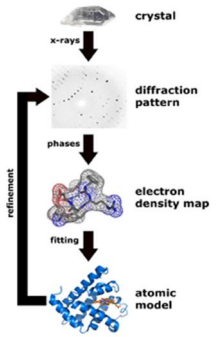
Protein Structure prediction

Review of Proteins

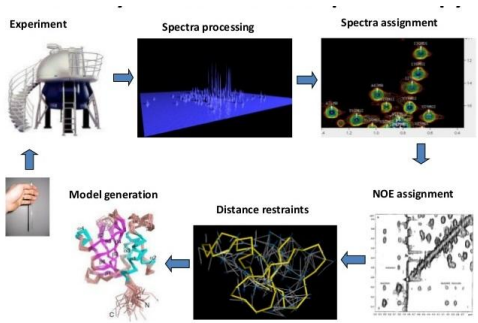
- ❑Proteins: Polypeptides with a three dimensional structure
- ❑Protein primary structure- Sequence of amino acids constituting polypeptide chain
- ❑Protein secondary structure- Local organization of polypeptide chain into secondary structures such as α helix, β sheet, β turns, Ω loops, 3/10 helices 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

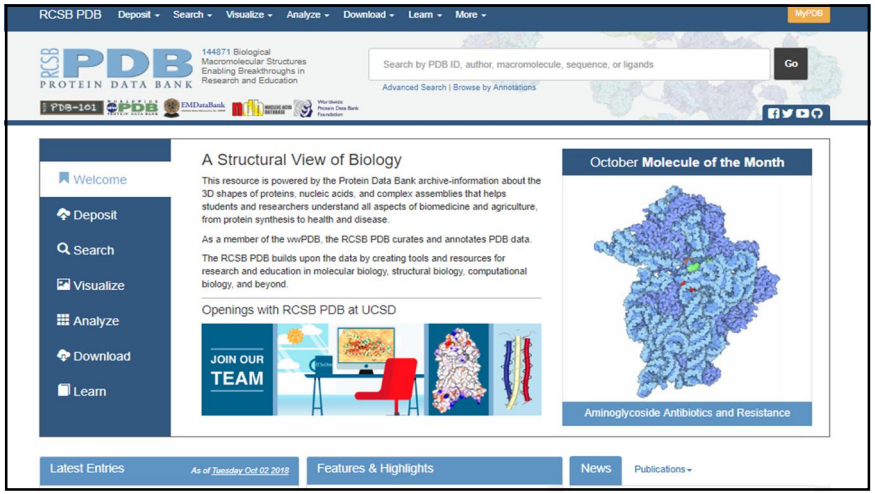
How Protein Structures are Determined?

- X-ray crystallography



- NMR spectroscopy





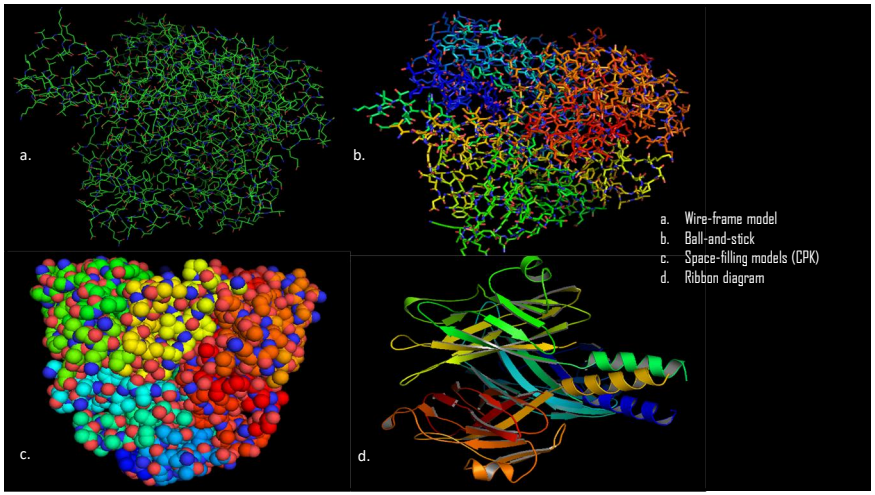
Other Structure Databases

- 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)
- PDBSum(http://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/pdbsum/GetPage.pl?pdbcode=index.html)
- TargetDB(http://europepmc.org/abstract/MED/10498778)

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 wed.

