



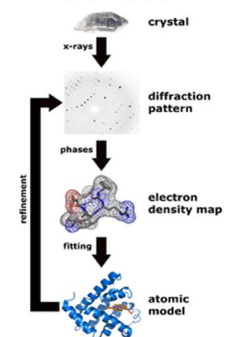
## Protein Structure prediction

## Review of Proteins

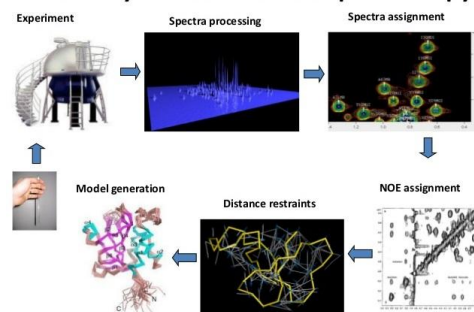
- ❑ Proteins: Polypeptides with a three dimensional structure
- ❑ Protein primary structure- Sequence of amino acids constituting polypeptide chain
- ❑ Protein secondary structure- Local organization of polypeptide chain into secondary structures such as  $\alpha$  helix,  $\beta$  sheet,  $\beta$  turns,  $\Omega$  loops, 3/10 helices etc.
- ❑ Protein Tertiary structure- Three dimensional arrangements of amino acids as they react to one another due to polarity and interactions between side chains.
- ❑ Protein Quaternary structure: Interaction of several protein subunit

## How Protein Structures are Determined?

### ▪ X-ray crystallography



### ▪ NMR spectroscopy



RCSB PDB Deposit Search Visualize Analyze Download Learn More

**RCSB PDB** 144571 Biological Macromolecular Structures Enabling Breakthroughs in Research and Education

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

Advanced Search | Browse by Annotations

[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.

Openings with RCSB PDB at UCSD

**JOIN OUR TEAM**

**October Molecule of the Month**

**Aminoglycoside Antibiotics and Resistance**

Latest Entries As of Tuesday Oct 02 2018 Features & Highlights News Publications

RCSB PDB Deposit Search Visualize Analyze Download Learn More **WYPR**

**RCSB PDB** 144871 Biological Macromolecular Structures Enabling Breakthroughs in Research and Education

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

Advanced Search | Browse by Annotations

**Structure Summary** 3D View Annotations Sequence Sequence Similarity Structure Similarity Experiment

**2TRX**

CRYSTAL STRUCTURE OF THIOREDOXIN FROM ESCHERICHIA COLI AT 1.68 ANGSTROMS RESOLUTION

DOI: 10.2210/pdb2TRX/pdb

Classification: [ELECTRON TRANSPORT](#)

Organism(s): [Escherichia coli \(strain K12\)](#)

Deposited: 1990-03-19 Released: 1991-10-15

Deposition Author(s): [Katti, S.K.](#), [Lemaster, D.M.](#), [Eklund, H.](#)

Experimental Data Snapshot

Method: X-RAY DIFFRACTION

Resolution: 1.68 Å

wwPDB Validation **3D Report** **Full Report**

Metric	Percentile Ranks	Value
Clashscore		5
Ramachandran outliers		0.9%
Sidechain outliers		1.7%

Experimental Data Snapshot

Method: X-RAY DIFFRACTION

Resolution: 1.68 Å

3D View: [Structure](#) | [Ligand Interaction](#)

Standalone Viewers

[Protein Workshop](#) | [Ligand Explorer](#)

This is version 1.3 of the entry. See complete [history](#).

