tool for (A)

   a) Multiple sequence alignment
   b) Pair wise sequence alignment
   c) a & b
   d) None

14. Two homologous sequences are direct descendents of a sequence in the common ancestor called (B)

   a) Paralogous
   b) orthologous
   c) a & b
   d) heterogonous

15 .Indel is (A)

a) An insertion or deletion in sequence alignment
b) Only insertion
c) Only deletion
d) Mismatch.

16. One of the following is the protein classification database (D)

a) CATH
b)TrEMBL
c) SWISS-prot
d)SCOP

17. BLOSSUM matrices are used for (D)

a.) Multiple sequence alignment
b.)Pair wise alignment
c.)Phylogenetic tree construction
d.)All of the above

18. BLAST X program is used for (C)

a) translate protein sequence
b) translate DNA databse
c) translate input sequence
d) none of these

19. Local alignments are more used when (A)

a) There are totally similar and equal length sequences
b) Dissimilar sequences are suspected to contain regions of similarity
c) Similar sequence motif with larger sequence context
d) Partially similar, different length and conserved region containing sequences

20. Each record in a database is called (A)

a) Entry
b)File
c) Record
d) Ticket

GOVT. DEGREE COLLEGE (W) GAJWEL, (SIDDIPET. DT.)
TELANGANA, INDIA

Certificate of Appreciation

This is to certify that Ms.D.Sai Deepthi, B,Sc BT.Z.C, II Yrs has successfully completed the certificate course on "Protein Engineering and Basics of Proteomics" offered by Dept of Biotechnology for the academic year 2021-22.

Mrs.K.Geethanjali
Asst. Professor in Biotechnology

Dr.P.V Uma Sasi
Principal

www.pngborders.org

DEPARTMENT OF Biotechnology

Certificate course in Protein structure prediction using Bioinformatics tools - Students work sheet

2021-22

```
>sp|Q15054|DPOD3_HUMAN DNA polymerase delta subunit 3 OS=Homo sapiens
GN=POLD3 PE=1 SV=2
MADQLYLENIDEFVTDQNKIVTYKWLSYTLGVHVNQAKQMLYDYVERKRKENSGAQLHVT
YLVSGSLIQNGHSCHKVAVVREDKLEAVKSKLAVTASIHVYSIQKAMLKDSGPLFNTDYD
ILKSNLQNCSKFSAIQCAAAVPRAPAESSSSSKKFEQSHLHMSSETQANNELTTNGHGPP
ASKQVSQQPKGIMGMFASKAAAKTQETNKETKTEAKEVTNASAAGNKAPGKGNMMSNFFG
KAAMNKFKVNLDSEQAVKEEKIVEQPTVSVTEPKLATPAGLKKSSKKAEPVKVLQKEKKR
GKRVALSDDETKETENMRKKRRRIKLPESDSSEDEVFPDSPGAYEAESPSPPPPPSPPLE
PVPKTEPEPPSVKSSSGENKRKRKRVLKSKTYLDGEGCIVTEKVYESESCTDSEEELNMK
TSSVHRPPAMTVKKEPREERKGPKKGTAALGKANRQVSITGFFQRK
```

ProtParam

Home | **Contact**

ProtParam

User-provided sequence:

```
        10         20         30         40         50         60
MLKDSGPLFN TDYDILKSNL QNCSKFSAIQ CAAAVPRAPA ESSSSSKKFE QSHLHMSSET

        70         80         90        100        110        120
QANNELTTNG HGPPASKQVS QQPKGIMGMF ASKAAAKTQE TNKETKTEAK EVTNASAAGN

       130        140        150        160        170        180
KAPGKGNMMS NFFGKAAMNK FKVNLDSEQA VKEEKIVEQP TVSVTEPKLA TPAGLKKSSK

       190        200        210        220        230        240
KAEPVKVLQK EKKRGKRVAL SDDETKETEN MRKKRRRIKL PESDSSEDEV FPDSPGAYEA
```

```
       250        260        270        280        290        300
ESPSPPPPPS PPLEPVPKTE PEPPSVKSSS GENKRKRKRV LKSKTYLDGE GCIVTEKVYE

       310        320        330        340        350        360
SESCTDSEEE LNMKTSSVHR PPAMTVKKEP REERKGPKKG TAALGKANRQ VSITGFFQRK
```

References and documentation are available.

