**BlastP**

**Aim:**

To perform database similarity search for the given query sequence by using BlastP.

**Description:**

Basic Local Alignment Search Tool, or BLAST, is an algorithm for comparing primary biological sequence information, such as the amino-acid sequences of different proteins or the nucleotides of DNA sequences. BLAST uses a heuristic algorithm that seeks local as opposed to global alignments and is therefore able to detect relationships among sequences that share only isolated regions of similarity. BlastP is a pair wise sequence comparison tool developed by NCBI and the programme compares an amino acid query sequence of a protein with amino acid sequence of protein data base. It takes amino acid sequences and compares them against the NCBI protein databases. The program allows to discover the structures and functions of proteins. BlastP uses the BLAST algorithm to compare an amino acid query sequence against a protein sequence database.

**Procedure:**

1. Open the BlastP program page from NCBI (www.ncbi.nlm.nih.gov)
2. Retrieve the protein sequence of interest from Uniprot.
3. Paste the retrieved protein sequence in FASTA format in the field titled "Enter Query Sequence” or Click on “Browse” and select the sequence.
4. Enter a Job Title, if necessary.
5. In the “Choose Search Set” section, change the database to “Protein Data Bank”.
6. Under “Program Selection”, select “blastp” (Protein-Protein blast).
7. Check the box “Show results in a new window” next to the “BLAST” button and click “BLAST”.
8. BLAST will now open a new window and shows it is working on your search.
9. Once the results are computed they will be presented in the window.

**BlastN**

**Aim:**

To perform database similarity search for the given query sequence by using BlastN.

**Description:**

BlastN is a pair wise sequence comparison tool developed by NCBI and the programme compares a nucleotide query sequence with nucleotide sequence database. It takes nucleotides sequences and compares them against the NCBI nucleotide databases. It is better at finding sequences similar, but not identical, to your query.

**Procedure:**

1. Open the BlastN program page from NCBI (www.ncbi.nlm.nih.gov)
2. Retrieve the nucleotide sequence from the NCBI.
3. Paste the retrieved nucleotide sequence in FASTA format in the field titled "Enter Query Sequence” or Click on “Browse” and select the sequence.
4. Enter a Job Title, if necessary.
5. In the “Choose Search Set” section, change the database to “Reference mRNA

sequences (refseq\_rna)”.

1. Under “Program Selection”, select “Somewhat similar sequences (blastn)”
2. Check the box “Show results in a new window” next to the “BLAST” button and click “BLAST”.
3. BLAST will now open a new window and shows it is working on your search.
4. Once the results are computed they will be presented in the window.

**ClustalW**

**Aim:**

To perform multiple sequence alignment using ClustalW.

**Description:**

Multiple sequence alignment (MSA) is a sequence alignment of three or more biological sequences, generally protein, DNA, or RNA. ClustalW is a widely used system for aligning any number of homologous nucleotide or protein sequences. ClustalW is a weighted variant to which access is provided by a large number of web portals including GenomeNet, EBI, and EMBNet. ClustalW helps in improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. ClustalW is used extensively for phylogenetic tree construction, which can be seen via. viewing cladogram and phylogram. From the resulting output, sequence homology can be inferred and phylogenetic analysis can be conducted to assess the sequences' shared evolutionary origins.

**Procedure:**

1. Open the ClustalW program page from GenomeNet tool.
2. Retrieve more than 2 sequences (either DNA or Protein) from NCBI database (www.ncbi.nlm.nih.gov) in FASTA format.
3. Paste the retrieved sequences onto program query box of ClustalW with labels or upload the file containing sequence.
4. Click the label Protein or DNA according to the sequence of interest.
5. Then click the requisite option in different places as per our requirement. Otherwise leave as such, the programme will take all default option.
6. Run the ClustalW.

**Modeller and SWISS-MODEL**

**Aim:**

To perform protein modelling using Modeller Software.

**Description:**

MODELLER is used for Homology or comparative modelling of protein three-dimensional structures. An aligned sequence to be modelled with the known related structures and it automatically calculates the model containing all the atoms. Modeller is equipped with the comparative protein structure modelling through the satisfaction of spatial restraints and it can also perform many additional tasks namely the de novo modelling of loops in protein structures.

**Procedure:**

1. Retrieve the sequence from UniProt.

2. Create a folder and save the sequence in the folder in the following format in text file (Notepad) and give a suitable filename.