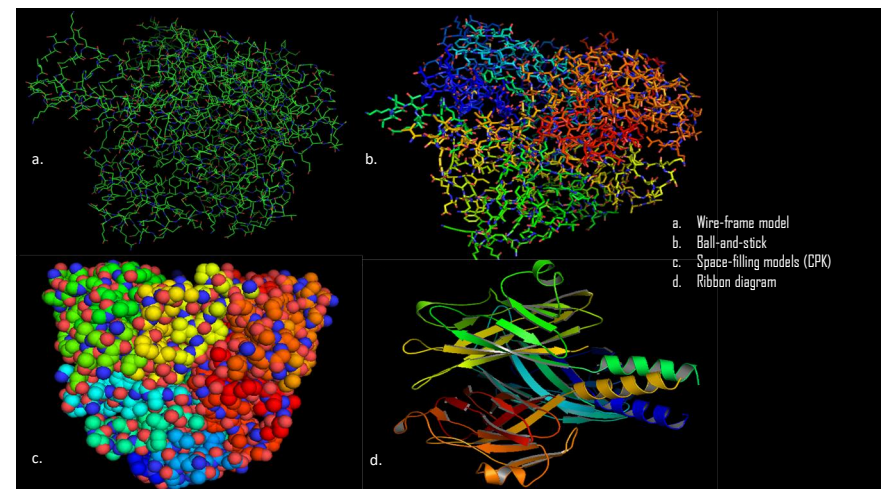
## Other Structure Databases

- MSD (<https://www.ebi.ac.uk/training/online/glossary/macromolecular-structure-database>)
- MMDB ([https://www.ncbi.nlm.nih.gov/Structure/MMDB/docs/mmdb\\_help.html](https://www.ncbi.nlm.nih.gov/Structure/MMDB/docs/mmdb_help.html))
- PDBSum (<http://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/pdbsum/GetPage.pl?pdbcode=index.html>)
- TargetDB (<http://europepmc.org/abstract/MED/10488778>)

## Three-Dimensional Visualization of Proteins

Typically protein structure can be drawn in just four standard ways:

- ❖ Wire-frame model
- ❖ Ball-and-stick
- ❖ Space-filling models (CPK)
- ❖ Ribbon diagram



## Tools for 3D Visualization of Proteins

There are perhaps more than two dozen freely available macromolecular visualization programs that can be found on the web.

- RasMol
- PyMol
- WebMol
- Cn3D
- SwissPDB-viewer/ DeepView
- MICE
- VRML
- Protein Explorer etc.

## Protein Structure Prediction

### Three different methods:

- Homology (or comparative) modeling
- Threading
- ab initio

## Ab initio Structure prediction

- Prediction from the beginning
- Predict 3D structure without the knowledge of any related 3D structure
- This method is still very experimental and quite unreliable

**Programs:** ROSETTA (not publicly available)

## Threading

- Predicting the structure, or recognizing a common fold in proteins having essentially no sequence homology, to any protein in PDB
- Picks up where homology modeling leaves off
- Limited to generating very approximate folds

**Programs:** SAMt99, three-dimensional-PSSM, FUGUE etc.

## Homology Modeling

- Most powerful and accurate approach
- Based on the coordinates of known homologs found in PDB
- Quality of the model strongly depends on the degree of similarity between the query sequence and the matching database sequence
- Highest degree of similarity being modeled best
- Actual structure drops by approximately 0.5 Å for each 10% reduction in sequence identity

## Steps of Homology Modeling

Homology modeling can be decomposed into five different steps:

1. Aligning the query or unknown protein sequence to the sequence of a known structure
2. Using the alignment to select and replace backbone segments (usually loops that are contained in a special loop library) that need to be altered because of insertions or deletions
3. Replacing side chains that have been changed due to the alignment or loop insertion and deletion process
4. Refining the model using energy minimization to relieve collisions or steric strains
5. Validating the model using visual inspection and software validation tools

## Homology Modeling

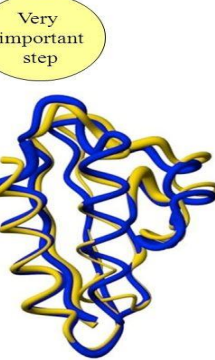
Identify homologous protein structures

Align query sequence with template sequence

Build a model for the query sequence  
Core modeling, side chain modeling  
Loop modeling

Model evaluation

Model refinement



Very important step

Most of the steps can be automated!!