Number of amino acids: 360

Molecular weight: 39320.47

Theoretical pI: 9.47

Amino acid composition: CSV format

Ala (A)	30	8.3%
Arg (R)	15	4.2%
Asn (N)	17	4.7%
Asp (D)	11	3.1%
Cys (C)	4	1.1%
Gln (Q)	13	3.6%
Glu (E)	37	10.3%
Gly (G)	19	5.3%
His (H)	4	1.1%
Ile (I)	7	1.9%
Leu (L)	17	4.7%
Lys (K)	49	13.6%
Met (M)	10	2.8%
Phe (F)	10	2.8%
Pro (P)	31	8.6%
Ser (S)	40	11.1%
Thr (T)	22	6.1%
Trp (W)	0	0.0%
Tyr (Y)	4	1.1%
Val (V)	20	5.6%
Pyl (O)	0	0.0%
Sec (U)	0	0.0%
(B)	0	0.0%
(Z)	0	0.0%
(X)	0	0.0%

Total number of negatively charged residues (Asp + Glu): 48
Total number of positively charged residues (Arg + Lys): 64

Atomic composition:

Carbon	C	1693
Hydrogen	H	2776
Nitrogen	N	492
Oxygen	O	553
Sulfur	S	14

Formula: $C_{1693}H_{2776}N_{492}O_{553}S_{14}$
Total number of atoms: 5528

Extinction coefficients:

This protein does not contain any Trp residues. Experience shows that this could result in more than 10% error in the computed extinction coefficient.

Extinction coefficients are in units of M^{-1} cm^{-1}, at 280 nm measured in water.

Ext. coefficient 6210
Abs 0.1% (=1 g/l) 0.158, assuming all pairs of Cys residues form cystines

Ext. coefficient 5960
Abs 0.1% (=1 g/l) 0.152, assuming all Cys residues are reduced

Estimated half-life:

The N-terminal of the sequence considered is M (Met).

The estimated half-life is: 30 hours (mammalian reticulocytes, in vitro).
>20 hours (yeast, in vivo).
>10 hours (Escherichia coli, in vivo).

Instability index:

The instability index (II) is computed to be 63.44
This classifies the protein as unstable.

Aliphatic index: 50.44

Grand average of hydropathicity (GRAVY): -1.009

prabi
Pôle Rhône-Alpes de Bioinformatique

PRABI-GERLAND
RHONE-ALPES BIOINFORMATIC POLE GERLAND SITE

- Institute of Biology and Protein Chemistry

Job **SOPMA** (ID: 2189f351740b) is running on **NPS@** server (started on 20180219-070250).
Results will be shown below. **Please wait and don't go back.**

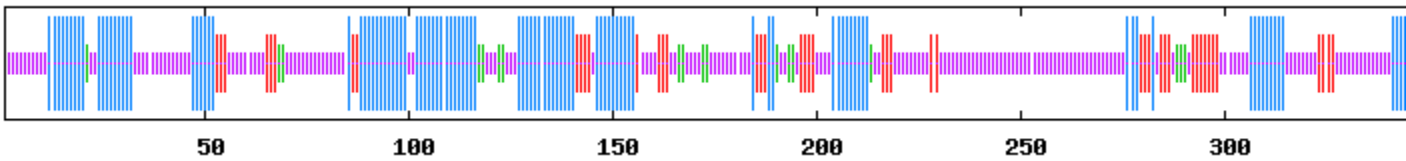
In your publication cite :
NPS@: Network Protein Sequence Analysis
TIBS 2000 March Vol. 25, No 3 [291]:147-150
Combet C., Blanchet C., Geourjon C. and Deléage G.