>P1 ; <file name>

Sequence:<file name>:::::::0.00:0.00

Finally, the sequences should end with \*

The note pad in which the sequence is kept should be saved in “.ali” format.

3. Blast the sequence and save the top ranked target in the same folder in “.pdb” format.

4. Copy the Scripts like “align2d.py”, “build\_model.py” and “model-single.py” and save in the same folder.

5. Make the necessary changes in the above scripts by opening in notepad and save the changes.

6. Keep the folder in any drive and Open modeller through command prompt cd Enter directory File Name

7. The following commands can be used to execute the model development.

>mod 9.23 align2d.py

>mod 9.23 model-single.py

8. In the log file of model-single, the summary of all built models were created and check the dope score to select the best model with the lowest value.

9. The best model can be used for further refinement.

**Validation parameters:**

1. Ramachandran Plot analysis using PDBsum Generate from EMBL-EBI.

2. ProSA-web - Protein Structure Analysis

**CASTp server**

**Aim:**

To identify and measure surface accessible pockets as well as interior inaccessible cavities, for proteins molecules using CASTp server.

**Description:**

Computed Atlas of Surface Topography of proteins (CASTp) provides an online resource for locating, delineating and measuring concave surface regions on three-dimensional structures of proteins. These include pockets located on protein surfaces and voids buried in the interior of proteins. CASTp uses both the solvent accessible surface model (Richards' surface) and molecular surface model (Connolly's surface) to analytically measure the area and volume of each void and pocket. Besides this, CASTp also measures the size of each pockets and mouth openings. This enables to determine the accessibility of binding sites to different ligands and substrates. It can be used to study surface features, annotated functional information of proteins and specific roles of key residues of proteins.

**Procedure:**

1. Open the CASTp home page (http://cast.engr.uic.edu)
2. A graphic user interface allows users to request a CASTp calculation.
3. The user can provide the input in either of the two ways. One either can upload the structure of the protein or can type the four-letter PDB code of a protein structure if it is available from the Protein Data Bank.
4. In the “Choose File” section, upload the structure of the molecule in PDB format and set the default value for probe radius as 1.4 Ȧ.
5. Then click “submit” button to run the calculation.
6. Once the results are computed, they will be appeared on the new page.
7. User can also get result through the email in a text file with all the information and the file. poc can be viewed through Chimera visualization tool.

**Mol Inspiration and SWISS-ADME**

**Aim:**

To calculate the molecular properties, bioactivity and conversion of 2D structure to 3D structure using mol inspiration online server.

**Description:**

Mol inspiration is an online tool focused on the visualization of the drug molecules or compounds to support the modern Cheminformatics techniques. This tool helps in the manipulation of the molecule, generation of tautomers, molecular fragmentation, calculation of various molecular properties and prediction of the bioactivity. This also helps in the generation of the 2D structure into the 3D structure. It helps in the substructure searching to get the similar types of molecules.

**Procedure:**

1. Use the web address www.molinspiration.com to open the tool.
2. Select the ligand molecule and choose the SMILES format to use for the calculation of the properties.
3. Click the Calculation of Molecular Properties and prediction of Bioactivity
4. A new page will be opened and the ligand molecule can be drawn or the SMILES representation can be given. Then click calculate option to retrieve the properties.
5. Click Galaxy 3D Structure Generator to convert 2 dimensional structures into 3 dimensional structures.
6. Select Molecular Database- Substructure and Similarity search to retrieve similar

molecules from the web source.

**Gromacs**

**Aim:**

To perform molecular dynamics simulations using Gromacs.

**Description:**

Write by own.

**Procedure for Desmond:**

**System Builder**

1. Open Maestro, go to Applications or Task.
2. Select Desmond and choose System builder.
3. System builder consists of two tab: solvation and ions.
4. In solvation tab, click calculate option and minimize option.
5. Set the appropriate membrane (if required).
6. Select appropriate explicit water model.
7. Set appropriate boundary box with correct distance (Default: Orthorhombic).
8. In ions tab, add salt at 0.15 M concentration.
9. Click run and monitor the job.

**Molecular Dynamics Simulation**

Once the system is solvated properly, we can perform molecular dynamics simulations by following steps:

1. Select Desmond suite and choose Molecular dynamics simulations.
2. Load the solvated system in the panel.
3. Set the NVT, NPT, temperature and relaxation protocol.
4. Set the simulation time and trajectory interval.
5. Click run and monitor the job.
6. Once the job is completed, calculate the RMSD, RMSF and Radius of Gyration for the protein.

**Results:**