## Tools for Homology Modeling

Those can be downloaded and installed on the Unix and Windows platforms:

- Modeller
- DeepView
- WHATIF

Web servers are also available for homology modeling:


- SWISS-MODEL server(<https://swissmodel.expasy.org>)
- CPH Models server(<http://www.cbs.dtu.dk/services/CPHmodels/>)
- SDSCl server(<http://pdg.cnb.uam.es/eva/cm/doc/sdsccl.html>)

**SWISS-MODEL** Source of life. Models to Research in Science Modeling Repository Tools Documentation Log in Create Account

Welcome to SWISS-MODEL

SWISS-MODEL is a fully automated protein structure homology-modelling server, accessible via the E-PRISy web server, or from the program DeepView (Swiss Pdb-Viewer). The purpose of this server is to make protein modeling accessible to all life science researchers worldwide


[Start Modelling](#)



Celebrating 25 years of SWISS-MODEL!  
Thursday, 18th October in Basel  
[Details and registration](#)

Protein Structure Bioinformatics Group  
c/o Prof. Tordis Schwede  
Swiss Institute of Bioinformatics  
Biozentrum, University of Basel  
Klingelbergstrasse 50/70  
CH-4055 Basel | Switzerland  
[help-swissmodel@unibas.ch](mailto:help-swissmodel@unibas.ch)

**BIOZENTRUM**  
The Center for  
Molecular Life Sciences



**SWISS-MODEL** Source of life. Models to Research in Science Modeling Repository Tools Documentation Log in Create Account

Start a New Modelling Project

Target Sequence:  
(Format must be FASTA, GCG, plain string, or a valid UniProt AC)

[Upload Target Sequence File...](#) [Validate](#)

Project Title:

Email:

[Search For Templates](#) [Build Model](#)

Supported Inputs

- Sequence(s)
- Target-Template Alignment
- User Template
- DeepView Project

By using the SWISS-MODEL server, you agree to comply with the following [terms of use](#) and to cite the corresponding [articles](#).

You are currently not logged in - to take advantage of the workspace, please [log in](#) or [create an account](#).  
(There is no requirement to create an account to use any part of SWISS-MODEL, however you will gain the benefit of seeing a list of your previous modelling projects here.)

**SWISS-MODEL** Source of life. Models to Research in Science Modeling Repository Tools Documentation Log in Create Account

Start a New Modelling Project

Target Sequence:  
(Format must be FASTA, GCG, plain string, or a valid UniProt AC)

[Upload Target Sequence File...](#) [Validate](#)

Project Title:

Email:

[Search For Templates](#) [Build Model](#)

Supported Inputs

- Sequence(s)
- Target-Template Alignment
- User Template
- DeepView Project



By using the SWISS-MODEL server, you agree to comply with the following [terms of use](#) and to cite the corresponding [articles](#).

You are currently not logged in - to take advantage of the workspace, please [log in](#) or [create an account](#).  
(There is no requirement to create an account to use any part of SWISS-MODEL, however you will gain the benefit of seeing a list of your previous modelling projects here.)

**SWISS-MODEL** Source of life. Models to Research in Science Modeling Repository Tools Documentation Log in Create Account