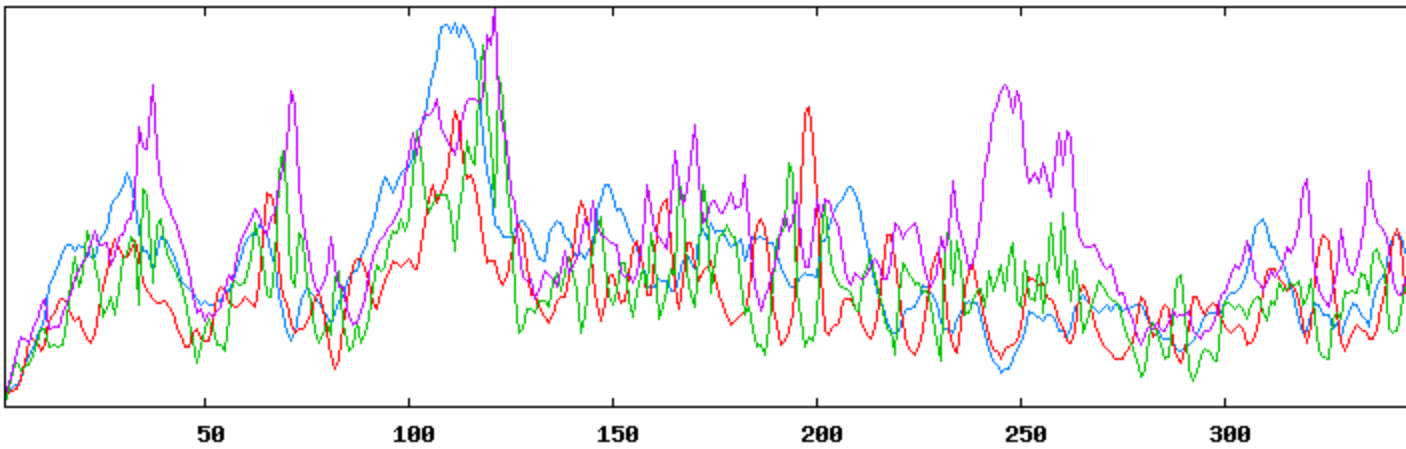
SOPMA result for : UNK_575060

Abstract Geourjon, C. & Deléage, G., SOPMA: Significant improvement in protein secondary structure prediction by consensus prediction from multiple alignments., Cabios (1995) 11, 681-684
View SOPMA in: [AnTheProt (PC) , Download...] [HELP]

```
        10        20        30        40        50        60        70
         |         |         |         |         |         |         |
MLKDSGPLFNTDYDILKSNLQNCSKFSAIQCAAAVPRAPAESSSSSKKFEQSHLHMSSETQANNELTTNG
eeccccccccchhhhhhhhtcchhhhhhhhccccccccccccchhhhheeeccccccccceeett
HGPPASKQVSQQPKGIMGMFASKAAAKTQETNKETKTEAKEVTNASAAGNKAPGKGNMMSNFFGKAAMNK
cccccccccccccccheehhhhhhhhhhhccchhhhhhhhhhhhhhhttccctttcccchhhhhhhhhhhh
FKVNLDSEQAVKEEKIVEQPTVSVTEPKLATPAGLKKSSKKAEPVKVLQKEKKRGKRVALSDDETKETEN
heeeechhhhhhhhhheccccecccttcccctttcccccccccheeehhtcctttceeeecccchhhhhh
MRKKRRRIKLPESDSSEDEVFPDSPGAYEAESPSPPPPPSPPLEPVPKTEPEPPSVKSSSGENKRKRKRV
hhhtcceeeccccccccceeccccccccccccccccccccccccccccccccccccccccccccchhhe
LKSKTYLDGEGCIVTEKVYESESCTDSEEELNMKTSSVHRPPAMTVKKEPREERKGPKKGTAALGKANRQ
eehceeecttttceeeeeeecccccccchhhhhhhhhcccccccceeeecccccccccccccchhhhhttcc
VSITGFFQRK
```

```
eeeeeeehhc
Sequence length :   360
SOPMA :
   Alpha helix      (Hh) :   108 is  30.00%
   3₁₀ helix        (Gg) :     0 is   0.00%
   Pi helix         (Ii) :     0 is   0.00%
   Beta bridge      (Bb) :     0 is   0.00%
   Extended strand (Ee) :    54 is  15.00%
   Beta turn        (Tt) :    20 is   5.56%
   Bend region      (Ss) :     0 is   0.00%
   Random coil      (Cc) :   178 is  49.44%
   Ambiguous states (?)  :     0 is   0.00%
   Other states          :     0 is   0.00%
```






```
Parameters :
   Window width          :  17
   Similarity threshold :   8
   Number of states      :   4
```

Prediction result file (text): [SOPMA]
Intermediate result file (text): [BLASTP on NRPROT]

User : *public@157.48.211.84.* ***Last modification time :*** *Mon Feb 19 07:03:09 2018.* ***Current time :*** *Mon Feb 19 07:03:09 2018*