-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

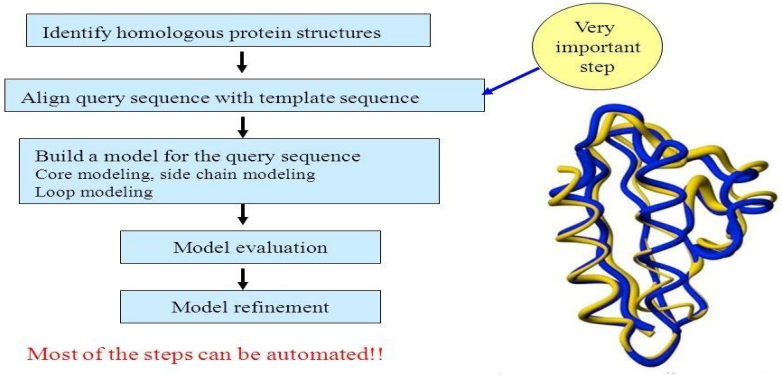
-Actual structure drops by approximately 0.5 Å for each 10% reduction in sequence identity

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 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



Tools for Homology Modeling

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

-Modeller

-DeepVieW

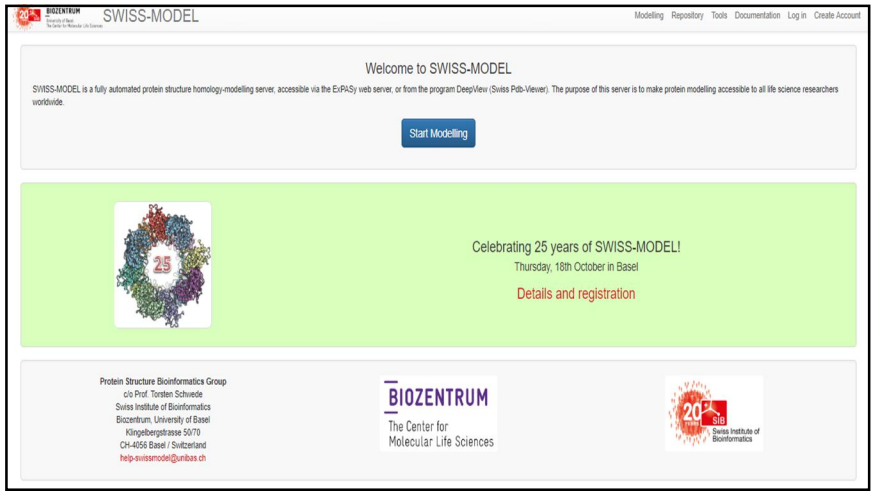
-WHATIF

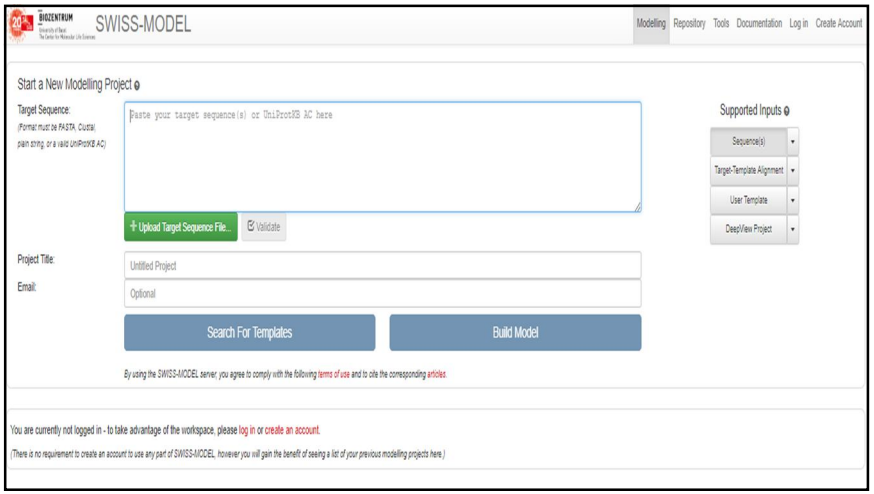
Web servers are also available for homology modeling:

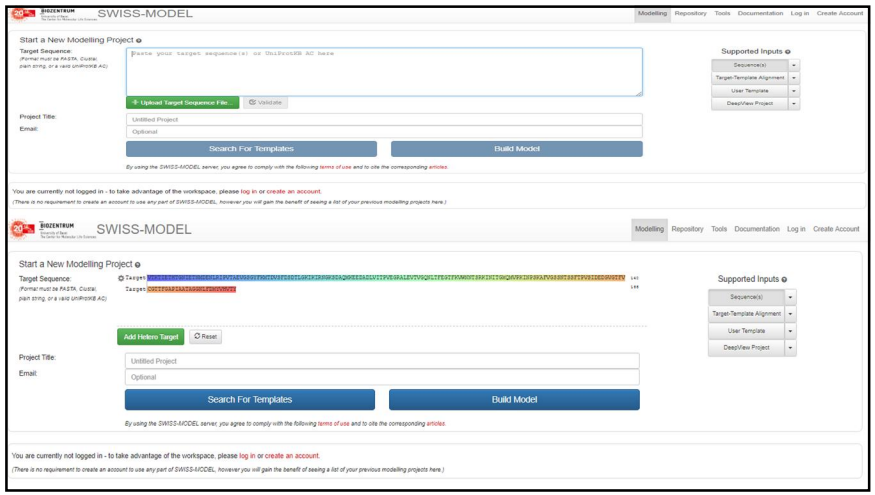
-SWISS-MODEL server(https://swissmodel.expasy.org)

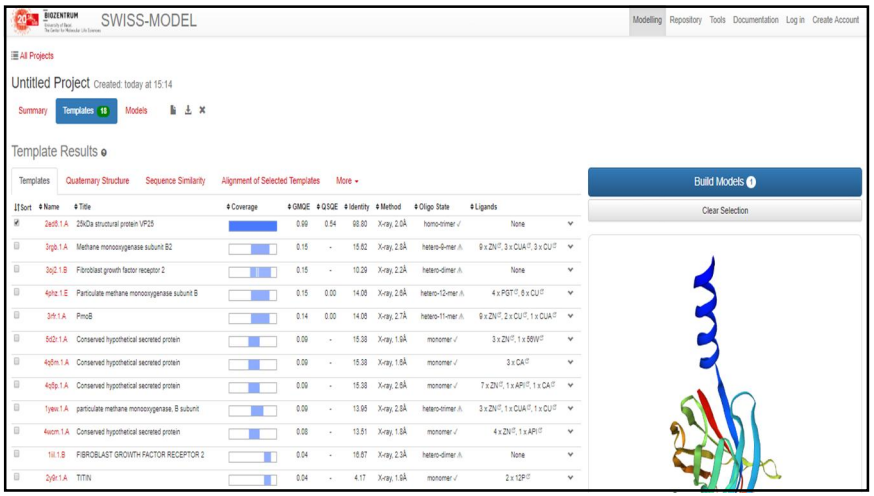
-CPH Models server(http://www.cbs.dtu.dk/services/CPHmodels/)

-SDSCI server(http://pdg.cnb.uam.es/eva/cm/doc/sdscl.html)


SWISS-MODEL
Modelling 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 ExPASy web server, or from the program DeepView (Swiss Pdb-Viewer). The purpose of this server is to make protein modelling 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. Torsten Schwede
Swiss Institute of Bioinformatics
Biozentrum, University of Basel
Klingelbergstrasse 50/70
CH-4056 Basel / Switzerland
help-swissmodel@unibas.ch
BIOZENTRUM
The Center for Molecular Life Sciences
Swiss Institute of Bioinformatics


SWISS-MODEL
Modelling Repository Tools Documentation Log in Create Account
Start a New Modelling Project
Target Sequence:
(Format must be FASTA, Clustal, plain string, or a valid UniProtKB AC)
Paste your target sequence(s) or UniProtKB AC here
Upload Target Sequence File...
Validate
Supported Inputs
Sequence(s)
Target-Template Alignment
User Template
DeepView Project
Project Title:
Untitled Project
Email:
Optional
Search For Templates
Build Model
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
Start a New Modelling Project
Target Sequence:
Paste your target sequence(s) or UniProtKB AC here
Upload Target Sequence File...
Validate
Supported Inputs
Project Title:
Untitled Project
Email:
Optional
Search For Templates
Build Model
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
Start a New Modelling Project
Target Sequence:
Add Hetero Target
Reset
Project Title:
Untitled Project
Email:
Optional
Search For Templates
Build Model
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
Modelling Repository Tools Documentation Log in Create Account
All Projects
Untitled Project Created: today at 15:14
Summary
Templates 18
Models
Template Results
Templates
Quaternary Structure
Sequence Similarity
Alignment of Selected Templates
More
Build Models
Clear Selection
Sort
Name
Title
Coverage
GMQE
QSQE
Identity
Method
Oligo State
Ligands
2ad5.1.A 25kDa structural protein VP26
3rgb.1.A Methane monooxygenase subunit B2
3oj2.1.B Fibroblast growth factor receptor 2
4phz.1.E Particulate methane monooxygenase subunit B
3rfr.1.A PmoB
5d2r.1.A Conserved hypothetical secreted protein
4p5m.1.A Conserved hypothetical secreted protein
4p5g.1.A Conserved hypothetical secreted protein
1yew.1.A particulate methane monooxygenase, B subunit
4wcm.1.A Conserved hypothetical secreted protein
1il.1.B FIBROBLAST GROWTH FACTOR RECEPTOR 2
2y9r.1.A TITIN