All Projects

Untitled Project Created: today at 15:14


Summary **Templates: 19** Models  

Template Results

IT Sort	Name	Title	Coverage	QMQE	QIQE	Identity	Method	Oligo State	Ligands
1	2ztl1.A	25kDa structural protein VP25	0.00	0.04	00.00	X-ray 2.0Å	homodimer	/	None
2	3jg1.1.B	Methane monooxygenase subunit B2	0.15	-	10.92	X-ray 2.0Å	hetero-9-mer	9 x ZN <sup>2+</sup> , 3 x CUA <sup>2+</sup> , 2 x CU <sup>2+</sup>	
3	3jg2.1.B	Fibroblast growth factor receptor 2	0.15	-	10.09	X-ray 2.3Å	hetero-dimer	/	None
4	4jkl.1.E	Particulate methane monooxygenase subunit B	0.15	0.00	14.06	X-ray 2.0Å	hetero-12-mer	4 x PGD <sup>2+</sup> , 6 x CU <sup>2+</sup>	
5	3jkl.1.A	ProB	0.14	0.00	14.06	X-ray 2.1Å	hetero-11-mer	6 x ZN <sup>2+</sup> , 2 x CU <sup>2+</sup> , 1 x CUA <sup>2+</sup>	
6	5jld.1.A	Conserved hypothetical secreted protein	0.00	-	15.38	X-ray 1.6Å	monomer	/	3 x ZN <sup>2+</sup> , 1 x B6W <sup>2+</sup>
7	4qfn.1.A	Conserved hypothetical secreted protein	0.00	-	15.38	X-ray 1.6Å	monomer	/	3 x CA <sup>2+</sup>
8	4kfs.1.A	Conserved hypothetical secreted protein	0.00	-	15.38	X-ray 2.0Å	monomer	/	7 x ZN <sup>2+</sup> , 1 x APT <sup>2+</sup> , 1 x CA <sup>2+</sup>
9	7yee.1.A	particulate methane monooxygenase, B subunit	0.00	-	13.66	X-ray 2.0Å	hetero-trimer	3 x ZN <sup>2+</sup> , 1 x CUA <sup>2+</sup> , 1 x CU <sup>2+</sup>	
10	4uon.1.A	Conserved hypothetical secreted protein	0.00	-	13.01	X-ray 1.6Å	monomer	/	4 x ZN <sup>2+</sup> , 1 x APT <sup>2+</sup>
11	1u11.B	FIBROBLAST GROWTH FACTOR RECEPTOR 2	0.04	-	10.87	X-ray 2.3Å	hetero-dimer	/	None
12	2jkl.1.A	TITN	0.04	-	4.17	X-ray 1.6Å	monomer	/	2 x 12P <sup>2+</sup>

[Build Models](#) 

[Clear Selection](#)



## Protein Structure Evaluation

Good protein structure (here we are primarily referring to water-soluble proteins) should:

1. Minimize the number of torsion angles in disallowed regions of the Ramachandran plot
2. Maximize the number of hydrogen bonds
3. Minimize the number of exposed hydrophobic residues
4. Maximize the number of exposed polar or charged residues
5. Minimize the number of interstitial cavities or packing defects
6. Minimize the number of number of nonbonded atoms within 2.6 Å
7. Minimize the standard deviation in hydrogen bond energies
8. Minimize the standard deviation in dihedral angles for helices
9. Have a low R factor (<0.20 for X-ray structures) or a low backbone RMSD value (< 0.8 Å for NMR structure ensembles)

## Tools for Protein Structure Evaluations

-DSSP(<http://2struc.cryst.bbk.ac.uk/about/>)

-PROCHECK(<https://www.ebi.ac.uk/thornton-srv/software/PROCHECK/download.html>)

-VADAR(<http://vadar.wishartlab.com/>)

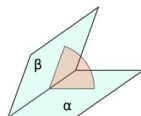
-Verify3D([http://services.mbi.ucla.edu/Verify\\_3D/](http://services.mbi.ucla.edu/Verify_3D/))

-Rampage(<http://mordred.bioc.cam.ac.uk/~rapper/rampage.php>)

## Ramachandran Plot Analysis

-A **dihedral angle** is the angle between two intersecting planes.