RCSB PDB PROTEIN DATA BANK

137692 Biological Macromolecular Structures Enabling Breakthroughs in Research and Education

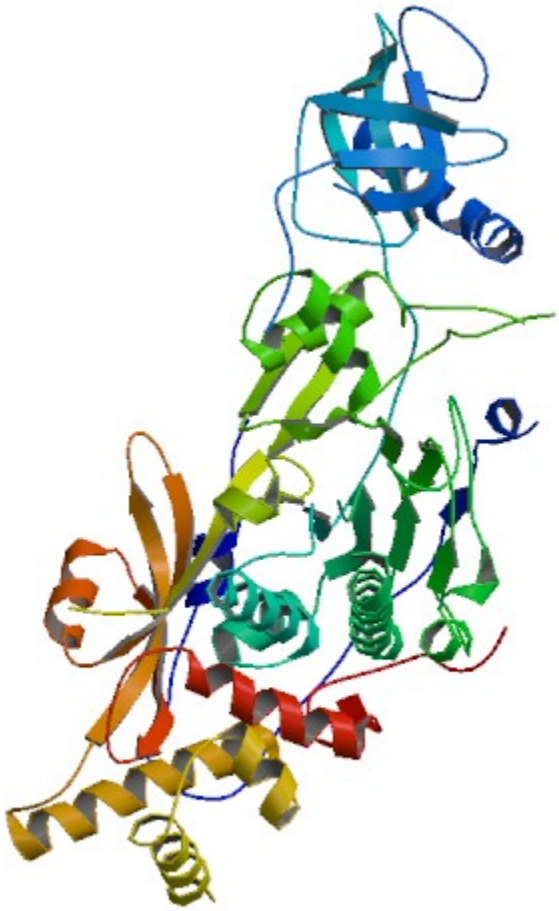
Go

Advanced Search | Browse by Annotations

RCSB PDB-101

WORLDWIDE PDB PROTEIN DATA BANK

Biological Assembly 1



3D View: Structure I Electron Density

Standalone Viewers
Protein Workshop I Ligand Explorer

Global Symmetry: Asymmetric - C1
Global Stoichiometry: Hetero 2-mer - AB

Biological assembly 1 assigned by authors and generated by PISA (software)

PreviousNext

Macromolecule Content

- Total Structure Weight: 273830.72
- Atom Count: 17092
- Residue Count: 2480
- Unique protein chains: 2

Display Files
Download Files

3E0J

X-ray structure of the complex of regulatory subunits of human DNA polymerase delta

- **DOI:** 10.2210/pdb3E0J/pdb

- **Classification:** TRANSFERASE
- **Organism(s):** Homo sapiens

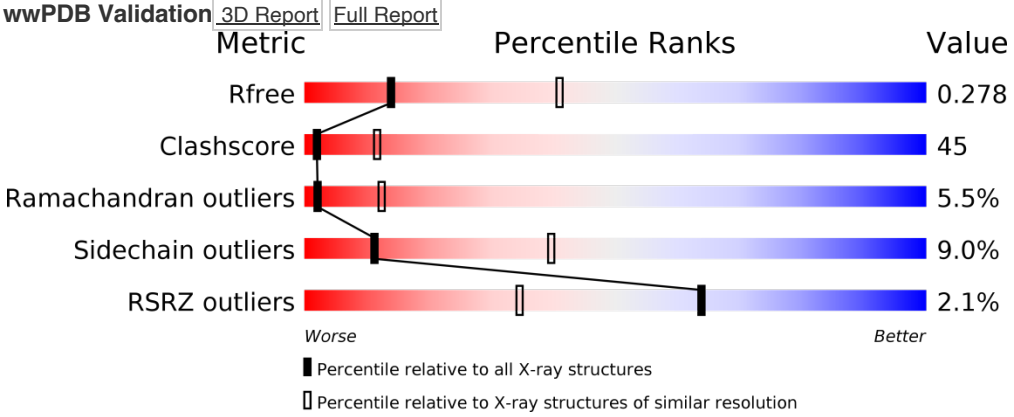
- **Deposited:** 2008-07-31 **Released:** 2008-12-16

- **Deposition Author(s):** Baranovskiy, A.G., Babayeva, N.D., Pavlov, Y.I., Vassylyev, D.G., Tahirov, T.H.

Experimental Data Snapshot

- **Method:** X-RAY DIFFRACTION
- **Resolution:** 3 Å
- **R-Value Free:** 0.281
- **R-Value Work:** 0.257

wwPDB Validation 3D Report Full Report



This is version 1.1 of the entry. See complete history.

Literature

Download Primary Citation

X-ray structure of the complex of regulatory subunits of human DNA polymerase delta.

Baranovskiy, A.G., Babayeva, N.D., Liston, V.G., Rogozin, I.B., Koonin, E.V., Pavlov, Y.I., Vassylyev, D.G., Tahirov, T.H.