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/)

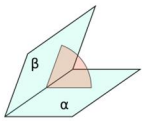
-Verify3D(http://services.mbi.ucla.edu/Verify_3D/)

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

Ramachandran Plat Analysis

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

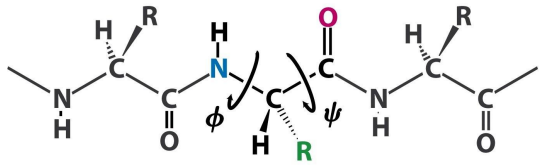
-In chemistry, a **torsion angle** is defined as a particular example of a dihedral angle, describing the geometric relation of two parts of a molecule joined by a chemical bond.

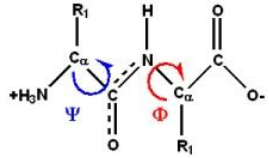


Ramachandran Plat Analysis cont.

-Dihedral angle along the N-Cα bond are called the ϕ (phi) angle

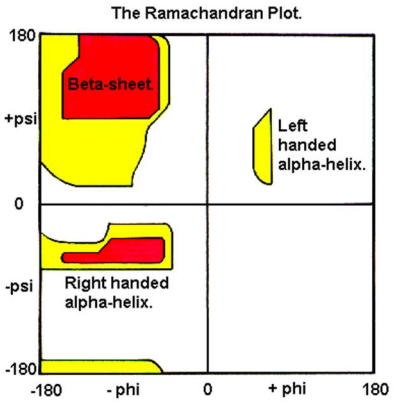
-Dihedral angle along the Cα-C bond are called the ψ (psi) angle

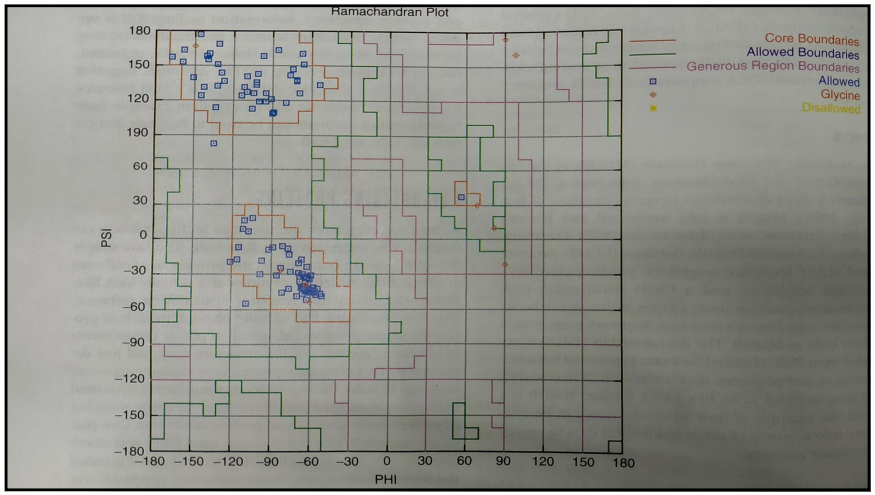




Ramachandran Plot Analysis Cont.

- By plotting the backbone torsion and dihedral angles of each residue in a simple two dimensional scatter plot
- Φ angles on x-axis, ψ angles on y-axis
- On β sheets, Φ, ψ coordinates (-120°, 120°)
- On α helix, Φ, ψ coordinates (-60°, -40°)
- Approximately 75% of the rest of the Ramachandran plat 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 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