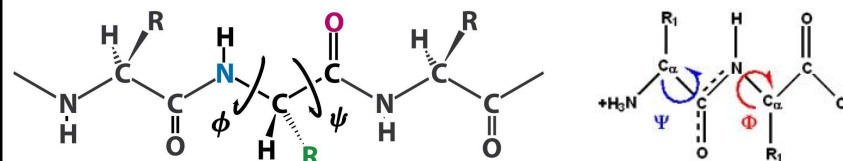
-In chemistry, a **torsion angle** is defined as a particular example of a dihedral angle, describing the geometric relation of two parts of a molecule joined by a chemical bond.



## Ramachandran Plot Analysis cont.

-Dihedral angle along the N-C $\alpha$  bond are called the  $\phi$  (phi) angle

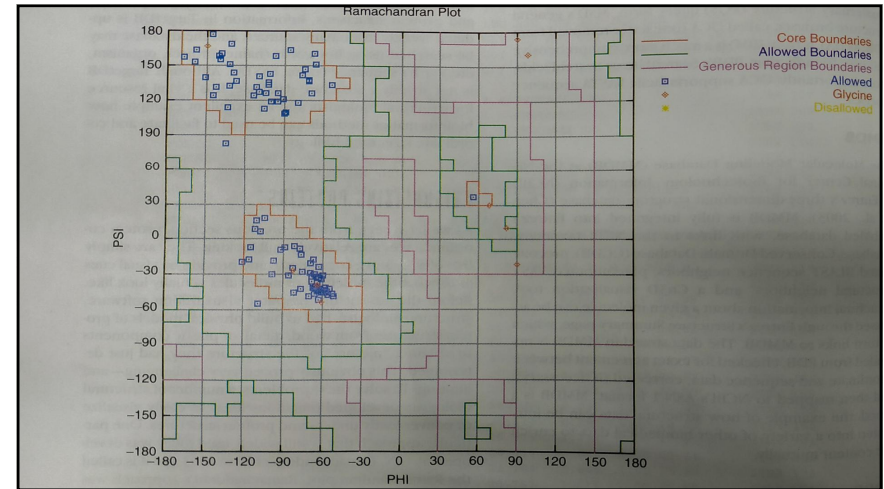
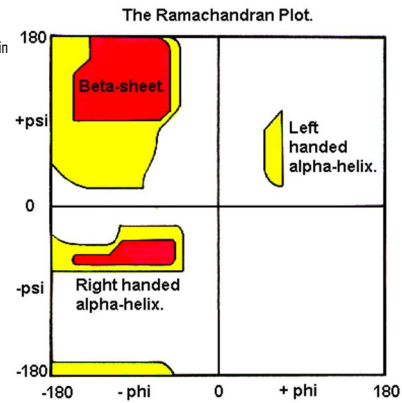
-Dihedral angle along the C $\alpha$ -C bond are called the  $\psi$  (psi) angle



## Ramachandran Plot Analysis Cont.

- By plotting the backbone torsion and dihedral angles of each residue in a simple two dimensional scatter plot

- $\Phi$  angles on x-axis,  $\Psi$  angles on y-axis
- On  $\beta$  sheets,  $\Phi, \Psi$  coordinates  $(-120^\circ, 120^\circ)$
- On  $\alpha$  helix,  $\Phi, \Psi$  coordinates  $(-60^\circ, -40^\circ)$
- Approximately 75% of the rest of the Ramachandran plot is empty



## Ramachandran Plot Analysis Cont.

- The "Core Boundaries" or red line on the plot delineate the region in the Ramachandran plot where approximately 85% of residues should be found in good quality structures.
- The "Allowed Boundaries" (green line) delineate the region where approximately 10% of residues should be found.
- Residues falling in the "Generally Allowed Boundaries" (yellow line) or outside this region indicate residues that have serious steric problems.
- Glycine residues are the exception as they can appear anywhere in the plot.
- Protein structures that are found to have a high percentage (>15%) of nonglycine residues in disallowed regions inevitably are found to be poor-quality structures.

## Example of Ramachandran plot