(2008) Cell Cycle **7**: 3026-3036

- **PubMed**: 18818516
- **DOI:** 10.4161/cc.7.19.6720

- **PubMed Abstract:**

The eukaryotic DNA polymerase delta (Pol delta) participates in genome replication, homologous recombination, DNA repair and damage tolerance. Regulation of the plethora of Pol delta functions depends on the interaction between the second (p50) and t ...

Macromolecules

Find similar proteins by: Sequence | Structure

Entity ID: 1

Molecule	Chains	Sequence Length	Organism	D
DNA polymerase subunit delta-2	A, C, E, G	476	Homo sapiens	G E

Find proteins for P49005 *(Homo sapiens)*

Go to Gene View: POLD2

- **Protein Feature View**

P49005P49005 - DPOD2_HUMAN - DNA polymerase delta subunit 2Molec. ProcessingDNA polymerase delta subunit 2UI
Validation3E0J.ATTTTTTT3E0J.CTTTTTTT3E0J.ETTTTTTT3E0J.GTTTTTTTPDB+

Entity ID: 2

Molecule	Chains	Sequence Length	Organism	D
DNA polymerase subunit delta-3	B, D, F, H	144	Homo sapiens	G

Find proteins for Q15054 *(Homo sapiens)*

Go to Gene View: POLD3

- **Protein Feature View**

Experimental Data & Validation

Experimental Data

- **Method:** X-RAY DIFFRACTION
- **Resolution:** 3 Å
- **R-Value Free:** 0.281
- **R-Value Work:** 0.257
- **Space Group:** P 1 2_1 1

Unit Cell:

Length (Å)	Angle (°)
a = 95.128	α = 90.00
b = 248.535	β = 106.94
c = 103.461	γ = 90.00

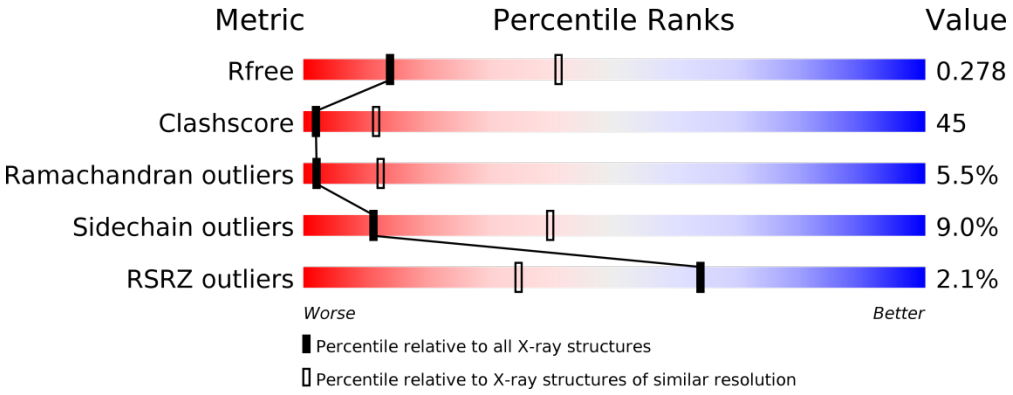
Software Package:

Software Name	Purpose
CrystalClear	data collection

Software Name	Purpose
DM	phasing
SCALEPACK	data scaling
CNS	refinement
DENZO	data reduction

Structure Validation

View Full Validation Report or Ramachandran Plots



View more in-depth experimental data

Entry History

Deposition Data

- **Deposited Date:** 2008-07-31
- **Released Date:** 2008-12-16
- **Deposition Author(s):** Baranovskiy, A.G., Babayeva, N.D., Pavlov, Y.I., Vassylyev, D.G., Tahirov, T.H.

Revision History

- **Version 1.0: 2008-12-16**
  Type: Initial release
- **Version 1.1: 2011-07-13**
  Type: Version format compliance

RCSB PDB (citation) is managed by two members of the
Research Collaboratory for Structural Bioinformatics:
Rutgers and UCSD/SDSC

RUTGERS | UC San Diego SDSC

RCSB PDB is a member of the

WORLDWIDE PDB PROTEIN DATA BANK

EMDataBank
Unified Data Resource for 3DEM

The RCSB PDB is funded by a grant (DBI-1338415) from the National Science Foundation, the National Institutes of Health, and the US Department of Energy.

Contact Us